WESTERN SYDNEY
UNIVERSITY

Molecular Biomarkers in

Type 2 Diabetes and Colorectal Cancer

Ehsan Alvandi

Primary Supervisors:

A/Prof Kevin J. Spring and Prof Anandwardhan A. Hardikar

This thesis is submitted as a fulfilment of the requirements for the degree of:

Doctor of Philosophy

School of Medicine, Western Sydney University, NSW, Australia

02 April 2023

Acknowledgements

First, I would like to express my special thanks and gratitude to Professor Anandwardhan A. Hardikar and Associate Professor Kevin J. Spring for all their kind help and support. I would never forget how they guided my growth during my PhD journey.

I am also grateful for the kind help and support of Dr Mugdha Joglekar, Dr Wilson Wong, and Dr Ho Pham form Diabetes and Islet Biology Group, School of Medicine, Western Sydney University, in various aspects of my laboratory experiments and data analysis.

I would like to acknowledge the special help and support of Associate Professor Kerry Hitos for training me to analyse the Concord colorectal surgery database and Associate Professor James Toh for his clinical insights. I would also like to thank the Concord Institute of Academic Surgery team, Professor Pierre Chapuis, Associate Professor Matthew Rickard, Dr Kheng-Seong Ng, and Ms Gael Sinclair for their guidance in using the colorectal surgery registry.

I would like to thank Dr Rossana Rosa Porto for mouse behavioural test training and the WSU Behavioural Neuroscience team, Professor Tim Karl, and Dr Rose Chesworth Vieyra for their support. Also, I would like to acknowledge the help of the WSU animal facility staff, Miss Nikola Mills and Miss Amy Woodley.

I would like to acknowledge Professor Ronald Ma and colleagues for the Hong Kong diabetes cohort samples. I am also thankful to cadaveric pancreas donors and islet isolation teams at the Westmead islet isolation centre, in Sydney, and the St. Vincent's Institute for Medical Research islet isolation centre in Melbourne. Also, I would like to thank Professor Roger Reddel and colleagues at the Children's Medical Research Institute (CMRI).

My special thanks to Western Sydney University for providing my scholarship support and the infrastructure support that made this PhD possible.

Finally, I really do not know how to thank my wife, Mrs Ameneh Najdi, for all her help and being a spiritual guide throughout my PhD journey. I am also grateful of being the father of two lovely daughters and for their patience and understanding.

Challenges in Thesis Completion

My PhD journey was affected by three major factors.

i) **Change of supervisory panel and PhD project**: At the end of year one of my PhD, I had to leave my (former) supervisory team as well as my former research work/project. Then, with the help of the Deputy Dean and School of Medicine, and the Graduate Research School (GRS) support, I was able to find a new (current) supervisory team and start with a new project in January 2021 in a field of research that was very different from my previous area. Considering my international student status and the need to seek employment whilst being the primary breadwinner to a family of four, I had under three years to complete the research presented in this thesis.

ii) **UN Compliance**: Shortly after moving to my new supervisory team, I was informed by the GRS to wait for UN Compliance Agreement approval before I could start my new project. For the first three months of 2021, I had no access to the lab but spent time in online meetings, workshops, training, and in reviewing the literature, as well as upskilling in techniques related to data analyses/interpretation for conducting the systematic review and meta-analysis.

iii) **COVID restrictions**: Due to the state-wide lockdowns in 2021 I was not able to access the laboratory.

Despite being a challenging journey, I am happy to have completed my doctoral study and remain thankful to the Deputy Dean, School of Medicine, Graduate Research School, my supervisors, their laboratory team members, and collaborators, for their support, time, and guidance.

Statement of Authentication

The work presented in this thesis is, to the best of my knowledge and belief, original except as acknowledged in the text. I hereby declare that I have not submitted this material, either in full or in part, for a degree at this or any other institution.

02 April 2023

Statement of Contribution

I hereby declare that I performed the majority of the work presented in this thesis. I performed all of the literature review, cell and molecular experiments, gene selection for senescence panel, animal ethics draft preparation, mouse dietary intervention/monitoring, mouse behavioural tests, data pre-processing, analyses and presentations, tables and figures generation, manuscript draft preparation and addressing any revision. The contribution of others involved in the studies which make up the results chapters is stated below.

The telomere length and senescence in type 2 diabetes study described in Chapter 2 was conceptualized by Prof. Hardikar, along with the provision of mentoring and feedback by Associate Prof. Spring. Dr. Joglekar provided practical advice on cellular and molecular techniques and experimental planning for the study. Dr. Wong provided the ddPCR training. Professor Ma kindly provided the clinical study samples.

For Chapter 3, the dietary oscillation in type 2 diabetes study was conceptualized by Prof. Hardikar. The animal ethics was finalised by Prof. Hardikar. Dr. Joglekar, Dr. Wong, and Prof. Hardikar contributed to mouse blood glucose measurement along with me. Prof. Karl provided expert instruction on the mouse behavioural tests while Dr. Porto provided training for those tests. Associate Prof. Spring and Prof. Hardikar provided mentoring and feedback.

For the systematic review described in Chapter 4, Profs Hardikar and Spring conceptualized the study and provided mentoring, troubleshooting and finalisation of the manuscript. Dr. Wong validated my data extraction and analyses. Dr. Wong and Dr. Joglekar provided feedback for the manuscript, along with Prof. Hardikar and Associate Prof. Spring.

The Early onset colorectal cancer study described in Chapter 5 was conceptualised by Associate Prof. Spring. The database was established by Prof. Chapuis. Assistance for database generation, and variable selection was provided by Ms. Sinclair, Profs Chapuis, Toh, and Spring. Data analysis training was provided by Associate Prof. Hitos. Profs Spring, Hitos, Toh, and Hardikar provided feedback on the study analyses and in preparing the manuscript draft.

All contributors recognize my contributions to be significant and appropriate for inclusion in my thesis.

List of Publications and Presentations Arising from This Thesis

Journal Article:

Alvandi E, Wong WKM, Joglekar MV, Spring KJ, Hardikar AA. Short-chain fatty acid concentrations in the incidence and risk-stratification of colorectal cancer: a systematic review and meta-analysis. *BMC Med* 2022 Oct 3;20(1):323 (doi: 10.1186/s12916-022-02529-4).

Manuscript finalised for submission:

Alvandi E, Hitos K, Toh JWT, Ng KS, Rickard MJFX, Lim SH, Wong WKM, Hardikar AA, Chapuis P, Spring KJ. Increased risk of rectal cancer and aggressive disease in sporadic early-onset colorectal cancer: a single site study of 3609 consecutive cases in NSW over 26 years from 1995 to 2020.

Conference and Seminar Presentations:

- NSW Cancer Conference, poster presentation (Sep 2022)
- South Western Sydney Research Grand Rounds (Aug 2022)
- School of Medicine Seminar, Western Sydney University (Jun 2022)
- Junior Researchers Conference, Western Sydney University (May 2022)
    - Granted the best presenter award
- 3MT (Three Minute Thesis), School of Medicine, Western Sydney University (Jul 2021)

Contents

List of Figures

List of Tables

List of Appendices

List of Abbreviations

ACEC	Animal Care and Ethics Committee
ADA	American Diabetes Association
C57BL/6	C57 black 6
CCK	Cholecystokinin
CDK	Cyclin-Dependent Kinase
CDKi	Cyclin-Dependent Kinase Inhibitor
CMRL	Connaught Medical Research Laboratories
CMS	Consensus Molecular Subtype
CRA	Colorectal Adenoma
CRC	Colorectal Cancer
CSS	Cancer-Specific Survival
Ct	Cycle threshold
CTR	Control
CV	Coefficient of Variation
CVD	Cardiovascular Disease
ddPCR	digital droplet Polymerase Chain Reaction
DMEM	Dulbecco's Modified Eagle Medium
EASD	European Association for the Study of Diabetes
EEC	Enteroendocrine Cell
EOCRC	Early-Onset Colorectal Cancer
FBG	Fasting Blood Glucose
FPG	Fasting Plasma Glucose
FBS	Foetal Bovine Serum
FFAR	Free Fatty Acid Receptor
FIT	Faecal Immunochemical Test
FoD	Fold over Detectable
GBD	Global Burden of Disease
gFOBT	Guaiac Faecal Occult Blood Test
GIP	Gastric Inhibitory Polypeptide

GLOBOCAN	Global Cancer Incidence, Mortality and Prevalence
GLP-1	Glucagon-Like Peptide-1
GPCR	G Protein-Coupled Receptor
GTT	Glucose Tolerance Test
HbA1c	Glycated Haemoglobin A1c
hBG	human Beta Globulin
HC	Healthy Control
HDAC	Histone Deacetylase
HDACi	Histone Deacetylase Inhibition
hEGF	human Epidermal Growth Factor
HFD	High Fat Diet
HG	High Glucose
hIPC	human Islet-derived Progenitor Cell
HKG	Housekeeping Gene
IBD	Inflammatory Bowel Disease
IBS	Irritable Bowel Syndrome
IDF	International Diabetes Federation
JBI	Joanna Briggs Institute
lncRNA	Long Non-coding RNA
LOCRC	Late-Onset Colorectal Cancer
Log_2	Logarithm to the Base 2
mM	millimolar
NDC	Nondigestible Carbohydrate
NG	Normal Glucose
NOS	Newcastle-Ottawa Scale
OR	Odds Ratio
OGTT	Oral Glucose Tolerance Test
OS	Overall Survival
OSC	Oscillating
PenStrep	Penicillin-Streptomycin
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses

PYY	Peptide YY
qPCR	quantitative Polymerase Chain Reaction
RevMan	Review Manager
rLTL	relative Leukocyte Telomere Length
RS	Resistance Starch
rTL	relative Telomere Length
SCFA	Short-Chain Fatty Acid
sEOCRC	sporadic Early-Onset Colorectal Cancer
SMD	Standardized Mean Difference
T2D	Type 2 Diabetes (Mellitus)
TERC	Telomerase RNA Component
TERT	Telomerase Reverse Transcriptase
TERF	Telomeric Repeat-Binding Factor
TNM	Tumour Nodes Metastasis

Synopsis

Globally, Type 2 Diabetes (T2D) and Colorectal Cancer (CRC) are among the most prevalent metabolic diseases and cancers, respectively. T2D is a progressive disease encompassing two underlying progressive conditions, insulin resistance and pancreas functional β-cell loss. CRC mostly arise from the aberrant proliferation of colonic epithelial cells in the form of colorectal polyps. T2D and CRC are chronic diseases and understanding the underlying molecular mechanism to identify biomarkers of their progression could help prevent or delay the course of the two diseases. Additionally, obesity is the major risk factor of T2D, and they are both considered as CRC risk factors. Therefore, there are common factors involved in the progression of T2D and CRC. In this regard, my thesis aimed to investigate T2D and CRC progression in four sections.

Telomere length shortening is one of the hallmarks of cellular senescence. Shorter leukocyte telomere length in T2D has been reported previously. Firstly, I aimed to investigate cellular senescence and assess telomere biology using a cell model of human pancreas islet-derived progenitor cells (hIPCs). Significant shortening of telomere length was identified in successive replication, as well as a significant increase in *TERC* level. *TERC* is the RNA component of the telomerase enzyme that compensates for telomere shortening by adding telomeric repeat sequences to telomeric ends. The next set of experiments on hIPCs in three different glucose conditions (normal, high, and oscillating normal and high levels) showed similar telomere shortening rates among these conditions, but a significant rise of *TERC* expression only in the normal glucose. This could be due to impairment in *TERC* expression in cells exposed to high and oscillating conditions. Also, by analysing human samples, plasma *TERC* level was found

significantly lower in patients with T2D than in healthy individuals, confirming the results obtained using the hIPC cell model.

Additionally, decreased abundance of *TERF1* (telomere-repeat binding factor, and one of the main components of Shelterin complex protecting the telomeres from degradation) transcript was also significant only in hIPCs exposed to oscillating conditions. Besides, analysing cellular senescence biomarkers revealed the increased transcript abundance of *GATA4* (as one of the master regulators of senescence-associated secretory phenotype) and *IL-8* (the main chemokine released by the senescent cells) gene only in the oscillating condition. Therefore, a different molecular profile was identified with respect to telomere biology and senescence markers in the oscillating condition modelling a similar environment of islet cells being exposed to high fluctuations of glucose level in T2D progression.

The second section studied was on how T2D progression was related to gut hormones regulating the insulin secretion from pancreas islets. These hormones are secreted when specialised gut cells sense short-chain fatty acid (SCFA) molecules in the gut lumen. SCFAs are generated by the gut microbiome from the fermentation of dietary fibre. SCFAs have several beneficial effects on both the colonic environment and metabolic health. I studied the underlying molecular alterations during T2D progression with respect to weight cycling and yoyo diet, i.e. oscillation of healthy (SCFA-rich) and unhealthy (SCFA-poor) diets *in vitro*. The gene expression level of gut hormones (GLP-1, GIP, CCK, PYY, and ghrelin) was measured longitudinally in two colonic cell lines (HT-29 and T84) while being exposed to the oscillating high and low levels of butyrate as the main SCFA.

A decreasing amplitude of change in the gene expression of these hormones was identified, suggesting that yoyo dieting may not be a healthy strategy because eventually, the level of these hormones may not increase in proportion to the level of fibre in food.

In addition, the oscillating SCFA was re-evaluated *in vivo* utilising a 22-week dietary intervention using a mouse model - normal chow control (CTR) representing healthy food (SCFA-rich), high-fat diet (HFD) representing unhealthy food (SCFA-poor) and oscillating normal and high-fat diet. Body weight, food intake, and fasting blood glucose (FBG) were measured during the intervention program. Surprisingly, insulin resistance appeared mid-way through the dietary intervention in the oscillating group earlier than the HFD group. Also, at the end of the intervention, the glucose tolerance test profile of the oscillating group was identical to the HFD group. Furthermore, behavioural tests were performed at three different time points (start, mid-way, and endpoint) providing some initial evidence of diabetic neuropathy in the form of weak neuromuscular function in the oscillating and HFD groups. The olfactory test showed impaired olfactory sensing in the oscillating group as well.

Therefore, the yoyo diet and subsequent weight cycling may not be a good approach to tackle obesity as it may result in eventually low secretion of the beneficial gut hormones regulating the insulin level, as well as insulin resistance progression as the two underlying T2D mechanisms.

Thirdly, SCFA was studied in relation to CRC progression. By conducting a systematic review and meta-analysis, a link between faecal SCFA level and the risk and incidence of CRC was established. The combined faecal concentration of the three major SCFA molecules (acetate, propionate, and butyrate) was found to be significantly lower not only in CRC patients compared to healthy controls, but also in high-risk CRC individuals than those at low risk.

Therefore, gut SCFA level was inversely associated with CRC and incidence, and faecal SCFA concentration could be considered as a potential biomarker to identify the CRC progression. In this regard, measuring faecal SCFA levels is suggested as a non-invasive and complementary method to the current CRC screening procedures.

Overall, SCFA molecules are beneficial in creating a healthy gut environment against CRC progression and regulating insulin secretion against T2D progression. This could be another aspect of the relationship between T2D and CRC, and the importance of a healthy diet to prevent these two medical conditions.

Finally, in the fourth section, the concerning rising incidence of early-onset CRC (EOCRC: CRC diagnosed earlier than 50 years of age) led us to investigate this further by analysing a consecutive series of 3609 patients that underwent colorectal surgery over a 26-year period at Concord Hospital (NSW, Australia). The focus was on the incidence trend, clinicopathological features and survival in this age group. No rising trend was identified; however, rectal cancer was significantly more prevalent among patients with EOCRC. Besides, lymph node and distant metastasis, as well as poorly differentiated tumours were more prevalent in this age group compared to their older counterparts. Patients with EOCRC had a better five-year survival rate than older patients, but with only a marginal cancer-specific survival time of one year on average.

The current National Bowel Cancer Screening Program commences at the age of 50 years. Our findings including a significantly higher percentage of metastasis in EOCRC, as well as significantly higher risk of developing an advanced tumour in this age group provides supporting evidence for lowering the commencement age for CRC screening.

In conclusion, this thesis has aimed to help improve our understanding of the relationship between diet, obesity, T2D and CRC by identifying the potential risk factors and biomarkers involved in T2D and CRC progression.

The conceptual link between the four data chapters of my thesis is shown in **Figure 1**. Chapter 2 (yellow) discusses the pancreas islet cell senescence data. The effect of diet and SCFA molecules on obesity and T2D progression is discussed in Chapter 3 (blue). A link between SCFAs and CRC risk is assessed in Chapter 4 (pink). Finally, EOCRC data that I analysed to identify the features that are associated with the younger CRC patients (orange), is on the right side. It is important to note that whilst the link between SCFAs, obesity, and T2D have been demonstrated in Diabetes and Islet Biology Group (Nhan et. al. 2023 *Nutrition Reviews*) and by others, more common links between CRC, obesity and T2D remain to be explored and established.

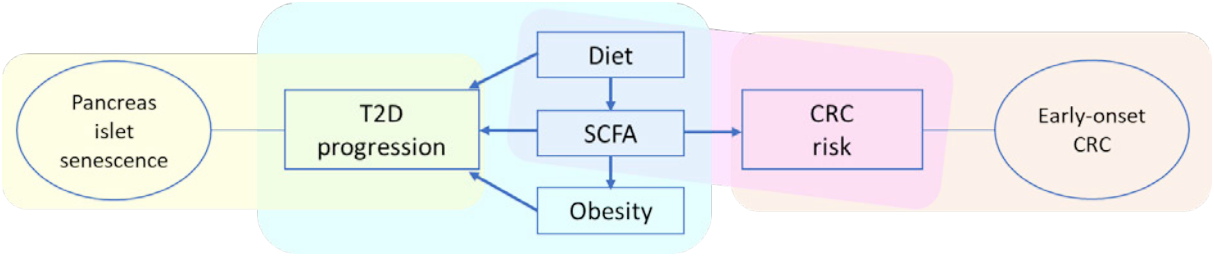


Figure 1. Schematic of the conceptual link between T2D and CRC
Chapter 2, 3, 4, and 5 are shown by yellow, blue, pink, and orange, respectively.

Chapter 1 - Introduction

Based on the 2019 report for Global Burden of Disease (GBD), diabetes and cancer are among the most prevalent health challenges[1]. The International Diabetes Federation (IDF) (10th edition) has estimated that the number of individuals with diabetes will increase to 783 million by 2045[2]. Type 2 diabetes (T2D) constitutes ~90% of these cases[3,4] and T2D numbers are reaching pandemic-like levels[4,5]. T2D rising incidence is accompanied by an increased risk of cardiovascular diseases and cancers, which in total imposes a great burden on the society and health system[5,6].

Colorectal cancer (CRC) is globally third leading cause of cancer death globally, encompassing 10% of all diagnosed cancers[7]. It has been estimated that overall, the risk of CRC in all age groups will increase worldwide by 60% by 2030 leading to more than 1.1 million deaths and 2.2 million new cases[8].

In addition, obesity, which is a major risk factor for T2D[9], is positively associated with CRC risk[10-13]. T2D is associated with the increased risk of several cancers[14,15], and several meta-analyses reported the association between T2D and CRC, as well as the higher rate of CRC in individuals with T2D, compared to healthy individuals[16-22]. All these studies have revealed the link between obesity, T2D, and CRC[10-22]. Since T2D and obesity are reaching pandemic levels[6,9], this increasing number could further increase CRC incidence, and therefore identification of associated or mechanistically-related obesity and T2D risk factors and biomarkers could also help predict T2D progression and improve the prevention and management of both T2D and CRC.

1.1 Type 2 Diabetes

Type 2 Diabetes is an escalating global health concern characterized by hyperglycaemia[5,23]. The two known main factors involved in T2D pathogenesis are insulin resistance - in the liver, muscle, and adipose tissue - and impaired insulin secretion from pancreatic islet β-cells, i.e., functional β-cell loss[5,24]. The prevalence of T2D has surged to pandemic levels[23,25], prompting substantial attention from healthcare professionals, policymakers, and researchers around the world[23,26]. Understanding the factors contributing to this increased prevalence and the widespread implications of T2D is crucial for effective diagnosis, management, and prevention efforts[23,26]. This introductory section discusses the global T2D trends, risk factors, pathophysiology, complications, management, challenges, and the need to identify readily available biomarkers of T2D progression. It then explores two novel aspects of T2D research in detail: 1) the telomere length and telomerase RNA component as potential T2D biomarkers; 2) the impact of oscillating healthy and unhealthy diet on glycaemic and incretin factors *in vitro* and *in vivo*.

1.1.1 Global Prevalence and Trends

According to the International Diabetes Federation (IDF), in 2021, approximately 537 million adults (20-79 years) were living with diabetes[2]. This number is projected to escalate even further to 783 million by 2045 if current trends persist[2]. Approximately 75% of individuals with diabetes are living in low- and middle-income countries[2] and around 6.7 million people die due to diabetes annually[2].

Type 2 diabetes (T2D) constitutes ~90% of these cases[3,4] and T2D numbers are reaching pandemic-like levels[4,5]. T2D rising incidence is accompanied by an increased risk of cardiovascular diseases and cancers, which in total imposes a great burden on the society and

health system[5,6]. It is important to note that one in three adults with diabetes are undiagnosed[2], and thus they present with severe diabetes and complications at diagnosis[23]. Therefore, one of the major challenges in T2D management is diagnosing the disease at the early stages of progression[23].

1.1.1.1 Contributing Risk Factors to T2D Epidemic

Several interconnected factors have contributed to the widespread prevalence of T2D[23,24,27] as listed below.

1. **Unhealthy Diets:** The modern shift toward diets rich in refined sugars, unhealthy fats, and low in fibre has been closely associated with obesity, which in turn is the primary risk factor for T2D[23,28]. Excessive caloric intake, especially from sugary beverages and processed foods, contributes to weight gain and insulin resistance[29,30].
2. **Obesity:** The obesity epidemic has significantly fuelled the rise of T2D[9]. Excess body fat, particularly visceral fat, creates a chronic inflammatory state in the body and contributes to metabolic abnormalities such as insulin resistance that underlies T2D[5,24]. Adipose tissue is not just a fat storage site; it also plays a crucial role in regulating metabolism[5,24]. In obesity, especially abdominal obesity, adipose tissue can become dysfunctional, leading to the release of various adipokines that contribute to insulin resistance and inflammation[5,31]. Additionally, immune cells in adipose tissue can contribute to inflammation and metabolic dysfunction[24,31]. Moreover, increased levels of free fatty acids in the bloodstream, can contribute to insulin resistance[5,31].
3. **Genetic and Ethnic Factors:** Certain ethnic groups, such as South Asians and African Americans, have a higher risk of developing T2D due to genetic factors that interact with

environmental influences[32,33]. The genetic factors influence insulin sensitivity, beta-cell function, and glucose metabolism[32,33].

4. **Physical Inactivity:** Sedentary lifestyles have become increasingly common due to technological advancements, urbanization, and changes in occupation[34,35]. Reduced physical activity leads to weight gain and negatively impacts insulin sensitivity, heightening the risk of T2D[34,35].
5. **Urbanization and Lifestyle Transitions:** All the above factors are included in the rapid urbanization which often leads to shifts in dietary patterns and decreased physical activity[23,26]. Traditional diets rich in whole foods are replaced by energy-dense, nutrient-poor diets, contributing to obesity and T2D[28,30].
6. **Aging Population:** The global increase in life expectancy has led to an aging population, and age is a major risk factor for T2D[27,36]. Aging is associated with decreased insulin sensitivity and impaired pancreatic function[23,36].

1.1.1.2 Health and Economic Impact of T2D

The health consequences of T2D are multifaceted and severe[23,26]. Poorly managed T2D can lead to a range of complications, including cardiovascular diseases, kidney disease, neuropathy, retinopathy and non-healing ulcers that may result in amputations[23,26]. In addition, T2D significantly diminishes individuals' quality of life, affecting daily activities, mental well-being and overall life satisfaction[23,26]. The burden of constant blood sugar monitoring, medication and potential complications can be overwhelming[23,26]. Moreover, the cost of medical care, medications, hospitalizations and management of complications put a huge burden on family and health system[23,26]. These expenses could potentially divert resources from other essential areas considering the fact that T2D is preventable[23,26].

The rapid escalation of T2D's prevalence draws parallels to a pandemic[25]. It affects diverse geographical, cultural, and socioeconomic populations[25]. However, it is important to note that T2D is largely preventable through lifestyle modifications[23,37]. Therefore, promoting awareness about healthy dietary habits, regular physical activity, weight management, and routine health screenings is paramount for prevention[23,26]. Nonetheless, low- and middle-income countries are particularly vulnerable due to limited access to healthcare, education, and resources for diabetes prevention and management[23,26].

1.1.2 Pathophysiology of T2D

The underlying mechanisms of T2D involve insulin resistance and an initial increase in insulin production by pancreatic β cells, followed by a gradual reduction in their ability to generate insulin[5,24]. When blood glucose level increases due to insulin resistance, the compensatory rise in insulin production by β-cells eventually leads to their exhaustion, premature senescence, apoptosis, and death, ultimately resulting in lowered insulin secretion[24,38,39]. The intricate combination of β-cell dysfunction and insulin resistance is responsible for the complexity of T2D[23,24]. In this regard, a model centred around β-cells has been suggested, highlighting abnormal β-cell function as the primary anomaly in T2D[24,40].

It has been proposed that there are other underlying causes that together with insulin resistance collectively result in T2D progression, including dysregulation of incretin hormones, increased breakdown of fats, elevated glucagon levels, enhanced glucose reabsorption in the kidneys, and disrupted appetite regulation all contribute to the pathophysiology of T2D[23]. This perspective was later expanded to include malfunctions in interconnected pathways, with compromised β-cell function acting as the central factor linking these pathways[23]. These pathways are shown in **Figure 1.1**. Therefore, pancreas (β and

α cells), liver, skeletal muscle, kidneys, brain, small intestine, and adipose tissue are the organs involved in T2D progression[23,24]. The incretin effect, gut microbiota dysbiosis, and inflammation are among the most important pathophysiological factors[24,40].

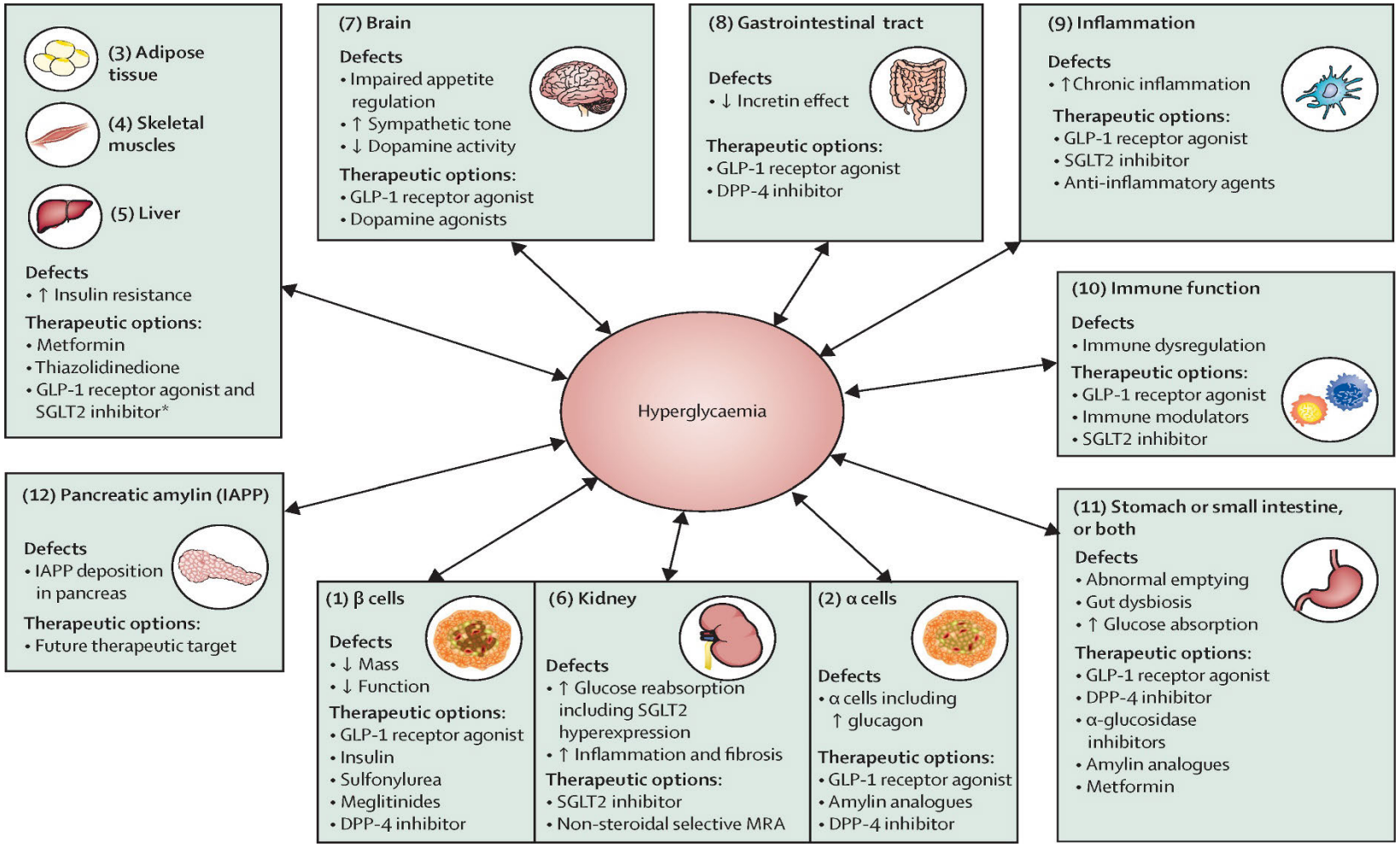


Figure 1.1 The deleterious dozen in T2D progression

The β-cell functional loss in T2D is attributed to 12 pathophysiological abnormalities. The potential therapeutic approach for each pathway is also stated. ↑=increase, ↓=decrease, DPP-4=dipeptidyl peptidase 4, GLP-1=glucagon-like peptide-1, IAPP=islet amyloid polypeptide, MRA=mineralocorticoid receptor antagonist, SGLT2=sodium-glucose cotransporter 2, *not the primary mechanism of action. From: Ahmad E et al. *Lancet* 2022.

1.1.3 Downstream complications of T2D

While the primary features of T2D is the disturbances in glucose metabolism, it can have significant downstream complications that affect various organ systems in the body[5,23,27]. These complications arise due to prolonged periods of elevated blood glucose levels and associated metabolic abnormalities[23,27,41]. These complications can be broadly categorized into two main groups of macrovascular and microvascular complications, affecting larger and small blood vessels, respectively[42,43].

1.1.3.1 Macrovascular complications

Cardiovascular diseases (CVD): Prolonged hyperglycaemia leads to endothelial dysfunction, inflammation, and oxidative stress, promoting the development of atherosclerosis in major arteries[27,43]. Atherosclerotic plaques can narrow and obstruct blood flow, leading to angina or myocardial infarction[43]. Additionally, these plaques can rupture, triggering clot formation and further impeding blood flow, potentially causing strokes[27,43].

Patients with T2D face an increased risk of other CVDs such as cardiomyopathy that impacts the heart's structural integrity, arrhythmias leading to irregular heartbeats and potential sudden death, as well as cerebrovascular disease affecting blood vessels in the brain, and peripheral artery disease that impedes blood flow to limbs[23,43]. Notably, cardiovascular disease stands out as the primary cause of mortality among patients with T2D[23,43].

1.1.3.2 Microvascular complications

Microvascular complications primarily involve damage to small blood vessels, affecting various parts of the body[27,42]. Around a quarter of individuals diagnosed with T2D experience the impact of diabetic nephropathy and/or retinopathy[42].

Diabetic Neuropathy: This group of medical conditions can be categorized into peripheral neuropathy and autonomic neuropathy[44]. Peripheral neuropathy leads to numbness, tingling, and pain in the extremities[45]. Autonomic neuropathy affects the autonomic nervous system, causing gastrointestinal, cardiovascular, and urinary dysfunction[46]. Neuropathy and poor circulation in the lower limbs lead to reduced sensation and delayed wound healing[47]. These issues, combined with the formation of atherosclerotic plaques in the peripheral arteries, increase the risk of non-healing ulcers and gangrene[45,47]. In severe cases, the reduced blood supply necessitates amputation[45,47].

Diabetic Nephropathy: Elevated blood glucose levels lead to the accumulation of advanced glycation end-products, which provoke inflammation and oxidative stress in the kidneys[48,49]. Inflammatory pathways and oxidative stress contribute to kidney damage, eventually progressing to end-stage renal disease[48,49]. Additionally, if hypertension coexists in patients with T2D, its impact on the renal vasculature further exacerbates kidney damage[48,49]. The glomerular filtration barrier is damaged due to prolonged hyperglycaemia, leading to increased permeability, which in turn allows proteins, particularly albumin, to leak into the urine and proteinuria[48,49]. Diabetic nephropathy is the primary factor for the development of end-stage renal disease globally[48,49].

Diabetic Retinopathy (DR): This medical condition stands as a primary contributor to vision impairment among patients with T2D in their middle-aged and elderly years[50,51]. Approximately one-third of people afflicted by diabetes are affected by DR[50,51]. Advanced stages of DR are characterized by the abnormal growth of novel blood vessels within the retina, and diabetic macular oedema, denoting the accumulation of fluid and swelling within the retinal central region[50,51]. The escalated number of free radicals, advanced glycosylation

end products, and inflammatory agents lead to DR which is also found to involve neurodegeneration within the retina[50,51].

Skin and Wound Complications: Skin complications in T2D involve microvascular changes that impair wound healing and increase the risk of infections[42]. Poor circulation due to microvascular damage reduces the ability of tissues to repair themselves, leading to slow-healing ulcers and an increased susceptibility to infections[41,42].

Cognitive Dysfunction: Microvascular changes in the brain can lead to cognitive dysfunction in T2D which is now considered as a risk factor of Alzheimer disease and vascular dementia[52,53]. Chronic hyperglycaemia contributes to inflammation, oxidative stress, and reduced blood flow to the brain, potentially increasing the risk of cognitive decline and neurodegenerative disorders[52,53].

It is important to note that apart from the well-known macro- and microvascular complications of T2D mentioned above, there also emerging number of evidence which associates T2D with cancer, infections, liver disease and affective disorders[41]. This will inevitably add additional burden for the management of T2D[41].

1.1.4 Clinical management of T2D

Functional β-cell loss is the key component of T2D pathophysiology[23,24]. Even though a significant loss of β-cell function has already occurred by the time of T2D diagnosis[24,40], effective control of blood sugar or the possibility of its temporary reduction can lead to a substantial recovery of functional β-cell mass[24,40]. This recovery might be attributed to the reactivation or regeneration of existing β-cells, as well as the transformation and renewal of other cell types within the pancreas[23,24].

According to the consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD)[26], T2D management involved the management of blood glucose levels, body weight, cardiovascular risk factors, and T2D complications[26]. Based on the consensus, T2D management is recommended to follow a holistic approach encompassing those four segments[26] as depicted in **Figure 1.2**. It is important to note that the first-line management is lifestyle management mainly involve the body weight control[37] to improve the glycaemic condition and delay the progression from pre-diabetes before functional β-cell loss reaches significant levels[23,26]. This includes a balanced diet as the primary component of the lifestyle change as well as routine exercise[23,26].

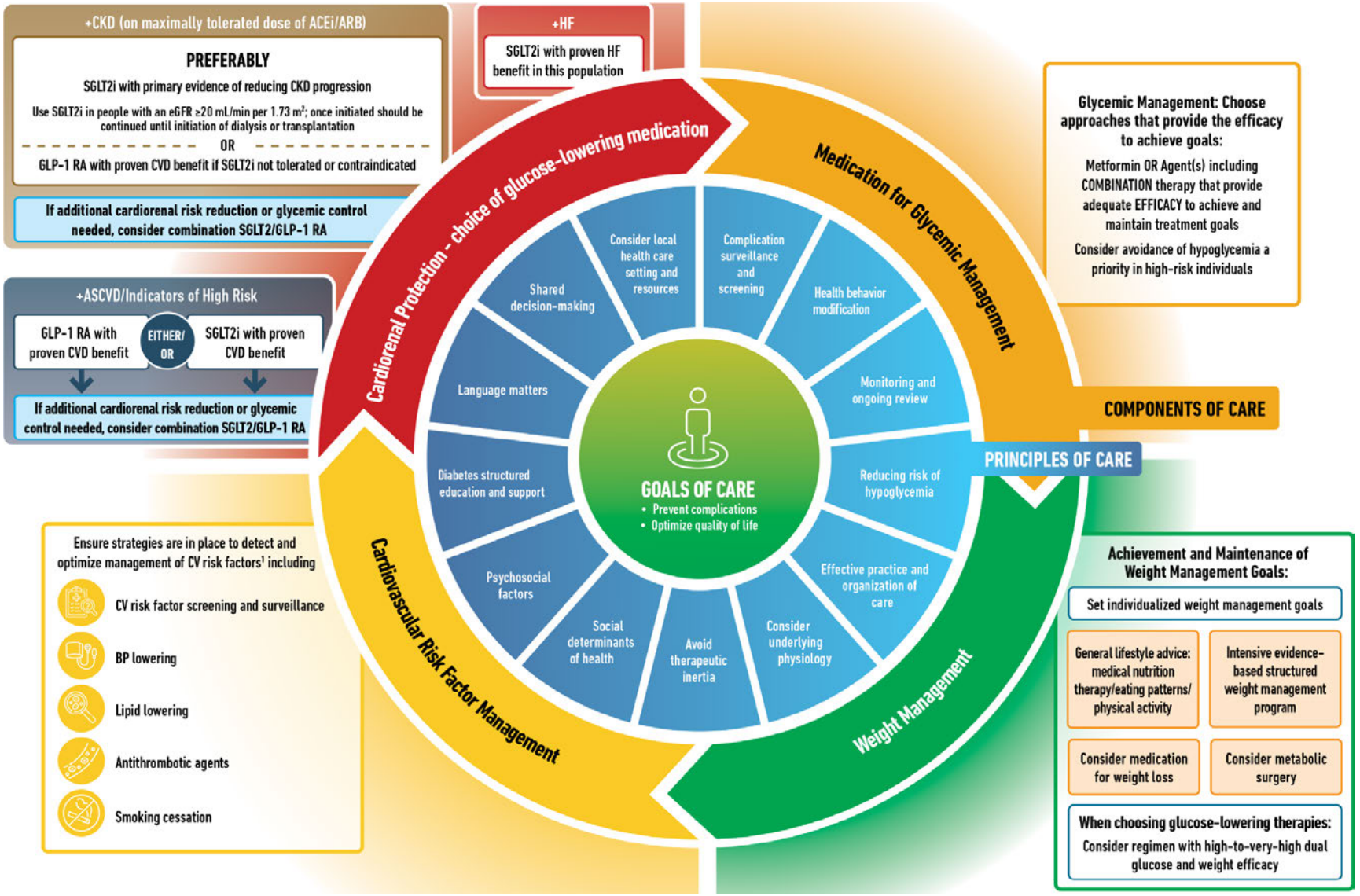


Figure 1.2 The management of T2D based on ADA and EASD 2022

The multifaceted management including multiple factors and pharmacological treatments with the aim of preventing or delaying the downstream complications and maintaining the quality of life. From: Davies MJ et al. *Diabetes Care* 2022.

1.1.5 Challenges in diagnosis and management of T2D

The major challenge in T2D diagnosis stems from the chronic and silent progression of the disease[5,23]. The progression is largely asymptomatic for an extended period[5,29], and many individuals may not experience noticeable symptoms until the disease has progressed significantly[23,29]. This delays diagnosis and can lead to more severe complications[24,29] and is worse in individuals at high risk such as those with genetic susceptibilities and high BMI[9,29]. Therefore, identifying individuals with prediabetes is crucial for preventing the progression to full-blown diabetes[23,29]. These apparently healthy individuals have higher than normal blood sugar levels but not high enough to be classified as diabetes and thus remain undiagnosed[23,29]. All these factors can result in delayed diagnosis and missed opportunities for early intervention since many individuals may not seek medical attention until the downstream complications of T2D arise[5,24].

Apart from diagnosis challenges, T2D management is faced with several obstacles[5,37]. Lifestyle modifications are introduced as the best strategy to tackle T2D[26,37], however, sustaining these changes in the long term can be difficult for many individuals[23,26]. Many patients require medications to control blood sugar levels[23,26]. Adherence to medication regimens can be challenging due to factors such as complexity of dosing, side effects, and cost[23,26]. The ADA and EASD consensus recommended the personalised approach as the optimal T2D management strategy[26] (**Figure 1.2**). Nonetheless, finding the right combination of medications and interventions for not only T2D but also its downstream complications per each patient can be a complex process[23,26]. In addition, regular monitoring of blood sugar levels, HbA1c (glycated haemoglobin), and other relevant markers is essential for adjusting treatment plans[23,54]. However, maintaining consistent monitoring can be logistically and

economically challenging for some individuals[23,54]. Other aspects can contribute to the complexity of T2D management such as behavioural factors, access to healthcare, patient education, and continuous support[26,54].

All these challenges in the diagnosis and management of T2D warrant the identification of new biomarkers to detect the early progression of T2D prior to the development of full-blown disease[23,54].

1.1.6 Biomarkers of T2D progression

A biomarker is a naturally occurring, measurable, biological trait or molecule representing normal or pathophysiological indicators of a biological process, pathogenic response, or changes to therapeutic intervention[55]. Based on "Standards of Medical Care in Diabetes" published by ADA in 2020, fasting plasma glucose (FPG), two-hour plasma glucose during a 75g oral glucose tolerance test (OGTT), and glycated haemoglobin A1c (HbA1c) concentration are the current established biomarkers for T2D diagnosis (**Table 1.1**)[56].

Table 1.1 The ADA reference values for diagnosing prediabetes and T2D

	FPG (mg/dL)	HbA1c %	OGTT, 2h (mg/dL)
Prediabetes	100–125	5.7–6.4	140–199
Diabetes	≥ 126	≥ 6.5	≥ 200

However, none of these markers are ideal. For example, FPG demands a minimum fasting period of 8 hours, displays significant biological and daily fluctuations, and the samples used encounter stability challenges[54]; OGTT can be costly and has low reproducibility in some cases with limitations in logistics and analysis due to its complicated design[54]; HbA1c is not capable of detecting temporary increases in blood sugar levels, and is affected by age, race, ethnicity,

and medical conditions which influence the lifespan of red blood cells or haemoglobin levels independently of glucose levels[54].

Advocating a universal diabetes screening is not advised due to limited evidence supporting its cost-effectiveness or positive impact on health results[23]. It is also recommended that screening to be limited to only those individuals with elevated risk factors such as overweight or obesity, and a significant family history of T2D or prior gestational diabetes[36,37]. To date, no biomarker has proven to be the ideal marker for all individuals with type 2 diabetes under all circumstances[54]. Nonetheless, novel plasma biomarkers such as fructosamine, glycated albumin, and 1,5-anhydroglucitol are suggested to enhance the predictive efficacy of the standard markers[54].

The quest for identifying novel T2D biomarkers started with genome-wide association studies (GWAS) to identify the susceptible polymorphisms[32,33]. Many of the identified variants were also involved in obesity and T2D comorbidities such as macro- and microvascular diseases, which shows common molecular changes underlying all these conditions[32,33]. Apart from the gene sequences, there are other mechanisms involved in gene expression that could be altered during T2D progression[57,58]. This include the recently studied epigenetic changes in different tissues, which is affected by the environmental factors[58]. This includes the DNA methylation and histone modifications in pancreatic islets, adipose, liver, and skeletal muscles[32,33]. Interestingly, the epigenetic changes in islet cells are for those genes involved in mature β cell function, insulin secretion, glucagon receptors, and glucose transporters[32,33]. In muscle cells, the same genes involved in oxidative phosphorylation which are normally downregulated in aging are also downregulated in T2D[32,33]. Additionally, distinctive changes in gene methylation were observed in the genes responsible for encoding inflammatory

proteins in both visceral and subcutaneous adipose tissue in T2D[32,33]. All these data reveal the complexity and heterogeneity of T2D progression and why it is considered a chronic disease that develops asymptomatically[5,23].

Metabolomics is another approach with great potential to identify deregulated metabolites of diverse metabolic pathways since T2D is fundamentally a chronic metabolic disorder[59,60]. Recent advancements in high-throughput technologies have enabled metabolomics and proteomics to detect potential biomarkers and pathways involved in T2D progression[59,60]. Whether as cause or consequence, it was found that the circulatory levels of branch chain amino acids are higher in individuals with obesity or insulin resistance than healthy individuals[59,60]; the same increased level was also observed in deoxycholic acid level – a bile acid[59,60]. Increased levels of several fatty acids such as diacylglycerols, triacylglycerols, and ceramides is also associated with insulin resistance[59,60].

By integrating metabolomics and proteomics with genomics in multi-omics approaches, a comprehensive analytical framework could emerge to start unravelling causal relationships[23,27], as well as early detection of T2D considering the fact that one in three adults with diabetes are undiagnosed[2].

1.1.7 Telomere length in T2D

Telomeres are known as the protecting factor of the tips of chromosomes ends[61]. Telomeres also play a role in regulating cellular senescence[62] - the irreversible cell cycle arrest[62], and telomere shortening is one of the hallmarks of the cellular senescence[63,64]. With each cell division, the telomeric ends gradually shorten[61], and upon reaching a critical length known as the Hayflick limit, the cell enters a state of senescence[61]. Telomeres also serve as indicators of DNA damage[65], a commonly used biomarker for measuring the biological aging[66].

Telomeres exhibit susceptibility to DNA damage due to the presence of 3' single-stranded telomeric overhangs[61], rendering them vulnerable to single-stranded breaks, especially when confronted with oxidative stress conditions[67]. Numerous investigations have highlighted that diminished telomere length serves as a reliable indicator of oxidative stress[68], a factor that significantly expedites the process of telomere attrition[67]. Within the human context, there exists a notable connection between the length of telomeres and biomarkers indicative of oxidative stress[67], a relationship that frequently exhibits heightened levels in individuals affected by diabetes[68].

In the context of high glucose levels observed *in vitro*, there's a notable tendency for cells to undergo irreversible progression to senescence[62]. However, this adverse effect can be mitigated by the overexpression of telomerase, an enzyme that maintains telomere length[61]. These findings shed light on the intricate mechanisms linking elevated blood sugar levels to the biological process of senescence, particularly in the context of diabetes[68]. Moreover, several recent studies have shown the association of their accelerated shortening with conditions such as cardiovascular disease, diabetes, and cancer[69-71]. Therefore, the attrition rate of telomeres length could be a potential biomarker of T2D progression and its downstream complications[68].

Structurally, telomeres are multiple tandem repeats of the TTAGGG sequence at the end of each chromosome arm[72,73] (**Figure 1.3**). Human telomeres consist of intricate complexes formed by the interaction between telomeric DNA, and a protein complex named shelterin[61,73] (**Figure 1.3**). Among the shelterin constituents, TRF1 and TRF2 stand out as they confer the capability to bind and recognize double-stranded telomeric DNA[61,73]. Other shelterin proteins such as POT1 and TPP1 envelops the single-stranded telomeric overhang –

the chromosomal termini[61,73]. This protective layer created by shelterin serves to cap the ends of chromosomes, serves as a shield against undesired processes such as degradation and fusion events that can compromise the integrity of telomeric DNA[61]. Other proteins such as TIN2 and RAP1 facilitate the integrity of shelterin complex and its binding to DNA[61] (**Figure 1.3**)

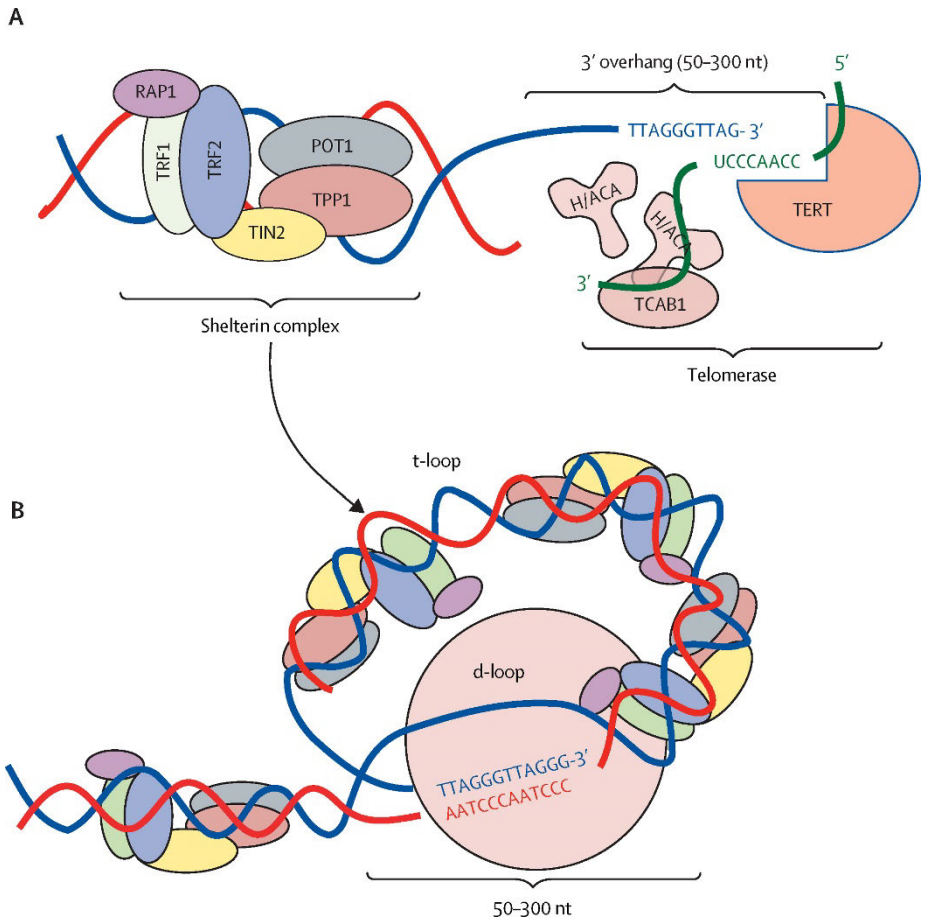


Figure 1.3 Telomere, shelterin complex, and telomerase enzyme

A) The six proteins of the shelterin complex (TRF1, TRF2, POT1, TPP1, RAP1, and TIN2) protect the telomere (left). Telomerase, shown on the right, contains the template RNA (*TERC*, green), and the reverse transcriptase protein (TERT). B) Several shelterin complexes form the T-loop to protect the telomere, shown in double-strand DNA with blue and red. The TTAGGG telomeric repeat is also shown. From: Chen F et al., *Lancet Diabetes & endocrinology* 2021.

Due to the nature of DNA replication, telomere shortening occurs with each cell cycle and is maintained by the telomerase enzyme[61,73] (**Figure 1.3**). However, this elongation mechanism

is reduced or even absent in many differentiated cells[62,66], leading to telomere shortening, and eventually the irreversible state of cellular senescence[72,74].

In human studies, the leukocyte telomere length (LTL) is measured as a proxy of biological ageing in age-related diseases, such as metabolic and cardiovascular diseases[63,68]. In addition, meta-analyses have shown the association of shorter LTL to diabetes - irrespective of type[71], T2D exclusively[75], as well as cardiometabolic outcomes[76]. However, the studies on telomerase in T2D is limited (**Table 1.2**).

Table 1.2 The summary of studies of telomerase gene or enzyme related to diabetes

Study	Type	Results
Blazer et al. *Biochem Biophys Res Commun* 2002	*in vitro*	Telomerase overexpression circumvented the effects of hyperglycaemia on replicative capacity in hyperglycaemia-induced replicative senescence in human skin fibroblasts[77]
Matthews et al. *Circ Res* 2006	*in vitro*	Telomerase expression alone rescued plaque vascular smooth muscle cells (VSMCs) senescence despite short telomeres[78]
Kuhlow et al. *Aging* 2010	*in vivo*	Young adult mice which are deficient for the *TERC* subunit of telomerase exhibit impaired glucose tolerance whilst normal insulin sensitivity[79]
Gutmajster et al. *J Appl Genet* 2018	Human study	Carriers of CC genotype in *TERT* rs2853669 polymorphism had the shortest rTL in the T2D group[80]
Wang et al. *Front Pharmacol* 2021	*in vitro*	*TERT* gene expression decreased after 48hrs exposure of human umbilical vein endothelial cells (HUVECs) to 40mM glucose[81]
Opstad et al. *Biomedicines* 2022	Human study	*TERT* rs7705526 polymorphism increased the risk two-fold for acute myocardial infarction in male patients, and stroke in female patients[82]

1.1.7.1 Telomere length in β-cells

β-cell proliferation is observed in the early neonatal life of humans, and is nearly absent in adults, with the pool of β-cells established before the age of thirty[83,84]. Although evidence suggests that new β-cells may form by self-replication or neogenesis even under specific conditions (e.g. in pregnancy[85]), it remains likely that most β-cells present in adult human individuals essentially last for a lifetime[86,87]. Nonetheless, the accelerated β-cell exhaustion and loss during progression to T2D cannot be compensated by their proliferation rate[88,89].

So far, there are only a few studies on β-cells telomere length[90,91]. Tamura et al. showed that β-cell TL is shorter in T2D patients compared to healthy individuals[91]. Later, they reported the age-dependent β-cell telomere attrition by plotting TL against age (0-100 years) in otherwise healthy individuals, despite not observing a significant decrease between the age of 18 and 100 years[90].

1.1.7.2 Knowledge gap: establishing a biomarker for telomerase level in addition to LTL measurements to monitor T2D progression

Telomere length is maintained by the activity of telomerase (**Figure 1.3**), a reverse transcriptase enzyme that adds the telomere repeat sequence to the end of telomeres[61,73]. This ribonucleoprotein enzyme is made up of an RNA component (Telomerase RNA Component or *TERC*), which serves as a template to synthesize the telomere repeats[61,73] (**Figure 1.3**). Telomerase activity is absent in non-dividing cells, and these cells enter the senescence phase when their telomeres reach a certain threshold to stop further proliferation[61,92].

It is worth noting that despite all the previous investigations on the association of shorter TL with T2D and its complications, assessing cellular senescence solely through TL measurement is not comprehensive by itself. Other aspects of telomere biology may improve the overall picture of disturbed telomere biology during T2D progression. This could include measuring telomerase level or activity which by itself is challenging, particularly when dealing with archived or biobanked plasma samples. Thus, there's a strong need for a dependable substitute that can reflect telomerase level or activity in plasma. This is crucial for discerning whether telomere shortening correlates with telomerase level or activity.

The potential to detect variations in telomerase level or activity could be achieved by assessing its long non-coding RNA component, *TERC*. This component could hold promise as a surrogate marker for telomerase activity. The primary objective was to delve into the biology of telomeres during T2D progression. Therefore, to comprehensively examine all three facets of telomere biology - telomere length, telomerase activity, and *TERC* levels, human Islet-derived Precursor Cells (hIPC)[93] is selected as an *in vitro* model for pancreatic islet cells. The novelty of this investigation was to include the measurement of *TERC* levels for the first time as a component of telomere biology when exposing hIPCs to high and normal glucose level *in vitro*, while considering an additional condition of oscillating high and normal glucose to mimic the daily rapid fluctuation in T2D patients.

This novel work also compared plasma TERC level between T2D patients and healthy individuals to estimate its impact as an additional factor of the telomere biology. Therefore, in addition to telomere length, evaluating the telomerase activity via TERC level measurement could provide a better evaluation to monitor T2D progression and CVD complications, as a predictive biomarker before the clinical diagnosis, as discussed in detail in Chapter 2.

1.1.8 Diet, short-chain fatty acids, and incretin hormones in T2D

Among the main risk factors of T2D, obesity is the leading one, and a high-fat diet is a primary contributor to obesity[3,94]. The nutrient content of digested food is absorbed by the enterocytes, whilst stimulating the enteroendocrine cells (EECs)[95,96]. These specialized cells, located in the inner surface of small and large intestines collectively form the biggest endocrine organ in the body[95,96]. As their name suggests they have endocrine functions, and they are traditionally divided into several groups of cells based on the type of hormones they produce[95,96]. EECs have specialized nutrient-sensing receptors that once stimulated, activate

downstream signalling pathways resulting in hormonal secretion[97]. These hormones include GLP-1 (Glucagon-like Peptide-1), GIP (Gastric Inhibitory Polypeptide), CCK (Cholecystokinin), PYY (Peptide YY), and Ghrelin[96]. GLP-1 and GIP are produced from L cells - in the ileum and colon - and K cells - in the duodenum, respectively[98,99]. These hormones are called incretin since they postprandially stimulate the pancreatic islets β-cells to promote insulin secretion to maintain glucose homeostasis[98,99]. They also play important roles in controlling the appetite, gastric emptying, and gut motility[98,99]. Apart from sensing carbohydrates, binding of short-chain fatty acid (SCFA) molecules to specific G protein-coupled receptors (GPCRs) on the luminal surface of L and K cells activate downstream signals which result in incretin secretion from these cells[98-101].

Short-chain fatty acids are fatty acids with two to six carbon atoms[102,103]. Acetic, propionic, and butyric acid constitute the majority of intestinal SCFA with a molar ratio of 60:20:20 and have 2-, 3-, and 4-carbon atoms respectively[102,103]. SCFAs act as signalling molecules to stimulate GPCRs, mainly GPR41 (FFAR3) and GPR43 (FFAR2), on the surface of EECs to activate signalling pathways that control the gene expression levels, mainly via epigenetic mechanisms such as histone deacetylase inhibition (HDACi) to promote the gene expression of incretins[104-106].

Further, SCFAs are among the main metabolites produced by the microorganisms residing in the digestive tract, which are collectively termed gut microbiota[107,108]. Although the proportion of microbiota has been debated (based on nucleated vs all human cells and bacteria vs all microbes), a recent estimation suggested a ratio of 1:1 for the number of human cells to bacteri[109]. Irrespective of the ratio, there is no doubt that bacteria, through their metabolites, have a significant impact on human physiology[108,110]. They collectively act

as an organ, mediating the environment and our body by producing metabolites that are in constant crosstalk with the gut epithelium[107,108,111]. The Gut microbiota constitutes the largest community of commensal microorganisms in the body[107,108]. Collectively, these microorganisms mainly reside in the lower small intestine and colon and can metabolize nutrients that cannot be processed by the digestive enzymes of the gastrointestinal (GI) tract[107,108]. Fibre, technically termed nondigestible carbohydrates (NDCs), is the main identified dietary component in this category and SCFAs are produced through the fermentation of NDCs[102-107,111-114]. Therefore, the amount of dietary fibre intake could be a determinant of glucose homeostasis through gut microbiota-SCFAs-incretin axis[96,98,99,108].

1.1.8.1 Knowledge gap: understanding the effect of dietary oscillation on T2D progression

Dietary oscillation, yo-yo dieting, or weight cycling is a consequence of attempts by many people to control their weight[115]. They turn to healthier foods when dissatisfied with their weight and revert to an unhealthy diet once they attain their desired weight[116,117]. These alternating shifts in dietary patterns have an impact on the gut microbiome composition and subsequently influence the production of metabolites such as SCFAs[118,119].

Oscillating SCFA levels could potentially influence the epigenetic suppression of incretin genes in EECs, leading to fluctuating incretin hormone secretion and subsequent intermittent effects on the host's glucose metabolism. In this context, butyrate can be considered as representative of SCFA molecules because the metabolic impact of SCFAs is predominantly attributed to butyrate[120,121]. Also, previously undisclosed findings from our research group have revealed that exposing T84 cells to butyrate for four days resulted in a substantial, dose-dependent increase in GLP1 expression, a phenomenon not observed with acetate and propionate exposure. Additionally, in another unpublished experiment by our team,

prolonged exposure (168 days) to high levels of butyrate (10mM) maintained elevated GLP1 gene expression. To build upon this line of investigation, this thesis intended to employ T84 and HT-29 cell lines as the well-known models for colon epithelial cells[122,123]. Therefore, subjecting these cells to alternating high and low butyrate concentrations over multiple cycles could provide additional insight into the expression of incretin hormone genes.

The effect of these intermittent changes can be investigated in more complex systems such as mouse models by subjecting mice to an alternating healthy and unhealthy diet. The combination of in vitro and in vivo experiments may also offer further understanding into the regulation of genes responsible for gene expression profiles of incretin hormones under different exposure conditions. Animal work can also provide an opportunity to examine potential instances of induced insulin resistance during dietary oscillations to study the impact of yo-yo dieting on T2D progression. The effects of the dietary oscillation can be compared with that of normal healthy diet as well as chronic unhealthy high fat diet to evaluate the diet-SCFA-incretin axis in relation to insulin secretion, along with changes in systemic insulin tolerance – two key physiological disorders in T2D progression[5,23]. This is novel research since studying the dietary oscillation in an animal model has not been previously reported, in particular when considering a cell model to back it up.

Furthermore, several mouse behaviours that might be influenced or compromised by being exposed to dietary oscillation can be investigated. Comparing these changes with other exposure conditions can provide new insight into the effects of different diets on cognitive function and neuromuscular activities in mouse. Knowing that high-fat diets can have a negative impact on mice[124-127], the implications of alternating between healthy and unhealthy diets on mouse neuromuscular functions is an uncharted territory. These investigations are

worth studying based on previous reports in patients with T2D including reduced olfactory sensitivity[128], as well as peripheral neuropathy and nerve damage which impact the muscle strength and reflexes[129,130].

Dietary oscillation may be important in explaining the rising epidemic of obesity and T2D[118,119]. Mechanisms linking diet to incretin production could play a central role in the long-term consequences of weight cycling. All these analyses could improve our understanding of the potential effects of dietary oscillation on incretin secretion as one of the deleterious dozen in T2D progression (**Figure 1.1**). Chapter 3 discusses the effect of dietary oscillation in cell and animal models, focusing on factors involved in glucose homeostasis and T2D progression.

1.2 Colorectal Cancer

Colorectal cancer is a heterogeneous disease that develops from precursor lesions known as a colorectal polyp (either conventional adenomas or serrated polyps) which arise from the aberrant proliferation of colonic epithelial cells [the term colorectal adenoma (CRA) is also sometimes used][131-133]. Around 80% of all CRC cases are sporadic which highlights the role of environmental factors on its development[133,134].

Globally, CRC constitutes approximately 10% of all cancer cases[7]. The disease ranks as the third most frequently diagnosed cancer, and the second leading cause of cancer-related death worldwide (9.4%)[7]. It is the second most prevalent cancer in women after the breast cancer, and the third most common in men after prostate and lung cancer[133]. Women experience a lower incidence and mortality rate, around 25% less than men[133]. Projections indicate a global increase of 60% in the overall risk of CRC across all age groups by 2030, resulting in over 1.1

million deaths and 2.2 million new cases[8]. These rates display geographic disparities, and the most developed countries showed the highest rate[133]. Additionally, the global incidence of CRC is estimated to increase by 60% to 2.5 million new cases in 2035[133]. Nonetheless, there was a stable or decreasing CRC incidence in highly developed nations, which is mostly related to preventive CRC screening programs for individuals over the age of 50 years[8,133]. On the other hand, a rising trend of CRC before the threshold age of screening has been observed and has been called Early-onset CRC (EOCRC)[135-139] and in particular the incidence of rectal cancer in individuals younger than 50 years has been increasing[140].

1.2.1 Pathogenesis

Colorectal cancer typically originates from an abnormal crypt that evolves into a precursor lesion known as a polyp[133,134]. The entire progression to CRC spans about 10 to 15 years[133,134], and the main source of most CRC cases is thought to arise from stem cells or cells with stem-cell-like properties[134,141]. These cancer stem cells emerge due to gradual accumulation of genetic and epigenetic changes that deactivate tumour-suppressing genes while activating oncogenes[141,142].

Two tumorigenesis pathways are proposed for CRC progression based on the primary precursor lesion[134,141,142]: the conventional adenoma-carcinoma pathway (also known as the chromosomal instability sequence) contributes to 70-90% of CRC[134,141,142], and the serrated neoplasia pathway accounting for 10-20% of CRC[134,141,142]. These pathways represent distinct series of genetic and epigenetic events[134,141,142]. In the chromosomal instability scenario, genomic changes usually arise following an APC mutation, followed by RAS activation or loss of TP53 function[131,142]. In contrast, the serrated neoplasia pathway is linked to RAS and RAF

mutations, along with epigenetic modification characterized by CpG island methylation resulting in microsatellite unstable CRC[131,142].

In addition, there are distinct features associated with the cancers of right-sided (proximal) colon (cecum, ascending colon, hepatic flexure) and left-sided (distal) colon (cecum, ascending colon, hepatic flexure)[133,134]. For example, the sites of metastasis differ between these locations, as liver and lung metastasis originate mostly from left-sided colon while peritoneal metastasis arise mostly from right-sided colon[143]. Also, higher recurrence and lower survival rate is associated with the proximal than distal colon[144].

1.2.1.1 Consensus molecular phenotype

The Consensus Molecular Subtype (CMS) of CRC refers to a classification system that categorizes colorectal tumours based on their molecular characteristics and genetic profiles[145]. The CMP defines distinct subtypes of CRC based on shared molecular features, which can provide insights into prognosis, treatment response, and potential targeted therapies. There are four commonly identified molecular subtypes of CRC named as CMS1-4[131,145,146].

CMS1 (MSI Immune, 14%): This subtype is characterized by high levels of microsatellite instability (MSI-H) due to defective DNA mismatch repair mechanisms[145,146]. Tumours in this subtype often have a high mutation burden and are more responsive to immunotherapy[145,146].

CMS2 (Canonical, 37%): These tumours are characterized by traditional adenoma-like features and alterations in the Wnt signalling pathway[145,146]. This is the most common subtype and is associated with chromosomal instability (CIN) that leads to numerous chromosomal changes. These tumours are often associated with mutations in APC and KRAS genes[145,146].

CMS3 (Metabolic, 13%): This subtype is characterized by metabolic dysregulation and epithelial-to-mesenchymal transition (EMT)[145,146]. Tumours in this category exhibit distinct metabolic changes and are associated with poor outcomes[145,146].

CMS4 (Mesenchymal, 23%): These tumours show prominent stromal infiltration and a strong mesenchymal phenotype[145,146]. They are associated with higher likelihood of invasion to surrounding tissues and distant metastasis[145,146].

The remaining 13% of CRC tumours cannot be assigned to either of the four CMS categories due to mixed genomic signature and intratumor heterogeneity[147].

1.2.1.2 Hereditary syndromes

Up to 10% of all CRC cases are affected by hereditary syndromes[133,148]. The hereditary cases are either polyposis (FAP) or non-polyposis (Lynch) syndrome[133,148]. Familial Adenomatous Polyposis is characterized by the development of numerous benign adenomatous polyps in the colon and rectum with a high risk of becoming cancerous[148,149]. This autosomal dominant genetic disorder is caused by mutations in the tumour suppressor gene *APC* similar to CMS2 in sporadic CRC[148,149]. Lynch syndrome or hereditary non-polyposis colorectal cancer (HNPCC) is another autosomal dominant disease characterised by mutation in one of the alleles coding for DNA mismatch repair proteins[150,151]. These tumours are similar to CMS1 category[150,151]; however, these patients have an increased risk of developing other cancers since they carry a germ-line mutation[148,150,151].

Hereditary CRC often presents as early onset of CRC due to carrying a germ-line mutation[133,148]. Nonetheless, it is important to differentiate the hereditary syndromes from sporadic EOCRC, which is characterized by the CRC onset before the age of 50 years[133,134]. Sporadic EOCRC is diagnosed based on age without the presence of a known underlying

genetic cause[133,134], however, FAP and Lynch syndrome is caused by genetic mutations[148-150]. In addition, while family history can be a risk factor in EOCRC, it does not necessarily imply a hereditary genetic predisposition[133,141]. In contrast, Lynch Syndrome is often characterized by a strong family history of CRC and other associated cancers, signifying the hereditary component[150,151]. Moreover, individuals with Lynch syndrome require more intensive surveillance and preventive measures such as regular colonoscopies and screenings due to the higher CRC recurrence risk and the risk of other cancers[150,151].

1.2.2 Stages of CRC

Properly staging colorectal cancer is crucial for determining the extent of the disease, planning appropriate treatment strategies, predicting prognosis and facilitating communication among healthcare professionals[152]. The staging system provides a standardized framework to describe the size and extent of the primary tumour, lymph node involvement and the presence of distant metastases[133,153]. The most commonly used staging system for CRC is the American Joint Committee on Cancer (AJCC) tumour-nodes-metastasis (TNM) system, which classifies the disease into stages based on the following components [153].

The TNM system comprises three main parameters: T, N, and M. The tumour itself is defined by T (Tumour) which describes the size and extent of the primary tumour within the colon or rectum. The involvement of lymph nodes is defined by N (Nodes) indicating the tumour spread to the regional lymph nodes[152,153]. Lastly, M (Metastasis) reflects the presence or absence of distant metastases, indicating whether the cancer has spread to other parts of the body[152,153]. All these parameters are summarised in **Table 1.3**. Combining T, N, and M leads to the overall TNM stage grouping, summarized in **Table 1.4**. The extent of tumour growth in stage 0-IV is shown in **Figure 1.4**.

Table 1.3 The categories of TMN classification of CRC

Category	Description
T Category	
Tis	Carcinoma in situ: Cancer cells confined to inner lining.
T1	Invasion into submucosa (inner layer).
T2	Invasion into muscularis propria.
T3	Spread through muscularis propria into subserosa or non-peritonealised tissues.
T4a	Penetration to the surface of the visceral peritoneum.
T4b	Passed through the wall of the colon or rectum and is attached to or has grown into other nearby tissues or organs.
N Category	
N0	No regional lymph node involvement.
N1a	Spread to one regional lymph node.
N1b	Spread to 2-3 regional lymph nodes.
N1c	Tumour deposits in subserosa, mesentery, or non-peritonealised tissues without regional node involvement.
N2a	Spread to 4-6 regional lymph nodes.
N2b	Spread to 7 or more regional lymph nodes.
M Category	
M0	No distant metastasis.
M1a	Cancer spread to one distant site beyond the colon or rectum.
M1b	Cancer spread to more than one distant site beyond the colon or rectum.
M1c	Cancer spread to the peritoneal surface.

*This table is based on TNM Staging available in Cancer Research UK official website (https://www.cancerresearchuk.org/about-cancer/bowel-cancer/stages-types-and-grades/TNM-staging)

Table 1.4 The AJCC 8th edition of TMN classification of CRC

AJCC Stage	T Category	N Category	M Category	Description
0	Tis	N0	M0	Carcinoma in situ, localized within the inner lining of the colon or rectum.

AJCC Stage	T Category	N Category	M Category	Description
I	T1-T2	N0	M0	Limited growth into deeper layers (into the submucosa (T1), or muscularis propria (T2)) no lymph node involvement, and no distant metastasis.
IIA	T3	N0	M0	Deeper invasion but not beyond the bowel wall, no lymph node involvement, and no distant metastasis.
IIB	T4a	N0	M0	Grown through the bowel wall but has not grown into other nearby tissues or organs, no lymph node involvement, and no distant metastasis.
IIC	T4b	N0	M0	Grown through the wall of the bowel wall and into other nearby tissues, no lymph node involvement, and no distant metastasis.
IIIA	T1-T2	N1a-N1b	M0	Grown into the submucosa (T1) or the muscularis propria (T2), spread to 1 to 3 nearby lymph nodes (N1a or N1b), and no distant metastasis.
IIIB	T3-T4a	N1a-N1b	M0	Grown into the outermost layers bowel (T3) or through the visceral peritoneum (T4a), spread to 1 to 3 nearby lymph nodes (N1a or N1b), and no distant metastasis.
IIIC	T3-T4a	N2a-N2b	M0	Grown into the outermost layers bowel (T3) or through the visceral peritoneum (T4a), spread to 4 to 6 nearby lymph nodes (N2a) or 7 or more nearby lymph nodes (N2b), and no distant metastasis.
IVA	Any T	Any N	M1a	Distant metastasis to one organ.
IVB	Any T	Any N	M1b	Distant metastasis to multiple sites or the peritoneum.
IVC	Any T	Any N	M1b	Metastasis to distant parts of the peritoneum, may or may not have spread to distant organs.

*The highlighted blue rows represent the metastasized cancer. Based on CRC stages in the American Cancer Society official website (https://www.cancer.org/cancer/types/colon-rectal-cancer/detection-diagnosis-staging/staged.html)

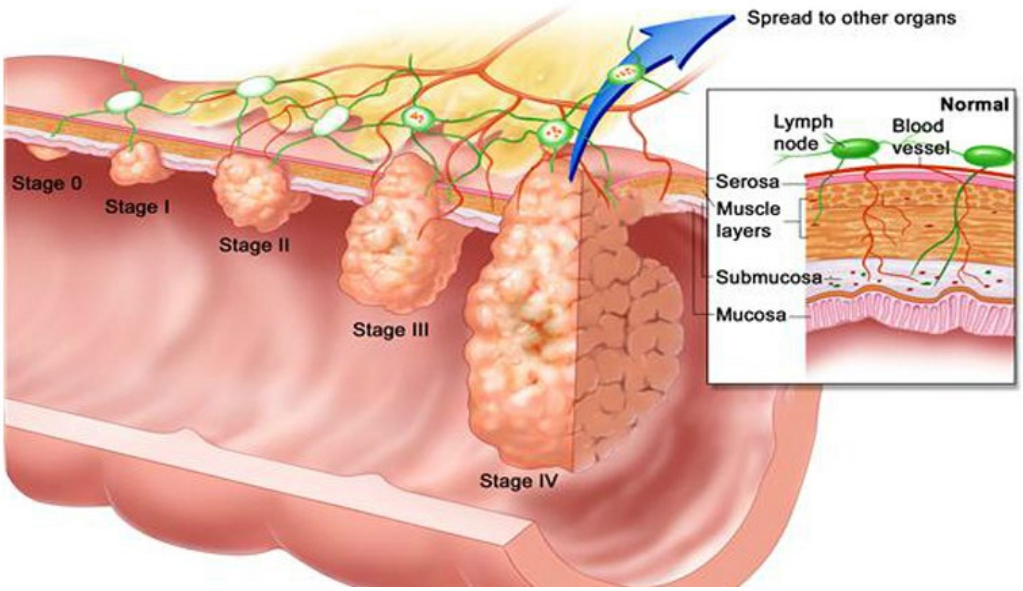


Figure 1.4 Schematic of CRC stage and progression

Schematic view of CRC stage 0-IV. (From: https://www.bowelcanceraustralia.org/bowel-cancer-staging)

1.2.3 Risk factors of CRC

One of the most significant risk factors for CRC is age[133,134], with the risk increasing notably as individuals advance beyond the age of 50[133,134]. Male sex is shown a strong association with the disease[133,134]. Other non-modifiable factors such as genetic predisposition plays a crucial role[133,141]. Positive family history exists in approximately 10–20% of all CRC patients[133,141]. These hereditary risk factors and genetic predispositions depends on the number and degree of affected relatives as well as their age at onset of CRC[133,134,141]. Individuals with a family history of certain hereditary condition such as familial adenomatous polyposis (FAP) and Lynch syndrome are at increased risk[134,141]. Moreover, a personal history of colorectal polyps, or other cancers such as ovarian, endometrial, or breast cancer can further increase the risk[133,134,141]. Inflammatory bowel disease (IBD) such as Crohn's disease and ulcerative colitis

elevates the risk due to their prolonged deleterious impact of inflammation to the bowel[133,141]. Furthermore, individuals with T2D have an elevated risk of developing CRC[21,154].

Among the modifiable risk factors, lifestyle play a major role[133,141]. Systematic reviews have shown that diets rich in red and processed meats[155,156] while lacking in fibre, fruits, and vegetables have been linked to an increased CRC risk[157-159]. Sedentary lifestyles and obesity further compound this risk[10,11]. Smoking tobacco and high consumption of alcohol have been established as other environmental risk factors[160]. It is important to note that the CRC development is influenced by an interplay of genetic and environmental factors[133,141], however, environmental factors play a substantial role in developing CRC which is reflected in the fact that 80-90% of all CRC cases are sporadic[133,161].

1.2.4 Diet and SCFAs in colorectal health

As stated earlier, the main CRC risk factors are family history, older age after the fifth decade of life, male sex, and lifestyle including diet, obesity, physical activity habits, smoking, and alcohol consumption[132,133]. Diet plays a substantial role among the environmental factors, and the dietary components can provide a healthy or unhealthy environment in the bowel lumen via gut microbiota composition and their produced metabolites[107,111,113].

The metabolites generated by the gut microbiota are in constant communication with colonocytes, with short-chain fatty acids (SCFAs) comprising a significant portion of these metabolites[107,112,113].

In addition to the role of diet and SCFAs on glucose homeostasis and T2D progression - discussed in section 1.1.8 and Chapter 3 - dietary fibres exhibit beneficial anti-inflammatory, anti-carcinogenic and pro-apoptosis effects on colonocytes, mainly mediated through the

production of SCFAs from specific gut microbiota[107,111,114]. Therefore, our diet can potentially contribute to the pathogenesis of colonic diseases such as IBDs as well as CRC[103,105,112].

Apart from acting as signalling molecules for EECs, SCFAs have many direct physiological effects on colonocytes[103,106,113]. These molecules are the energy source for colonocytes and help maintain colonic homeostasis through the integrity of the mucosal barrier, regulation of energy metabolism, and inhibition of gut pathogen growth via lowering the lumen pH[102,103,106]. Furthermore, *in vitro* and animal model studies revealed several anti-inflammatory, anti-carcinogenesis and anti-oxidative roles of SCFAs, as they affect colon epithelial or immune cell proliferation, and differentiation[104-106]. The beneficial properties of dietary fibres to protect against colorectal cancer, T2D, obesity, and IBDs appear to be mediated through SCFAs[106,107]. Among the three major gut microbiota-generated SCFAs – acetic, propionic, and butyric acid, although acetic acid is the most abundant gut SCFA, butyric acid is the main SCFA with varied physiological benefits for colonocytes[102,103,106,112]. Additionally, butyric acid is recognized as a primary energy source for colonocytes[102,103,107]. Consequently, colonic health could be affected by alteration in SCFA concentrations resulting in the abnormal growth of colonocytes and tumour development[105,106].

1.2.4.1 Knowledge gap: establishing a link between SCFA level and CRC progression

Several *in vitro* and *in vivo* studies have demonstrated the beneficial effects of SCFAs in CRC[102,104,106], however, human studies investigating the association between SCFA concentration and risk and/or incidence of CRC, have been inconclusive[162-167]. Other human studies have examined the SCFAs levels in faecal samples of patients with colorectal carcinoma or adenoma[162,163,165-179]. Another series of investigations were conducted to compare faecal SCFA concentrations among healthy individuals residing in different

geographical locations or ethnic backgrounds, particularly those with the highest and lowest CRC incidences[180-185].

Nevertheless, due to inconsistent findings, a conclusive assessment of SCFA profile in patients with CRC or at-risk individuals is currently absent. Due to the protective effect of dietary fibres against CRC progression[157-159], it was hypothesised that the faecal concentration of the three major SCFA molecules is associated with the CRC risk and progression. The results of all primary observational human studies can be analysed systematically to investigate the association between the three major SCFA molecules on the two categories of CRC risk and CRC incidence.

It would be of great benefit to divide these studies into two distinct categories of CRC incidence and CRC risk. The focus can be on comparing the patients with CRC, apparently healthy individuals with diagnoses colorectal adenoma, and healthy controls. Moreover, comparing the populations with the highest or lowest incidence of CRC can provide a valuable result about the concentration of faecal SCFAs in those regions. All these data can help identify a potential link between SCFA level and CRC.

1.2.5 Diagnosis and management of CRC

Colorectal cancer often remains without noticeable symptoms until it becomes an advanced disease[133,134]. The CRC symptoms such as abdominal pain and change in bowel habits are not CRC specific[133,134]. However, rectal bleeding, although not specific to CRC, can be considered as a symptom[133,134]. Rectal bleeding in combination with other risk factors such as positive family history of CRC and weight loss warrant an in-depth examination by colonoscopy[133,134]. Colonoscopy is the gold standard for CRC diagnosis which can identify tumour location with high accuracy[133,186]. It involves inserting a flexible tube with a camera into the rectum and

colon to examine its lining to detect various conditions such as polyps, cancer, IBDs, diverticulosis, and bleeding[133,186].

1.2.5.1 Endoscopic treatment

Apart from diagnosis, colonoscopy is used to remove polyps and collect biopsy samples at the same time for diagnosis[133,134]. The diagnostic biopsy is used for histological validation and for molecular profiling[133,134]. Notably, colonoscopy remains the sole screening method that delivers both diagnostic and therapeutic benefits[133,186]. By removing neoplastic polyps, it effectively prevents the tumour formation and possible metastasis leading to treatment and the patient survival[133,134]. The efficacy of colonoscopy in reducing CRC incidence and mortality was first demonstrated in the US population[187,188].

1.2.5.2 Surgical treatment and post-surgical survival

Surgery is the fundamental curative intervention for individuals diagnosed with non-metastasized CRC[133,134]. In more advanced tumours, neoadjuvant therapy (such as preoperative chemotherapy and chemoradiotherapy or radiotherapy for locally advanced cancer can enhance the likelihood of a successful resection[133,134].

Post-treatment long-term survival of CRC patients depend primarily on the cancer stage[8]. Patients diagnosed with CRC in its early stages show the average 5-year survival rates of around 90%[134], whereas the survival rates for those with advanced-stage metastatic disease is as low as 10%[134]. Also, the survival rate of CRC at earlier stages was lower in elderly patients compared to their younger counterparts[133,134]. The high incidence of CRC, low survival rate for the metastatic disease and the high costs of treatment, together with the fact that CRC is mostly an asymptomatic disease until late stage has prompted the adoption of CRC screening programmes in many countries in all individuals with 50–75 years of age[133,134,186,189].

1.2.6 Prevention of CRC

Primary prevention of CRC includes controlling for the modifiable risk factors by adopting a healthier lifestyle and measures such as regular exercise, active lifestyle, quit smoking, and most importantly improving the diet by for example higher intake of fibre-rich foods[107,133]. Several meta-analyses have shown the association between high CRC risk and high consumption of processed and unprocessed meat[155,156], as well as the role of high fibre intake as a protective factor against CRC progression[157-159]. The potential chemopreventive effects of vitamin D, folate, and calcium is also proposed[190,191].

1.2.6.1 Secondary prevention

Colonoscopy is the primary method to prevent CRC with high sensitivity and specificity[133,134,186,189]. The procedure is invasive; however, it provides the advantage of directly excising the colorectal polyps and precursor abnormalities[133,134,186,189]. Therefore, those at heightened risk, such as individuals with hereditary or familial predisposition, or with the prior history of adenomas or CRC are advised to undergo periodic colonoscopy surveillance[133,134,186,189].

There are other non-invasive methods available for the population screening of CRC[186,189]. The guaiac faecal occult blood test (gFOBT) and faecal immunochemical test (FIT) are non-invasive stool-based methods[192-194]. These two stool-based screening tests are much cheaper than colonoscopy, however, with lower sensitivity and specificity since the presence of blood in the stool could be related to other underlying conditions such as haemorrhoids, Diverticulosis, ulcerative colitis, Crohn's disease, and other gastrointestinal bleedings[186,189,192,193]. Therefore, a positive gFOBT or FIT test should be followed up by colonoscopy.

1.2.7 Early-Onset Colorectal Cancer

Globally, the CRC incidence has either declined or remained stable over the past two decades[135,195,196]. This trend can largely be attributed to the implementation of CRC screening programs for individuals aged 50 and above[197-199]. Age is one of the main risk factors of CRC is, and the risk of developing CRC significantly increases after the age of 50 years[132,133]. In western countries, population screening of CRC is performed from the age of 50 years[133,200]. These preventive methods have improved the detection and removal of initial polyps before transforming into carcinoma and therefore has lowered the number of CRC cases[133,200].

However, the rate of the diagnosis of CRC in before the age of 50 years, i.e., EOCRC, is increasing in western countries which is exacerbated by a lack of screening in this age group[138,200-204]. In 2010, 4.8% of colon cancers and 9.5% of rectal cancers were detected in younger individuals before the age of 50 years[205]. It is projected that by 2030, these figures will rise significantly to comprise 10.9% of colon cancers and 22.9% of rectal cancers[205]. This doubling within two decades is an alarming rise in EOCRC incidence[197,205]. Sporadic EOCRC now comprises approximately 10% of all CRC cases[136]. Importantly, EOCRC typically involves the rectum or distal colon and often manifests at more advanced stages, possibly due to delayed diagnosis and the absence of screening for this age group[198,199].

This escalating trend has been observed in cancer registries spanning 20 European countries, collecting data between 1990 and 2016[203]. Within the United States, an examination of the population-based Surveillance, Epidemiology, and End Results (SEER) database from 1980 to 2016 revealed a rise in EOCRC cases and an increased incidence of rectal cancer in this demographic[206]. Furthermore, the rising prevalence of EOCRC within this age bracket was also evident in the US Cancer Statistics report covering the years 2001 to 2017[202]. Intriguingly, a

comparative analysis of EOCRC incidence between the United States and Europe unveiled a significantly higher occurrence rate in the US, potentially linked to the higher prevalence of obesity in the US population[207]. These investigations collectively highlight the disconcerting increase in EOCRC cases across both the US and Europe, a pattern similarly found in UK[195], and Canada[208].

In Australia, the incidence of CRC stands among the highest in developed nations[209]. While the overall incidence has been decreasing for individuals aged over 50 since the mid-1990s, a notable increase in the occurrence of both colon and rectal cancer has been reported in individuals under the age of 50 over the past twenty years[210]. However, two population-based studies conducted in the two most populous states of the country reported stable incidence rates of EOCRC[211,212]. Similarly, other studies in the UK and Japan also found no evidence of an increase in the proportion of EOCRC cases[213,214].

This highlights the need for targeted research and interventions to address the unique challenges posed by EOCRC[204,215] and necessitates a deeper understanding of its aetiology and implications for clinical practice[199].

1.2.7.1 Current challenges in EOCRC management

The same risk factors for CRC also contribute to sporadic EOCRC progression[216,217]. Apart from hereditary CRC with younger age onset, genetic predisposition may play a role in sporadic EOCRC[218,219]. It means that tumour may develop earlier in susceptible individuals[217,219]. Molecular analyses of EOCRC cases have revealed some distinct features compared to later-onset CRC[220,221]. There were also some attempts to classify EOCRC based on the tumour location[221]. However, EOCRC is marked by significant heterogeneity in terms of clinical presentation and molecular features[222,223]. Therefore, a lack of standardized diagnostic

criteria, molecular subtyping, and staging systems specific to EOCRC makes it challenging to compare results across studies and establish consistent management guidelines[199,204].

In addition, while genetic predisposition is recognized as a major factor in EOCRC, identifying all relevant genetic variants and understanding their functional significance is complex[216,217,219]. The contribution of polygenic risk factors, rare mutations, and their interactions is not fully elucidated, making it difficult to implement effective genetic screening and counselling strategies[199,204]. However, one of the most challenging fact about sporadic EOCRC is related to the diagnosis of the disease at the younger age[199,204]. Sporadic EOCRC often presents with non-specific symptoms and therefore at more advanced stages due to delayed diagnosis[138,224]. This is exacerbated by the lack of awareness among clinicians as they mostly expect to see the disease at later age[225,226].

Tailoring treatment strategies for EOCRC based on molecular characteristics is a promising avenue, but challenges arise in identifying actionable mutations and developing targeted therapies[222,223]. Clinical trials specific for treatment of sporadic EOCRC is another challenge since patient recruitment for these trials can be difficult due to the rarity of the disease[227]. Furthermore, the lack of reliable animal models that mimic EOCRC development hinders the investigation of underlying molecular mechanisms and the testing of potential therapeutic interventions in a controlled environment.

Currently, most existing research on EOCRC is retrospective in nature, however, prospective studies are needed to gather comprehensive, real-time data on EOCRC patients[204]. This may allow for a better understanding of risk factors, disease progression, treatment response, and outcomes[204]. There is also the need for standardized data collection and sharing protocols to

enhance collaborative efforts and achieve a comprehensive understanding of sporadic EOCRC[228].

It is important to note that patients with sporadic EOCRC appear to have a better overall survival compared to their older counterparts[229]. However, since they suffer from the disease at younger age, the long-term survival is still not clear compared to late-onset CRC patients[227]. The long-term physical, psychological, and social impacts of EOCRC and its treatments on this younger age group are not well understood[225]. Addressing survivorship concerns, fertility preservation, and quality of life issues requires specialized support and research efforts[227].

The rising incidence of sporadic EOCRC could also be related to other comorbidities as risk factors of CRC[216,219]. For example, the age at onset of IBDs is around 40 years[230], which can result in CRC progression at the younger age considering the fact that environmental risk factors such as unhealthy diet and sedentary lifestyle are more prevalent[231].

Nonetheless, lifestyle factors, such as high consumption of processed meats, low fibre intake, and sedentary behaviour, are implicated in sporadic EOCRC cases[216,217]. Obesity as another CRC risk factor is becoming more prevalent among young individuals over the past recent decades[232]. Considering obesity as one of the contributing risk factors for CRC[233], the rising tide of obesity and sedentary lifestyle among young people could be a possible reason for the earlier age at diagnosis of CRC[200]. Obesity by itself is the main risk factor of T2D[37], which in turn is associated with higher incidence of CRC[13,233]. Taken together, all these environmental factors could contribute to the rising incidence of sporadic EOCRC over the past two decades[234].

Considering all the challenges discussed here as well as the concerning rise in EOCRC incidence warrant the development of reliable and cost-effective screening methods for

younger populations, especially those without a strong family history[234]. Determining the appropriate age to initiate screening remains a topic of debate[197,234].

1.2.7.2 Knowledge gap: exploring features associated with sporadic EOCRC

Australia is among the countries with the highest prevalence of CRC[235,236]. Additionally, several studies have reported either a rising trend or a steady trend in the incidence of EOCRC in the Australian population[139,211,212,237,238]. Given the disparities in various population-based studies within Australian states and the perceived lack of understanding regarding the implications of EOCRC in the local context, further cohorts need to be analysed. The focus needs to be on sporadic EOCRC by excluding the hereditary cases since the underlying mechanisms of sporadic and hereditary cases are distinct[138,234]. Other outcomes such as the trend of sporadic EOCRC incidence as well as clinicopathological features can be studied using a longitudinal approach. These features could include the primary tumour location (proximal, distal colon, or rectum), tumour histology (adenocarcinoma, mucinous, signet ring), tumour stage, metastasis (lymph node or distant), and pattern of metastasis. Comparing EOCRC with their older counterparts could be beneficial to understand the features that specifically associated with patients in the younger age group. In addition, examining the survival between the two age groups can provide valuable information. Furthermore, all these factors can be compared between men and women with sporadic EOCRC to identify the possible differences between them.

All these potential results may provide additional evidence and support for reducing the commencing age of the current Australian National Bowel Cancer Screening Program to identify more young individuals at risk of diagnosis with advanced EOCRC. These results could challenge the traditional screening recommendations that typically target older

populations[234]. This could prompt the discussions about the need for earlier screening guidelines for individuals at higher risk. Chapter 5 of my thesis is dedicated to such in detail analyses.

1.3 Hypotheses

In this thesis, four hypotheses were explored, each contributing to a comprehensive understanding of T2D and CRC progression whilst laying the foundation for further research in the field. Regarding T2D, the first hypothesis centres on the potential of *TERC*, the RNA component of telomerase enzyme, as a complementary biomarker to the telomere length attrition during T2D progression. In another aspect of T2D, it is hypothesised that dietary oscillation may adversely affect glucose metabolism via changing the SCFA content in the colon and the subsequent imbalance in incretin hormones. The linking hypothesis between T2D and colorectal cancer is a relationship between the concentration of SCFAs and increased risk of CRC. Lastly, with regards to CRC at younger age, the hypothesis is that there are specific tumour features and risk factors that associated with that age group.

1.4 Aims

In this thesis, a set of four objectives was formulated to address the hypotheses above. The enumerated list of these objectives is as follows.

1) To employ hIPCs as a model of human islet cells to investigate the telomere length attrition rate, as well as changes in *TERC* expression level and senescence markers by exposing the cells to different glucose conditions.

2) To study the potentially deleterious effects of dietary oscillation on incretin hormone levels by A) Analysing the changes in their gene expression level in human colonic cell lines exposed to intermittent SCFA concentrations; B) Performing an animal study including exposing the mice to alternating healthy and unhealthy diet to investigate the downstream effects on obesity, glucose metabolism, and behavioural changes.
3) To conduct a systematic review and meta-analyses of the evidence to assess the association between faecal SCFA levels and the risk or incidence of CRC.
4) To investigate sporadic EOCRC incidence and to identify potential incidence trends, risk factors as well as clinicopathological and survival features associated with this age group by analysing a large cohort including 3609 CRC cases collected over a 26-year period.

Chapter 2 - Telomere Length Preservation and Cellular Senescence in Type 2 Diabetes Progression

2.1 Background

Type 2 diabetes is a chronic health condition and telomere attrition has been associated with T2D progression and accompanied by clinical complications[5,6,63,68]. The pathophysiology of T2D encompasses decreased β-cell mass due to the increased activity of these cells to produce insulin to compensate for insulin resistance-derived hyperglycemia[24,39]. The relative leukocyte telomere length (rLTL) is commonly measured in readily available plasma samples as an estimation of biological aging and cellular senescence in chronic diseases[63,239]. Several meta-analyses have reported shorter rLTL in T2D patients[71,75,76], which in general depends on two factors: telomeric attrition during each cell division and the capacity of telomerase to repair this shortening[68,72,73].

Telomerase is a ribonucleoprotein enzyme with a lncRNA (Long non-coding RNA) component, called *TERC*, which does not have a catalytic function but serves as a template for the enzyme to add telomere repeats[61,72,73].

It is important to note that telomere length measurement by itself is not fully representative of cellular senescence, and measuring telomerase activity is a big challenge, especially in stored/biobank plasma samples. Therefore, A reliable surrogate for telomerase activity in plasma is highly desired to understand if the telomere shortening is associated with telomerase activity.

We hypothesized that changes in telomerase activity could be measured through assessment of its lncRNA component, *TERC*, which can potentially provide a surrogate of telomerase activity. Since the overall aim was to investigate the telomeres biology during T2D progression, we decided to evaluate all three aspects of telomere biology - telomere length, telomerase activity, and *TERC* level - in human Islet-derived Precursor Cells (hIPC)[93] as the representative model of pancreatic islet cells.

We also hypothesized that different glucose conditions could affect the β-cells senescence and telomeres differently, and thus aimed to investigate the telomere shortening rate, and *TERC* level as well as the expression levels of senescence markers in glucose-exposed hIPCs.

It is important to note that telomeric length shortening is one of the hallmarks of cellular senescence[62-64]. This is the permanent state in which the cells stop dividing i.e. cell cycle arrest[240,241]. The cell cycle is regulated by an intricate network of proteins including cyclin, cyclin-dependant kinase (CDK), CDK inhibitor (CDKi), and checkpoint molecules[242]. As the negative regulators of cell divisions, CDKi molecules are among the main inducers of cellular senescence, of which $p21^{CIP}$ (hereafter p21) and $p16^{INK4a}$ (hereafter p16) and expressed by CDKN1A and CDKN2A, respectively, are considered as the main inducers of cell cycle arrest[62,240,241]. p21 and p16 are well-known and specific markers for establishing and maintaining cellular senescence[62,240,243].

The two well-known senescence pathways are p53/p21 and p16/pRB[240,241,244]. The former is induced by DNA damage caused by telomere shortening or other sources of nuclear stress[240,241]. DNA damage activates key proteins involved in senescence such as 53BP1 (p53-binding protein 1), ATM (Ataxia Telangiectasia Mutated protein) and ATR (ataxia-telangiectasia and Rad3-related protein)[240,241]. ATM and ATR phosphorylate the downstream

checkpoint kinases CHK2 and CHK1, respectively, which in turn lead to p53 phosphorylation[240,241]. The phosphorylated p53 is more stable and stimulates the p21 expression[62,240,241]. p21, together with p27 and p57, can inhibit several CDK molecules including CDK2 and CDK4 causing the cell to arrest at the G1 stage[240,241,243].

p16 is another CDKi molecule playing a key role in maintaining the senescent state of the cell[62,240,241,243-245]. Downstream of p16 is the RB family of tumour suppressors (pRB, p107, and p130)[244,246]. These proteins inhibit the cell cycle progression when bound to E2F proteins, the activator of the genes involved in cell cycle progression (G1 to S phase)[241,244]. Hyperphosphorylation of the RB protein family by CDK molecules (CDK2, CDK4, and CDK6) release RB proteins resulting in cell cycle progression (G1 to S phase)[244,246]. Inhibiting the CDKs by CDKi molecules such as p16 and p15 lead to the binding of RB proteins to E2Fs and subsequent cell cycle arrest[241,243].

Apart from well-known molecules regulating cellular senescence, there are several factors involved in the cell cycle. Among them, cyclin molecules such as A2, B1, D1, D2, D3, E1, and E2, as well as CDK1, p18, and p19 play important roles in either G1 to S or G2 to mitosis phase of the cell cycle[242,247]. Therefore, activation or inhibition of these molecules could potentially affect cellular senescence. Besides, proteins that protect at telomere ends such as telomeric repeat binding factors 1 and 2 (TRF1 and TRF2) are worth being investigated for studying the telomeres biology in cellular senescence[92,248].

Other factors are also considered the classic markers of cellular senescence. Lamin B1 is a scaffolding component of the nuclear envelope, and its decreased level is studied in senescent cells [62,65]. Galactosidase β1 (GLB1) is a lysosomal enzyme accumulating in senescent cells which are used as another classic marker for senescence[65,245]. Proliferating cell nuclear

antigen (PCNA) is also used to distinguish senescent from proliferating cells[240,245], and GSK3B (glycogen synthase kinase 3 beta) is involved in the regulation of the metabolic phenotype in senescent cells[249,250]. Another molecule of interest is BCL-W, a well-known antiapoptotic protein upregulated to maintain the senescence state[240,246]. On the contrary, sirtuin proteins are shown to be involved in delaying cellular senescence and aging[65,251].

Interestingly, senescent cells can affect their microenvironment by secreting a complex of proteins called senescence-associated secretory phenotype (SASP) as another hallmark of cellular senescence[240,241,243,244,252,253]. This SASP signature constitutes proinflammatory cytokines and chemokines (notably interleukin 6 and 8), matrix metalloproteinases (degrading the extracellular matrix, such as MMP1), insulin-like growth factor binding proteins (IGFBPs), and serine/cysteine proteinase inhibitors (SERPINs)[252-255]. SASP factors are activated at the transcription level, notably by NF-κB (nuclear factor-κB) and CCAAT/enhancer-binding protein-β (C/EBPβ), and GATA Binding Protein 4 (GATA4) transcription factors[62,240,241,246,253,255]. p38MAPK (mitogen-activated protein kinase 14, or p38) is a MAP kinase family which acts as a modulator of SASP factors [244,246], and is related to DNA damage and p53[241].

In summary, there is no single marker identified for cellular senescence and this phenomenon is not studied in all the tissues[240,245,252]. Nonetheless, a set of senescence markers is a better option to study the changes in the differentiated cells with replicative arrest[240,245,252]. Therefore, a customised panel of senescence markers was designed to investigate the expression of 50 known senescence-associated molecules in hIPC samples aiming to evaluate the potential changes in the expression of these biomarkers during their cell passages till their replication arrest, while being exposed to the three aforementioned glucose conditions. The

results could provide us with a better understanding of gene expression during the senescence of islet cells as a potential therapeutic option to slow down their senescence during T2D progression.

2.2 Methods

2.2.1 Cell culture

The human cadaveric samples of pancreas islets used in this study were obtained through human research ethics committee (HREC) approvals X16-0289 (previously X12-0176) and the HREC/12/RPAH/282 as well as MQ5201300330 protocols and Western Sydney University human ethics number H14213.

2.2.1.1 Phase 1

The overall procedure of Phase 1 is shown in **Figure 2.1**. Four hIPC samples were cultured in CMRL media, customised by adding 10% FBS, 1% Pen-Strep, 1% Glutamax, and 0.01% hEGF at 37°C with 5% CO_2. As adherent cells, hIPCs obtained >90% confluency after passaging for four days. After each passage, the cells were harvested and counted, from which 0.5x10^6 cells were seeded in a total of 4ml in a T25 flask for the next passage. Also, 0.2x10^6 and 0.1x10^6 cells were pelleted for RNA and DNA isolation, respectively. The hIPC passaging continued till senescence in which the cells were not proliferating. The detailed cell culture protocol is available in **Appendix A**.

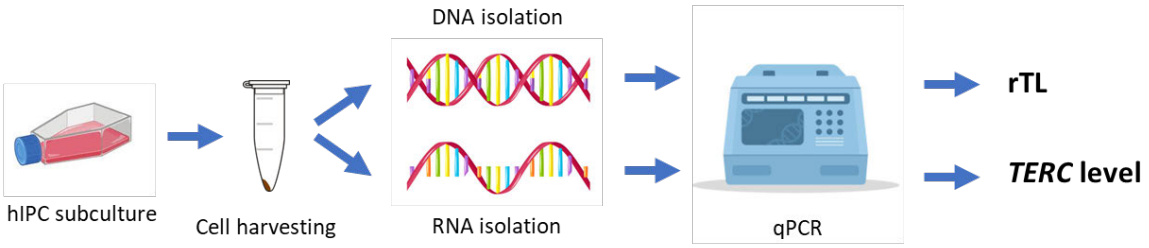


Figure 2.1 Workflow of Phase 1

Four hIPC samples were passaged every four days in parallel till senescence. The cells were harvested at the end of each passage and pelleted for DNA and RNA isolation to be used for rTL and *TERC* level measurement, respectively, using qPCR. The schematic icons were downloaded from the BioRender website.

2.2.1.2 Phase 2

In Phase 2 of the study, as depicted in **Figure 2.2**, 12 hIPC samples were cultured in the customised CMRL medium, and passaged for four days while being exposed to glucose in three different conditions (four hIPC samples per condition):

1) Chronic high-glucose condition (HG) containing 20mM D-glucose. Since CMRL itself has 5.5mM D-glucose, the concentration increased by 14.5mM to obtain the target 20mM concentration. This was done by adding 58ul of 1M D-glucose to 4ml CMRL in the T25 flask.
2) Chronic normal glucose condition (NG) containing 5.5mM D-glucose and 14.5mM L-glucose. In this condition, the total glucose concentration is 20mM. L-glucose cannot be digested by the cells, and it is applied to give the same osmolality as D-glucose to CMRL to rule out the effect of osmolality in damaging the cell membranes. This media is made separately by adding 725ul of 1M L-glucose to 50ml CMRL media.
3) Oscillating glucose condition (OSC), with changing the media every 24 hours (Days 1 and 3 in NG, and days 2 and 4 in HG condition).

The normal glucose condition (NG) containing 5.5mM D-glucose which is approximately the normal glucose plasma level in healthy individuals[5]. The chronic high-glucose condition (HG) in the experiments contained 20mM D-glucose which is within the normal range (10-25mM) used in many other T2D *in vitro* studies[256]. D- and L-glucose (cat. No. G7021, and G5500, respectively) were purchased from Merck (Merck & Co, Whitehouse Station, NJ, USA).

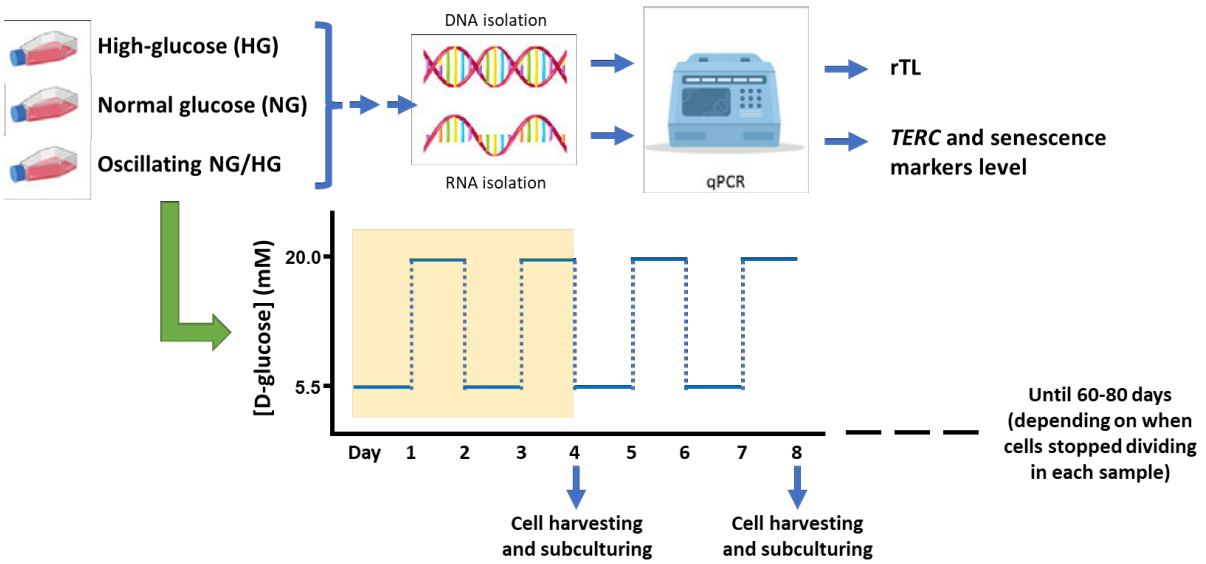


Figure 2.2 Workflow of Phase 2

In total, 12 hIPC samples were subcultured every four days in parallel till senescence, four samples per each of three exposure groups: HG, NG, and oscillating NG/HG (OSC). In the OSC group, the cells were exposed to alternate high (20mM) and normal (5.5mM) concentrations of D-glucose periodically. The highlighted yellow section shows two cycles of OSC in one four-day period. The cells were harvested at the end of each passage and pelleted for DNA and RNA isolation to be used for rTL and *TERC* level measurement, respectively, using qPCR. The schematic icons were downloaded from the BioRender website.

2.2.2 Molecular analysis

DNA was isolated from hIPC samples using QIAmp DNA Blood Mini Kit (Qiagen, Chadstone, Australia), with minor protocol revision (**Appendix B**), and quantified using NanoDrop™ Lite

Spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA). The relative telomere length (rTL) was measured by real-time quantitative PCR (qPCR) method using Fast SYBR Green master-mix and ViiA 7 Real-Time PCR System (ThermoFisher Scientific, Waltham, MA, USA). This method was previously optimised in the team by Dr Mugdha Joglekar, and published recently[257]. The human β-globin (*hBG*) gene was utilised for normalisation as the single-copy gene (per haploid genome). The detailed step-by-step protocol is available in **Appendix C**.

RNA isolations from hIPC biological replicates were performed manually using a protocol that has been well-established in our lab (**Appendix D**) using TRIzol reagent (ThermoFisher Scientific, Waltham, MA, USA). RNA concentration was measured with the NanoDrop machine, followed by cDNA synthesis using High-Capacity cDNA Reverse Transcription Kit (ThermoFisher Scientific, Waltham, MA, USA). TaqMan gene expression assay was used to measure *TERC* relative abundance, whilst 18s rRNA was used as the housekeeping gene using TaqMan Fast Universal PCR Master Mix and ViiA 7 Real-Time PCR System (ThermoFisher Scientific, Waltham, MA, USA). The details of these two steps are stated in **Appendix E and F**.

The droplet digital polymerase chain reaction (ddPCR) method was used to measure the *TERC* level in the human plasma samples, using the *TERC* TaqMan assay (Hs03454202_s1). This technique utilised oil-based droplets of similar size to segregate the sample into single molecules per droplet which also contains the PCR reagents. The Bio-Rad 2x digital PCR Supermix for probes (cat. No. 1863024), Automated Droplet Generator, C1000 thermal cycler, and QX200 Droplet Reader (Bio-Rad, Hercules, USA) were used to perform the ddPCR. **Appendix G** explained the detailed protocol used for ddPCR.

2.2.2.1 Senescence panel

A new customised panel for senescence markers was developed to include the genes that have been reported in previous studies to be enriched in cellular senescence. The panel constituted the genes involved in cell cycle and division (cyclins, cyclin-dependent kinases (CDKs), and CDK inhibitors (CDKi)), telomere biology (TERF1 and TERF2), and DNA damage response factors (TP53BP1, ATM, ATR, CHEK1, CHEK2, TP53, and RB family), as well as secretory markers and their regulation (NFKB1, CEBPB, and GATA4). **Table 2.1** provides the list of 50 markers used to study the senescence in hIPC samples. The detailed OpenArray custom panel protocol is available in **Appendix H**. The analysis pipeline for calculating the correlation and regression is shown as a flowchart in **Figure 2.3**.

Table 2.1 Markers of cellular senescence

Gene name	Gene ID	Protein name
Cyclins:		
CCNA2	ENSG00000145386	Cyclin A2
CCNB1	ENSG00000134057	Cyclin B1
CCND1	ENSG00000110092	Cyclin D1
CCND2	ENSG00000118971	Cyclin D2
CCND3	ENSG00000112576	Cyclin D3
CCNE1	ENSG00000105173	Cyclin E1
CCNE2	ENSG00000175305	Cyclin E2
CDK:		
CDK1	ENSG00000170312	Cdk1
CDK2	ENSG00000123374	Cdk2
CDK4	ENSG00000135446	Cdk4
CDK6	ENSG00000105810	Cdk6
CDKi:		
CDKN1A	ENSG00000124762	p21
CDKN2A	ENSG00000147889	p16
CDKN2B	ENSG00000147883	p15
CDKN1B	ENSG00000111276	p27
CDKN1C	ENSG00000129757	p57
CDKN2C	ENSG00000123080	p18
CDKN2D	ENSG00000129355	p19

Gene name	Gene ID	Protein name
DDR:		
ATM	ENSG00000149311	ATM
ATR	ENSG00000175054	ATR
CHEK1	ENSG00000149554	CHK1
CHEK2	ENSG00000183765	CHK2
CDC25C	ENSG00000158402	Cdc25c
TP53	ENSG00000141510	p53
TP53BP1	ENSG00000067369	53BP1
MDM2	ENSG00000135679	MDM2
RB1	ENSG00000139687	pRb
RBL1	ENSG00000080839	p107
RBL2	ENSG00000103479	p130
E2F1	ENSG00000101412	E2F1
E2F3	ENSG00000112242	E2F3
MAPK14	ENSG00000112062	p38MAPK
Telomeres-related:		
TERF1	ENSG00000147601	TRF1
TERF2	ENSG00000132604	TRF2
Classic markers:		
LMNB1	ENSG00000113368	Lamin B1
GLB1	ENSG00000170266	Glb1
PCNA	ENSG00000132646	PCNA
PLAU	ENSG00000122861	PLAU
GSK3B	ENSG00000082701	GSK3B
BCL2L2	ENSG00000129473	Bcl-W
SIRT1	ENSG00000096717	Sirtuin 1
SASP regulation:		
NFKB1	ENSG00000109320	NF-κB
CEBPB	ENSG00000172216	C/EBPβ
GATA4	ENSG00000136574	GATA4
SASP factors:		
IGFBP3	ENSG00000146674	IGFBP3
IL6	ENSG00000136244	IL-6
CXCL8	ENSG00000169429	IL-8
MMP1	ENSG00000196611	Matrix metallopeptidase 1
SERPINE1	ENSG00000106366	Serpin E1

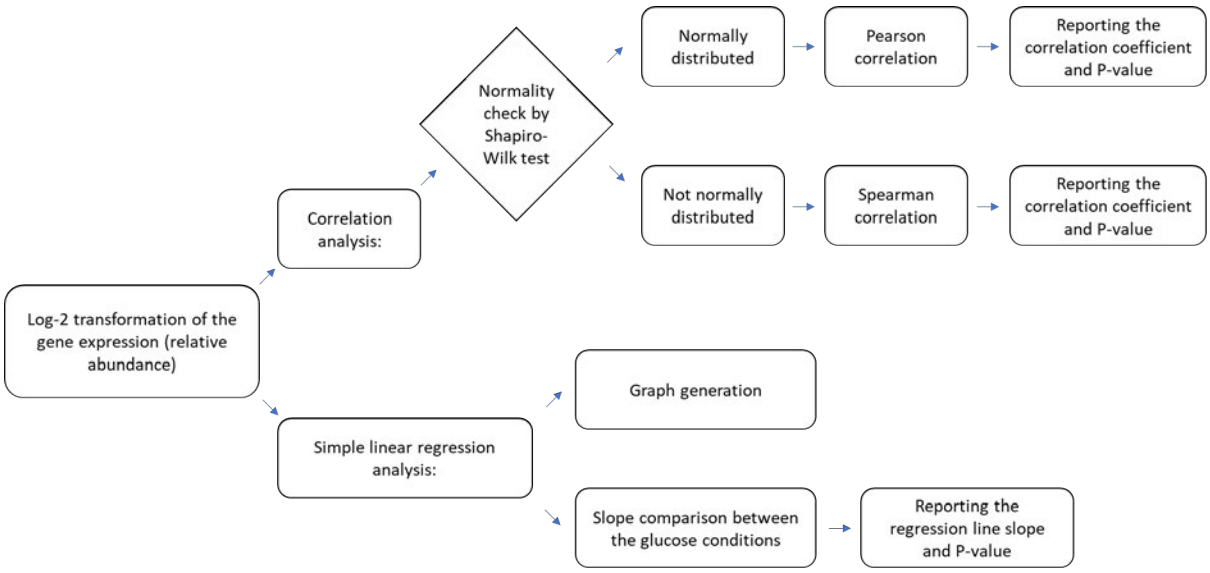


Figure 2.3 Flowchart showing the steps for analysing the senescence marker data

2.3 Results

2.3.1 Phase 1 (no glucose exposure)

2.3.1.1 Cell growth

Four hIPC samples from cadaveric islet donors (**Table 2.2**) were sub-cultured till senescence. The cell samples reached a senescence state after sub-culturing for 55 to 83 days, as shown in **Figure 2.4**.

Table 2.2 Demographic data of the four islet donors

hIPC sample	Age (Y)	Sex	BMI
A	48	F	23.4
B	45	M	23.6
C	47	M	33.2
E	35	M	33.2

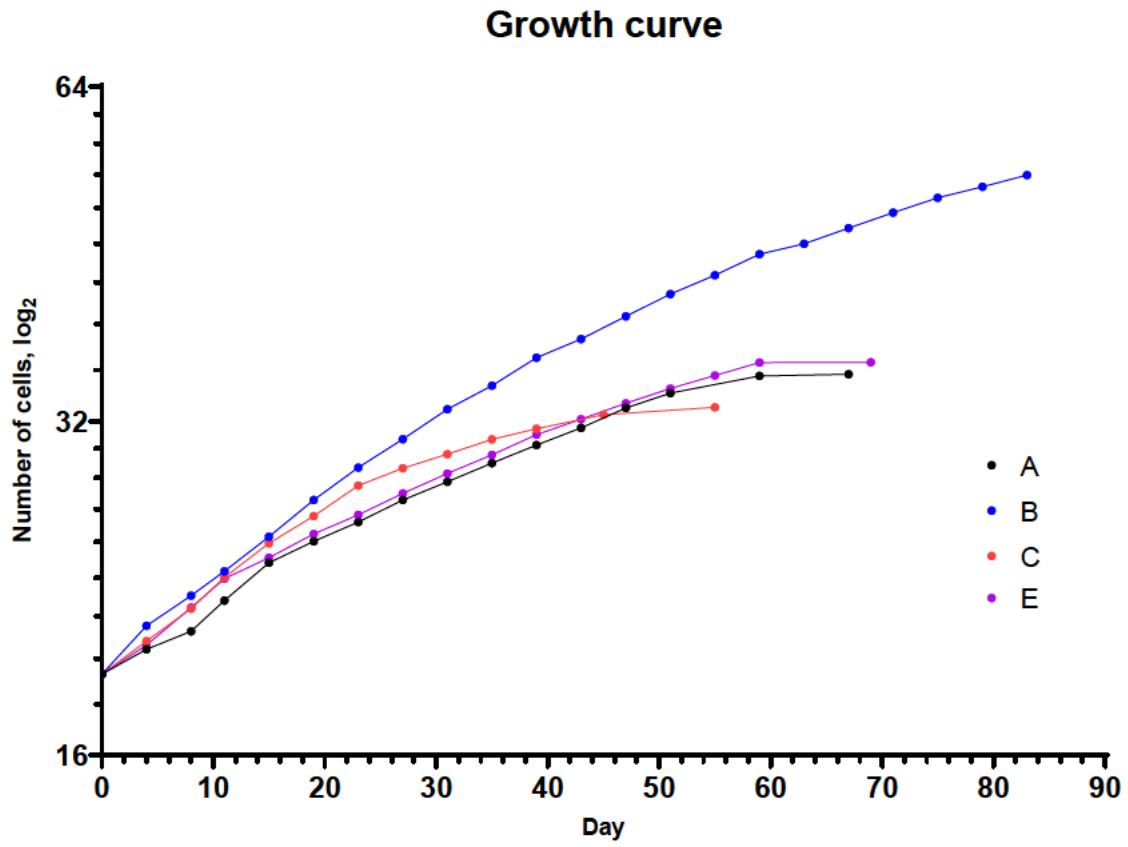


Figure 2.4 Growth curve of four hIPC samples

The four hIPC samples (shown in four colours) reached senescence at different time points.

2.3.1.2 Telomere length, *TERC* level, and telomerase activity

The relative telomere length (rTL) was measured for each hIPC sample collected at the end of every cell passage. The rTL was calculated in two steps as explained previously[257]. First, for each timepoint, the qPCR Ct value of the telomeres was normalised by subtracting the Ct value of *hBG* (single-copy gene used in the experiment)[257]. This normalisation was performed to calculate the difference between single-copy gene and telomere Ct values (ΔCt)[257]. Next, for each hIPC sample, the ΔCt of each timepoint was normalised by subtracting the ΔCt value of the last timepoint, assuming that the last time point would have the shortest telomere

because telomeres shorten gradually in every cell division[61,63,73]. The resulting ΔΔCt was the relative telomere length (rTL).

Figure 2.5A demonstrates the rTL of four hIPC samples over time (days). There was a significant negative correlation between the rTL and time (days) (Pearson r: -0.75, P<0.0001), which was also reflected in the regression line (slope: -0.016, P<0.0001). The *TERC* relative abundance was also calculated per each timepoint per hIPC sample, as shown in **Figure 2.5B**. Significantly increased expression level of *TERC* was identified through time (Pearson r: 0.62, P = 0.0012), also reflected in the regression line (slope: 0.011, P = 0.0018). Additionally, comparing the slope of regression line of each individual sample showed no significant difference for rTL (0.23); but the slopes were significantly different regarding *TERC* relative abundance (P = 0.0004).

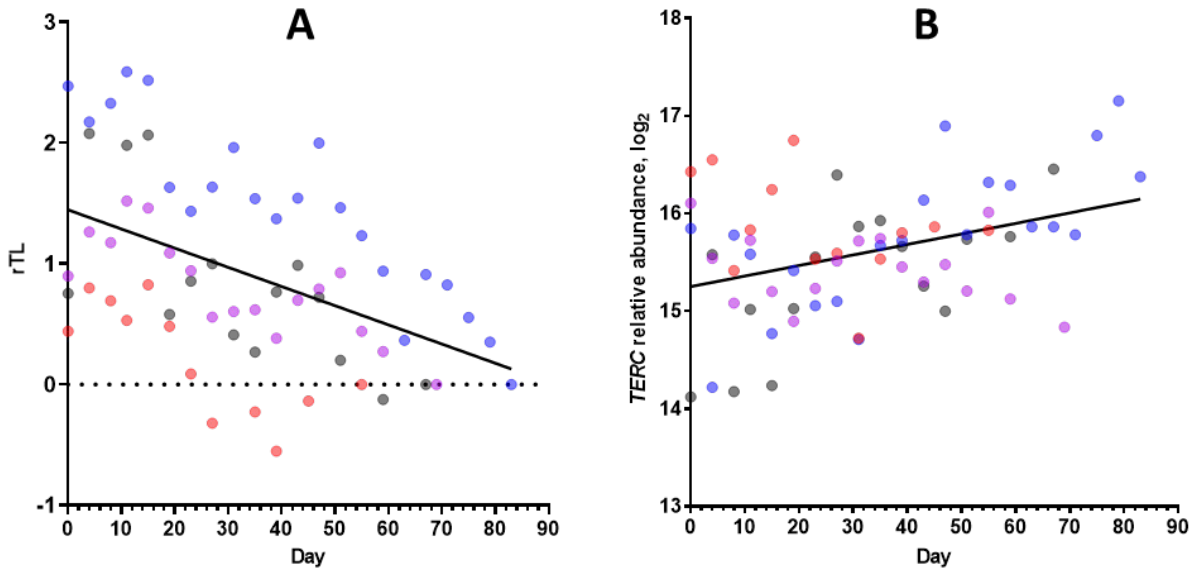


Figure 2.5 The rTL and *TERC* level through time

The four hIPC samples are shown in four colours. **A)** The association between rTL and subculturing days of hIPC samples shows a significant telomere shortening through time (66 data points, slope: -0.016, Pearson r: -0.75, P<0.0001). **B)** The association between *TERC* relative abundance and subculturing days of hIPC samples shows a significant increase in *TERC* level through time (68 data points, slope: 0.011, Pearson r: 0.62, P = 0.0012).

Also, the correlation between *TERC* relative abundance with rTL was measured to identify the association between these two factors related to telomere biology (**Figure 2.6**). A significant negative correlation was observed (Pearson r: -0.28, P = 0.0212) indicating higher *TERC* increased expression is associated with shorter telomere length.

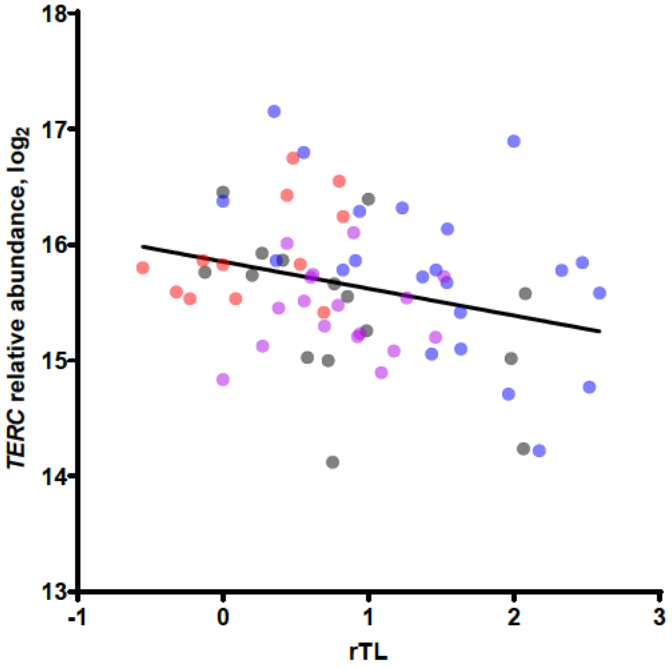


Figure 2.6 Correlation between *TERC* level and rTL

The graph shows the negative correlation between *TERC* relative abundance and rTL in hIPC samples (66 data points, slope: -0.23, Pearson r: -0.28, P = 0.0212), indicating *TERC* level increase is associated with telomere shortening. The four samples are shown in four colours.

Telomerase activity measurement was impossible in hIPC cells. The samples were sent to Professor Roger Reddel's laboratory at Children's Medical Research Institute, Westmead, University of Sydney, NSW, Australia. Professor Reddel informed us they were unable to detect the telomerase in hIPC samples, as this enzyme is mostly only detectable in stem cells, but not in differentiated cells (email communications between Prof. Reddel and Prof. Hardikar).

2.3.2 Phase 2 (glucose exposure)

2.3.2.1 Cell growth

Four hIPC samples were cultured in three different glucose conditions in parallel, as explained in the Methods section. Three of these samples were in common with Phase 1 of the study (samples B, C, and E), and another sample was from a male cadaveric islet donor of 57 years of age and a BMI of 27.3. **Figure 2.7** shows the growth curves of the four hIPC samples exposed to different glucose conditions. Similar to Phase 1, the cells were passaged until they stopped dividing. Surprisingly, the cells in the NG condition reached senescence a few passages earlier than the other two conditions.

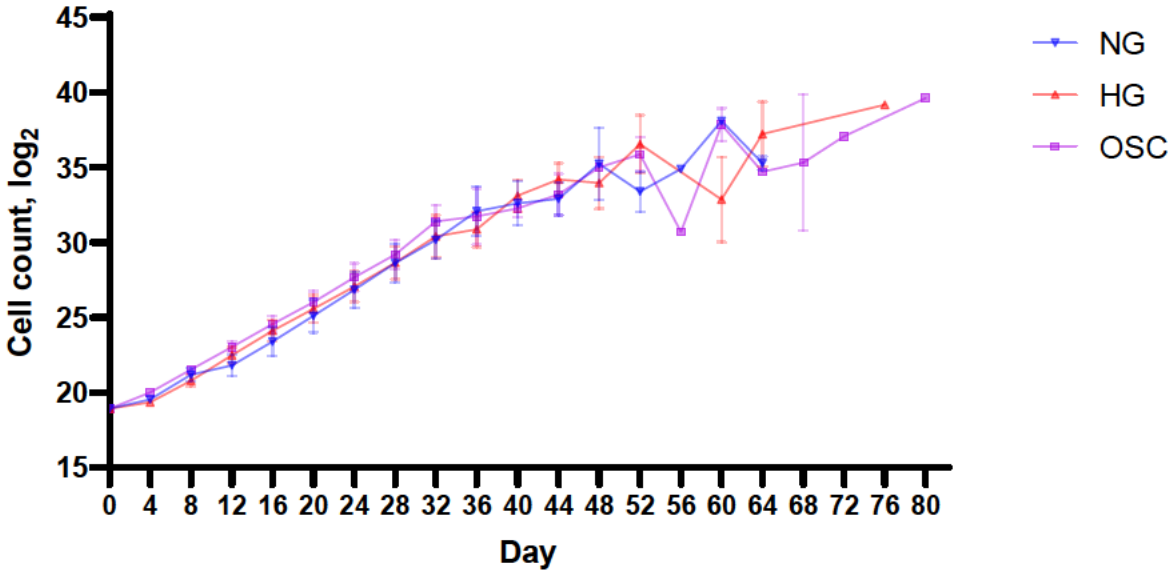


Figure 2.7 Growth curve of hIPC samples exposed to three glucose conditions

The growth of cells in NG, HG, and OSC glucose-exposure conditions till senescence. Four hIPC samples were included per each condition (mean±SEM). Since the cells were proliferating at different rates in the later stage (Day 56-80), the dots without error bars represent the cell count of one sample, not the mean. The senescence state of samples ranges from 52-64 days in NG, 60-76 days in HG, and 64-80 days in OSC condition. NG: normal glucose, HG: high glucose, OSC: oscillating condition

2.3.2.2 Telomere length and *TERC* level

A comparable telomere shortening rate was observed across different glucose conditions. In each condition, there was a significant correlation between rTL and time (days) (Pearson r: NG: -0.83, HG: -0.844, OSC: -0.9; P <0.0001 in all three conditions) (**Figure 2.8**). The slope of regression lines was also similar between the conditions (NG: -0.02; HG: -0.018; OSC: -0.021; P <0.0001 in all three conditions). No significant difference between the slopes was identified (P=0.9).

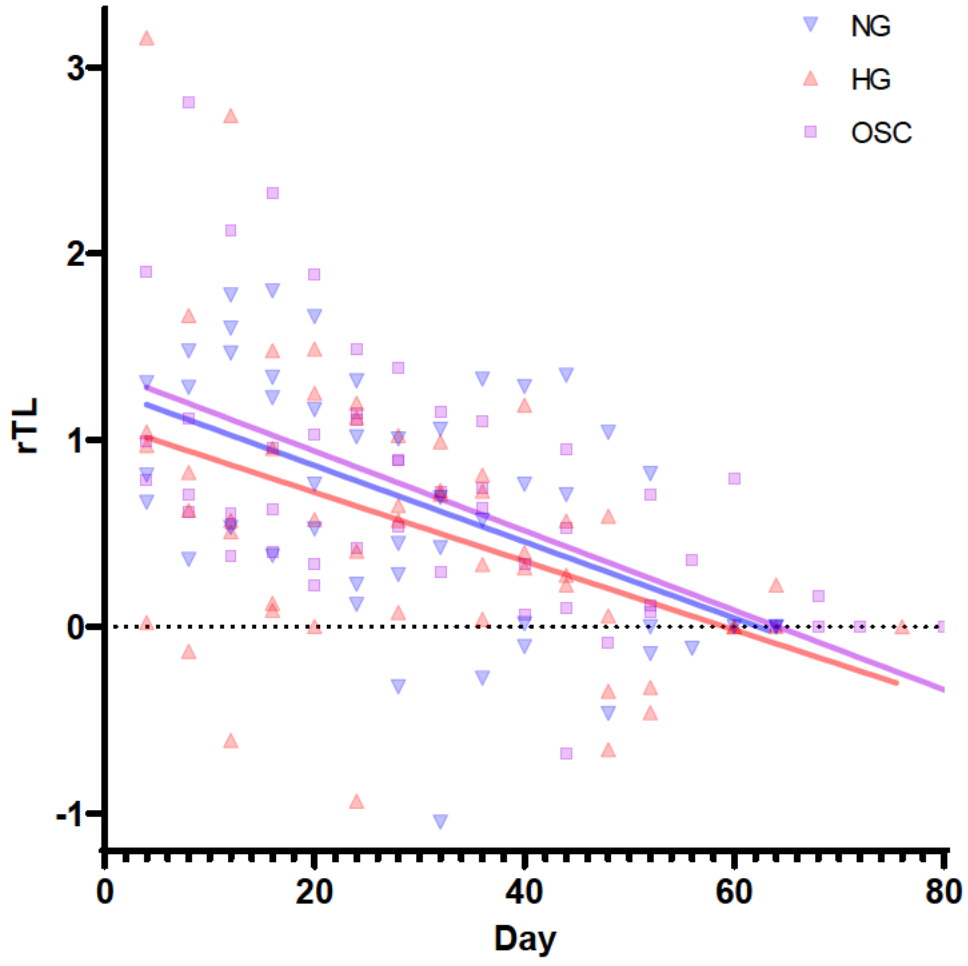


Figure 2.8 rTL through time, in three glucose-exposure conditions

The association between rTL and subculturing days of hIPC samples shows a similar telomere shortening trend through time between the three glucose-exposure conditions, although being significant in each condition (normal glucose condition: 51 data points, slope: -0.019,

Pearson r: -0.88; high glucose condition: 48 data points, slope: -0.022, Pearson r: -0.84; oscillating normal and high glucose condition: 51 data points, slope: -0.018, Pearson r: -0.85; P <0.0001 in all three conditions). The three colours indicate the three glucose conditions, and the data of four hIPC samples per condition is shown. NG: normal glucose, HG: high glucose, OSC: oscillating condition

However, the changes in *TERC* expression were not similar between the three glucose conditions (**Figure 2.9**). The *TERC* level increased significantly during the cell passages in NG (Pearson r: 0.53, P = 0.0269), but not in HG (Pearson r: 0.037, P = 0.89), and OSC (Pearson r: 0.16, P = 0.51). Similarly, the regression line slope was positive in NG (0.0085, P = 0.0073), but was almost flat in HG (0.0014, P = 0.7), and OSC (0.0033, P = 0.75). There was no significant difference between the slopes of the regression line in NG vs HG (P=0.14), but the slopes of NG vs OSC were significantly different (P = 0.039).

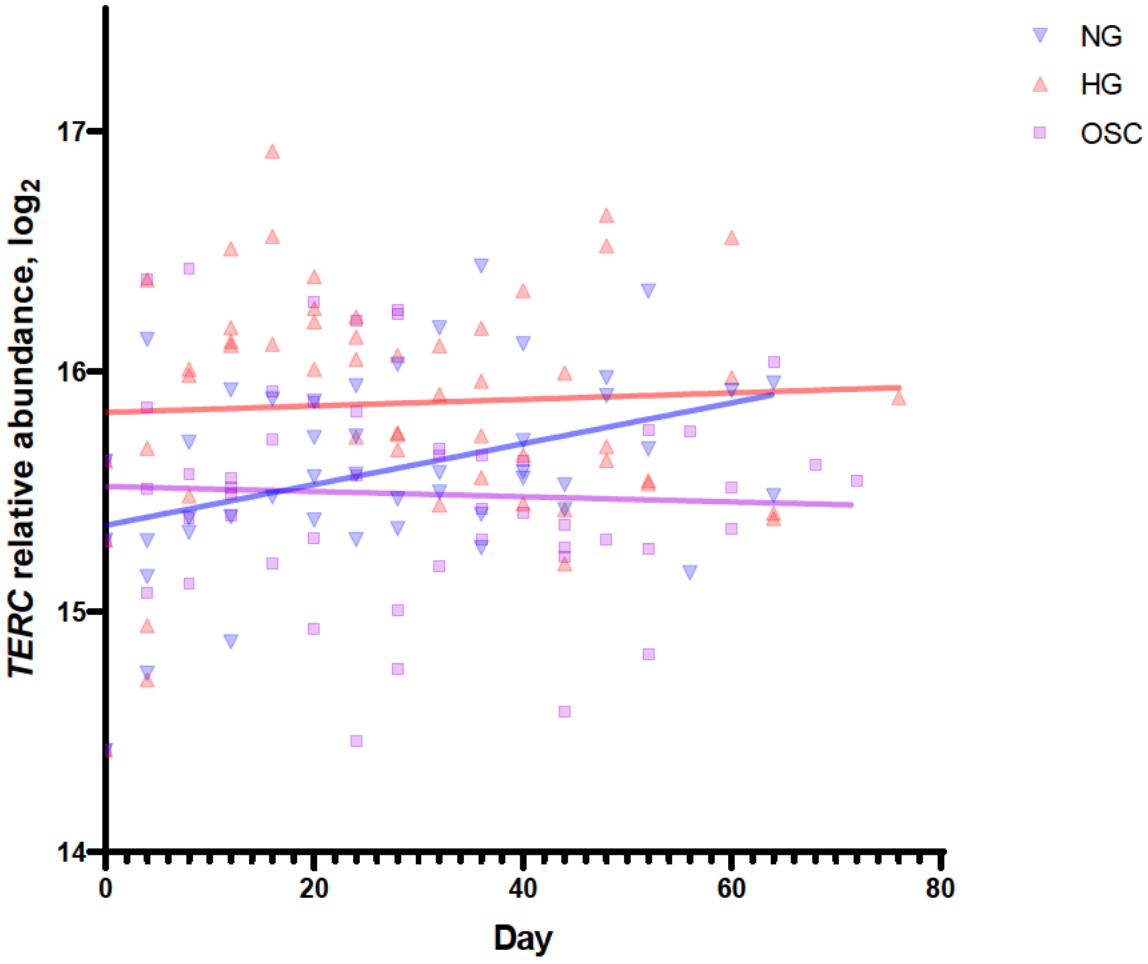


Figure 2.9 *TERC* level through time, in three glucose-exposure conditions

The association between *TERC* relative abundance and hIPC samples subculturing days revealed a significant increase of *TERC* level in NG (46 data points, slope: 0.0085, Pearson r: 0.53, P = 0.0269), but not in HG (53 data points, slope: 0.0014, Pearson r: 0.037, P = 0.89), and OSC condition (52 data points, slope: -0.001, Pearson r: 0.16, P = 0.51). All the data points of four hIPC samples per condition are shown. NG: normal glucose, HG: high glucose, OSC: oscillating condition

2.3.2.3 Senescence markers

The trend of relative gene expression level of several senescence markers was assessed in hIPC samples exposed to the three glucose conditions (NG, HG, and OSC). The results of all the correlation and trend analyses are summarised in **Table 2.3**.

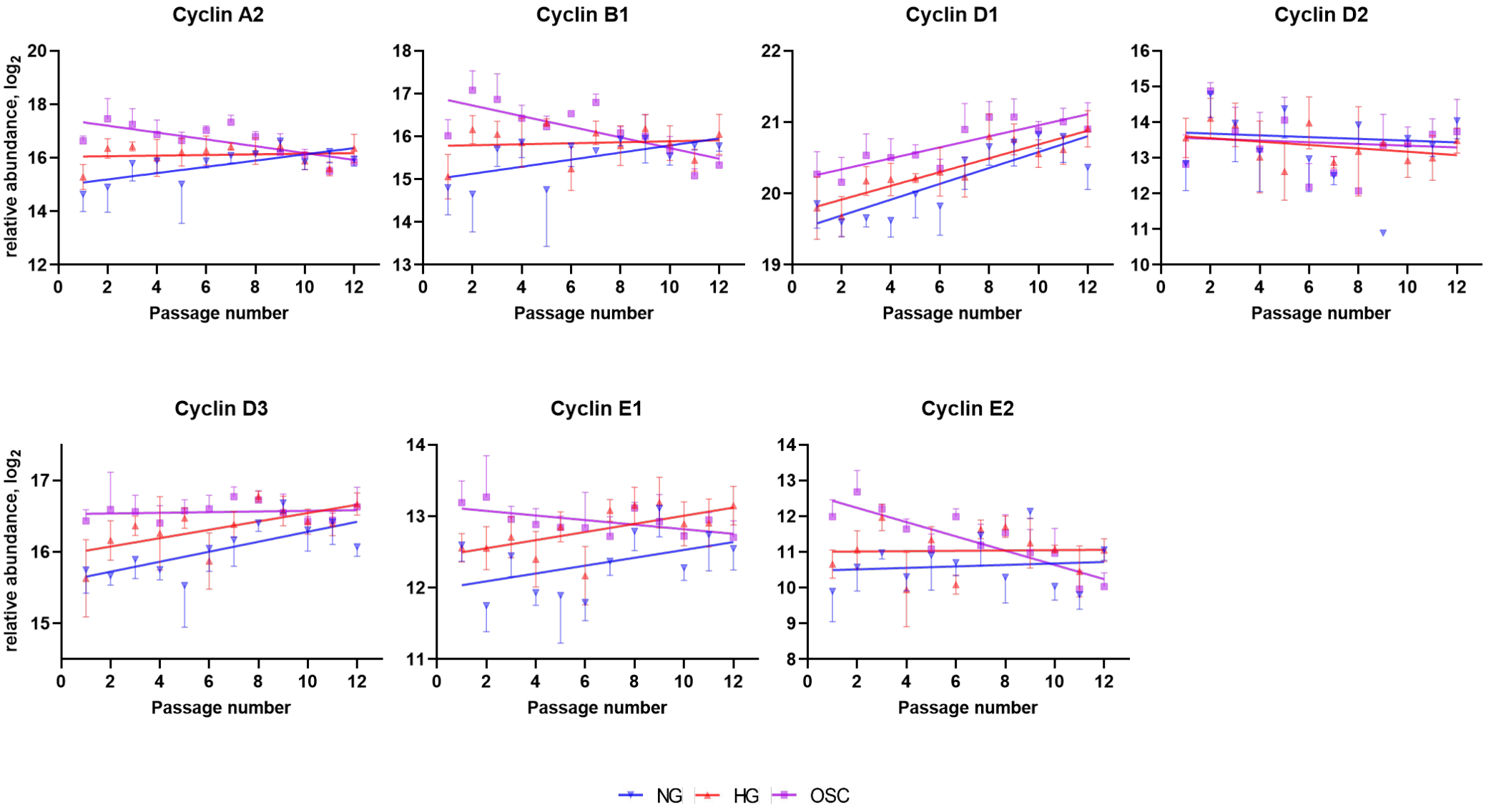
Table 2.3 Summary of the correlation and trend analysis of the senescence markers

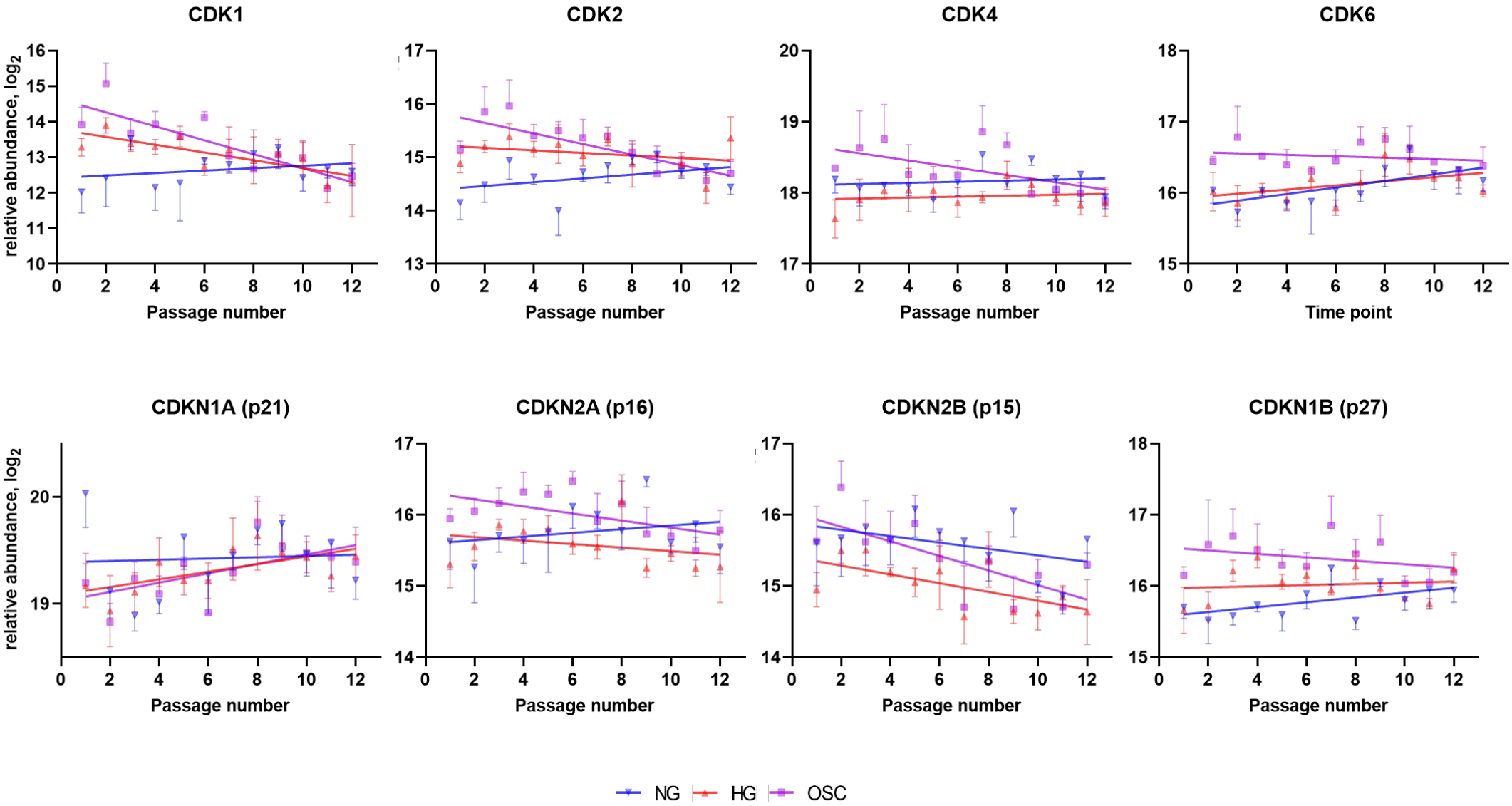
Pearson or Spearman correlation test was performed for parametric or nonparametric data, respectively. Significant P-values are shown in red. Slope equality p-value denotes if there was a significant difference between the slopes of regression lines between the three glucose conditions. NG: normal glucose, HG: high glucose, OSC: oscillating, P: p-value

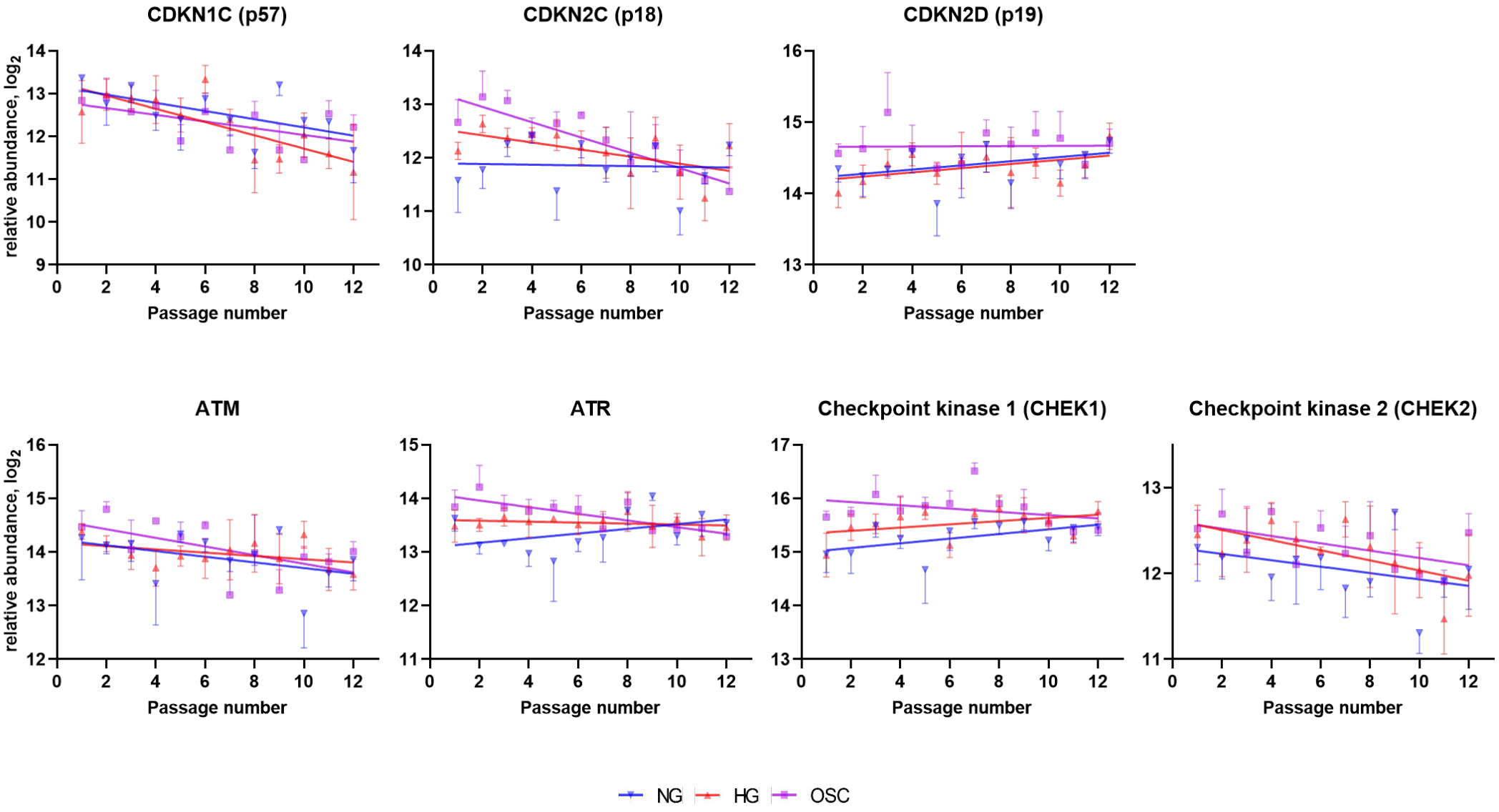
		Correlation							Linear regression			
Gene	**Protein**	**Test**	**r NG**	**r HG**	**r OSC**	**P NG**	**P HG**	**P OSC**	**Slope NG**	**Slope HG**	**Slope OSC**	**Slope equality P**
CCNA2	Cyclin A2	Spearman	0.7483	0.01399	-0.6853	**0.007**	0.9739	**0.017**	0.117	0.01126	-0.1277	**<0.0001**
CCNB1	Cyclin B1	Spearman	0.5175	-0.03497	-0.7343	0.0888	0.921	**0.0087**	0.08257	0.01163	-0.1244	**0.0002**
CCND1	Cyclin D1	Pearson	0.8452	0.9079	0.8506	**0.0005**	**<0.0001**	**0.0005**	0.1111	0.0966	0.07719	0.55
CCND2	Cyclin D2	Pearson	-0.1966	-0.3991	-0.1321	0.5404	0.1988	0.6823	-0.02466	-0.0473	-0.02415	0.96
CCND3	Cyclin D3	Pearson	0.7372	0.6634	0.1373	**0.0062**	**0.0187**	0.6706	0.06985	0.05857	0.004663	0.079
CCNE1	Cyclin E1	Pearson	0.4732	0.6424	-0.6136	0.1202	**0.0243**	**0.0338**	0.05491	0.05689	-0.03207	**0.0198**
CCNE2	Cyclin E2	Pearson	0.133	0.03706	-0.8768	0.6802	0.909	**0.0002**	0.02074	0.004773	-0.1995	**0.0007**
CDK1	Cdk1	Pearson	0.2529	-0.8035	-0.8503	0.4277	**0.0016**	**0.0005**	0.0341	-0.1096	-0.196	**0.0005**
CDK2	Cdk2	Pearson	0.3866	-0.3067	-0.8119	0.2145	0.3322	**0.0013**	0.03545	-0.02368	-0.09913	**0.0003**
CDK4	Cdk4	Pearson	0.1842	0.1686	-0.5817	0.5667	0.6003	**0.0472**	0.007767	0.00662	-0.05103	**0.0465**
CDK6	Cdk6	Pearson	0.6864	0.4854	-0.2107	**0.0137**	0.1097	0.511	0.04617	0.02919	-0.01009	0.064
CDKN1A	p21	Pearson	0.1217	0.6507	0.6282	0.7063	**0.0219**	**0.0287**	0.005911	0.03568	0.04385	0.37
CDKN2A	p16	Pearson	0.3281	-0.3175	-0.6013	0.2977	0.3146	**0.0386**	0.02592	-0.02453	-0.04961	0.076
CDKN2B	p15	Pearson	-0.4327	-0.6473	-0.7118	0.16	**0.0229**	**0.0094**	-0.04487	-0.06168	-0.1023	0.37
CDKN1B	p27	Pearson	0.5417	0.1143	-0.3481	0.0689	0.7236	0.2675	0.034	0.008184	-0.02421	0.123
CDKN1C	p57	Pearson	-0.6221	-0.798	-0.5688	**0.0308**	**0.0019**	0.0536	-0.09589	-0.1548	-0.0783	0.364
CDKN2C	p18	Pearson	-0.06155	-0.5963	-0.9026	0.8493	**0.0407**	**<0.0001**	-0.00661	-0.06634	-0.143	**0.013**
CDKN2D	p19	Pearson	0.4208	0.5014	-0.0227	0.1731	0.0968	0.9442	0.02945	0.02952	0.001319	0.57
ATM	ATM	Pearson	-0.4135	-0.4368	-0.6034	0.1816	0.1556	**0.0378**	-0.05289	-0.03052	-0.08058	0.56
ATR	ATR	Pearson	0.4729	-0.2711	-0.801	0.1205	0.3941	**0.0017**	0.0435	-0.00884	-0.06219	**0.0041**
CHEK1	CHK1	Spearman	0.4336	0.4406	-0.3636	0.1616	0.1542	0.2464	0.04341	0.03009	-0.03034	**0.0451**
CHEK2	CHK2	Pearson	-0.3744	-0.6705	-0.5515	0.2305	**0.017**	0.063	-0.03727	-0.05912	-0.04272	0.82
CDC25C	Cdc25c	Pearson	-0.1133	0.116	-0.7537	0.7258	0.7195	**0.0046**	-0.02066	0.009492	-0.1231	**0.0286**
TP53	p53	Pearson	0.05267	0.4179	-0.09426	0.8709	0.1765	0.7708	-0.00046	0.03785	-0.00793	0.31
TP53BP1	53BP1	Pearson	0.004474	0.1452	-0.6996	0.989	0.6524	**0.0113**	-0.00256	0.00687	-0.04254	0.162

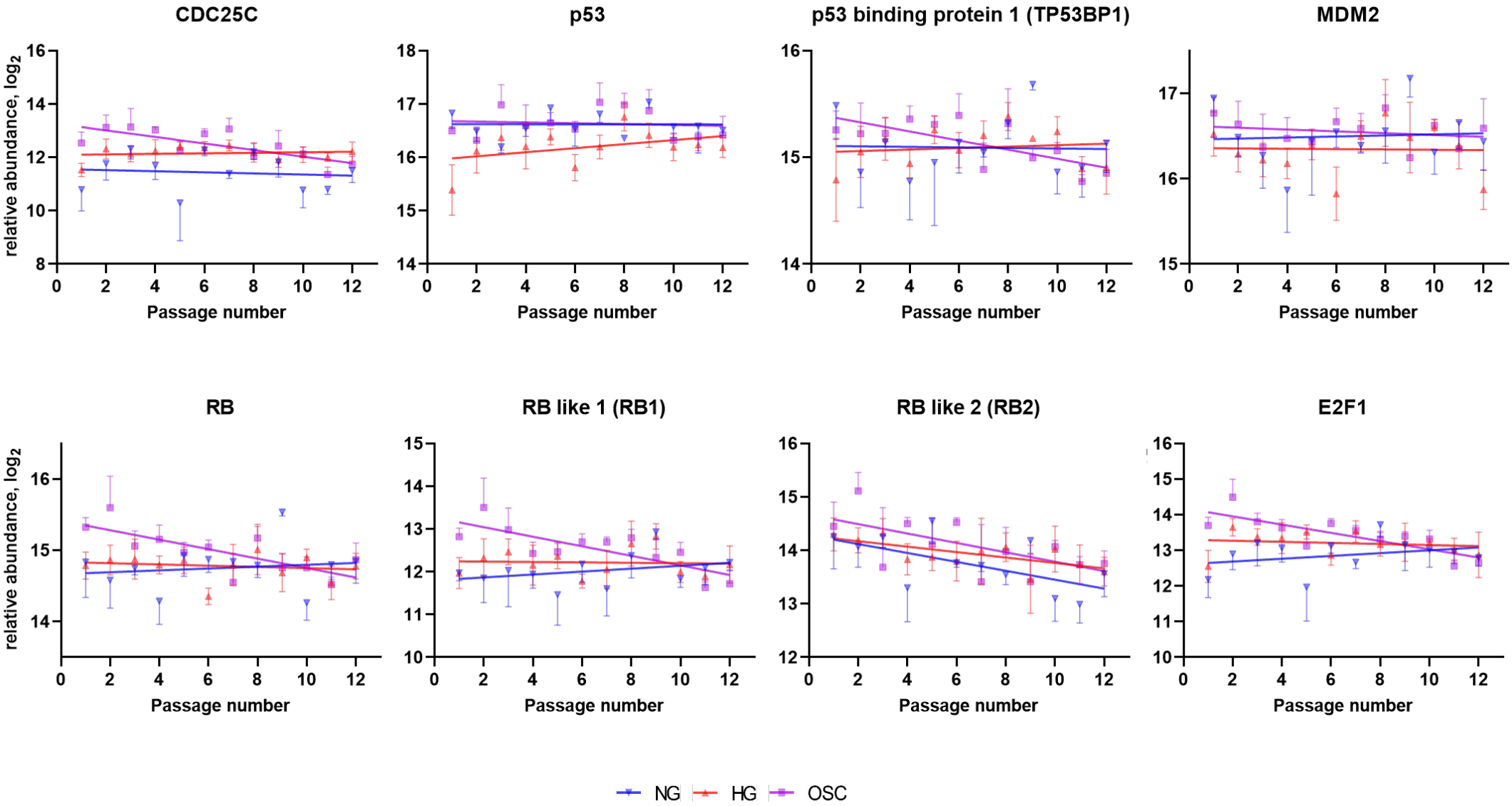
		Correlation							Linear regression			
Gene	**Protein**	**Test**	**r NG**	**r HG**	**r OSC**	**P NG**	**P HG**	**P OSC**	**Slope NG**	**Slope HG**	**Slope OSC**	**Slope equality P**
MDM2	MDM2	Pearson	0.1127	-0.01125	-0.1995	0.7273	0.9723	0.5341	0.006161	-0.00196	-0.01068	0.88
RB1	pRb	Pearson	0.1731	-0.1766	-0.75	0.5906	0.583	**0.005**	0.0131	-0.00863	-0.06634	0.223
RBL1	p107	Pearson	0.3542	-0.01814	-0.7907	0.2586	0.9554	**0.0022**	0.03414	-0.00469	-0.1118	**0.0035**
RBL2	p130	Spearman	-0.6294	-0.6923	-0.5035	**0.0323**	**0.0155**	0.0989	-0.08344	-0.05079	-0.08798	0.67
E2F1	E2F1	Pearson	0.3033	-0.1627	-0.8003	0.338	0.6134	**0.0018**	0.03956	-0.01503	-0.1156	**0.0029**
E2F3	E2F3	Pearson	0.6306	0.3126	-0.4278	**0.0279**	0.3225	0.1654	0.05396	0.01983	-0.0221	**0.0259**
MAPK14	p38MAPK	Spearman	0.5315	0.08392	-0.1608	0.0794	0.8004	0.6192	0.03058	0.003823	-0.00754	0.147
LMNB1	Lamin B1	Pearson	0.4936	-0.1117	-0.82	0.1029	0.7297	**0.0011**	0.07059	-0.01648	-0.1676	**<0.0001**
GLB1	Glb1	Pearson	0.2175	-0.03244	-0.4799	0.4971	0.9203	0.1143	0.01158	-0.00077	-0.02452	0.346
PCNA	PCNA	Spearman	0.3147	-0.1329	-0.8182	0.3194	0.6832	**0.0019**	0.04055	-0.00739	-0.1121	**0.0006**
PLAU	PLAU	Pearson	0.5916	0.5449	0.582	**0.0427**	0.0669	**0.0471**	0.1417	0.07983	0.1368	0.73
GSK3B	GSK3B	Pearson	0.03481	0.2032	-0.7098	0.9145	0.5264	**0.0097**	-0.00159	0.00953	-0.02479	0.26
BCL2L2	Bcl-W	Pearson	0.2644	0.6712	0.1824	0.4063	**0.0169**	0.5705	0.01374	0.03663	0.006611	0.33
SIRT1	Sirtuin 1	Pearson	0.05706	-0.7167	-0.7731	0.8602	**0.0087**	**0.0032**	0.003279	-0.0613	-0.1161	**0.0058**
NFKB1	NF-κB	Pearson	0.3277	-0.02114	-0.5082	0.2984	0.948	0.0916	0.02653	-0.00169	-0.02053	0.319
CEBPB	C/EBPβ	Pearson	-0.00109	-0.6397	-0.6032	0.9973	**0.0251**	**0.0379**	-0.00091	-0.05768	-0.06017	0.3
GATA4	GATA4	Pearson	0.7347	0.5795	0.6038	**0.0065**	**0.0483**	**0.0376**	0.13	0.1477	0.1022	0.91
IGFBP3	IGFBP3	Pearson	0.04646	0.3586	0.2912	0.886	0.2523	0.3585	0.006642	0.03302	0.02866	0.88
IL6	IL-6	Pearson	0.1181	-0.2258	0.1702	0.7147	0.4803	0.5968	0.009558	-0.01664	0.01326	0.84
CXCL8	IL-8	Pearson	0.1149	-0.09933	0.3258	0.7221	0.7587	0.3014	0.02388	-0.01729	0.1054	0.57
MMP1	Matrix metallopeptidase 1	Spearman	0.4755	0.2378	0.08392	0.1215	0.4573	0.8004	0.1466	0.05968	0.06534	0.82
SERPINE1	Serpin E1	Pearson	0.181	0.011	-0.2738	0.5735	0.9729	0.3892	0.01699	0.001341	-0.02	0.57
TERF1	TRF1	Pearson	-0.08733	0.14	-0.6239	0.7873	0.6644	**0.0301**	-0.01472	0.008347	-0.04426	0.45
TERF2	TRF2	Pearson	0.418	0.3821	-0.0806	0.1763	0.2203	0.8034	0.03183	0.0214	-0.00374	0.4

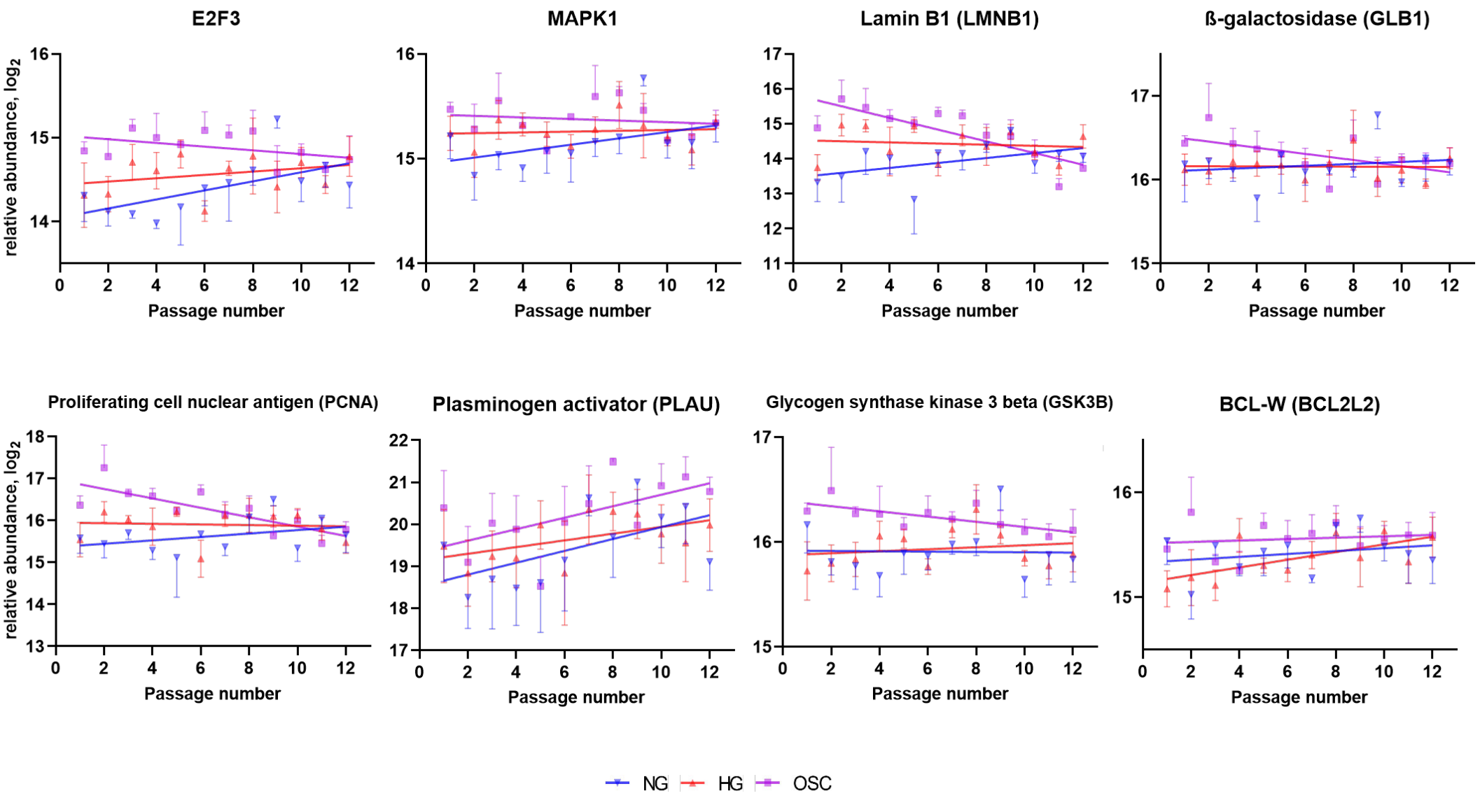
The changes in the gene transcript levels of senescence markers revealed different dynamics among the three glucose conditions (**Figure 2.10**). Interestingly, the temporal changes of several markers were different in the OSC condition compared to NG and HG. For instance, the expression level of cyclin A2, B1, E1, E2, CDK1, CDK2, CDK4, p16, p18, ATR, CDC25C, RB, RBL1, E2F1, LMNB1, PCNA, GSK3B, CEBPB, and SIRT1 was decreasing through time in OSC condition (which was significant for cyclin A2, B1, E1, CDK1, p16, p18, ATR, CDC25C, RBL1, LMNB1, PCNA, and SIRT1) but not in NG and HG (**Table 2.3**). In addition, the regression line showed a rising slope for cyclin D3 and CDK6, p27, and CHEK1 in both NG and HG conditions, however with a flat pattern in the OSC condition (**Figure 2.10**). In addition, the changes in gene expression levels of the senescence markers are shown as a heat map (**Figure 2.11**).

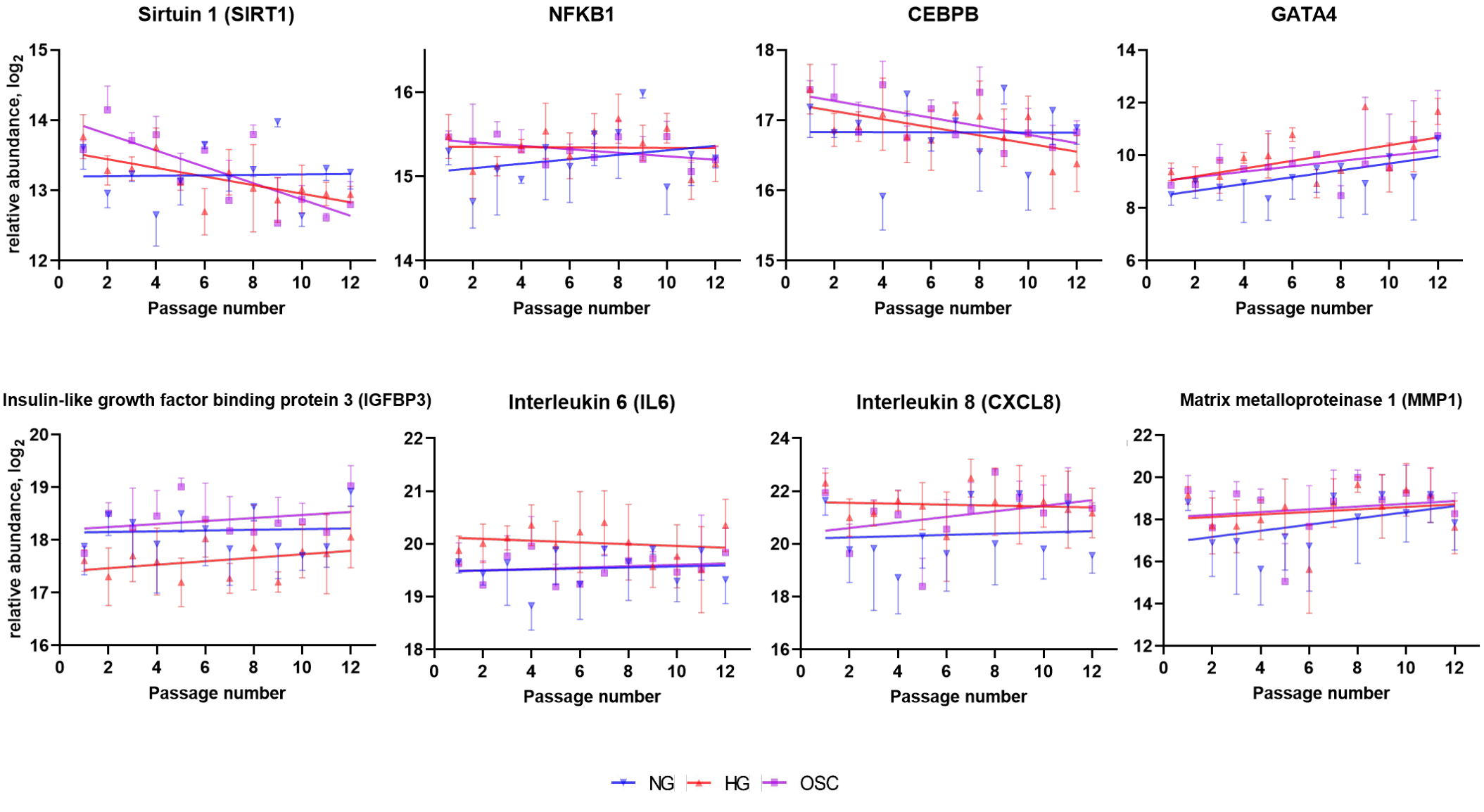

Cyclin A2
Cyclin B1
Cyclin D1
Cyclin D2
Cyclin D3
Cyclin E1
Cyclin E2
relative abundance, log2
Passage number
NG
HG
OSC


CDK1
CDK2
CDK4
CDK6
CDKN1A (p21)
CDKN2A (p16)
CDKN2B (p15)
CDKN1B (p27)
relative abundance, log2
Passage number
Time point
NG
HG
OSC


CDKN1C (p57)
CDKN2C (p18)
CDKN2D (p19)
ATM
ATR
Checkpoint kinase 1 (CHEK1)
Checkpoint kinase 2 (CHEK2)
relative abundance, log2
Passage number
NG
HG
OSC


CDC25C
p53
p53 binding protein 1 (TP53BP1)
MDM2
RB
RB like 1 (RB1)
RB like 2 (RB2)
E2F1
relative abundance, log2
Passage number
NG
HG
OSC


E2F3
MAPK1
Lamin B1 (LMNB1)
ß-galactosidase (GLB1)
Proliferating cell nuclear antigen (PCNA)
Plasminogen activator (PLAU)
Glycogen synthase kinase 3 beta (GSK3B)
BCL-W (BCL2L2)
relative abundance, log2
Passage number
NG
HG
OSC


Sirtuin 1 (SIRT1)
NFKB1
CEBPB
GATA4
Insulin-like growth factor binding protein 3 (IGFBP3)
Interleukin 6 (IL6)
Interleukin 8 (CXCL8)
Matrix metalloproteinase 1 (MMP1)
relative abundance, log2
Passage number
NG
HG
OSC

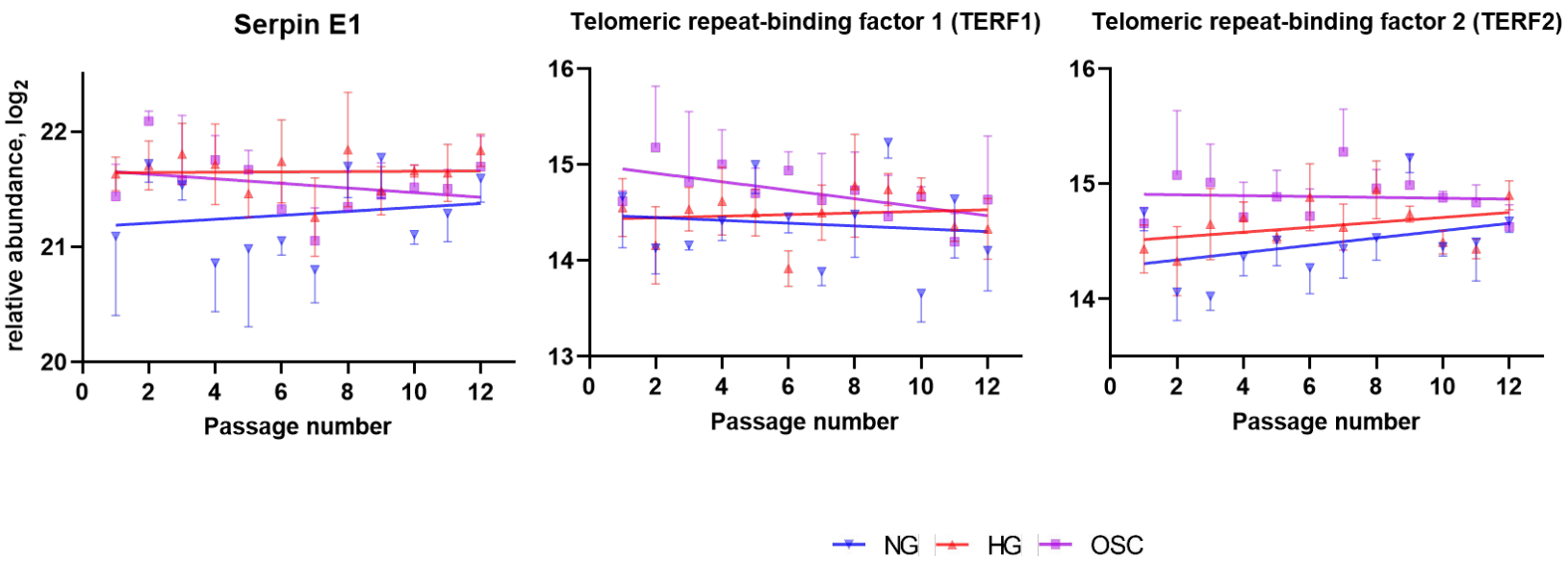


Figure 2.10 Trend lines showing the dynamic changes of senescence markers

The gene expression level of senescence markers in three glucose conditions (NG, HG, and OSC, shown with blue, red, and purple, respectively). There were four hIPC samples per condition, with 12 successive passages till senescence. The average period of days per cell passage was 4.92, 5.06, and 5.51 days in NG, HG, and OSC condition, respectively. The data points on the graph are shown as mean±SEM. NG, HG, and OSC condition are shown by blue, red, and purple colour, respectively.

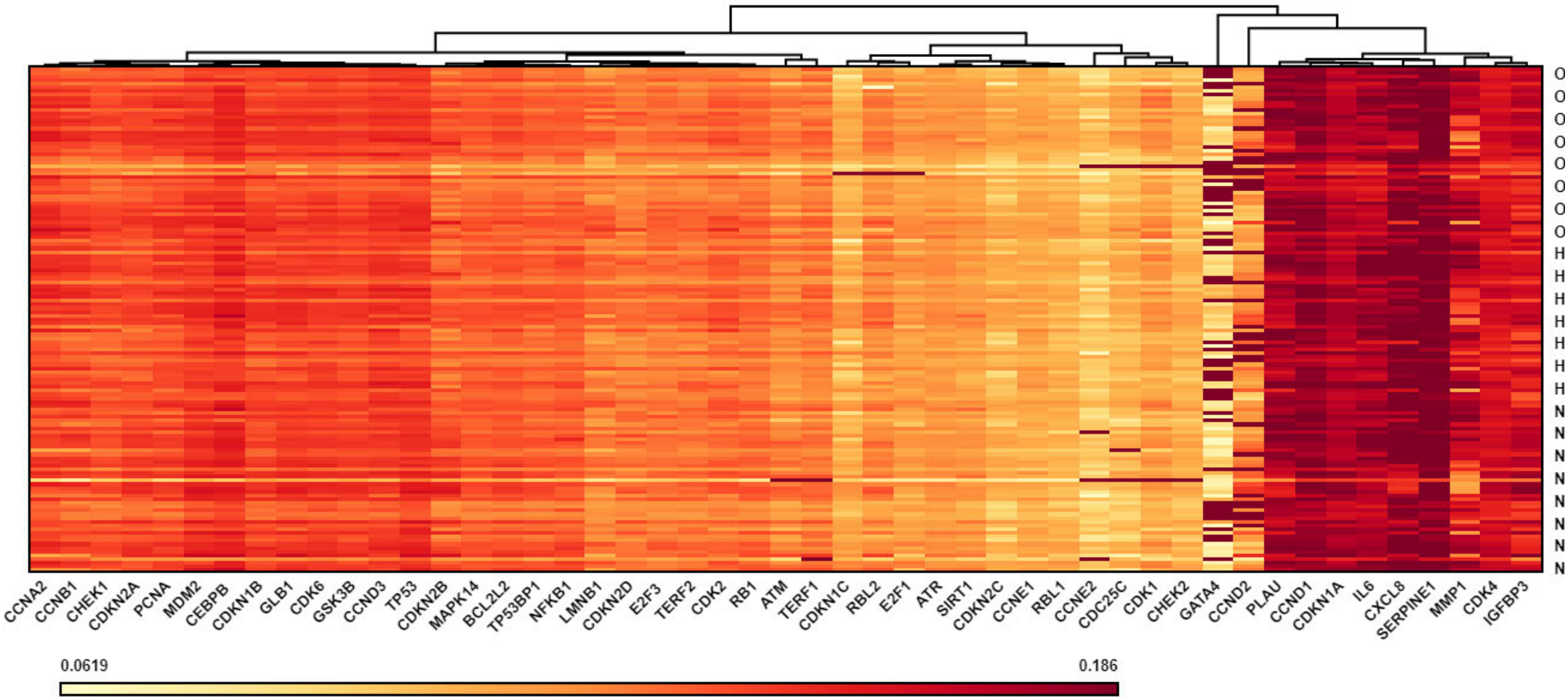


Figure 2.11 Heatmap showing the gene expression levels of the senescence markers

NG, HG, and OSC condition are shown by N, H, and O, respectively.

2.3.3 Quality control of the molecular analyses

2.3.3.1 Telomere length measurement qPCR reactions

The Ct value of one control DNA sample used throughout the qPCR reactions (Phase 1 and 2) for both hBG and telomeres is plotted in **Figure 2.12**. In addition, the inter-assay coefficient of variation (CV) for the Ct values of the control sample was calculated for all the nine qPCR reactions for hBG and telomeres[257]. The CV was 0.47% and 1.86% for hBG and telomeres, respectively, which were well below the threshold of 4% indicated in the guideline[257]. Furthermore, in each qPCR reaction, the intra-assay CV of the control sample for both hBG and telomeres was below the 2% threshold stated in the guideline[257].

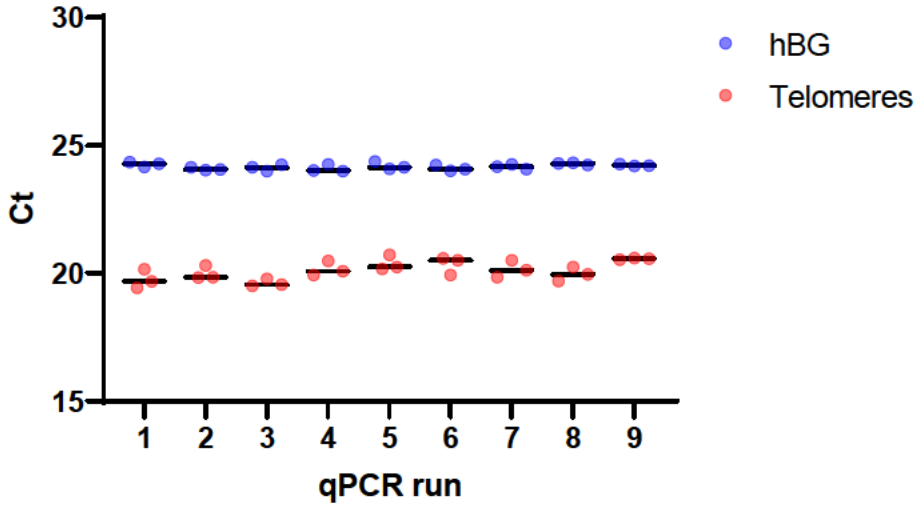


Figure 2.12 Quality control of the qPCR experiments for rTL measurement

The qPCR Ct values of one control sample used in nine qPCR runs for both hBG gene (blue) and telomeres (red) reactions in both phases 1 and 2 of the study. The qPCR runs were in triplicate, hence the tree dots per gene per run.

2.3.3.2 TERC level measurement qPCR reactions

The *18s* was used as the housekeeping gene in all the qPCR reactions. Therefore, the *18s* Ct values was checked for quality control of the qPCR reactions for *TERC* relative abundance

calculations. These values ranged mostly between 6 to 9 across the qPCR reactions for all hIPC samples in Phase 1 and 2 indicating similar expression patterns between the samples and high-quality qPCR runs as shown in **Figure 2.13**.

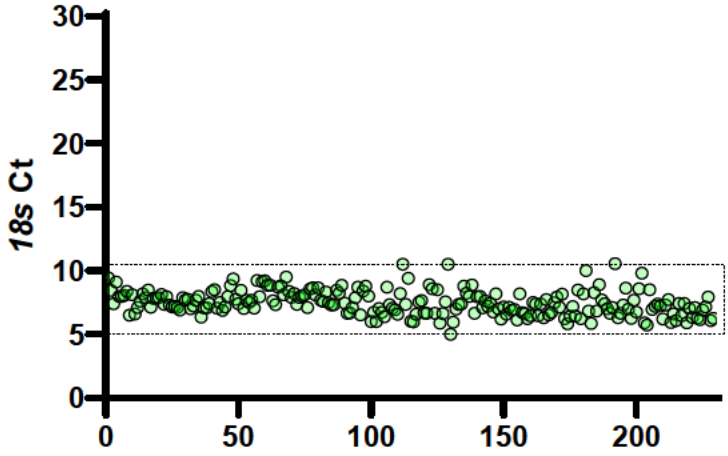


Figure 2.13 The qPCR Ct values of the *18s* rRNA housekeeping gene
The graph includes 229 data points across all hIPC samples in Phases 1 and 2.

2.3.3.3 Senescence panel

To check the quality of the reactions in the panel, two different housekeeping genes (*18s* and *GAPDH*) were included in our customised panel. This allowed us to compare the reaction of these two genes for each sample. **Figure 2.14** shows the scatter plot of the Ct value of both housekeeping genes for each sample. A very similar pattern was found for *18s* and *GAPDH* among the samples indicating the high quality of the panel and the experiment.

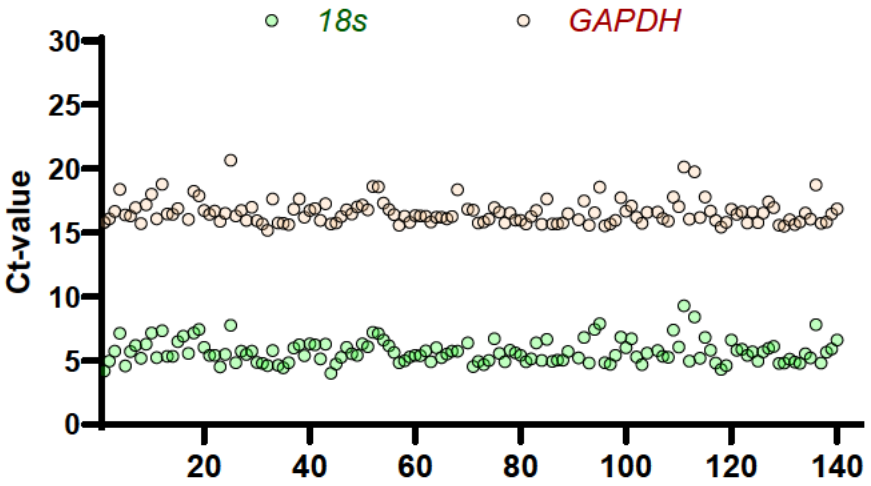


Figure 2.14 Scatter plot of the two housekeeping genes used in the panel

Green and orange dots represent the Ct value of the corresponding reaction for *18s* and *GAPDH* per each sample, respectively. The graph includes 144 data points per housekeeping gene.

As an additional quality control evaluation, we designed our customised panel to contain the same TaqMan assay for one of the genes (*CDK4*) in two separate wells. Therefore, the *CDK4* Ct values of the two wells sample must be close to identical for each sample. **Figure 2.15** shows the correlation of two Ct values for each sample (Spearman r = 0.98, P<0.0001).

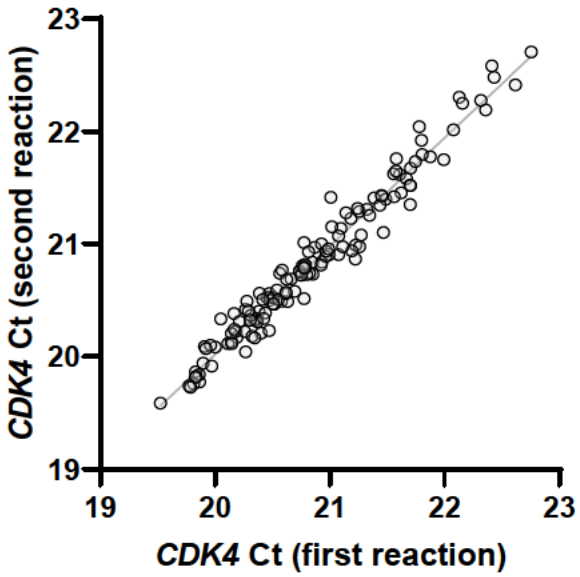


Figure 2.15 Correlation plot of the *CDK4* gene Ct values

2.3.4 Human samples

The *TERC* relative abundance was also measured in the plasma samples of individuals with T2D, compared to healthy controls (HC). Initially, real-time qPCR was used, however, because *TERC* plasma level was found to be below the technique sensitivity, we optimised the ddPCR reaction to achieve the results (**Figure 2.16**). The ddPCR reaction setup for measuring the *TERC* plasma level is stated in **Appendix G**. We found that *TERC* plasma level is significantly lower in individuals with T2D (median: 1.7 (T2D) vs 4.03 (HC) copies per 20ul ddPCR reaction, P = 0.0012).

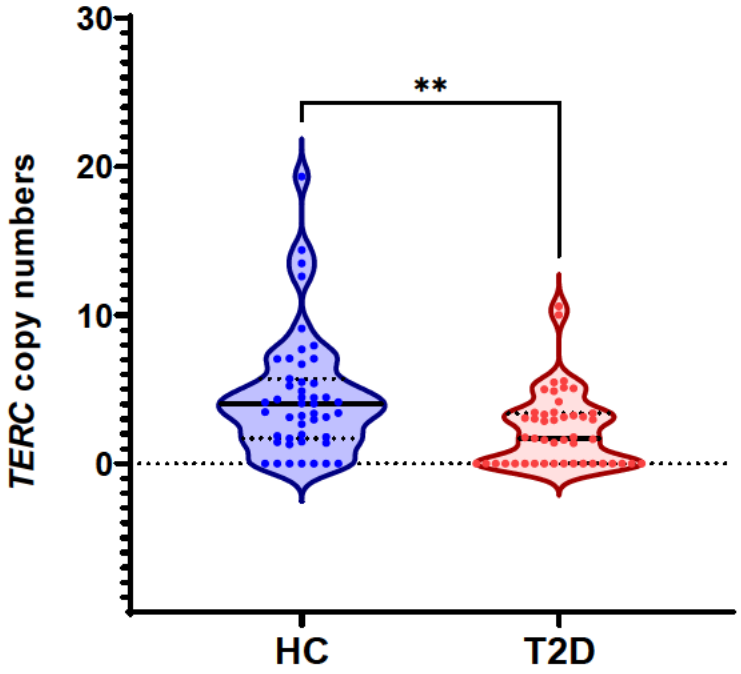


Figure 2.16 *TERC* plasma level in patients with T2D vs healthy controls

ddPCR results show significantly different *TERC* plasma levels between 47 patients with T2D and 47 healthy controls (HC) (P = 0.0012). The data is shown as copy number per 20µl reaction.

2.4 Discussion

We conducted an *in vitro* study to investigate telomere biology in hIPCs as a cell model for pancreatic islets and T2D progression. Apart from several studies on leukocyte telomere length in T2D[258-263], there were few studies on the telomere length in pancreas islet cells[90,91]. Tamura et al. compared the pancreas β- and α-cells obtained from an autopsy of patients with T2D with that of healthy individuals and reported rTL attrition of 27%±25% and 15%±27% for β- and α-cells, respectively, in T2D than the control group ($P < 0.01$ for both)[91]. They also found a significantly higher rTL shortening rate in β- than α-cells in the T2D group ($P < 0.01$)[91]. In this thesis, the rTL was measured in hIPCs which are derived from the pancreas of the cadaveric donors[93] and found a significant decrease of rTL through cell passages till

senescence. However, this rate was similar between the three glucose exposure conditions. One possible reason for this finding could be related to our choice of cells. It is believed that results may have been different for endothelial cells (for example), but since islet cells are used to seeing high concentrations of glucose throughout the lifetime, an 80-day exposure to different glucose concentrations did not seem to impact significantly across the experimental groups.

Besides islet cells, several studies have shown shorter rLTL (of leukocytes) in individuals with T2D than in healthy controls[258-263]. For instance, analysis from the Hong Kong diabetes register has shown a significant inverse correlation between rLTL and T2D duration ($P < 0.001$)[259]. A more recent study on the same cohort revealed significantly shorter baseline rLTL in patients with T2D who developed insulin requirement compared to the other patients with T2D (4.43 ± 1.16 vs. 4.69 ± 1.20; $P < 0.001$), resulting in finding a significantly higher risk of glycaemic progression (HR: 1.10 [95% CI: 1.06-1.14], $P<0.001$) per each unit decrease of rLTL [to ~0.2 kilobases][258]. Despite meta-analyses showing shorter rLTL in T2D patients[75,76], one study on US National Health and Nutrition Examination Survey could not identify rLTL to be associated with T2D status[264]. Therefore, measuring the telomere length in leukocytes by itself may not be sufficient to predict T2D progression, and another telomere biology-related marker could improve this predictability[61,73]. In this regard, the *TERC* relative abundance was studied in our cell model and found a significant increase in the relative *TERC* expression level in hIPC passages till their senescence only in normal glucose exposure conditions. This trend was not observed in HG and OSC conditions, suggesting a potential effect of abnormal glucose level on *TERC* expression, and possibly telomerase function, that to our surprise was not reflected in the telomere length.

As explained, initially we decided to study telomere biology in three aspects of telomere length, telomerase activity, and *TERC* expression level. However, despite Professor Roger Reddel's team efforts, measuring telomerase activity was not possible in hIPCs, and thus we were not able to correlate the telomerase activity with its lncRNA, *TERC*. In addition, the plasma *TERC* level was below the level of qPCR detection in most cases, nonetheless, it was measured using ddPCR. We found that *TERC* level was significantly lower in plasma samples of patients with T2D than in healthy individuals (P = 0.0012). This finding was in line with the *in vitro* experiments on hIPC samples in which *TERC* level was not increasing during the time in HG and OSC conditions, as representative models of T2D. To our best knowledge, our experiment - although obtained from a small sample size - was the first reported for measuring *TERC* as a biomarker, in these plasma samples. *TERC* is the fundamental part of the telomerase enzyme, and previous studies from the Diabetes and Islet Biology Group suggested that the plasma copies of *TERC* may relate to telomerase activity, which in turn may result in accelerated telomere shortening[63,68,72].

While comparing the hIPC samples from four different donors, some differences was identified between the samples. In **Figure 2.5B**, sample A and B (grey and blue dots, respectively) are related to donors with healthy BMI (**Table 2.2**). The data distribution for *TERC* relative abundance of these two samples showed a steeper increase, as opposed to the samples C and E from donors with obesity (**Figure 2.5B**). The possible reason may be related to an underlying relation between higher increase rate in *TERC* relative abundance in cells from healthy BMI donors, compared to cells from donors with obesity. The current sample size does not allow for a definite conclusion; however, this can be followed up in larger sample size as well as in other cell types such as vascular endothelial cells.

The telomeric repeat binding factor 1 and 2 (TERF1 and 2) are part of the Shelterin complex which protect the integrity of telomeric ends[61,92]. A significant negative correlation was found between *TERF1* expression level and the passage number in hIPC samples exposed to oscillating glucose (P = 0.0301). This trend was not observed in NG or HG conditions, suggesting a potentially detrimental effect of the oscillating glucose on telomere biology. Curiously, the TERF2 level was slightly rising in NG and HG, however, a flat trend was observed in the OSC condition. This could be another piece of evidence for a different molecular change profile in the oscillating glucose condition.

In addition, different trends of gene expression levels were found in several senescence-associated markers in the oscillating condition, compared to the normal and high glucose exposure conditions. The decreasing trend of several cyclin molecules such as cyclin A2, B1, E1, and E2 was only observed in the oscillating condition with a significant negative correlation (**Table 2.3**). A similar trend and negative correlation were also observed for CDK1, 2, and 4. These molecules are directly involved in cell proliferation[242,247] and their decreasing level could be associated with the observed longer mean number of days per passage in the oscillating condition (5.51 days), compared to NG and HG (4.92, and 5.06 respectively).

Among the molecules involved directly in the cellular senescence, the level of p21 was increasing in all three glucose exposure conditions, although with significant correlation only in HG and OSC (P = 0.0219, and 0.0287, respectively). p21 is a well-known CDKi and a classic senescence marker that is activated early in the cellular senescence process[240,241,244]. As expected, the rising trend of p21 gene expression was identified in our samples. However, the same trend was not reflected in other CDKi molecules, such as p16 and p15. These are involved in maintaining the senescence phenotype [240,241,244], and based on our experimental

design, the hIPCs were harvested after only eight days of senescence in which the increased level may not have occurred. Moreover, p21 is known to be activated downstream of several molecules involved in DNA damage response (DDR) including ATM, ATR, CHEK1, CHEK2, CDC25C, and p53 among others[62,240,243,252]. We did not observe a rising trend in any of these molecules. One possible reason could be the lack of activation of DDR in our samples.

The same reason (early harvest of hIPCs after eight days of senescence) could be applied to the lack of an increasing trend in the level of SASP markers. Because these molecules are activated in the later stage of senescence to disseminate the senescent phenotype to the adjacent cells[241,252,253,255]. Nonetheless, a significant increase in *GATA4* expression level was observed in all three conditions. GATA4 is a transcription factor involved in the gene expression of several SASP factors in the senescent cells[240,241,246]. *GATA4* increased transcript abundance could be a potential sign of islet cells adopting the SASP during senescence. Also, the increasing trend in *IL-8* level was observed only in the oscillating condition, however, this did not reach the significance level. IL-8 is the main chemokine (and SASP molecule) promoting inflammation in the surrounding cells during senescence[241,253,255]. Increased abundance of *IL-8* transcript in the oscillating condition may be another evidence for the detrimental effect of fast changes in glucose levels for the islet cells.

Cellular senescence is a dynamic process that involves several molecular pathways comprising tens of different molecules[62,243-245,252]. This phenomenon is different from one cell type to another and defining a set of universal senescence markers is a highly challenging[243,245,252]. There are several studies on human islet senescence indicating the declining insulin secretion in ageing and their decreasing proliferation rate in response to higher metabolic demands[265,266]. However, to the best of my knowledge, the data on understanding

senescence markers in human islet-derived progenitor cells is scarce. Telomere biology was also the focus to improve our understanding of replicative senescence in pancreatic islet cells during T2D progression.

One of the limitations of our study was the sample size. Processing and obtaining viable islet cells from cadaveric donors is challenging since research-consented islet donor availability is rare (especially during a pandemic) and that islets constitute only 1-2% of the pancreas tissue[267] (fewer numbers of islets received). Another limitation was related to the small number of available human serum samples for *TERC* measurement. Despite our small sample size, assessing the correlation between *TERC* level and rLTL could provide a better understanding of telomere status associated with T2D progression. Analysing the plasma *TERC* level in a larger sample size and identifying its correlation with the telomere length could provide a potential biomarker for T2D progression in future studies. The RNA samples of these subjects were part of a collaboration with Professor Ronald Ma and colleagues for the Hong Kong diabetes cohort, and currently the information regarding the details of these subjects is not available. More samples will be included in future studies to analyse rLTL, *TERC* relative abundance, and their associations with age, sex, and BMI among other factors.

We have measured the relative telomere length, relative *TERC* abundance, and the relative gene expression levels of several known senescence markers in human islet-derived cells as the cell model of pancreas islets. Similar *in vitro* investigation on vascular endothelial cells can be conducted as future studies to identify the impact on telomere length in these cells since they are one of the largest tissues that is exposed to abnormally high levels of fluctuating blood glucose in patients with diabetes.

Chapter 3 - Role of Dietary Oscillation in Type 2 Diabetes Progression

3.1 Background

Type 2 diabetes is a progressive medical condition and pancreatic β-cell dysfunction and insulin resistance in the target organs occur through time[5,6,39]. Diet composition is one of the main environmental factors affecting several metabolic diseases including T2D[5,6]. Resistant starch and dietary fibres are considered healthy food since they become fermented by the gut microbiome to produce several beneficial metabolites including SCFA molecules[102,103,105]. In addition to the anti-inflammatory, anti-carcinogenesis, and anti-oxidative roles of SCFAs, these molecules can also bind to their specific receptors such as FFAR2 and FFAR3 on the luminal surface of enteroendocrine cells (EECs)[104-106]. Among the two SCFA receptors, butyrate mainly binds to FFAR3[268,269]. This binding triggers a signalling pathway that epigenetically affects the gene expression and secretion of gut hormones including GLP1, GIP, PYY, and CCK among others[96,98,99]. Apart from their role in appetite regulation, these hormones can stimulate the pancreatic islet β-cell to produce insulin[98,99,101]. Therefore, SCFAs could serve as signalling molecules to modify the hormonal release of EECs and subsequently influence the glucose metabolism of the host[96,108]. Furthermore, several studies have shown that faecal transplantation to germ-free mice could change their weight and insulin sensitivity levels[270,271]. This reflects the importance of gut microbiota-generated metabolites of the donor, including its SCFA content, to infer changes in the receiver phenotype[272,273].

Dietary oscillation is an approach adopted by many individuals, in which people tend to healthy food after being disappointed with their weight and shift back to an unhealthy diet

after their reach their ideal weight[116,117]. These alternative changes in the diets affect the composition of the gut microbiome and consequently the level of their produced metabolites such as SCFAs[118,119]. We hypothesised that the oscillating changes in the SCFA level could affect the epigenetic silencing of incretin genes in EECs, the subsequent incretin hormone secretion, and the host glucose metabolism intermittently. To test our hypothesis, two well-established colon cell lines, T84 and HT-29, were utilised as the model for cells that are exposed to SCFA in the colon lumen[122,123].

We considered butyrate as the representative of SCFA molecules since the effect of SCFA on the host metabolism is mostly attributed to butyrate[102-105]. Also, previously unpublished results in our team have shown that four-day exposure of T84 cells to butyrate resulted in significantly increased expression of GLP1 in a dose-dependent manner, and similar changes were not observed by exposing the cells to acetate and propionate. In another unpublished experiment in our team, chronic exposure (168 days) to high butyrate (10mM) maintained higher levels of GLP1 gene expression. We aimed to continue that line of work employing T84 and HT-29 cell lines since they are used extensively as a model for the colon epithelial cells[122,123]. Therefore, to test our hypothesis, we planned to culture these two cell lines and expose them alternatively with high and low amounts of butyrate over several cycles assessing the changes in the expression level of incretin hormone genes.

Additionally, to validate our observations from cell lines, we aimed to replicate these experiments using a mouse model to be exposed to an intermittent healthy and unhealthy diet. Together the results of *in vitro* and *in vivo* experiments could provide us with a better understanding of the changes in the gene expression profile of incretin hormones in different exposure conditions. In addition, the possible induced insulin resistance during the dietary

oscillation in mice could help us understand the possible effect of yoyo dieting on T2D progression while comparing it to chronic exposure of the animals to an unhealthy and healthy diet as the two control groups. Overall, the diet-SCFA-incretin axis will be evaluated with respect to incretin hormones and their effect on insulin secretion together with changes in systemic insulin tolerance as the two underlying physiological disorders of T2D.

Furthermore, in the animal study, we aimed to assess several behaviours, which may be affected or impaired in diet-induced obese (DIO) mice related to their olfactory abilities, and neuromuscular reflexes and function[274]. The aim was to evaluate the effects of the oscillation diet on the cognitive and neuromuscular functions while comparing the obtained data with the control group and the mice exposed to HFD only. To the best of our knowledge, it is well-established that a high-fat diet can impact behaviour by reducing cognition, increasing anxiety, and impairing the sensorimotor gating[124-126,275], however, the impact of healthy and unhealthy dietary oscillation on mouse neuromuscular functions has not been evaluated before. My focus was on the aspects of behaviours which can be impaired during T2D progression. For instance, the sense of smell is often reduced in individuals with T2D[128]. Also, muscular strength and inbuilt reflexes can be affected in these individuals since T2D can lead to peripheral neuropathy and nerve damage, leading to muscle weakness[129,130]. Studying the changes in behaviours related to the diet could add another layer to understanding the impact of diet on neuromuscular functions along with metabolism.

3.2 Methods

The study utilised both cell and animal models as described below.

3.2.1 Cell model

Two human colorectal adenocarcinoma cell lines, HT-29 and T84, were used for the *in vitro* experiments. The cells were exposed to sodium butyrate in three conditions: chronic low [0.1mM] for 30 days, chronic high [10mM] for 30 days, and oscillating conditions. The oscillation condition was performed in four cycles. Each cycle included these steps: 1) seeding the cells in normal medium, 2) changing the media and exposing the cells to high (10mM) sodium butyrate for four days, 3) trypsinisation and sample collection, 4) seeding the cells in normal medium, 5) changing the media and exposing the cells to low (0.1mM) sodium butyrate for four days, 6) trypsinisation and sample collection. After trypsinization, 75% of cells were seeded for the next step and 25% were pelleted for RNA isolation, as depicted in **Figure 3.1**. Three experimental replicates of each cell line were utilized in each condition (in total, six replicates per condition).

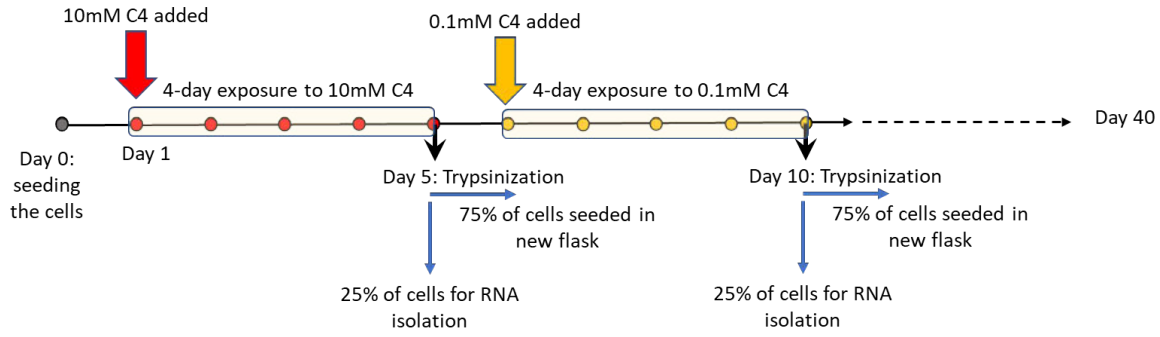


Figure 3.1 *in vitro* experimental design for oscillating condition

Three experimental replicates of each HT-29 and T84 cell line were exposed to 4-day exposure to high and 4-day exposure to low sodium butyrate. Each exposure is preceded with one day culture in normal medium. There were four cycles, and each contained 10 days. C4: sodium butyrate.

In the chronic high and low conditions, the cells were harvested when they reached 90-100% confluency level. Due to the harsh condition of 10mM exposure and high level of cell death, HT-29 cells were harvested on day 7 and at the end of the experiment (day 30), but T-84 cells were harvested only on day 30.

The DMEM/F-12 medium contained 10% FBS and 1% Pen-Strep was used as the base culture (ThermoFisher Scientific, Waltham, MA, USA). The same medium was used to prepare high and low sodium butyrate medium. The incubation condition for all conditions and cell lines was at 37°C with 5% CO_2.

3.2.1.1 Molecular analysis

RNA isolation and cDNA synthesis was performed as described in **Appendix D** and **E**, respectively. TaqMan gene expression assays were utilised to measure the relative abundance of the target genes as explained previously in **Section 2.2.2** and **Appendix F**.

3.2.2 Animal model

The animal study was performed in line with the cell study and similarly constituted the three exposure conditions in the form of mice diet. In total, 60 C57BL/6J mice were purchased at the age of six weeks old from Australian BioResources (ABR) (Moss Vale, NSW, Australia). The study was run in two cohorts to manage the workload and study plan more efficiently. Each cohort contained 30 mice, in which the mice were randomly assigned to three dietary groups, each with 10 mice (1:1 male-to-female ratio). Each cage contained either three or two mice of the same sex.

The mice were held in the Western Sydney University, School of Medicine, Campbelltown campus, Building 30, Animal Facility throughout the study. This study was approved by Western Sydney University (WSU) Animal Care and Ethics Committee (ACEC number A14718).

3.2.2.1 Dietary intervention

The mice were acclimatized for one week before the start of the dietary intervention program, and all 60 mice were fed with normal chow during that period. The dietary intervention plan spanned 22 weeks (Day 7-161) in three formats: 1) a healthy chow diet throughout the study (control group, CTR); 2) an "unhealthy" High Fat Diet (60%) throughout the study (HFD); and 3) an Oscillating (CTR and HFD diets) on a fortnightly basis (OSC). The fortnightly change in the mouse diet was selected because 100 years of human age is approximately equal to 3 years of age for mouse[276]. Therefore, two weeks of mouse age corresponds to about one year of human age – the period in which people normally stick to a healthy diet and then return to their previous unhealthy one[116,119]. The high-fat diet (SF02-006) is a semi-pure dietary formulation based on AIN-93G[277] (Specialty Feeds Pty Ltd., NSW Australia) with a matched CTR diet. The dietary intervention program is illustrated in **Figure 3.2**.

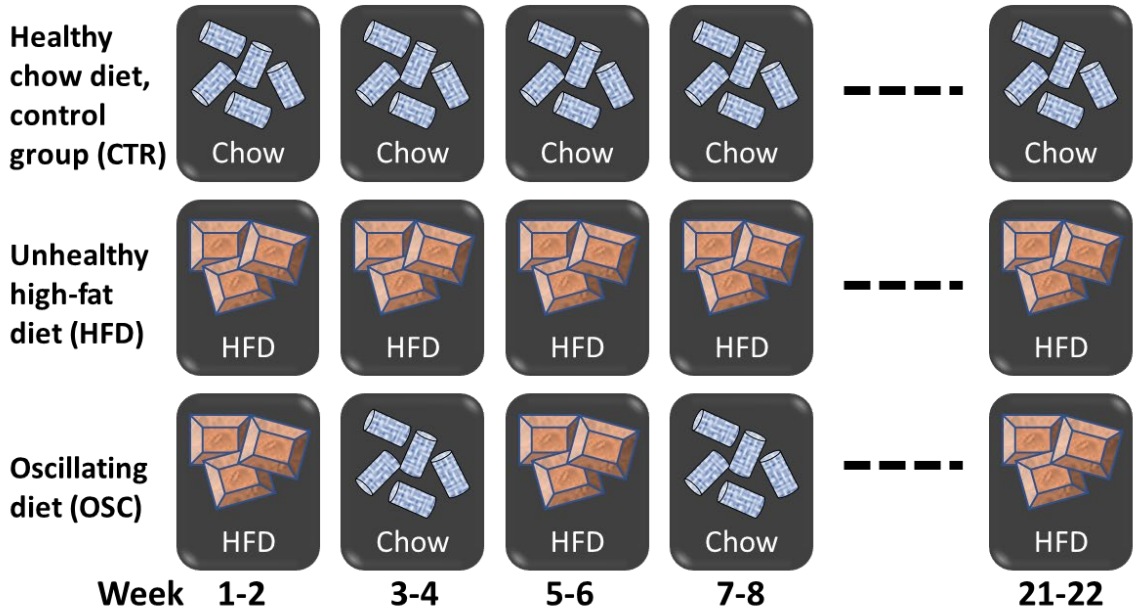


Figure 3.2 The dietary intervention program

Three groups, each containing 20 mice (10 males; 10 females), were provided with three diet plans over 22 weeks. The diet (Chow vs HFD) in the oscillating (OSC) group was swapped fortnightly. At the end of the study, the OSC group animals were on an High-fat chow diet.

3.2.2.2 Measurements

Several factors were measured throughout the 22-week dietary intervention. The body weight and dietary intake were measured weekly. For the dietary intake, the amount of food remainder was subtracted from the weight of the food given in the previous week. The shredded paper bedding allowed for easy measurement of remainder food from the bottom of the cage. The food and bedding were changed weekly throughout the study. Additionally, faecal samples were collected from each mouse every week using forceps. Each mouse was put in a separate cage in which they freely move and release faeces. The forceps were cleaned and sterilized with 80% v/v ethanol for each mouse. The faecal samples stored at -20°C for future experiments. Throughout the study, the mice were monitored at least twice per week and the monitoring results were recorded in the standard sheets provided by the animal facility.

The fasting blood glucose (FBG) was measured at the baseline of the study before the dietary intervention started (Day 7). FBG measurements were performed fortnightly during the first 10 weeks of dietary intervention (Day 21, 35, 49, 63). This was followed by weekly measurements on days 77, 84, 91, 98, 105, and 112, as well as two other measurements on days 140 and 161 (the endpoint). Fasting involved transferring mice to new cages with water bottles only and no food. After seven hours, the FBG was measured in the afternoon and then mice were fed with new food. Blood was taken from the tail tip using a lancet. The first drop of blood was discarded to avoid errors in measuring interstitial fluid and the second drop of

blood was taken to measure on a single-use glucose test strip using an ACCU-CHEK glucometer and test strips (Roche GmbH, Mannheim, Germany).

At the end of the dietary intervention (Day 161), the glucose tolerance test (GTT) was performed using the same glucometer and test strips. Mice were fasted for seven hours during the day (8am to 3pm). The glucose was injected intraperitoneally (IP) based on the weight of each mouse ($2.0 gkg^{-1}$). In practice, 200g/L D-glucose solution was prepared by dissolving 3g of D-glucose in 10mL of saline (9g NaCl dissolved in 1 litre water). Thus, the amount of D-glucose solution (µl) was 10 times more than the mouse weight (g). GTT started with FBG (considered as min 0), followed by glucose IP injection, and measuring the blood glucose at 30-, 60-, 90- and 120-minutes after glucose injection.

Animals were euthanized after GTT using the mobile isoflurane vaporizer unit at the animal facility according to SOP SSH296, approved by the WSU ACEC. After euthanasia, several tissues were collected from each mouse including blood, serum, eyes, brain, liver, gallbladder, pancreas, heart, kidney, adipose, ileum, cecum, and colon. The tissues were collected in 4% Paraformaldehyde solution for tissue fixation. For the ileum, cecum, and colon, first, the content was collected, followed by PBS wash twice. Next, the internal part was washed three times with TriZol (ThermoFisher Scientific, Waltham, MA, USA), first with a 1ml fresh TriZol, and then with the pass-through TriZol to completely release the epithelial cells into the TriZol. Both the content and the TriZol-containing epithelial cells were stored at -80°C. The blood samples were kept at room temperature for 30 minutes and then centrifuged at 3000rpm for 15 minutes for isolating serum from the red blood cells. The serum and blood were also stored at -80°C.

3.2.2.3 Behavioural tests

The changes in behaviours of the mice were assessed in all three diet groups before starting the intervention program (baseline), halfway through the intervention (during week 9), and towards the end of the study (during week 21). The OSC group was on HFD in weeks 9 and 21. The mice were not fasting in these three weeks to eliminate any effects of fasting on behavioural studies. The three behavioural tests were performed as explained below.

1) Righting reflex: The mouse was turned on its back and the time it took for the mouse to right itself to a normal position is measured (postural reflex). This test was performed three times per day while considering the maximum threshold as 10 seconds.

2) Wire hang: The mouse was placed on a wire mesh which then turned upside down approximately 50cm above the surface of soft bedding material. The mouse gripped the wire mesh and the latency to fall onto the bedding was recorded. This test was performed once per day while considering five minutes as the cut-off.

3) Olfaction: The test checked the ability of the animal to distinguish the same and different odours. The experiment was performed in two steps (training and test), each for 3 minutes. In both steps, the mouse was put in a cage where two cotton-tipped applicators hung from the rail roof in the right and left corners. In the training set, both pieces of cotton were soaked in the same odour (for example vanilla extract, diluted 1:100). In the test set, the left cotton had the same odour as the training set, but the right one had a new and different odour (for example, almond extract, diluted 1:100). The number of sniffing the pieces of cotton (source of odours) in both training and test steps were measured. The nosing index (%) was calculated from the test set using this formula: (number of sniffing the right cotton / total number of sniffing the right and left pieces of cotton)*100.

3.3 Results

3.3.1 Cell model

The relative gene expression level of seven molecules involved in gut hormone expression was measured in three sodium butyrate (C4) conditions for approximately five weeks: five incretin hormones (GLP-1, GIP, CCK, PYY, and Ghrelin), the main butyrate receptor on EEC cells (FFAR3), and HDAC2 as the main downstream histone deacetylase enzyme. *GCG* is the name of the gene that provides instruction for GLP-1 and glucagon hormone, and we have selected the TaqMan assay which targets the mRNA sequence that translated to GLP-1 hormone.

Gene expression analyses on the cells harvested before the conditioning (on day 0) showed similar expression levels between HT-29 and T84 cells per gene. Also, the gene expression pattern of all the seven studied genes was similar at all the time points. Therefore, we run the analyses while combining the expression level of both cell lines to have a greater amount of data per gene per timepoint to increase the statistical power.

We noticed the gradual cell loss during the 30 days of chronic exposure in both HT-29 and T84 cell lines due to the harsh condition of exposing the cells to high C4. The dead cells were being washed in each refreshing of the high C4 medium, however, by the end of the 30-day exposure, the viable cell confluency dropped to 10%. Therefore, we omitted the results of gene expression analyses on day 30 of exposure to high C4 since the expression level for all the seven genes were close to zero, even well below the level at day 0. This could be mainly related to the viability of the cells used for RNA isolation. We also observed cell death in the

oscillating condition in each 4-day exposure to high C4, but they were recovering to 90-100% confluency during the following 4-day exposure to low C4.

In the oscillating condition, for each gene we calculated the average amplitude of fold change in gene expression relative to the previous step. Next, the relative change through the 32 days of exposure to the oscillating high and low C4 was plotted per each gene (**Figure 3.3**). A similar pattern of gene expression was identified in all seven genes. In this pattern, the fold changes are greater at the first one and two cycles of exposure to high and low C4 and then became smaller towards the end on days 28 and 32 (**Figure 3.3**). *GIP* and *PYY* expression also showed a great drop on day 24. This pattern showed that the increase or decrease in gene transcript level is not only proportional to the C4 level, and in later cycles, the yo-yo pattern became less intense.

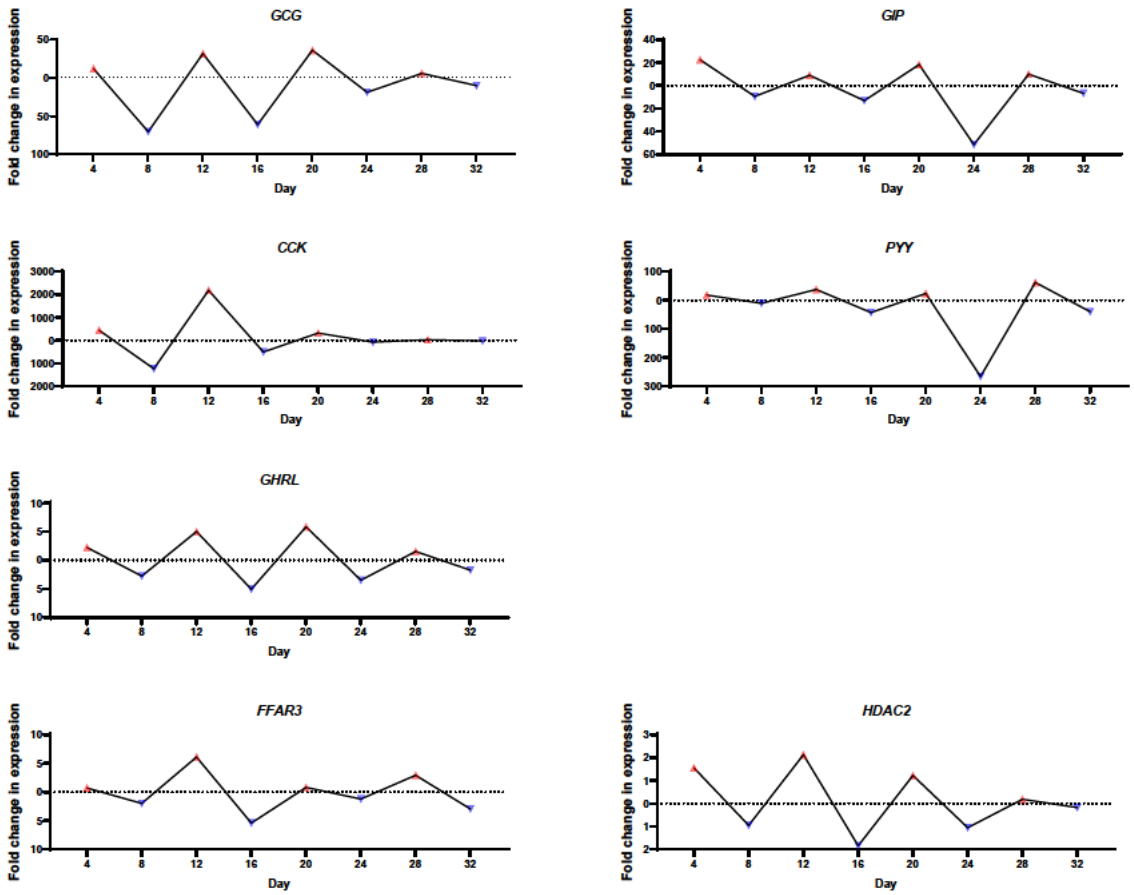


Figure 3.3 Average change in the expression amplitude in the oscillating condition

Each point is the average of gene expression fold change relative to the previous step. The data on day 4 is relative to day 0. The x-axis shows the number of days of exposure (4-day interval of exposure to high or low C4). The gene expression level related to high and low C4 levels is shown by the red and blue triangle, respectively.

The gene expression pattern in the oscillating condition was further analysed by calculating the Z-score in each timepoint relative to day 0 (**Figure 3.4**). These analyses showed a similar pattern of gene expression in the five studied gut hormones (*GCG*, *GIP*, *CCK*, *PYY*, and *GHRL* genes), for which the initial rising on days 4, 12, and 20 did not occur on days 28. *GCG* expression on days 4, 12, 20, and 28 (related to high C4 exposure) was decreasing in each step. *GIP* expression level was on the rise on days 0, 12, and 20 with a sudden drop on day 28. The gene expression pattern of *CCK*, *PYY* and *GHRL* was almost identical (**Figure 3.4**). The transcript level of these three genes increased from day-4 to -12 and then decreased by days 20 and 28. The Z-score analysis revealed lowered amplitude of changes in the gene expression level of *GCG*, *GIP*, *CCK*, *PYY*, and *GHRL* after 3 oscillations. The amplitude change in *GCG* expression level was significant in the first cycle of exposure to high and low C4 (**Figure 3.4**). This was significant for *GIP* in the third cycle, as well as *CCK* and *PYY* in the second cycle,

however, the changes in *GHRL* level were not significant (**Figure 3.4**).

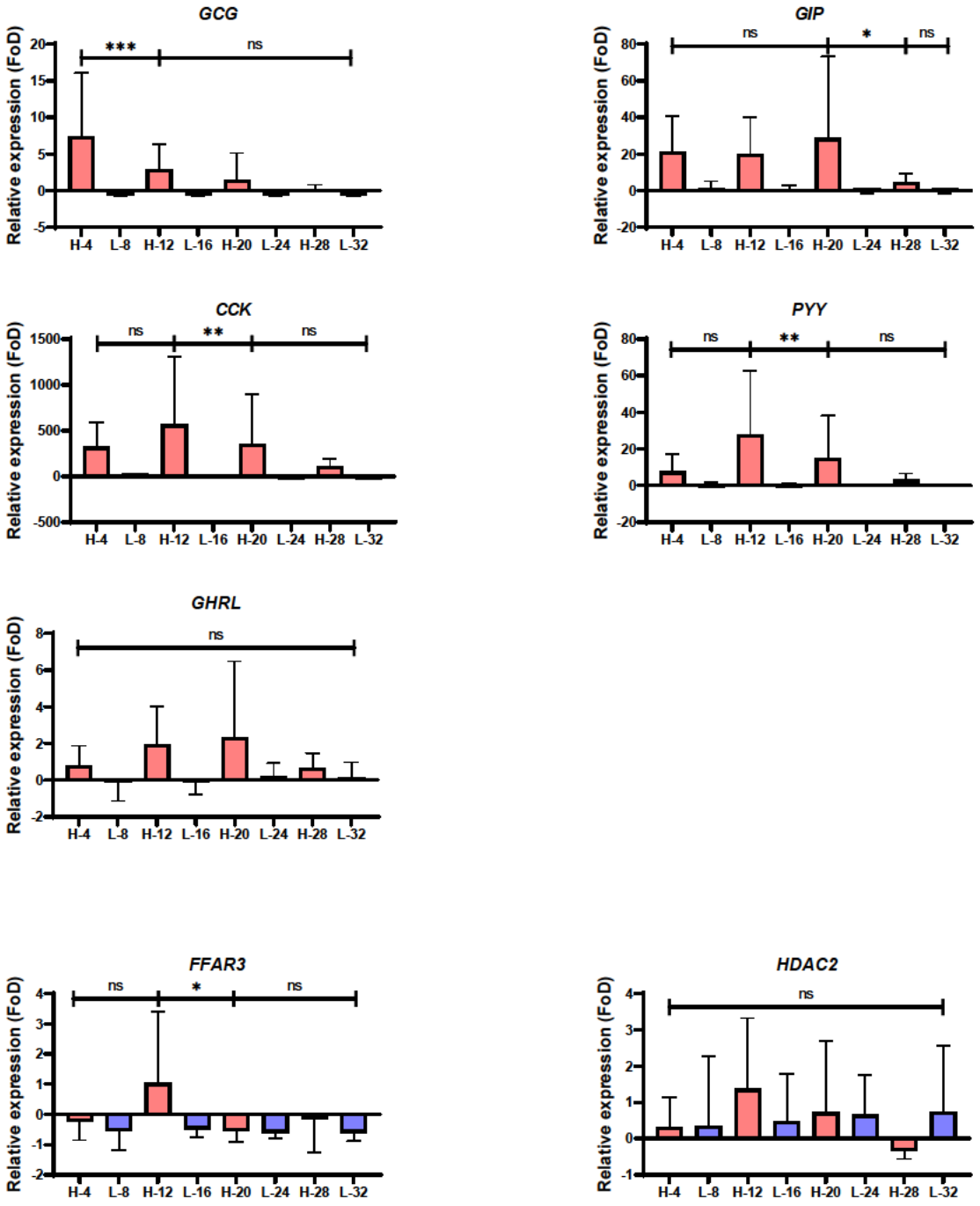


Figure 3.4 Gene expression level relative to day 0 in the oscillating condition

Each bar represents the mean and SD of calculated Z-scores of the gene expression in each time point relative to day 0. The time point after four days of exposure to high C4 is shown in red (days 4, 12, 20, and 28). The time point after four days of exposure to low C4 is shown in blue (days 8, 16, 24, and 32). The x-axis shows the number of days of exposure (4-day intervals of exposure to high and low C4). The two-way ANOVA followed by multiple comparisons was used to compare the gene expression level between high and low C4 in each timepoint. In

each graph, the P values represents the comparison of a high C4 condition with the next low C4. ns = not significant, *= P<0.05, **=P<0.01, ***=P<0.001, ****=P<0.0001

Interestingly, *FFAR3* and *HDAC2* did not show the same pattern of increase and decrease in transcript level seen in the gut hormones (**Figure 3.4**). *FFAR3* expression was lower than day 0 even when exposed to high C4 (except for day 12). The *HDAC2* transcript level was almost always higher than the day 0 whether in high or low C4 condition (except for day 28). In addition, the expression level of gut hormones genes in low cycles (day 8, 16, 24, and 32) was almost the same as day 0. However, *FFAR3* transcript level decreased in low C4 exposure, and on the contrary, *HDAC2* transcript level was increased when exposed to low C4 (**Figure 3.4**). This could be related to the functional role of these two molecules in the butyrate-gut hormone release axis. Whenever FFAR3 is more present, it activates downstream signalling pathways activating HDAC inhibitors to downregulate HDAC molecules such as HDAC2[98,120]. In addition, the amplitude change in *HDAC2* expression was not significant in different cycles of exposure to high and low C4, but *FFAR3* level changed significantly in the second cycle (**Figure 3.4**).

Furthermore, we compared the gene expression level at the last time points: day 30 for chronic low C4 (L), day 28 as the last time point of high C4 exposure in the oscillating condition (OSC-H), and day 32 as the last time point of low C4 exposure in the oscillating condition (OSC-L) (**Figure 3.5**). We utilized either a parametric or non-parametric one-way ANOVA test based on the normality of the data. No significant difference in the expression level of any studied genes was identified between L and OSC-L. the gene expression level of all five hormones was higher in OSC-H compared to L or OSC-L, and this was significant in *GCG* (vs OSC-L), *GIP* (vs L), *PYY* (vs L and OSC-L) (**Figure 3.5**). No significant difference was identified in *FFAR3* and *HDAC2*

gene expression levels between the three endpoints, confirming the Z-score finding for these two genes.

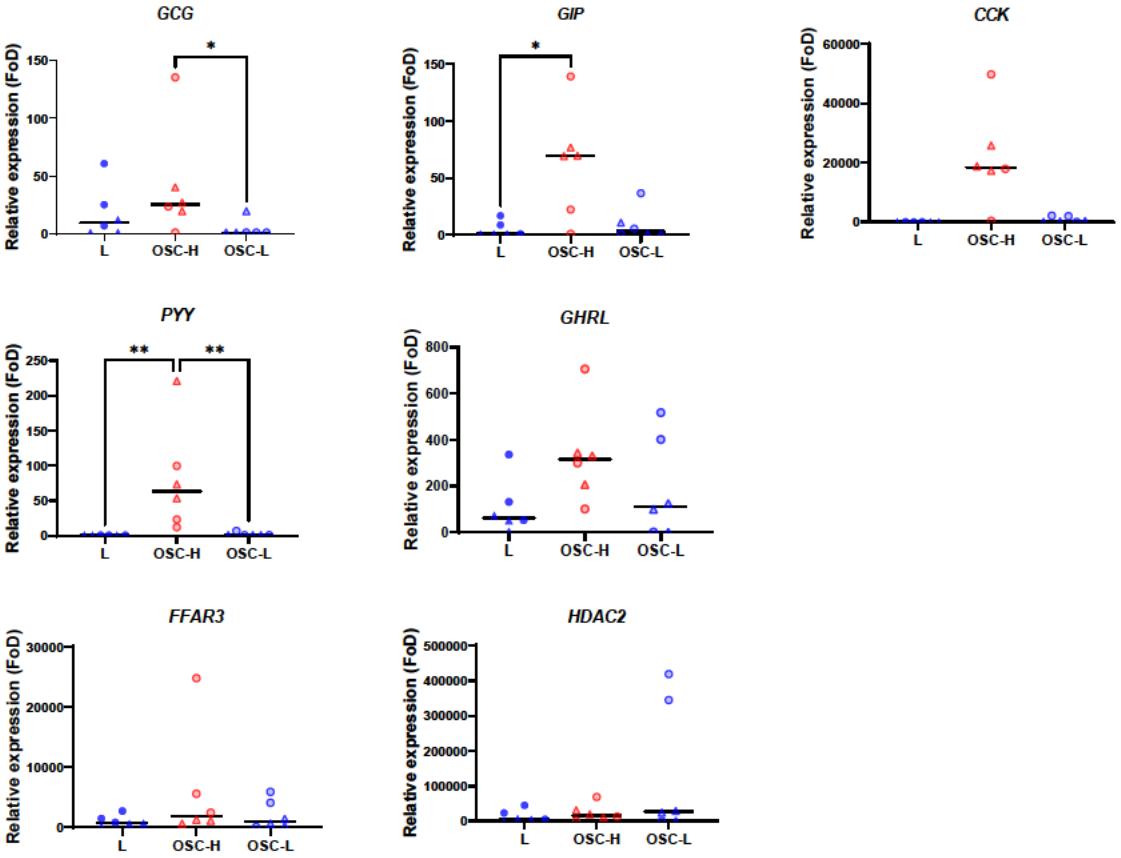


Figure 3.5 Endpoint comparison of gene expression levels in three C4 condition

Circles and triangles show the gene expression data of HT-29 and T84 cells, respectively. The median line is shown since the data was nonparametric in most cases. L: endpoint for chronic low C4 (day 30), OSC-H: endpoint of high C4 exposure in the oscillating condition (day 28), OSC-L: endpoint of low C4 exposure in the oscillating condition (day 32). *: P-value 0.01-0.05, **: P-value 0.001-0.01

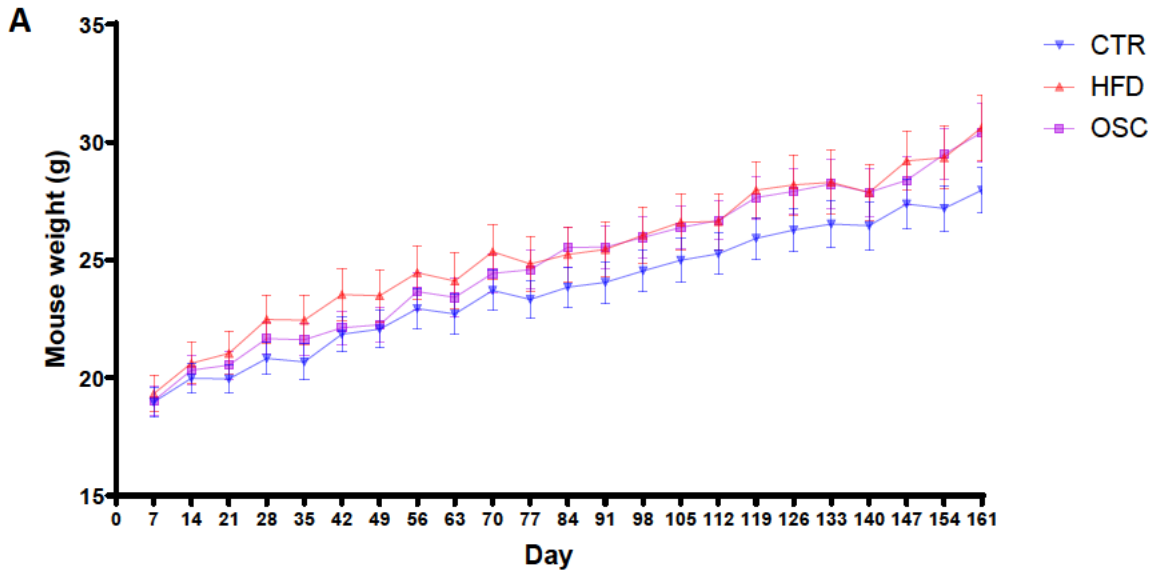
3.3.2 Mouse model

During the 22-week intervention program, four mice were lost. One male in the HFD group was killed during a cage fight (week 18), and three females in the OSC group died because of

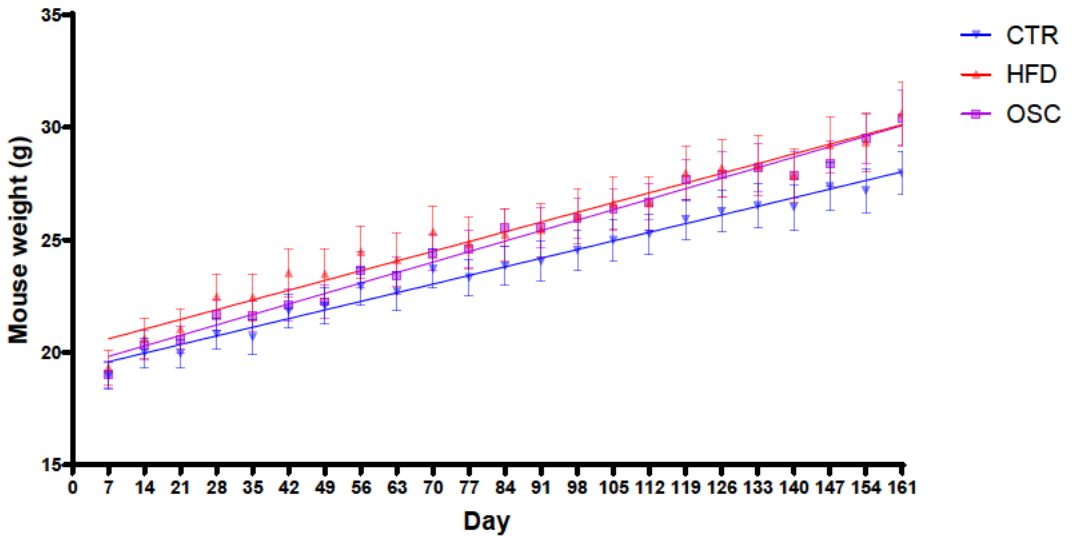
the water bottle drip and subsequent cage flooding (week 11). Therefore, the number of mice at the end of the diet program was 20, 19, and 17 in CTR, HFD, and OSC groups, respectively.

3.3.2.1 Body weight

Weekly recording of the mice's body weight showed a significant increase in all three dietary intervention groups (Pearson r: 0.99, 0.98, and 0.99 for CTR, HFD, and OSC group, respectively; P<0.0001 for all the three groups) (**Figure 3.6**). Throughout the study, gradually the mice in HFD and OSC groups gained more weight, however, no significant difference was observed in the mean body weight between the three groups (**Table 3.1**). However, the mixed-effects analysis showed an overall significant effect of diet (P = 0.0371) and time (P<0.0001) on the body weights. Interestingly, as seen in the OSC group trend line, the body weight in this group increased towards the end of the study to reach that of the HFD group (**Figure 3.6**).



B



C

	Day vs. CTR	Day vs. HFD	Day vs. OSC
r	0.9924	0.9836	0.9922
95% confidence interval	0.9818 to 0.9968	0.9610 to 0.9931	0.9814 to 0.9967
R squared	0.9848	0.9674	0.9845
P value			
P (two-tailed)	<0.0001	<0.0001	<0.0001
P value summary	****	****	****

	CTR	HFD	OSC
Slope	0.05482	0.06182	0.06656

	CTR	HFD	OSC
P value	<0.0001	<0.0001	<0.0001

Figure 3.6 Mouse body weight throughout the dietary intervention

A) The line graph of mean weight gain in the three groups. B) The weight gain trend was significant in the three intervention groups. No significant difference was observed between the slope of the trend lines (P = 0.148). Day 7 indicates the measurement at baseline before the dietary intervention starts. The data are shown as mean and SEM. C) The table summarises the results of correlation and linear regression analysis.

Table 3.1 Body weight (g) comparison between the three groups

The P value of the multiple comparison test with Benjamini, Krieger and Yekutieli correction is shown in the table. CTR: control group on chow diet, HFD: high-fat diet, OSC: oscillating Chow/HFD

Day	HFD vs CTR	OSC vs CTR	OSC vs HFD
7 (Baseline)	0.73	0.95	0.76
14	0.56	0.72	0.79
21	0.33	0.50	0.65
28	0.18	0.38	0.51

Day	HFD vs CTR	OSC vs CTR	OSC vs HFD
35	0.18	0.38	0.51
42	0.21	0.80	0.29
49	0.31	0.87	0.36
56	0.29	0.54	0.57
63	0.35	0.56	0.64
70	0.26	0.54	0.52
77	0.30	0.28	0.86
84	0.34	0.17	0.84
91	0.35	0.24	0.95
98	0.32	0.26	0.95
105	0.29	0.29	0.89
112	0.36	0.24	0.97
119	0.18	0.18	0.84
126	0.23	0.23	0.87
133	0.23	0.25	0.83
140	0.37	0.33	1.00
147	0.28	0.50	0.62
154	0.20	0.12	0.93
161	0.12	0.13	0.91

Moreover, one-way ANOVA analysis showed no significant difference in body weight (g) between the three diets on day 161 in either male (P = 0.16) or female mice (P = 0.28). The results of multiple comparison analysis were not statistically significant as well (**Table 3.2**).

Table 3.2 Body weight (g) comparison between the three diet groups on day 161

The P value of multiple comparison test with Benjamini, Krieger and Yekutieli correction is shown. CTR: control group on chow diet, HFD: high-fat diet, OSC: oscillating Chow/HFD

Comparison	Male	Female
HFD vs. CTR	0.07	0.19
OSC vs. CTR	0.72	0.16
OSC vs. HFD	0.14	0.83

This was further analysed using two-way ANOVA to analyse the effect of sex and diet on weight gain simultaneously. The results showed the significant effect of sex (P<0.0001) but

not diet (P=0.1) on the final body weight. This was followed by multiple comparisons test with Benjamini, Krieger and Yekutieli correction (**Table 3.3**) which showed a significantly lower body weight in female compared to male mice in all three diet groups (P=0.0002, P=0.001, and P=0.04 in CTR, HFD, and OSC, respectively).

Table 3.3 Body weight (g) comparison on day 161 in male and female mice

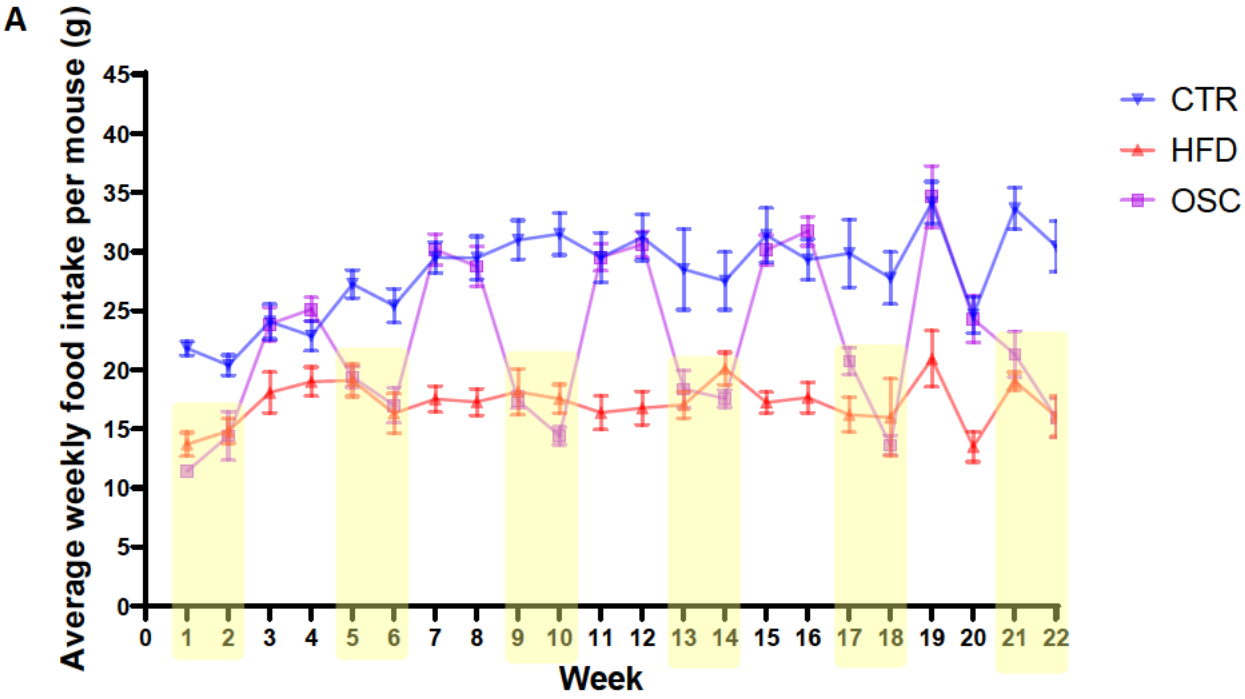
The significant P value of multiple comparison test with Benjamini, Krieger and Yekutieli correction is shown in red. CTR: control group on chow diet, HFD: high-fat diet, OSC: oscillating Chow/HFD

Comparison	P value
Male:CTR vs. Male:HFD	0.19
Male:CTR vs. Male:OSC	0.80
Male:CTR vs. Female:CTR	**0.0002**
Male:CTR vs. Female:HFD	**0.03**
Male:CTR vs. Female:OSC	0.07
Male:HFD vs. Male:OSC	0.28
Male:HFD vs. Female:CTR	**<0.0001**
Male:HFD vs. Female:HFD	**0.001**
Male:HFD vs. Female:OSC	**0.004**
Male:OSC vs. Female:CTR	**<0.0001**
Male:OSC vs. Female:HFD	**0.01**
Male:OSC vs. Female:OSC	**0.04**
Female:CTR vs. Female:HFD	0.10
Female:CTR vs. Female:OSC	0.08
Female:HFD vs. Female:OSC	0.79

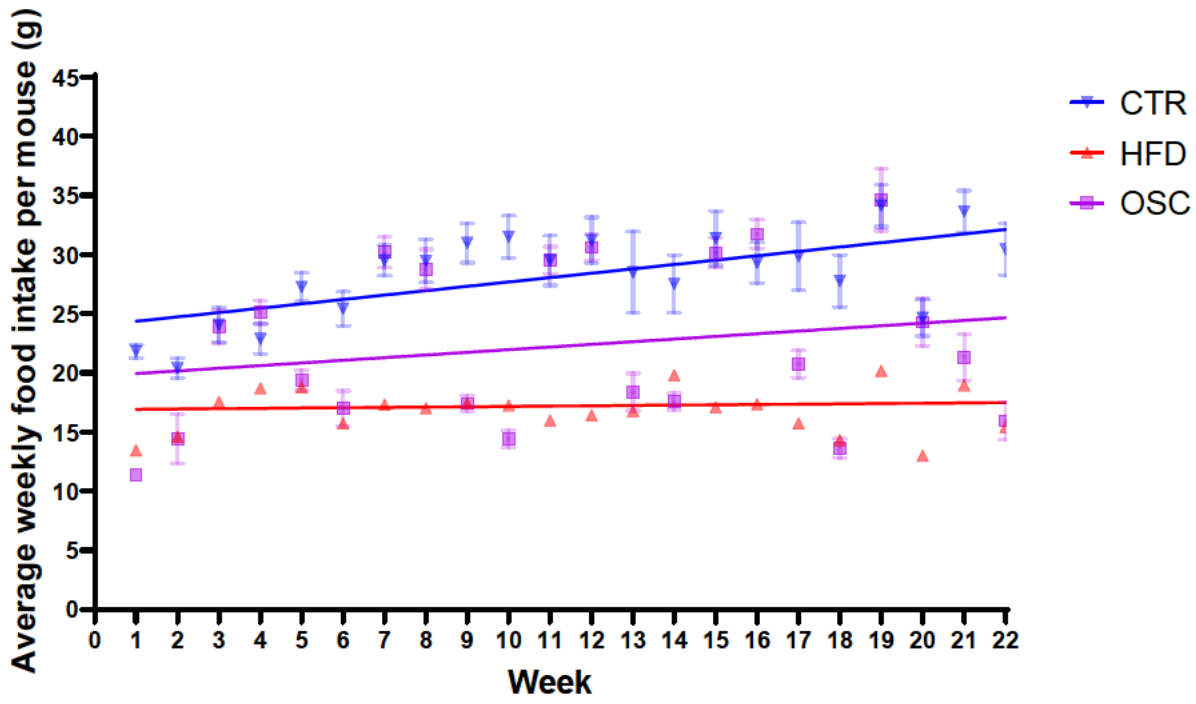
3.3.2.2 Diet and energy intake

The weekly dietary intake had also an overall positive trend in all three groups (**Figure 3.7**). Throughout the 22-week dietary intervention, the mice on HFD were eating significantly less than their counterparts on CTR (**Table 3.4**), as seen in the flat trend line of the HFD group (**Figure 3.7**). OSC group showed interesting dietary behaviour. When these mice were

exposed to HFD, they were eating significantly less that the CTR group, with no significant difference with the HFD. Similarly, when exposed to a chow diet, they were eating significantly higher than the HFD group, but with no significant difference from the CTR group. These comparisons are summarised in **Table 3.4**. Also, the line graph in **Figure 3.7** shows the differences in food intake of the OSC group.



B



C

	Week vs. CTR	Week vs. HFD	Week vs. OSC
Pearson r			
r	0.6504	0.09530	0.2022
95% confidence interval	0.3152 to 0.8413	-0.3400 to 0.4969	-0.2398 to 0.5748
R squared	0.4230	0.009083	0.04089
P value			
P (two-tailed)	0.0010	0.6731	0.3668
P value summary	**	ns	ns

	CTR	HFD	OSC
Slope	0.06865	0.05330	0.09253

	CTR	HFD	OSC
P value	<0.0001	0.6111	0.0163

Figure 3.7 Weekly food intake throughout the dietary intervention

A) Each dot point indicates the average food intake (g) per mouse. The highlighted yellow sections show the weeks that the OSC group were exposed to HFD. B) Only the trend line for the CTR group had a significant rise (P<0.0001), and a significant positive correlation (P = 0.001). There was a significant difference between the slope of the trend lines (P = 0.0038). The data are shown as mean and SEM. C) The table summarises the results of correlation and linear regression analysis.

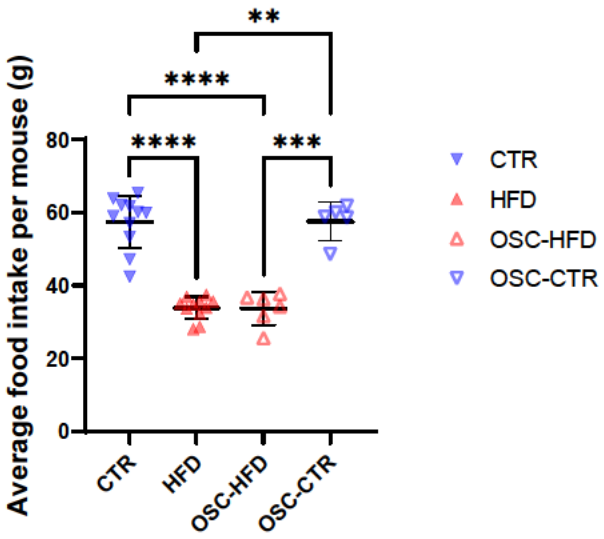
Table 3.4 Food intake (g) comparison between the three groups

The significant P value of the multiple comparison test with Benjamini, Krieger and Yekutieli correction is shown in red. The yellow-highlighted weeks are those in which the mice in the OSC group were exposed to HFD. CTR: control group on chow diet, HFD: high-fat diet, OSC: oscillating Chow/HFD

Week	HFD vs CTR	OSC vs CTR	OSC vs HFD
1	<0.0001	<0.0001	0.06
2	0.0011	0.025	0.86
3	0.02	0.92	0.02
4	0.05	0.19	0.002
5	0.0005	0.0001	0.86
6	0.001	0.001	0.77
7	<0.0001	0.72	<0.0001
8	0.0001	0.78	0.0001
9	0.0002	<0.0001	0.74
10	<0.0001	<0.0001	0.054
11	0.0002	0.99	<0.0001
12	<0.0001	0.8	<0.0001
13	0.012	0.02	0.51
14	0.02	0.0043	0.13
15	0.0003	0.65	<0.0001
16	0.0001	0.27	<0.0001
17	0.0015	0.02	0.03
18	0.01	0.002	0.51
19	0.0007	0.87	0.002
20	<0.0001	0.9	0.001
21	<0.0001	0.0005	0.31
22	0.0002	0.0002	0.97

Additionally, a significant difference was identified between the cumulative food intake (g) between the diet groups using one-way ANOVA (P<0.0001). This was followed by multiple comparison test revealing significance difference among the groups (P<0.0001, P<0.0001, and P=0.0005 in HFD vs CTR, OSC vs CTR, and OSC vs HFD, respectively). Further analyses in male and female mice showed that the significant difference in cumulative food intake between diets was mostly related to female (P=0.0006, P=0.024, and P=0.052 in HFD vs CTR, OSC vs CTR, and OSC vs HFD, respectively) than male mice (P=0.0017, P=0.12, and P=0.12 in HFD vs CTR, OSC vs CTR, and OSC vs HFD, respectively).

Overall, the oscillating OSC group were eating much less when being exposed to HFD than the chow diet (mean±SD: 16.8±2.98 vs 28.89±3.47, P<0.0001). Furthermore, the food intake of the OSC group while being exposed to chow (OSC-CTR) or high-fat diet (OSC-HFD) was compared to CTR and HFD group food intake. The OSC mice were eating almost the same amount of food when being exposed to chow and a high-fat diet compared to CTR and HFD groups, respectively (**Figure 3.8**).



Dunnett's T3 multiple comparisons test	Mean Diff.	95.00% CI of diff.	Below threshold?	Summary	Adjusted P Value
CTR vs. HFD	23.63	16.57 to 30.70	Yes	****	<0.0001
CTR vs. OSC-HFD	23.76	15.20 to 32.33	Yes	****	<0.0001
CTR vs. OSC-CTR	-0.1378	-10.04 to 9.760	No	ns	>0.9999
HFD vs. OSC-HFD	0.1289	-6.828 to 7.086	No	ns	>0.9999
HFD vs. OSC-CTR	-23.77	-33.53 to -14.01	Yes	**	0.0010
OSC-HFD vs. OSC-CTR	-23.90	-33.85 to -13.95	Yes	***	0.0002

Figure 3.8 Comparing the food intake of the OSC group with HFD and CTR
Data are shown as mean and SD.

Additionally, the Z-scores of food intake were calculated for the OSC group relative to CTR, and HFD (**Figure 3.9**). In the OSC group, the mice were eating much less when exposed to a high-fat diet (yellow bars), and almost the same as the CTR group when exposed to a chow

diet (green bars). This eating behaviour did not change during the 22-week dietary intervention program, and we did not observe a specific rising or declining trend in the OSC group when being exposed to either diet in this analysis.

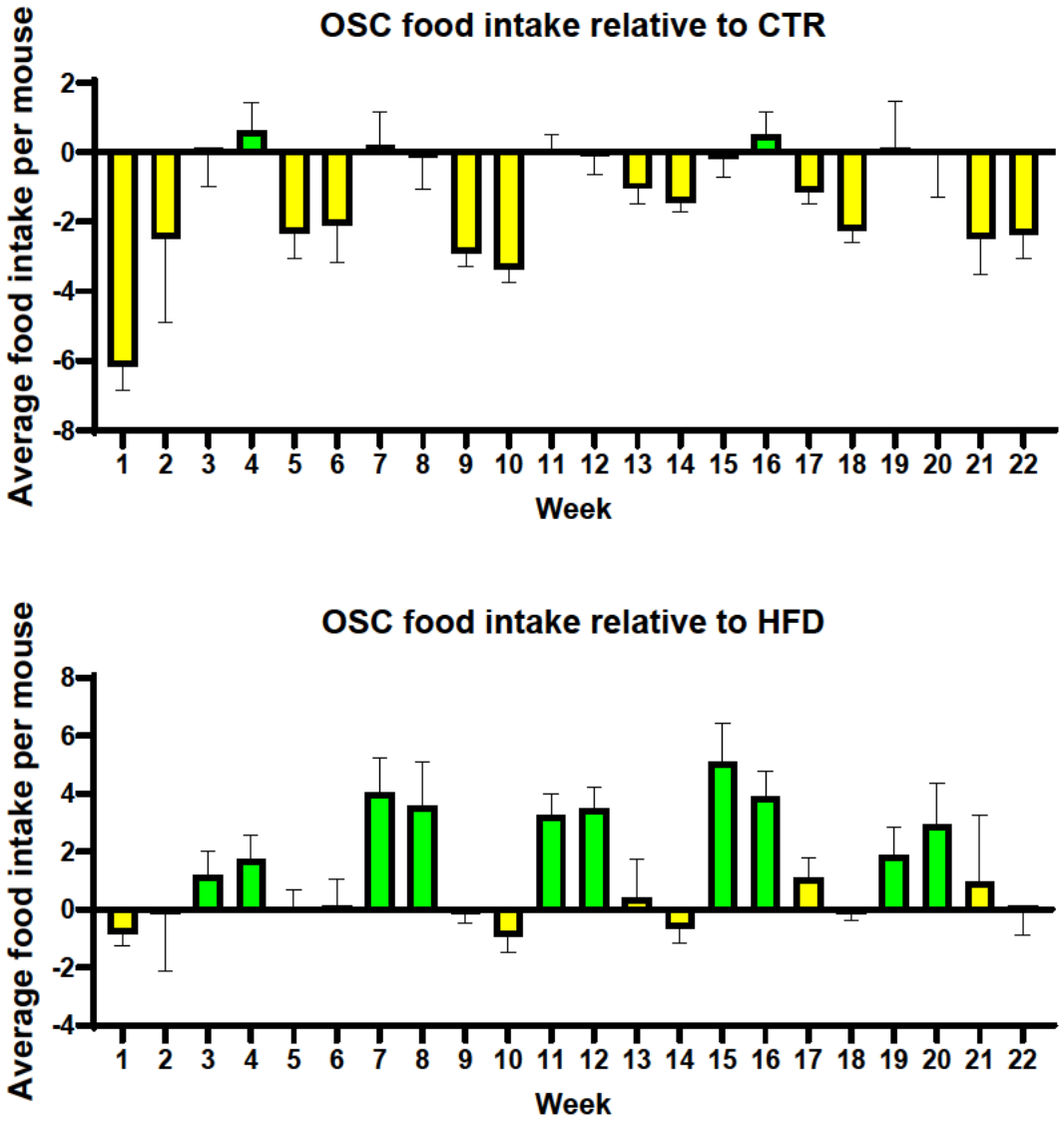


Figure 3.9 Food intake in OSC relative to the other two groups

The mice in the OSC group were exposed to high fat and chow diet fortnightly. The yellow and green bars show the food intake while being exposed to high fat or chow diet, respectively. The data are shown as mean and SD.

Furthermore, by considering each fortnightly diet in the OSC group as a cycle, another Z-score of food intake in each cycle was calculated relative to the previous cycle (**Figure 3.10**). The pattern of eating more while on a chow diet and less while on high fat was observed here as well. However, these differences began to increase till cycles 6 and 7 (weeks 11, 12, 13, and 14) and then started to decline towards the end of the study. It seems that the overall tendency to eat was reduced towards the end of the dietary program in the OSC group despite the steady weight gain in this period (**Figures 3.10** and **3.6**).

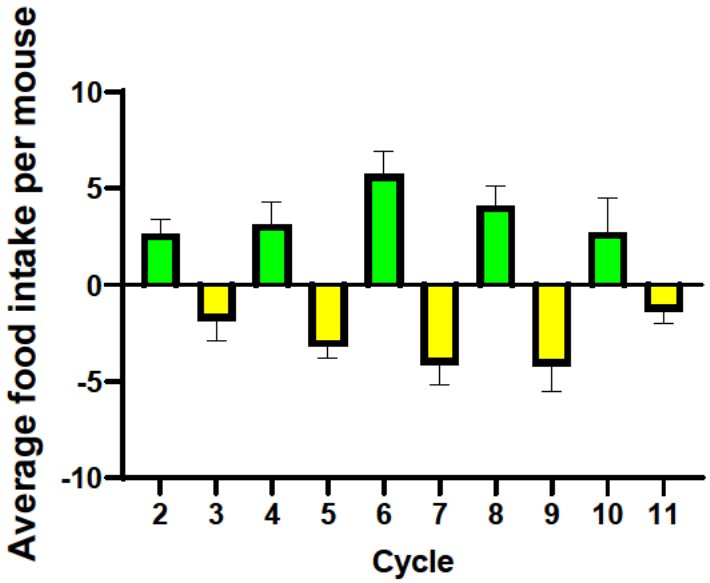
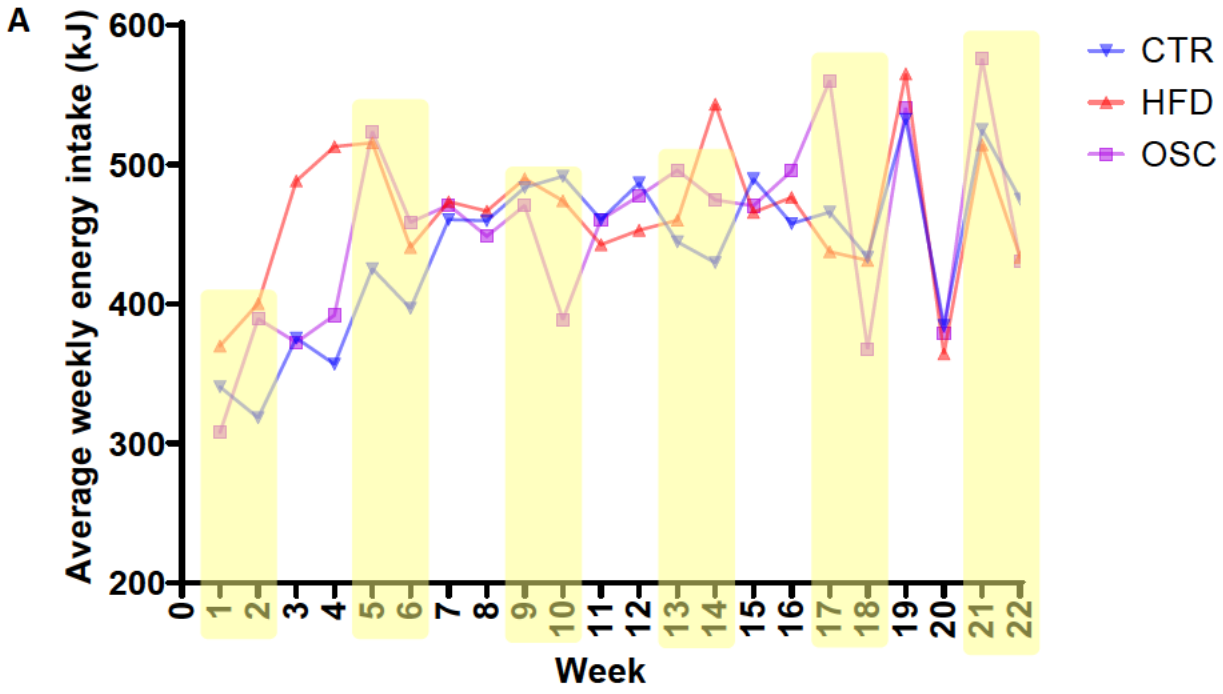


Figure 3.10 Food intake in the OSC group relative to the previous cycle

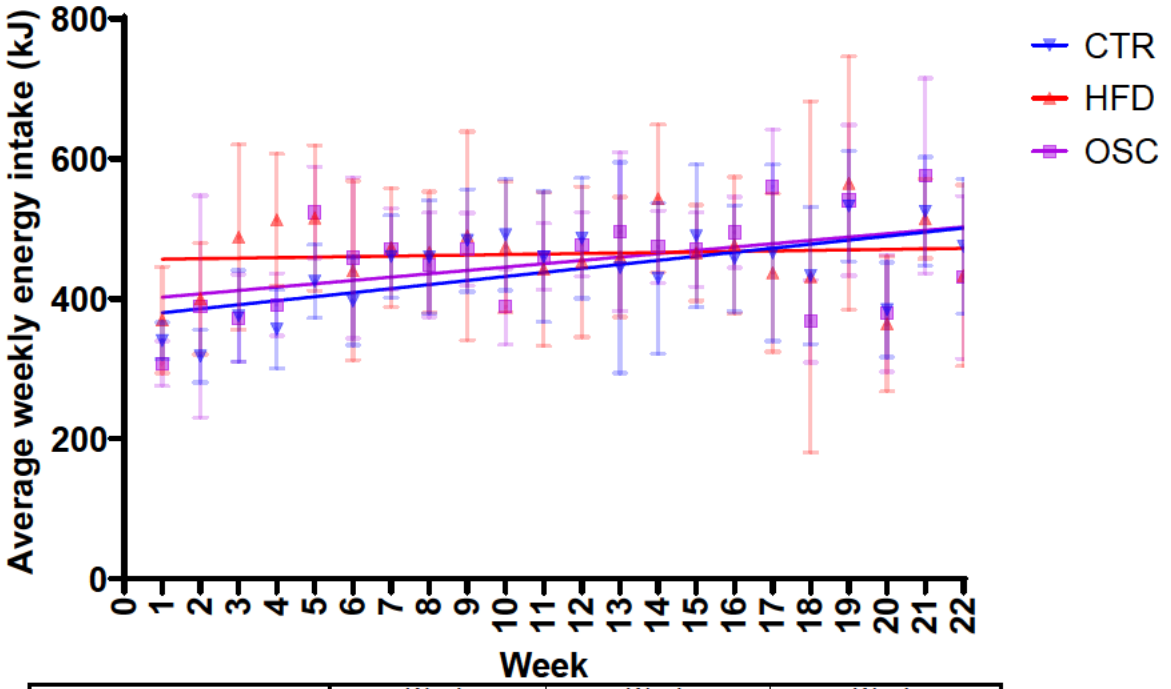
A cycle is defined as two-week exposure to the same diet in the OSC group. The green and yellow bars show the food intake while being exposed to a chow or high-fat diet, respectively. For instance, cycle 2 represents weeks 3-4 when the OSC group was exposed to the chow diet (green bar), and cycle 11 is related to weeks 21-22 for exposure to a high-fat diet (yellow bar). The data are shown as mean and SD.

In addition, the energy intake (kJ) was calculated based on the digestible energy of the chow and high-fat diet (15.6 and 27 kJ/g, respectively) as stated by the supplier (Specialty Feeds Pty Ltd., NSW Australia). No specific pattern was identified regarding the energy intake in either

group (**Figure 3.11A**). However, the overall trend was significantly positive in CTR (P=0.001) and OSC (P=0.037) group (**Figure 3.11B**). The trend was almost flat in HFD (P=0.67) (**Figure 3.11B**), similar to the dietary intake trend in this group (**Figure 3.7B**). Two-way ANOVA showed the significant effect of time (P<0.0001) but not the type of diet (P=0.52) on the energy intake. This analysis was followed by the multiple comparison test using Benjamini, Krieger and Yekutieli correction method to see the difference in each week (**Table 3.5**). Apart from few exceptions, there was no significant difference in the energy intake from different diets throughout the study (**Table 3.5**).



B



C

	Week vs. CTR	Week vs. HFD	Week vs. OSC
Pearson r			
r	0.6505	0.09560	0.4462
95% confidence interval	0.3153 to 0.8413	-0.3397 to 0.4972	0.03031 to 0.7304
R squared	0.4231	0.009140	0.1991
P value			
P (two-tailed)	0.0010	0.6721	0.0374
P value summary	**	ns	*

	CTR	HFD	OSC
Slope	1.071	1.439	1.203

	CTR	HFD	OSC
P value	<0.0001	0.6100	0.0001

Figure 3.11 Weekly energy intake throughout the dietary intervention

A) Each dot point indicates the average energy intake (kJ) per mouse. The highlighted yellow sections show the weeks that the OSC group were exposed to HFD. No specific pattern identified in any group. B) The positive trend in energy intake of CTR and OSC group was not observed in HFD. There was a significant difference between the slope of the trend lines (P = 0.0103). The data are shown as mean and SEM. C) The table summarises the results of correlation and linear regression analysis.

Table 3.5 Energy intake (kJ) comparison between the three groups

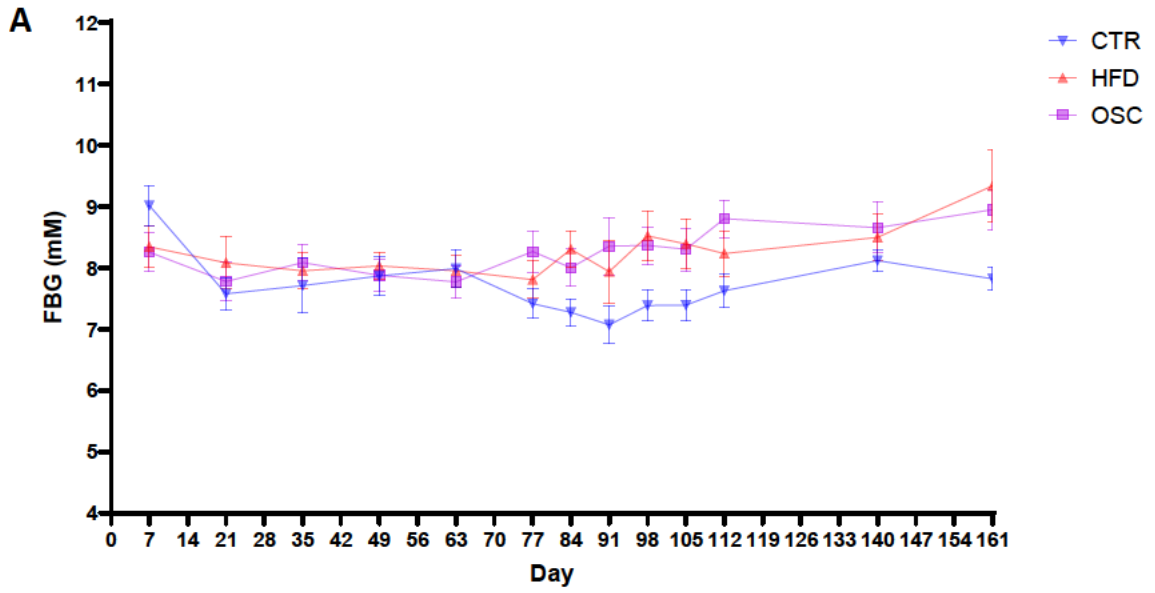
The significant P value of the multiple comparison test with Benjamini, Krieger and Yekutieli correction is shown in red. The yellow-highlighted weeks are those in which the mice in the OSC group were exposed to HFD. CTR: control group on chow diet, HFD: high-fat diet, OSC: oscillating Chow/HFD

Week	HFD vs CTR	OSC vs CTR	OSC vs HFD
1	0.33	**0.04**	0.06
2	**0.02**	0.25	0.86
3	0.06	0.92	**0.05**
4	**0.00**	0.19	**0.01**
5	0.05	**0.01**	0.86
6	0.41	0.21	0.77
7	0.73	0.72	0.95
8	0.87	0.78	0.66
9	0.91	0.69	0.74
10	0.69	**0.01**	0.05
11	0.73	0.99	0.68
12	0.50	0.80	0.57
13	0.81	0.47	0.51
14	0.05	0.32	0.13
15	0.59	0.65	0.88
16	0.67	0.27	0.64
17	0.64	0.11	**0.03**
18	0.98	0.14	0.51
19	0.65	0.87	0.75
20	0.64	0.90	0.75
21	0.76	0.42	0.31
22	0.48	0.44	0.97

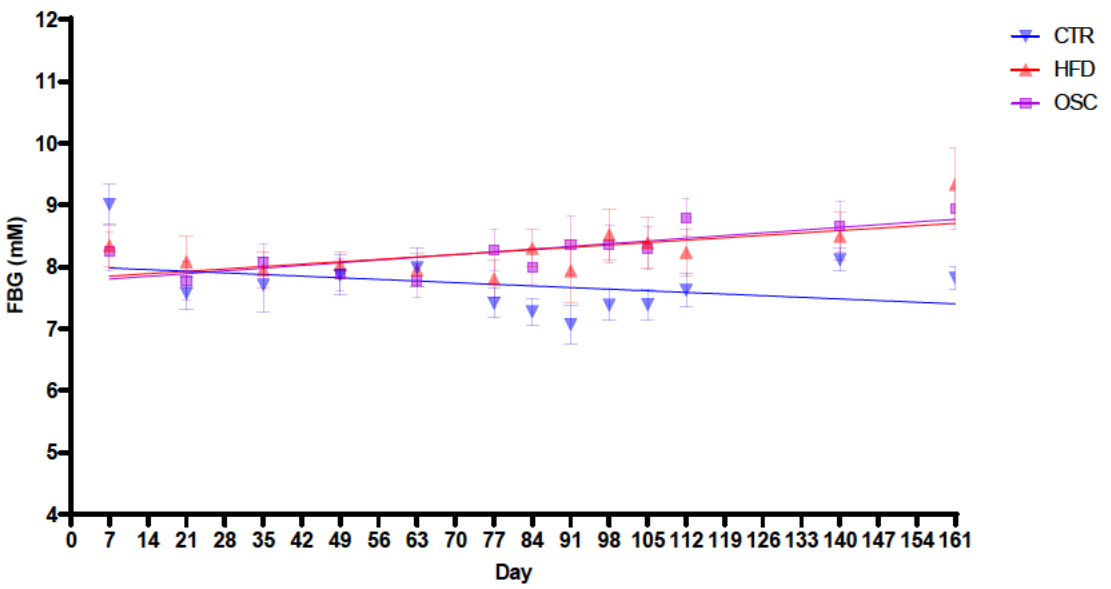
Furthermore, one-way ANOVA showed a non-significant difference between the cumulative energy intake (kJ) of the diet groups using (P=0.53). The following multiple comparison test also revealed no significance difference among the groups (P=0.27, P=0.63, and P=0.54 in HFD vs CTR, OSC vs CTR, and OSC vs HFD, respectively). This was further analysed in male and female mice without obtaining a significant difference (male: P=0.1, P=0.66, and P=0.15 in HFD vs CTR, OSC vs CTR, and OSC vs HFD, respectively) (female: P=0.9, P=0.95, and P=0.95 in HFD vs CTR, OSC vs CTR, and OSC vs HFD, respectively).

3.3.2.3 Blood glucose level

The fasting blood glucose (FBG) throughout the dietary intervention revealed distinctive patterns among the groups (**Figure 3.11**). The mean FBG values were similar in the three groups until day 63. The FBG values of the HFD and OSC groups started to cluster together from day 77 towards the end of the dietary intervention (**Figure 3.11**). Also, their trend lines were overlapping. This high level of blood glucose was indicative of possible insulin resistance in these two groups. FBG in HFD and OSC was significantly correlated with time (Day 7-161) (HFD: Pearson r: 0.64, P=0.0188; OSC: Pearson r: 0.78, P=0.0017), and not in the CTR group as expected (Pearson r: -0.33, P=0.27) since CTR group was being exposed only to the healthy chow diet.



B



C

	Day vs. CTR	Day vs. HFD	Day vs. OSC
Pearson r			
r	-0 3309	0.6387	0.7778
95% confidence interval	-0.7459 to 0.2692	0.1354 to 0.8800	0.3970 to 0.9302
R squared	0.1095	0.4080	0.6050
P value			
P (two-tailed)	0.2695	0 0188	0.0017
P value summary	ns	*	**

	CTR	HFD	OSC
Slope	-0.003765	0.005544	0.006288

	CTR	HFD	OSC
P value	0.0488	0.0238	0.0026

Figure 3.12 Fasting blood glucose throughout the dietary intervention

A) The line graph of mean FBG (mM) in the three groups. B) The trend line for HFD and OSC groups had a significant rising slope. There was a significant difference between the slope of the trend lines (P = 0.0011). The data are shown as mean and SEM. C) The table summarises the results of correlation and linear regression analysis.

Table 3.6 shows the FBG comparison between the three diets at baseline (Day 7), endpoint (Day 161), and interval time points throughout the intervention. The significant differences started to show almost mid-way through the dietary intervention on day 77. This was the first sign of insulin resistance observed in the OSC but not in the HFD group. Therefore, the significantly higher than normal blood glucose level appeared earlier in OSC than HFD group.

Also, towards the end of the program, the significant differences were seen more in OSC vs CTR, than in HFD vs CTR group (red highlighted values in **Table 3.6**).

Table 3.6 FBG (mM) comparison between the three intervention groups

The significant P value of the multiple comparison test with Benjamini, Krieger and Yekutieli correction is shown in red. The significant differences started from day 77. The yellow-highlighted weeks are those in which the mice in the OSC group were exposed to HFD. CTR: control group on chow diet, HFD: high-fat diet, OSC: oscillating Chow/HFD

Day	HFD vs CTR	OSC vs CTR	OSC vs HFD
7 (Baseline)	0.16	0.1	0.86
21	0.32	0.63	0.57
35	0.66	0.49	0.74
49	0.68	0.98	0.65
63	0.93	0.61	0.64
77	0.32	**0.048**	0.32
84	**0.01**	0.06	0.49
91	0.16	**0.03**	0.55
98	**0.02**	**0.02**	0.76
105	**0.048**	**0.04**	0.87
112	0.2	**0.01**	0.25
140	0.73	0.51	0.78
161	**0.02**	**0.01**	0.56

One-way ANOVA analysis showed an overall significant difference in FBG level (mM) on day 161 among the three diet groups (P=0.023). However, the difference was not significant in separate analysis for male (P = 0.26) and female mice (P = 0.087). The results of multiple comparison analysis are shown in **Table 3.7**. The only significant difference was related to higher FBG level in female mice exposed to HFD compared with CTR (P=0.03) (**Table 3.7**).

Table 3.7 FBG (mM) comparison between the three diet groups on day 161

The significant P value of the multiple comparison test with Benjamini, Krieger and Yekutieli correction is shown in red. CTR: control group on chow diet, HFD: high-fat diet, OSC: oscillating Chow/HFD

	Male	Female
HFD vs. CTR	0.13	**0.03**
OSC vs. CTR	0.21	0.18
OSC vs. HFD	0.76	0.49

This was further analysed using two-way ANOVA to check the effect of sex and diet on FBG. The results showed the significant effect of diet (P=0.023) but not sex (P=0.66) on the final FBG. This was followed by multiple comparisons test with Benjamini, Krieger and Yekutieli correction as shown in **Table 3.8**. The only significant difference within a specific sex was between female mice exposed to HFD compared to CTR (P=0.04).

Table 3.8 FBG (mM) comparison on day 161 in male and female mice

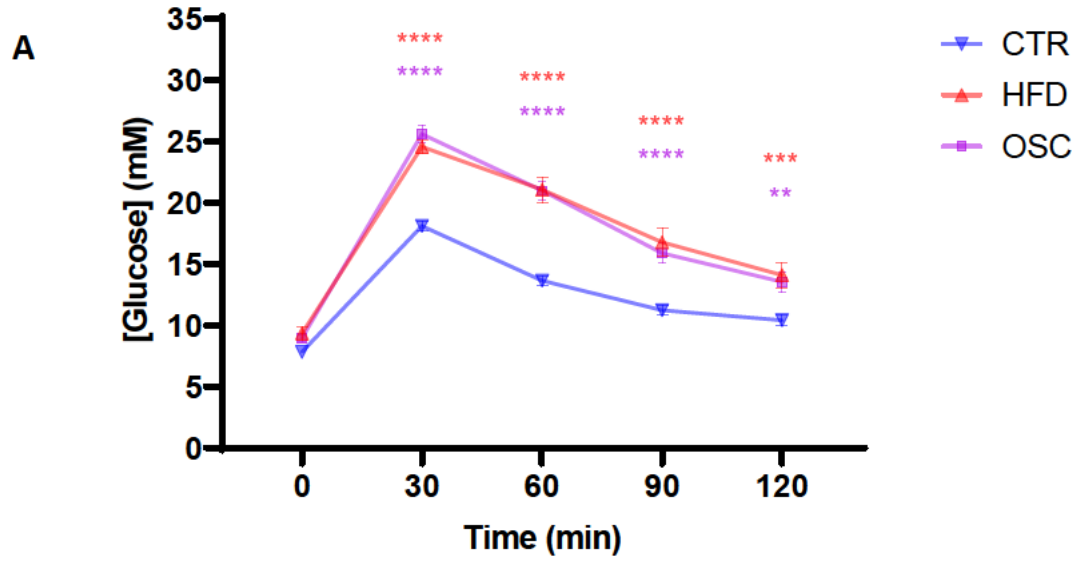
The significant P value of multiple comparison test with Benjamini, Krieger and Yekutieli correction is shown in red. CTR: control group on chow diet, HFD: high-fat diet, OSC: oscillating Chow/HFD

Comparison	P value
Male:CTR vs. Male:HFD	0.10
Male:CTR vs. Male:OSC	0.18
Male:CTR vs. Female:CTR	0.69
Male:CTR vs. Female:HFD	0.09
Male:CTR vs. Female:OSC	0.37
Male:HFD vs. Male:OSC	0.74
Male:HFD vs. Female:CTR	**0.04**
Male:HFD vs. Female:HFD	0.99
Male:HFD vs. Female:OSC	0.52
Male:OSC vs. Female:CTR	0.08
Male:OSC vs. Female:HFD	0.74
Male:OSC vs. Female:OSC	0.73
Female:CTR vs. Female:HFD	**0.04**
Female:CTR vs. Female:OSC	0.21
Female:HFD vs. Female:OSC	0.53

The glucose tolerance test (GTT) at day 161 provided supportive evidence for the presence of insulin resistance (**Figure 3.12**). The area under the curve (AUC) of GTT was 52.05, 74.08, and 73.69 for CTR, HFD, and OSC groups, respectively. The same AUC value for OSC and HFD

groups as well as similar patterns in their GTT curve were indicative of induced insulin resistance in the OSC group similar to HFD.

Two-way ANOVA showed that in all the four time points after glucose injection (30, 60, 90, and 120 min), there was a significant difference between the blood glucose level in CTR vs. HFD and CTR vs. OSC group, but not in HFD vs. OSC group comparison (**Figure 3.12**). Comparing the GTT at 0 and 120 min resulted in significant P values in all three diet groups (P=0.0034, P<0.0001, and P<0.0001 in CTR, HFD, and OSC group, respectively).



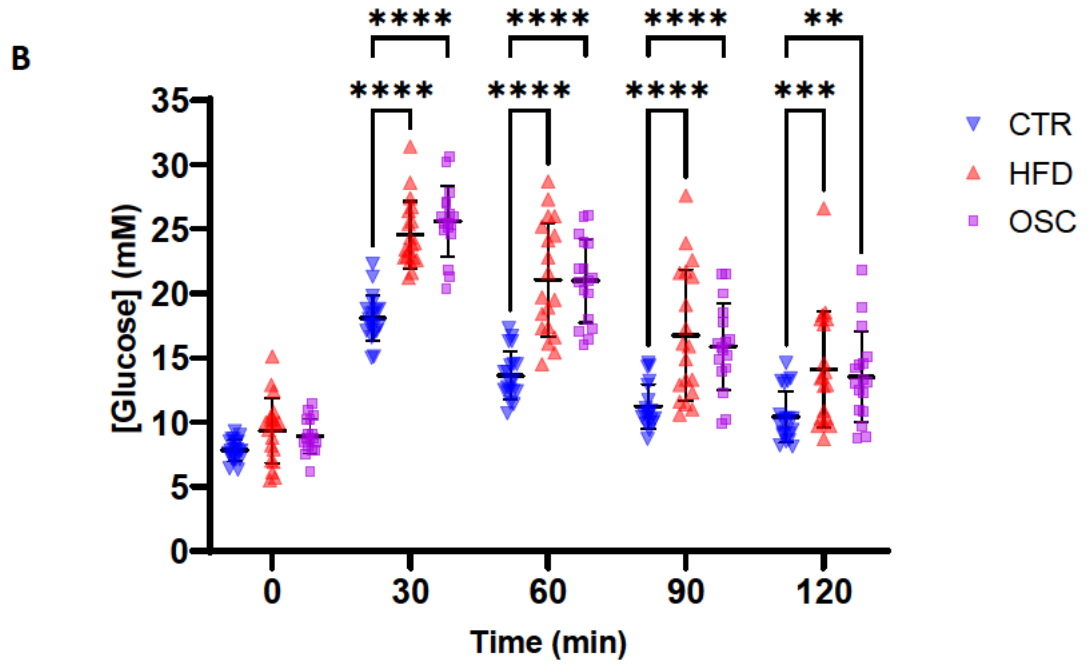


Figure 3.13 Glucose tolerance test on day 161

A) GTT curves in the three groups (mean±SEM). In each time point, the red and purple asterisks on the graph show the P-value for comparing CTR vs. HFD, or CTR vs. OSC group. B) Same GTT data grouped by time points (mean±SD).

3.3.2.4 Behavioural tests

The three behavioural tests (righting reflex, wire hang, and olfaction) were performed at baseline (before the dietary intervention started), during weeks 9, and 21. As per protocol, the data of each test at each time point was compared between male and female mice to investigate the possible effect of sex on the results.

There were no significant differences between male and female mice in either time point for the righting reflex test (P = 0.36, 0.1, and 0.8, for baseline, week 9, and 21, respectively). In addition, no significant difference was observed between the three dietary groups at the baseline and in week 9 (**Figure 3.13**). In week 21 however, mice in the HFD and OSC groups were significantly slower in reverting to the normal position, when compared to their

counterparts in the CTR group (P=0.0028, and P=0.0021 in HFD vs. CTR and OSC vs. CTR, respectively) (**Figure 3.13**). Also, in week 21, the mean reaction time was very close in HFD vs. OSC comparison (P=0.98). This indicated a similar disadvantageous effect of both the HFD and OSC diet on the neuromuscular function in mice. Additionally, comparing the time points in each dietary intervention group revealed significant differences between week 21 and the baseline in both HFD and OSC groups (**Figure 3.14**).

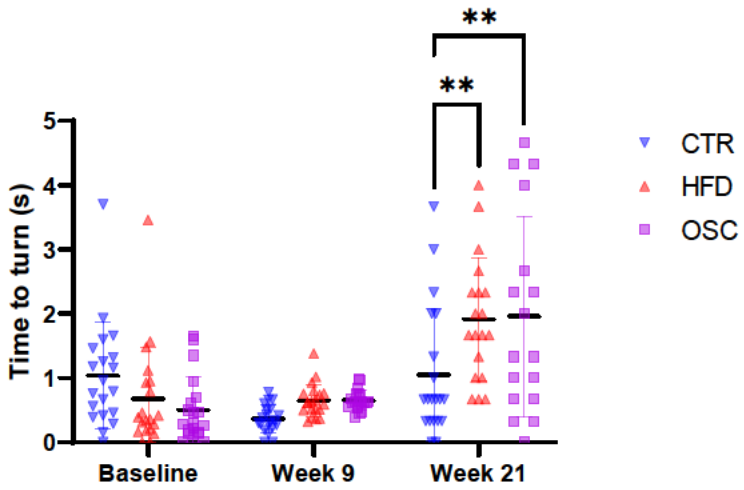


Figure 3.14 Righting reflex test

The only significant difference was observed in HFD vs. CTR and OSC vs. CTR in week 21. The data are shown as mean and SD.

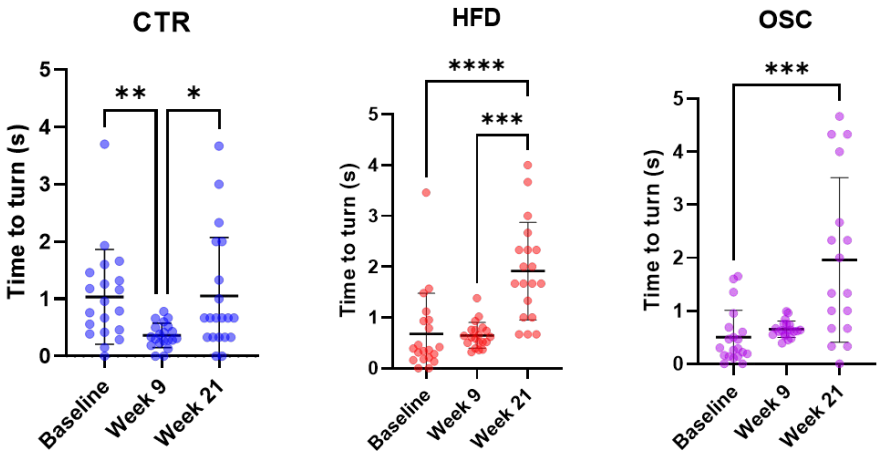


Figure 3.15 Righting reflex results in each dietary intervention group

Apart from significant differences observed in the HFD and OSC groups, the significant differences in the CTR group were related to high reaction time at the baseline and lower in week 9 in that group. The data are shown as mean and SD.

The wire-hang test data revealed significant differences between male and female mice at all time points (P<0.0001). Therefore, the analyses were performed in male and female mice separately (**Figure 3.15**). In female mice, no significant differences were found between the diet groups at either time point. Curiously, the male mice in the HFD group at the baseline (before the dietary intervention started) were significantly less resistant to falls. Nevertheless, at week 21, the mean level of grip time was almost equal between OSC and HFD groups (**Figure 3.15**). Moreover, analysing the results in each dietary intervention group revealed that male mice on the OSC diet became significantly weaker in wire hang grip in weeks 9 and 21, compared to the baseline (**Figure 3.16**). No other significant difference was observed in either sex in other diet groups.

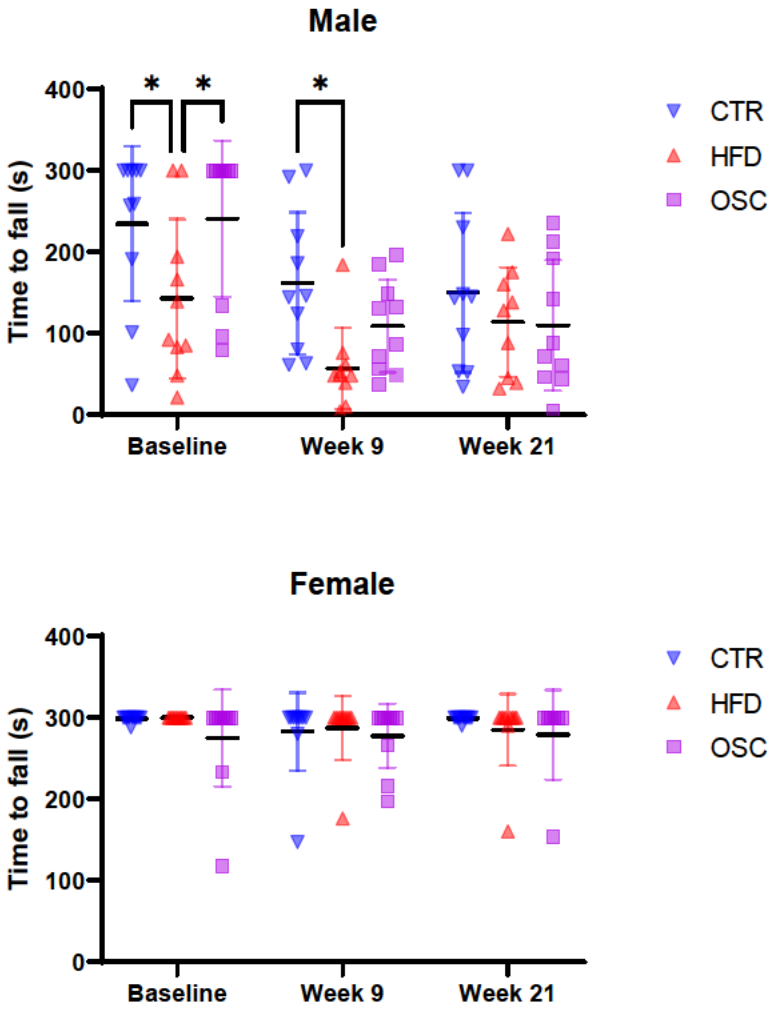


Figure 3.16 Wire hang test in male and female mice

The data are shown as mean and SD.

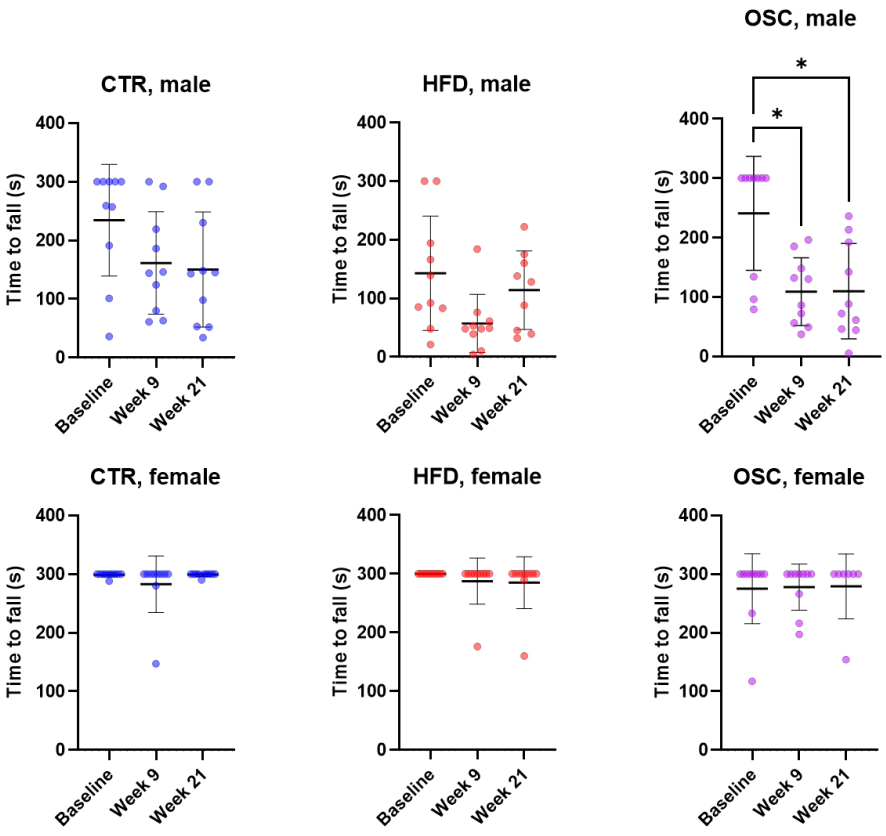


Figure 3.17 Wire hang test in each dietary intervention group
The data are shown as mean and SD.

The nosing index was not significantly different between male and female mice in either time point (P = 0.36, 0.068, and 0.91, for baseline, week 9, and 21, respectively). In week 21, the nosing index was lower in OSC compared to CTR and HFD, however, this difference was not significant (**Figure 3.17**). Further analyses per each diet group showed that the nosing index in mice exposed to OSC was significantly lower in week 21, compared to baseline, and week 9 (**Figure 3.18**). These differences were not observed in either CTR or HFD group, indicative of the disadvantageous effect of the oscillating diet on the olfaction system in mice.

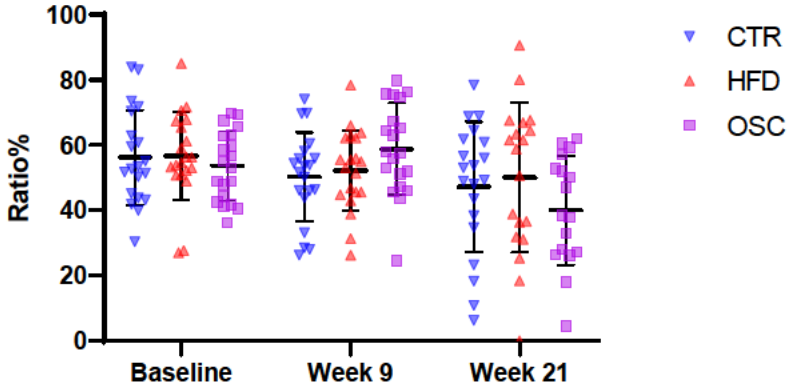


Figure 3.18 Nosing index as the olfaction representative
The data are shown as mean and SD.

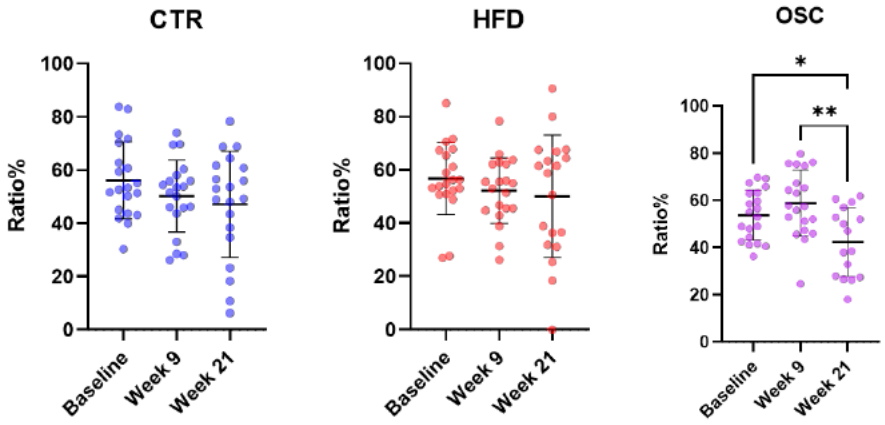


Figure 3.19 Nosing index in each dietary intervention group
The data are shown as mean and SD.

3.4 Discussion

The two main progressive conditions involved in Type 2 diabetes aetiology are insulin resistance and pancreatic β-cell dysfunction[5,6]. Diet is the main factor affecting these two conditions[29,30]. Unhealthy high fat is linked to insulin resistance via altered gut microbiome

composition and their generated metabolites[278-280]. On the contrary, a healthy diet constitutes rich sources of fibres which are fermented by the gut microbiome to generate SCFAs[281,282]. These molecules have a beneficial effect on metabolic health including insulin sensitivity and pancreatic function[283-285]. The SCFA-sensing on gut EECs leads to incretin hormone secretion, such as GLP-1, GIP, CCK, PYY, and ghrelin[285,286]. These hormones regulate the appetite and insulin secretion from pancreas islets[96,108].

Not following a healthy lifestyle is the main cause of the obesity epidemic worldwide[9,287]. This has propelled the increasing rate of metabolic disorders, such as T2D, and cardiovascular diseases[288-290]. Weight cycling caused by the yoyo diet is a new lifestyle adopted by many individuals to tackle obesity[117,119,291]. Many who are not happy with their body weight and/or shape go on a repeating cycle of weight gain and loss by going on a diet, losing weight, quitting the diet, regaining weight, and adopting the diet again several times[117,119,291]. To understand the effect of the yoyo diet on metabolic health, we employed *in vitro* and *in vivo* models. Several studies have identified the effect of SCFAs on incretin hormone levels[292-294], and sodium butyrate was utilised as the representative of a healthy diet in our cell study.

We found an interesting pattern in the gene expression level, which was not fully proportional to the butyrate level. The oscillating pattern of the gene expression level of *GCG* (producing GLP-1 hormone), *GIP*, *CCK*, *PYY*, and *GHRL* (producing ghrelin hormone) revealed a sharp rise and decline in the first cycles, however, the amplitude decreased over time. This was also reflected when we analysed our data relative to the pre-conditioning state before the exposure started.

These results suggested that in yoyo dieting, people may not obtain the beneficial metabolic effects of SCFAs in later cycles despite consuming a great amount of fibre. Therefore, the yoyo diet may not eventually be considered a beneficial strategy for metabolic health[118,119].

Replicating the oscillating alterations of healthy and unhealthy diets (using chow and high-fat diets, respectively) helped us assess the blood glucose level longitudinally as a metabolic health index. We surprisingly identified insulin resistance halfway through the dietary intervention in the OSC group, earlier than HFD. This occurred despite the lack of interest of mice in the OSC group to eat a high fat diet. Additionally, GTT results were almost identical between OSC and HFD groups indicating the harmful effect of yoyo dieting on metabolic health despite consuming a healthy diet in between.

These findings were also reflected in the animals' behavioural tests showing the malfunction in the neuromuscular weakness in OSC similar to the HFD group, but not in the controls. Surprisingly, the nosing index, as the representative of olfactory system function, dropped significantly only in the OSC group. These are indicative of possible diabetic neuropathy in the OSC group, similar to previous reports in DIO mice[125,295] and T2D mouse model[296]. Further analyses of the tissues collected from the mice after euthanasia could provide a better understanding of the impact of the yoyo diet on metabolic health and neuromuscular functions.

The current thesis chapter was part of a larger study already being carried out in Prof. Hardikar's group, to investigate the role of dietary oscillation in Type 2 Diabetes progression. To the best of my knowledge, attempts to assess the oscillating fashion of high and low butyrate levels on gut hormones levels, are not well-studied longitudinally. The same oscillating (cycling) approach was employed in our animal model of dietary intervention.

The *in vitro* experiments on two colonic cell lines (HT-29 and T84) and the mouse dietary intervention were replicated/conducted by me during this time. Also, several factors were measured during the mouse study including body weight, food intake, FBG, as well as behavioural tests. GTT was also measured at the endpoint. A large number of tissues were collected from mice at the end of the animal study to investigate the impact of dietary oscillation further, and to investigate their association with the *in vitro* findings. These form continuing studies within the Diabetes and Islet Biology Group and analyses on all of these are not part of my PhD. Future experiments on the mice samples aim to measure multiple variables including:

- The gene expression level of incretin hormones in the ileum, cecum, and colon epithelial cells
- Insulin and glucagon gene expression levels in the pancreas and insulin receptors in the liver
- Insulin level in the serum
- Gut microbiome composition and their generated metabolites in the faecal samples collected throughout the study and on the ileum, cecum, and colon content collected at the endpoint
- Gene expression and protein level of factors involved in neuromuscular function in several brain tissues, and identifying their potential correlation with the results of behavioural tests

Apart from the mouse study, the *in vitro* model studied by Dr. Joglekar's and Prof. Hardikar's teams analyses epigenetic changes and chromosomal accessibility/ATAC-seq at different time points of exposure to butyrate. Single-cell RNA sequencing studies are planned to be

performed to analyse the effect of oscillating SCFAs on gene sets. All the results obtained from the cell and animal model could combine to broaden our understanding of the potential disadvantageous effects of dietary oscillation on metabolic health.

Chapter 4 - The Link between Short-chain Fatty Acids Concentrations and Colorectal Cancer

This chapter is published as an article in *BMC Medicine* 2022 Oct 3;20(1):323 (PMID: 36184594, PMCID: PMC9528142, DOI: 10.1186/s12916-022-02529-4), titled: **Short-chain fatty acid concentrations in the incidence and risk-stratification of colorectal cancer: a systematic review and meta-analysis**

The format of the article is modified to conform to the thesis.

4.1 Background

According to the Global Cancer Incidence, Mortality and Prevalence (GLOBOCAN) 2020 report, colorectal cancer (CRC) is the third-most commonly diagnosed cancer (10% of all diagnosed cancers) and the second (9.4%) leading cause of cancer-related death [7]. It has been estimated that the overall risk of CRC in all age groups will increase 60% worldwide by 2030, leading to more than 1.1 million deaths and 2.2 million new cases [8]. Colorectal cancer develops from precursor lesions collectively known as colorectal adenomas (CRA), in the form of adenomatous polyps or to a lesser extent (10-20%) serrated polyps [131,133]. It is a heterogeneous disease and environmental factors have a potential impact on the development of CRC, among which diet is a risk factor [113,132,133]. According to several meta-analyses, high consumption of processed and unprocessed meat is related to high CRC risk

[155,156], and high fibre intake is suggested as a protective factor against CRC progression and incidence [157-159].

The effect of diet on colonic health is partly mediated through alteration of gut microbiota composition, diversity, and metabolism [107,113]. Gut microbiota constitutes the largest community of commensal microorganisms in the body, which mainly resides in the lower small intestine and colon [107,112,113]. The gut microbiota-derived metabolites are in constant crosstalk with colonocytes, and short-chain fatty acids (SCFAs) make up a large group of these metabolites [107,112,113].

Short-chain fatty acids are small molecules generated via the fermentation of dietary fibres by gut microbiota. Acetic, propionic and butyric acid constitutes the majority of colonic SCFA content [103,105] and the beneficial anti-inflammatory and anti-carcinogenic effects of dietary fibres on colonocytes are mediated through these SCFA molecules [102,106]. Among the three major SCFA molecules, butyric acid is also considered as one of the main energy sources for colonocytes [102,103,107]. Therefore, alteration in SCFA levels could impact the colonic health and predisposition of colonocytes to aberrant proliferation and tumour formation [105,106].

Several studies have assessed faecal SCFA concentration in patients with colorectal carcinoma or adenoma [162,163,165-179]. However, due to variable results, the conclusive evaluation of SCFA profiles from CRC patients versus healthy subjects is lacking. In addition, other studies have compared SCFA concentration within healthy individuals from various countries and ethnic groups with the highest and lowest prevalence of CRC; although with inconsistent results [180-185].

Therefore, systematic analyses designed to better understand the link between SCFA concentration in CRC risk and incidence is highly desired. We divided our analyses on the

available evidence into two broad categories; (1) CRC-risk and (2) incidence. We aimed to systematically analyse the results of all primary observational human studies, which measured faecal SCFA levels in "at-risk" individuals or in CRC patients. In the CRC risk category, the focus was on at-risk individuals, which was further sub-divided into two groups based on (1a) studies that analysed clinical data (presence of colorectal adenomas) or (1b) those that assigned CRC risk based on non-clinical evaluation of study participants (ethnic background or location). The CRC incidence category included studies that compared faecal SCFA levels in individuals with clinically diagnosed CRC and healthy individuals. Our results underline the potential association of the three major SCFA molecules (acetic, propionic, and butyric acid) with CRC risk and incidence.

4.2 Methods

We used Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guideline [297,298] to systematically search and extract data from primary human studies with SCFA measurement in CRC risk or incidence.

4.2.1 Database Search

The Medline, Embase, and Web of Science database search was performed for articles involving human subjects that are in English from database conception until 29st June 2022. The details of the search keywords and strategies utilized in Ovid and Web of Science are available in the **Supplementary Methods** section.

4.2.2 Eligibility Criteria

All the records, including abstracts, were imported to EndNote X9 (Clarivate Analytics, Toronto, Canada). Duplicate records were first removed. The records were then filtered using EndNote's built-in search tool for the following criteria: (i) searching for concentration*, level*, quanti*, measure*, assess*, evaluat*, estimat*, calculat*, mmol, and µmol as the inclusion criteria to capture studies which reported the SCFA measurement based on these terms. The asterisk symbol (*) applied was to include all the variations of the search terms, and (ii) searching for mouse, mice, murine, rats, conference, ethyl acetate (EtOAc), and phorbol as the exclusion criteria to exclude rodent studies, conference proceedings, and studies that have stated the use of any unrelated chemicals (such as EtOAc and 12-O-Tetradecanoylphorbol-13-acetate). The abstracts of the remaining records were then screened to exclude reviews, methodology, human studies not related to SCFAs in CRC or CRA, non-human studies (i.e. *in vitro* or other non-rodent animal studies), to identify the human studies on SCFA measurement in CRC or CRA. The full text of the remaining (n = 57) records were then screened to include only the observational studies which have measured faecal SCFA concentration. A final set of 23 observational studies qualified for further data extraction and quality assessment for meta-analysis.

4.2.3 Data Extraction and Quality Assessments

The data and additional details available for analysis (such as study subjects and SCFA levels) from the finalized primary studies were extracted and added to an Excel worksheet. The Newcastle-Ottawa Scale (NOS) [299] was used as a standard tool for quality assessment of 17 case-control studies in the selection, comparability and exposure categories, to provide a score range between 0-9 (≤ 6, 7-8, and 9 indicate high, medium, and low risk of bias,

respectively) [298]. Evaluation of six cross-sectional studies was performed using the Joanna Briggs Institute (JBI) Critical Appraisal Checklist tool [300], as recommended [301].

4.2.4 Statistical analyses

Review Manager (RevMan) software version 5.4 (Cochrane, Copenhagen, Denmark) was used to analyse the quantitative faecal SCFA concentration data, which were available in 10 of the final 23 observational studies (9 of 17 case-control, plus 1 of 6 cross-sectional studies). The faecal concentration of acetic, propionic, or butyric acid was considered as the subgroups. Before data entry, SEM or 95% CI upper and lower bound values were converted to SD. Due to variation in the reported SCFA concentration units between different papers, standardized mean difference (SMD) was selected as a measure of effect size for each study. The statistical heterogeneity among studies was calculated using Chi^2 and I^2 tests and a P-value of 0.05 was considered significant [302]. To normalise the use of different SCFA measurement methods, a random-effects model was applied to analyse the pooled effect size and P-value for each SCFA molecule in each subgroup. One overall effect size and P-value of combined acetic, propionic, and butyric acid were also calculated. In all analyses, the effect size was reported with 95% confidence intervals, and the P-value < 0.05 was considered significant. Furthermore, the fixed-effect model was also applied in the case of non-significant heterogeneity of $I^2 < 50$ [302]. All the data conversions, as well as qualitative and quantitative analyses, were validated by the second team member and confirmed by the senior authors.

One study [179] reported values that could not be converted to mean and SD for the meta-analysis, and was therefore only included in our qualitative analysis. Another study [180] reported the numeric values of butyric acid concentration and other SCFA molecules in graphs, hence was included in both quantitative (for butyric acid) and qualitative data (for

acetic acid, propionic acid and total SCFA). Therefore, in addition to studies in which the faecal SCFA concentration was presented using graphs (with no reported actual values), 14 of 23 studies were considered as qualitative studies (8 of 17 case-control, plus all 6 cross-sectional studies – including *Ocvirk et al. 2020*). The outcome of analyses from these qualitative studies was plotted as stacked bar charts, using Microsoft Excel (ver. 2016; Microsoft Corporation, Redmond, WA, USA).

4.3 Results

4.3.1 Study Selection and Quality Assessment

The workflow on the identification and stepwise selection of the observational studies is presented in **Figure 4.1**. Initially, a total of 2133 English language records obtained from searching through the three databases (Medline, Embase, and Web of Science) were imported to EndNote along with their abstracts. After removing duplicate records, the titles and abstracts of the remaining 1466 records were filtered and screened for eligibility as detailed in the Methods section. In total, 1409 records were excluded, of which most were *in vitro* studies. From the remaining 57 human studies, 34 studies were excluded. Of these 28 were interventional studies, three were observational studies on serum SCFA [303-305], two studies had indistinct grouping (one case-control study with the presence of individuals with adenomatous polyps in the healthy control group [164] and one cross-sectional study with no clear definition of CRC high- and low-risk group [306]), and one retracted observational study [307].

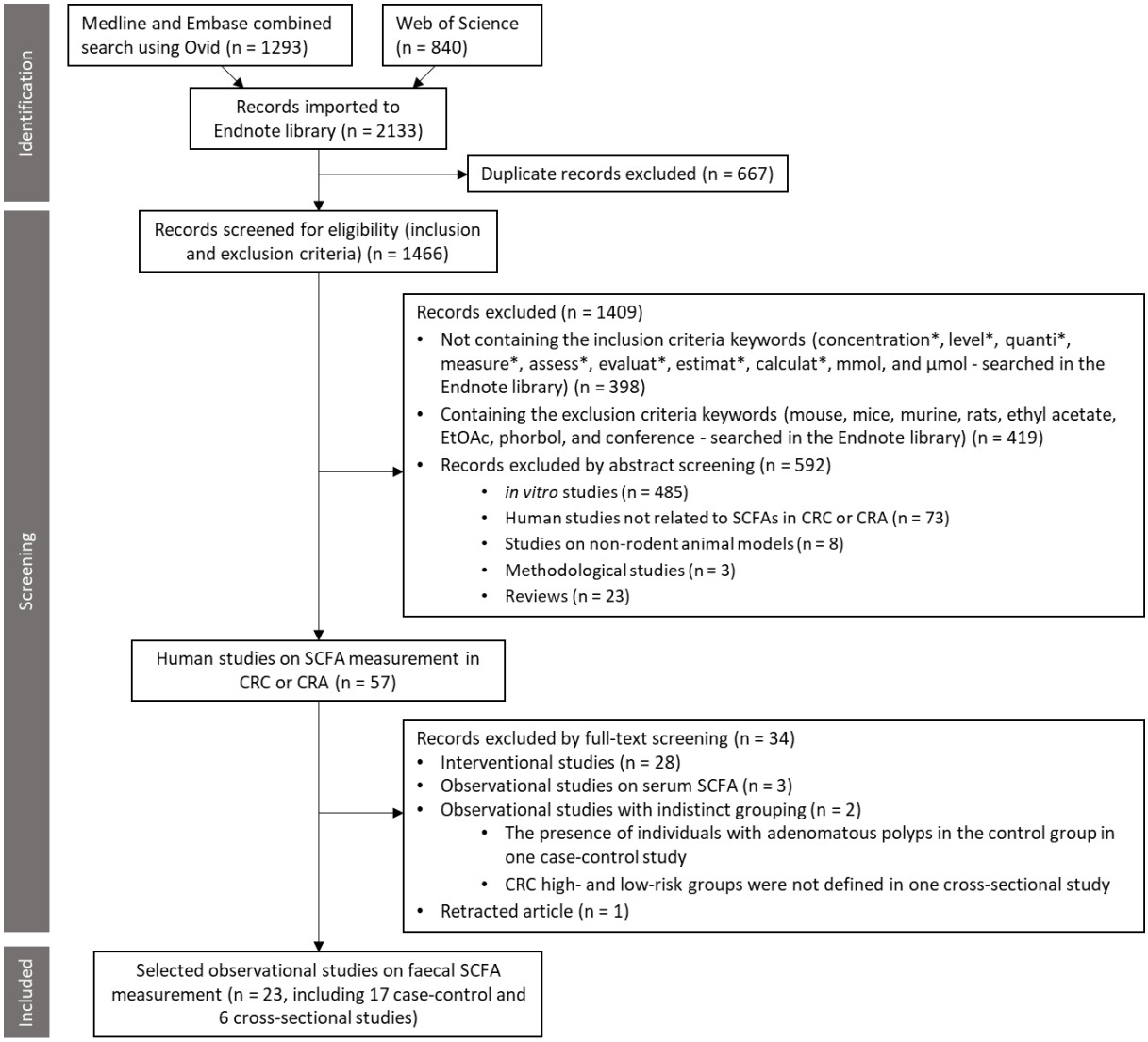


Figure 4.1 The PRISMA flowchart shows the selection process of the systematic review.

The abstracts of all the studies were imported into Endnote from the indicated databases. SCFA: short-chain fatty acid, CRC: colorectal cancer, CRA: colorectal adenoma.

Table 4.1 Characteristics of the selected studies.

Cross-sectional studies are highlighted in grey, and case-control studies are not highlighted. [¶]CRC: colorectal cancer, AP: adenomatous polyposis, CD: celiac disease, CRA: colorectal adenoma, HC: healthy controls, IBD: inflammatory bowel disease. [†]GC-MS: gas chromatography-mass spectrometry; HPLC: high-performance liquid chromatography; FL: fluorescence; 1H NMR: 1H nuclear magnetic resonance spectroscopy; UPLC-MS: ultra-performance liquid chromatography-tandem mass spectrometry; GLC: gas-liquid chromatography. [§]C2: acetic acid, C3: propionic acid, C4: butyric acid. [‡]Refer to the text for the definition of CRC risk and incidence category. [°]Values in this paper were measured on enema samples, not faeces. Therefore, they used in qualitative analysis. [*]More details are provided in the article. [‖]Removed from quantitative analysis as the reported SCFA values could not be converted to mean and SD. [¦]Removed from meta-analysis due to insufficient data on SCFA measurement method. [#]SCFAs were measured in only a subset of these subjects (n = 25 large/small adenoma and n = 23 adenoma-free). [Δ]Combined values of males and females.

	Study	Country	Study Population[¶] (No. of Subjects)	Age (y) Presented as Mean±SD, Median(IQR), Mean[min-max], Median[min-max], or [min-max]	Sex Male/Female	SCFA Measurement Technique[†] (Unit)	Measured SCFA[§]	Analysis Category[‡]
Qualitative data (values of SCFA levels not reported)	Sze et al. 2019	USA/Canada	CRC (120) vs CRA (198) vs HC (172)	[29-89] Median=60	-	HPLC (mmol/kg)	C2, C3, C4	Incidence and risk
	Lin et al. 2019	China	CRC (70) vs HC (70)	-	-	1H NMR (fold difference CRC vs HC)	C2, C3, C4	Incidence
	Lin et al. 2016	China	CRC (68) vs HC (32)	56±21 vs 57±23	36/32 vs 15/17	1H NMR	C2, C3, C4	Incidence
	Weir et al. 2013	USA	CRC (11) vs HC (10)	63.7±17.7 vs 40.7±14.6	8/2 vs 2/8	GC-MS	C2, C3, C4	Incidence
	Ohigashi et al. 2013	Japan	CRC (93) vs control adenoma (22) vs non-adenoma (27)	68.9±12.1 vs 66.6±9.2 vs 65.6±13.5	49/44 vs 11/11 vs 16/11	HPLC (µmol/g)	C2	Incidence and risk
	Monleon et al. 2009	Spain	CRC (21) vs HC (11)	63.2±12.6 vs 58.1±11.9	7/14 vs 2/9	1H NMR	C2, C3, C4	Incidence
	Weaver et al. 1988[*]	USA	Polyps and/or colon cancer [no diverticula (18) vs including diverticula (33)] vs diverticula [no polyps (39) vs including polyps and/or colon cancer (54)] vs colon cancer (including diverticula) (11) vs IBD (8) vs normal (35)	[62.5±13.9 vs 63.8±13.0] vs [62.9±13.0 vs 63.6±12.7] vs 67.7±7.5 vs 41.6±19.3 vs 54±12.5	-	GC (µmol/ml/day weight of enema sample)	C2, C3, C4, total	Incidence
	Ocvirk et al. 2020	USA/South Africa	Alaska native (32) vs Rural African (21)	51±50.3 vs 53.3±52.7	8/24 vs 9/12	GC (µmol/g)	C2, C3, total	Risk
	Katsidzira et al. 2019	USA/Zimbabwe	Urban (10) vs Rural (10) Zimbabweans	61.6±8.1 vs 65.3±10.0	5/5 vs 5/5	GC (µmol/g)	C2, C3, C4	Risk
	Hester et al. 2015	USA	African American (5) vs others (15) (American-Indian (5), Hispanic (5), White (5))	61.8[50-72] vs (59.4[50-75], 54.4[50-59], 63.8[57-74])	1/4 vs 2/3, 2/3, 1/4	GC (mg/ml)	C2, C3, C4, total	Risk
	Ou et al. 2013	USA/South Africa	African Americans (12) vs native Africans (12)	58±8.7 vs 57±6.6	3/9 vs 4/8	GC (µmol/g)	C2, C3, C4	Risk
	Ou et al. 2012	USA/South Africa	African-American (12) vs Caucasian-American (10) vs native African (13)	[50-60]	-	GC (µmol/g)	C2, C3, C4	Risk
	O Keefe et al. 2009	USA/South Africa	Caucasian-Americans (18) vs African-Americans (17) vs native Africans (17)	[50-65]	-	GC (mmol)	C2, C3, C4, total	Risk
Quantitative data (reported SCFA values)	Nannini et al. 2021[‖]	Italy	CRC (32) vs AP (16) vs HC (38)	72[36-85] vs 59[41-79] vs 47[27-68]	22/10 vs 9/7 vs 28/10	1H NMR (arbitrary unit)	C2, C3, C4	Incidence and risk
	Chen et al. 2021[¦]	China	Colorectal adenomatous polyps (30) vs HC (30)	53.23±10.14 vs 50.33±10.87	20/10 vs 13/17	Ion chromatography / UPLC-MS (mg/L)	C2, C3, C4	Risk
	Torii et al. 2019	Japan	CRC (15) vs HC (38)	73.13±4.49[66-82] vs [28-82]	8/7 vs 17/21	HPLC-FL (nmol/g)	C2, C3, C4	Incidence
	Niccolai et al. 2019	Italy	CRC (19) vs AP (9) vs CD (16) vs HC (16)	80(13.5) vs 68(28) vs 35.5(21) vs 46(9)	17/2 vs 5/4 vs 6/12 vs 14/2	GC-MS (µmol/g)	C2, C3, C4, total	Incidence and risk
	Yusuf et al. 2018	Indonesia	CRC (14) vs non-CRC (14)	53.8±13.3 vs 50±17.6	10/4 vs 9/5	GC-MS (µg/ml)	C2, C3, C4	Incidence
	Song et al. 2018	South Korea	CRC (26) vs HC (28)	59.7±12.2 vs 51.1±6.0	16/10 vs 22/6	GC-MS (µg/mg)	C2, C3, C4	Incidence
	Bridges et al. 2018[¦]	USA	Presence of AP (13) vs absence of AP (26)[*]	[50-75]	-	GC-MS (mg/ml)	C2, C3, C4, total	Risk
	Chen et al. 2013	China	Advanced-colorectal adenoma (344) vs without obvious abnormality (HC) (344)	[≥50][*]	174/170 vs 172/172	GC-MS (µg/L)	C2, C3, C4	Risk
	Boutron-Ruault et al. 2005	France	Adenoma (50 (large (18) and small (32) adenoma) vs adenoma-free (44)[#]	57.35±8.41 vs 52.5±8.77[Δ]	(14/4 and 21/11) vs 21/23	GC-MS (mmol/g)	C2, C3, C4, total	Risk
	Kashtan et al. 1992	Canada	Polyp (45) vs non-polyp (49)	61.3±10.1 vs 51.2±12.6	31/14 vs 26/23	HPLC (mmol/L)	C4, total	Risk
	Ocvirk et al. 2020	USA/South Africa	Alaska native (32) vs Rural African (21)	51±50.3 vs 53.3±52.7	8/24 vs 9/12	GC (µmol/g)	C4	Risk

Finally, 17 case-control and 6 cross-sectional studies were selected for data extraction and analysis. **Table 4.1** summarises the characteristics of these observational studies. The results of quality assessment using NOS and JBI tools on case-control and cross-sectional studies are provided in **Supplementary Tables 4.1** and **4.2**, respectively.

4.3.2 Stratifications Based on CRC Risk or Incidence

Studies listed in **Table 4.1** are presented based on the type of data provided (qualitative or quantitative) and CRC risk and/or incidence. Among the 17 case-control studies (not highlighted in **Table 4.1**), 8 studies comparing CRC cases and healthy control subjects were allocated to the CRC incidence category, 5 studies comparing individuals with CRA and healthy controls assigned to the CRC risk category, and the remaining 4 studies were included in both incidence and risk categories since they compared CRC patients, CRA individuals and healthy subjects. All 6 cross-sectional studies (highlighted grey in **Table 4.1**) comparing populations with high- versus low-risk of CRC were allocated to the risk category. Therefore, the CRC incidence and risk category included 12 and 15 studies, respectively (**Table 4.1**). For each study, the details of the measured SCFA and CRC risk and/or incidence grouping are provided in **Supplementary Table 4.3**. Some studies reported total SCFA concentration in addition to the individual (acetic, propionic, and butyric acid) SCFAs.

The primary studies analysed in this systematic review were performed in various countries and ethnic groups. Age was matched in some of the studies [168,173,174,183-185], although the male-to-female ratio was not similar between the study groups in most studies (**Table 4.1**). The SCFA concentrations were measured using different techniques, such as gas chromatography, liquid chromatography, gas-liquid chromatography and ^{1}H nuclear magnetic resonance spectroscopy.

4.3.3 Data Analyses

The meta-analysis of the quantitative data extracted from the 10 selected studies [166-172,174,178,180] are presented in **Figure 4.2**. In the risk category (**Figure 4.2A and B**), two studies [169,170] were excluded from the meta-analysis due to the lack of sufficient details of the methods used for SCFA measurement from stool samples. In CRC risk meta-analysis, the effect size of each of the three SCFAs was not statistically significant, however, their combined effect size was significantly higher in low-risk compared to high-risk CRC (SMD = 2.02, 95% CI 0.31 to 3.74, P = 0.02, **Figure 4.2A**). The effect size of total SCFA concentration was not statistically significant in the low- vs high-risk group **(Figure 4.2B**).

In the CRC incidence analysis (**Figure 4.2C**), the faecal concentrations of acetic acid (SMD = 0.61, 95% CI 0.09 to 1.13, P = 0.02) and butyric acid (SMD = 0.45, 95% CI 0.02 to 0.88, P = 0.04) were significantly higher in the healthy control compared to CRC cases. In addition, the combined effect size of acetic, propionic, and butyric acid remained significant between CRC cases and healthy controls (SMD = 0.45, 95% CI 0.19 to 0.72, P = 0.0009, **Figure 4.2C**).

Furthermore, the I^2 heterogeneity index was in the "moderate" range (30% to 60%) [302] for the meta-analysis of total SCFA concentration in CRC risk (**Figure 4.2B**) and butyric acid in CRC incidence (**Figure 4.2C**) category. Therefore, we performed another meta-analysis using the fixed-effect model on the same data instead of the random-effect model presented in **Table 4.2**. This resulted in a more pronounced difference in butyric acid concentration between CRC cases and healthy controls (SMD = 0.42, 95% CI 0.1 to 0.74, P = 0.009). The results of the fixed-effect model meta-analyses are presented in **Figures 4.3 and 4.4**, respectively, and the findings of all quantitative meta-analyses are summarised in **Table 4.2**.

Qualitative analysis was carried out on the studies which reported lower, higher or no changes to the concentration of SCFAs between high-risk CRC (for risk category) or CRC case (for incidence) and low-risk or control, respectively [162,163,165,173,175-177,179-185] (**Figure 4.5**). In the risk category, more studies (70.4%) reported significantly lower concentrations of faecal acetic, propionic, and butyric acid as well as total SCFA in individuals at high risk of CRC. In the incidence category, more studies (66.7%) reported significantly lower concentrations of faecal acetic and butyric acid in CRC patients compared to healthy controls. However, the number of studies reporting no significant difference in the propionic acid was the highest in the incidence category. Overall, our qualitative analysis (**Figure 4.5**) corroborates with the meta-analysis results (**Figure 4.2**).

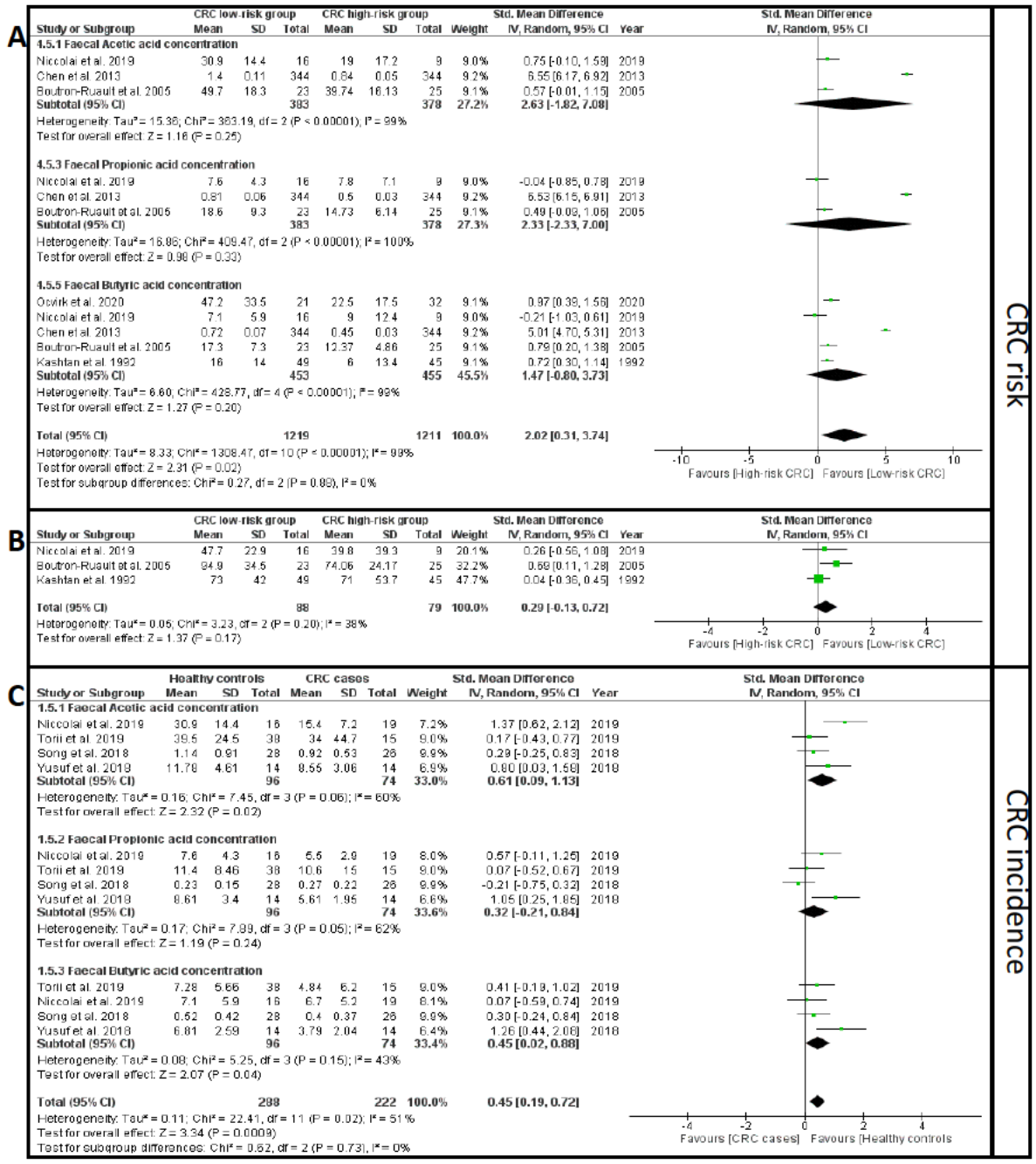


Figure 4.2 Forest plots representing the meta-analyses of the faecal SCFA concentrations.

A) acetic, propionic, and butyric acid in CRC risk category; **B)** total SCFA in CRC risk category; and **C)** acetic, propionic, and butyric acid in CRC incidence category. Note that in **B**, the total SCFA indicates the collection of all the SCFA molecules - not only acetic, propionic, and butyric acid.

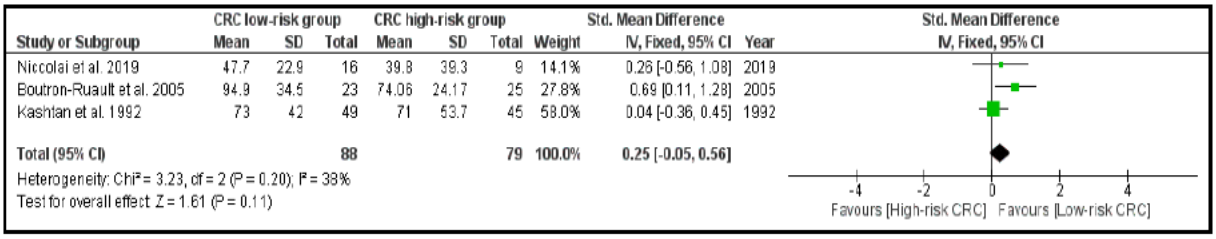


Figure 4.3 Fixed-effect model, CRC risk, total SCFA.

Forest plot representing the meta-analyses of the faecal total SCFA concentration in the CRC risk category using fixed-effect model.
Note that total SCFA indicates the collection of all the SCFA molecules - not only acetic, propionic, and butyric acid.

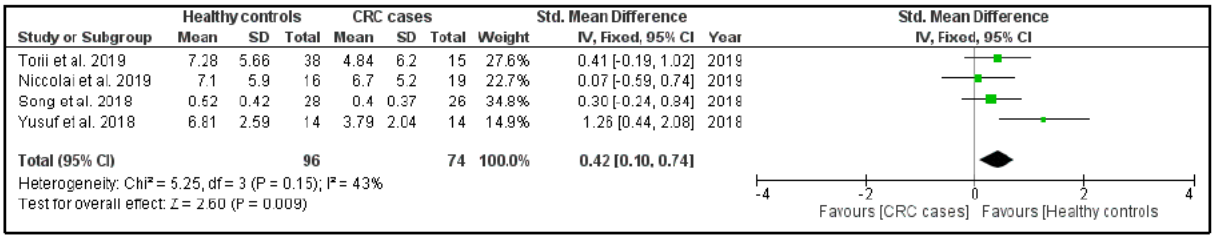


Figure 4.4 Fixed-effect model, CRC incidence, butyric acid.

Forest plot representing the meta-analyses of the faecal butyric acid concentration in the CRC incidence category using fixed-effect model.

Table 4.2 Summary of the outcomes of each meta-analysis.
Significant P values of the effect size are in bold. ¶Combined effect size of acetic, propionic, and butyric acid. Note that the total SCFA indicates the collection of all the SCFA molecules - not only acetic, propionic, and butyric acid.

	Measured SCFA	Number of Studies	Heterogeneity (I^2 %, P value)	Statistical Model	Effect size (SMD [95% CI], P value)
CRC risk	Acetic acid	3	99, < 0.00001	Random effect	2.63 [-1.82 to 7.08], 0.25
	Propionic acid	3	99, < 0.00001	Random effect	2.33 [-2.33 to 7.00], 0.33
	Butyric acid	5	99, < 0.00001	Random effect	1.47 [-0.80 to 3.73], 0.2
	Combined¶	11	99, < 0.00001	Random effect	2.02 [0.31 to 3.74], **0.02**
	Total SCFA	3	38, 0.2	Random effect	0.29 [-0.13 to 0.72], 0.17
	Total SCFA	3	38, 0.2	Fixed effect	0.25 [-0.05 to 0.56], 0.11

	Measured SCFA	Number of Studies	Heterogeneity (I^2 %, P value)	Statistical Model	Effect size (SMD [95% CI], P value)
CRC incidence	Acetic acid	4	60, 0.06	Random effect	0.61 [0.09 to 1.13], **0.02**
	Propionic acid	4	62, 0.05	Random effect	0.32 [-0.21 to 0.84], 0.24
	Butyric acid	4	43, 0.15	Random effect	0.45 [0.02 to 0.88], **0.04**
	Combined¶	12	51, 0.02	Random effect	0.45 [0.19 to 0.72], **0.0009**
	Butyric acid	4	43, 0.15	Fixed effect	0.42 [0.1 to 0.74], **0.009**

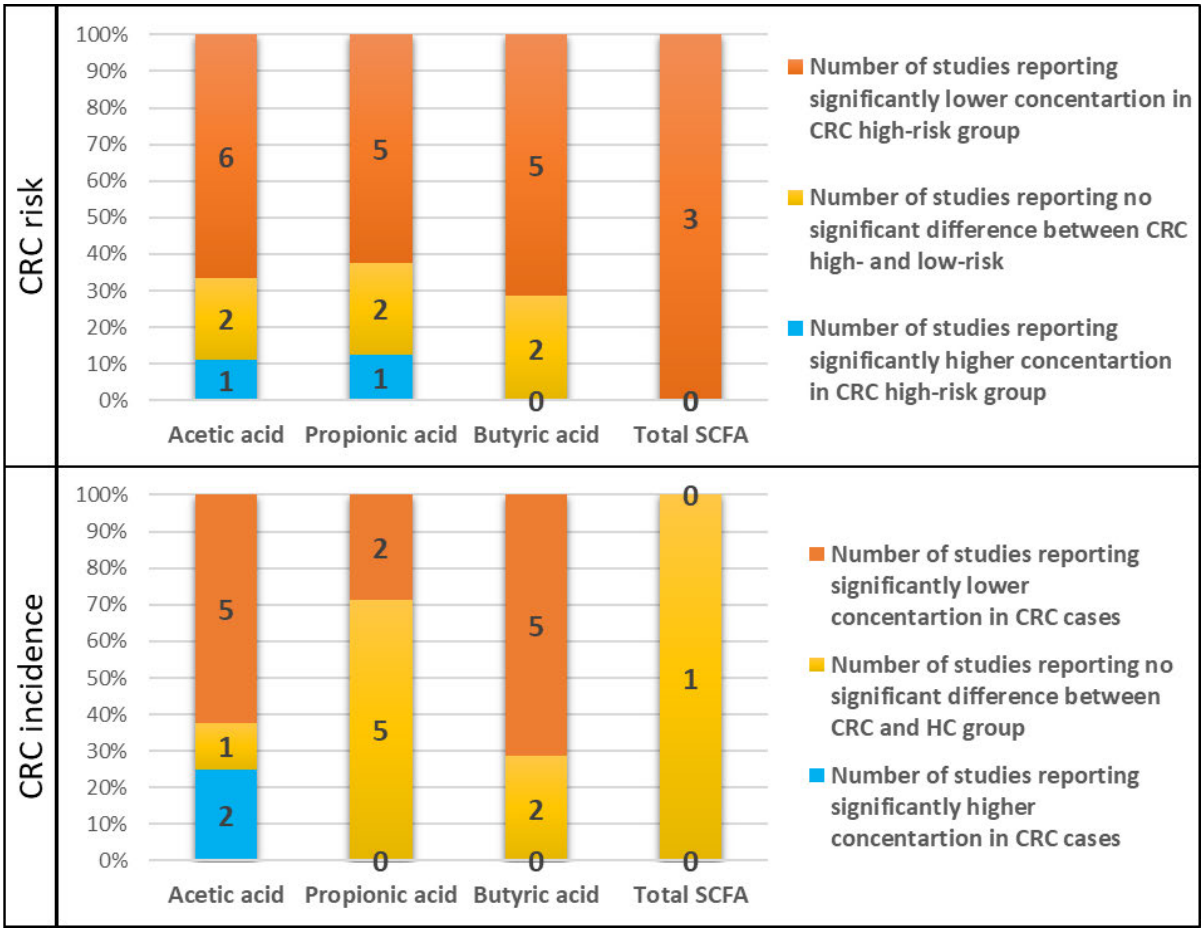


Figure 4.5 Graphical representation of faecal SCFA concentration.

Stacked bar charts summarizing the results of the qualitative data in CRC risk (9, 8, 7, and 3 studies have measured faecal concentration of acetic, propionic, and butyric acid, and total SCFA, respectively) and incidence categories (8, 7, 7, and 1 study have measured faecal concentration of acetic, propionic, and butyric acid, and total SCFA, respectively). HC: healthy controls.

4.4 Discussion

For more than three decades, *in vitro*, animal, and human studies have identified numerous potentially beneficial anti-inflammatory and anti-carcinogenic roles of SCFA molecules in gut health and colonic diseases [102,103,105,106,113,114]. In addition, several meta-analyses (**Supplementary Table 4.4**) have assessed the role of colonic microbiota [308], non-digestible carbohydrates [309] and dietary fibre in colorectal carcinoma [157,310] or adenoma [158,159] as well as the alteration of SCFAs in irritable bowel syndrome (IBS) [311], or inflammatory bowel disease (IBD) [312].

This systematic review and meta-analysis were conducted on 23 studies to better determine the potential association between faecal SCFA concentration and CRC risk and incidence. The combined mean difference of acetic, propionic, and butyric acid in the CRC risk category analysis revealed a significantly lower concentration of these SCFAs in individuals at risk of developing CRC compared to healthy subjects, indicating a potential association between these three major SCFA molecules and CRC development. This finding was further confirmed in the CRC incidence category analysis where the faecal levels of SCFAs in CRC patients were significantly lower compared to those in healthy subjects.

Our findings in CRC risk and incidence were consistent with the observations reported in other meta-analyses, which focused on the association between dietary fibre intake and the risk of colorectal adenoma [158,159], and carcinoma [157]. These systematic reviews suggested a protective effect of dietary fibre intake against CRA and CRC [157-159]. Since SCFAs are produced by gut-microbiota via the fermentation of dietary fibres [102,103,105,106], our meta-analysis of SCFA concentrations in CRC further confirms earlier observations and underlines the importance of dietary fibres/SCFAs in the risk and progression of CRC.

Another meta-analysis, which assessed the effect of non-digestible carbohydrate [resistance starch (RS)] or inulin supplementation on the risk of colorectal neoplasia, did not find significant increase in faecal total SCFA or butyric acid concentration and excretion before and after the intervention [309]. Many studies which investigated the effect of RS on healthy subjects or individuals with sporadic CRC or adenoma had a period of ≤ 4-week of intervention. A few studies reported 7- and 8-week intervention on adenoma or healthy individuals and the remaining studies were conducted on individuals with inherited CRC syndromes after > 2-year intervention [309]. The duration of intervention was longest (> 2 years) for studies involving hereditary CRC cases with reported germ-line mutations, which may have outweighed the effect of RS supplementation. While interventions involving sporadic cases or healthy subjects had much shorter periods of RS intervention (< 8 weeks) [309]. In our meta-analysis, we also did not observe a significant difference in total faecal SCFAs in the CRC risk category. This could be due to other SCFA molecules such as valeric, iso-butyric, and iso-valeric acid being included in total SCFA measurements; the latter two are the branched SCFAs mainly produced via fermentation of branched amino acids in the colon and not from non-digestible carbohydrates [104,313].

Another systematic review on the food-microorganism-SCFA axis, without any meta-analysis, concluded that most evidence demonstrated higher SCFA levels in individuals at risk of CRC compared to healthy individuals [314], which contrasts with findings in our systematic review which showed lower faecal SCFA concentration in at-risk individuals (**Figure 4.2**). In comparison to our systematic review, their search strategy restricted their analysis to only 8 of the final 23 studies that we analysed [163,169,176,177,182-185]. Therefore, their conclusion was based on a smaller subset of the primary studies available and was also not supported by a meta-analysis.

Both the quantitative and qualitative analyses of CRC risk identified comparable findings of significantly lower concentration of acetic and butyric acid in the high- versus low-risk CRC group. For the CRC incidence category, the quantitative meta-analysis of butyric acid was consistent with observations identified in most of the articles from the qualitative analysis, supporting the evidence of lower concentration of three SCFAs in CRC cases compared to healthy controls. The meta-analysis of propionic acid was not significantly different between cases and controls. Similarly, most of the studies (5 of 7) reported no significant difference in faecal propionic acid concentration between CRC and healthy control in the qualitative analysis. The meta-analysis on IBS revealed a significantly higher concentration of faecal propionic acid in these patients in comparison to healthy controls [311]. Therefore, further studies comparing SCFA profiles among multiple gut diseases could shed more light on the importance of these molecules in the development of varied medical conditions.

To our knowledge, this systematic review is the first to provide a comprehensive search and data collection on observational studies linking SCFA molecules with the CRC risk and incidence. A limitation of our analysis is the heterogeneity of the studies evaluated in this systematic review, which is very difficult to control for. One such factor was the age group assessed for CRC incidence and risk. The mean age of the group in the studies was greater than 50 years and faecal SCFA concentration was not measured in younger populations to provide a comparison with low-risk, young age individuals. Although CRC is most often diagnosed in individuals > 50 years, the incidence for early-onset CRC (EOCRC) in adults aged 20-49 years has increased over the past decade in the USA, Australia, and Europe [135,197,201,203]. It would be of interest in the future to study different age group populations for CRC risk and incidence. Family history [315], diet and lifestyle [316] are known factors contributing to CRC incidence. Only a few studies assessed in this systematic review provided information on the

dietary difference between groups [169,171,180,181,185]. There was also no information about the type of polyps (conventional vs serrated) in individuals with CRA.

One common limitation is related to the nature of observational studies. Case-control studies are inherently prone to recall bias, and appropriate matching of case and control groups [317]. Similarly, the results of the cross-sectional studies could be affected by bidirectional relationship [318] and confounding factors. We have assessed the effect of these inherent limitations on our analysis by undertaking appropriate quality checks such as the "comparability" category of NOS quality assessment (**Supplementary Table 4.1**) and the JBI tool (**Supplementary Table 4.2**) for all the studies included in our systematic review and meta-analysis.

Another limitation is the diversity in sample handling/storage workflows and the methodologies used to measure faecal SCFA across the studies (**Table 1**). We did not obtain enough studies to perform separate meta-analyses to understand the effect of each of these variables on the SCFA concentrations Whilst these factors could influence our interpretation, the levels of SCFAs were lower in high-risk as well as incident CRC cases, irrespective of the method used to measure SCFAs. Since these are well-known and established techniques for measuring SCFA concentrations, it appears that the standardised sample handling workflows and analytical methods had little impact on differences across study groups, so long as optimised procedures for SCFA assessment were followed. This systematic review did not include non-English records. To our knowledge, no longitudinal studies have reported faecal SCFA measurements at different time points during CRC progression, nonetheless, the 23 studies assessed in this systematic review and meta-analysis provide a comparison between CRC risk/incidence and respective controls from various countries and ethnic groups.

In addition to the SCFAs assessed in this systematic review, other metabolites such as bile acids were also measured in six of the selected studies [167,170,177,180,181,184]. Among the bile acids investigated, a significantly higher faecal concentration of deoxycholic acid in the CRC high- versus low-risk group was reported in three studies [180,181,184]. Dietary fibre and fat promote the production of SCFA and bile acid molecules in the gut, respectively, and the latter is associated with gastrointestinal carcinogenesis [319-321]. Measurement of faecal SCFAs and other gut metabolites (such as bile acids) in longitudinal studies comparing individuals with colorectal adenoma/risk and healthy subjects could strengthen their association with CRC progression.

This study supports further exploration into faecal concentration of SCFAs: acetic, propionic, and butyric acids, as biomarkers for CRC risk. Among the current CRC screening methods, colonoscopy is the gold standard [322] however, being invasive it presents some procedural risk [192]. The guaiac faecal occult blood test (gFOBT) and faecal immunochemical test (FIT) are other, in practice, non-invasive stool-based methods for CRC screening, which however require improvement, in particular for detection of CRA or early-stage colonic carcinogenesis [192-194]. Faecal SCFA could be considered as a potential non-invasive biomarker to be measured in combination with or as an alternative to the commonly used non-invasive and current CRC screening methods [192,323], to improve specificity and sensitivity of current screening, as well as for potential early detection of CRA.

In conclusion, gut microbiota dysbiosis and changes in their metabolites have been the focus of epidemiological studies aimed at uncovering associations with colonic inflammation and carcinogenesis. In line with the protective role of faecal SCFAs against the development of gut diseases [105,106], and the protective effect of dietary fibres against CRC risk and/or incidence

[157-159], we determined that the combined faecal concentration of the three major SCFA molecules was significantly lower not only in CRC patients compared to healthy controls, but also in high-risk CRC individuals. Gut SCFA concentrations are inversely associated with CRC risk as well as incidence, thus could be used as potential biomarkers for CRC-progression, as well as a drug target in future intervention studies aimed to retard or prevent CRC progression.

4.5 Supplementary Materials

Supplementary Methods. The details of the search strategy conducted on May 21, 2021.

1) Medline and Embase combined search, followed by deduplication, using Ovid search interface:

Search link:

https://ezproxy.uws.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&NEWS=N&PAGE=main&SHAREDSEARCHID=6oHhfLqbyqUT2KwL2FpUgcWyfohUdPU40kumYqdCTlI8X2Ibo6ZjVT3zsbLZCbIw7

Search history details:

Embase <1974 to 2021 May 20>

Ovid MEDLINE(R) ALL

1	exp Fatty Acids, Volatile/	113151
2	short chain fatty acid*.mp.	24264
3	short-chain fatty acid*.mp.	24264
4	SCFA*.mp.	10870
5	exp Acetates/	166777
6	acetate.mp.	427012
7	exp Propionates/	28315
8	propionate.mp.	56255
9	exp Butyrates/	90653
10	butyrate.mp.	35882
11	exp Colorectal Neoplasms/	240756

12 exp Colonic Neoplasms/ 418868

13 "colorectal cancer".mp. 324884

14 "colon cancer".mp. 157155

15 "colorectal carcinoma".mp. 48725

16 "colon carcinoma".mp. 39183

17 "colorectal neoplasm".mp. 2133

18 "colon neoplasm".mp. 294

19 "colorectal neoplasia".mp. 6269

20 "colon neoplasia".mp.499

21 "colo* cance*".mp. 450574

22 "colo* carcinom*".mp. 93299

23 "colo* neoplas*".mp. 185903

24 crc.mp. 94637

25 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 775397

26 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 650155

27 25 and 26 7563

28 ((Fatty Acids, Volatile or short chain fatty acid* or short-chain fatty acid* or SCFA* or Acetates or acetate or Propionates or propionate or Butyrates or butyrate) and (Colorectal Neoplasms or Colonic Neoplasms or "colorectal cancer" or "colon cancer" or "colorectal carcinoma" or "colon carcinoma" or "colorectal neoplasm" or "colon neoplasm" or "colorectal neoplasia" or "colon neoplasia" or "colo* cance*" or "colo* carcinom*" or "colo* neoplas*" or crc)).ti,ab. 4013

29 28 not (letter or news or comment or editorial or congresses or abstracts).pt. 4006

30 limit 29 to humans 3044

31 remove duplicates from 30 1876

32 limit 31 to english language 1811

Next, the search results were filtered by “Publication Type: Article”, resulted in 1217 records.

The search was repeated on June 29, 2022, only for articles published in 2021 and 2022, resulting in 76 new records (1217+76=1293 records in total).

2) Web of science:

ab=(("scfa*" OR "short chain fatty acid*" OR "short-chain fatty acid*" OR acetate OR propionate OR butyrate) AND ("colo* cance*" OR "colo* carcinom*" OR "colo* neoplas*" OR crc) AND (human* OR patient* OR subject* OR individual* OR participant* OR case* OR control* OR character* person* OR people) NOT (rats OR mouse OR mice OR murine))

Next, the search results were refined by “Document Types: Articles”, and “Languages: English”, resulted in 783 records.

The search was repeated on June 29, 2022, only for articles published in 2021 and 2022, resulting in 57 new records (783+57=840 records in total).

Supplementary Table 4.1 Quality assessment of the selected case-control studies (n = 17) using Newcastle-Ottawa Scale (NOS). According to NOS guideline, one star can be given to each sector in Selection or Exposure category, and two stars for Comparability category, one per one matched factor. ¶Refer to the text for the definition of CRC incidence or risk category. §Based on the definition of control in NOS guideline, the star was given only to studies that "explicitly stated that controls have no history of this outcome". †As explained in the text, the SCFA values in this study could not be converted to mean and SD, and thus were included in the qualitative analysis.

			Selection				Comparability	Exposure			
	Study	**Analysis Category¶**	**Adequate definition of cases**	**Representativeness of cases**	**Selection of controls**	**Definition of controls§**	**Control for important factor or additional factor**	**Ascertainment of exposure**	**Same method of ascertainment for cases and controls**	**Non-response rate**	**Total scores**
Qualitative data (values of SCFA levels not	Sze et al. 2019	Incidence and risk	★	★				★	★	★	5
	Lin et al. 2019	Incidence	★		★		★★	★	★	★	7
	Lin et al. 2016	Incidence	★	★	★			★	★	★	6
	Weir et al. 2013	Incidence	★					★	★	★	4
	Ohigashi et al. 2013	Incidence and risk	★	★	★			★	★	★	6
	Monleon et al. 2009	Incidence	★	★	★			★	★	★	6
	Weaver et al. 1988	Incidence	★	★	★	★		★	★	★	7
Quantitative data (reported SCFA values)	Nannini et al. 2021†	Incidence and risk	★	★	★			★	★	★	6
	Chen et al. 2021	Risk		★	★			★	★	★	5
	Torii et al. 2019	Incidence	★					★	★	★	4
	Niccolai et al. 2019	Incidence and risk	★	★	★			★	★	★	6
	Yusuf et al. 2018	Incidence	★	★	★			★	★	★	6
	Song et al. 2018	Incidence	★	★	★	★	★	★	★	★	8
	Bridges et al. 2018	Risk	★	★	★	★		★	★	★	7
	Chen et al. 2013	Risk	★	★	★			★	★	★	6
	Boutron-Ruault et al. 2005	Risk	★	★	★	★	★★	★	★	★	9
	Kashtan et al. 1992	Risk	★	★	★			★	★	★	6

Supplementary Table 4.2 Quality assessment of the included cross-sectional studies (n = 6) using Joanna Briggs Institute (JBI) Critical Appraisal tool. The studies are shown with alternate shading. ¶Matched for sex. §Matched for age and sex. †Matched for age, as recruited subjects aged 50-60 years old.

Ocvirk et al. 2020	Response options			
1. Were the criteria for inclusion in the sample clearly defined?	***Yes***	No	Unclear	Not applicable
2. Were the study subjects and the setting described in detail?	***Yes***	No	Unclear	Not applicable
3. Was the exposure measured in a valid and reliable way?	***Yes***	No	Unclear	Not applicable
4. Were objective, standard criteria used for measurement of the condition?	***Yes***	No	Unclear	Not applicable
5. Were confounding factors identified?	***Yes***	No	Unclear	Not applicable
6. Were strategies to deal with confounding factors stated?	Yes	***No***	Unclear	Not applicable
7. Were the outcomes measured in a valid and reliable way?	***Yes***	No	Unclear	Not applicable
8. Was appropriate statistical analysis used?	***Yes***	No	Unclear	Not applicable
Overall appraisal: Include ☑ Exclude □ Seek further info □				
Katsidzira et al. 2019	Response options			
1. Were the criteria for inclusion in the sample clearly defined?	***Yes***	No	Unclear	Not applicable
2. Were the study subjects and the setting described in detail?	***Yes***	No	Unclear	Not applicable
3. Was the exposure measured in a valid and reliable way?	***Yes***	No	Unclear	Not applicable
4. Were objective, standard criteria used for measurement of the condition?	***Yes***	No	Unclear	Not applicable
5. Were confounding factors identified?	***Yes***	No	Unclear	Not applicable
6. Were strategies to deal with confounding factors stated?	***Yes***¶	No	Unclear	Not applicable
7. Were the outcomes measured in a valid and reliable way?	***Yes***	No	Unclear	Not applicable
8. Was appropriate statistical analysis used?	***Yes***	No	Unclear	Not applicable
Overall appraisal: Include ☑ Exclude □ Seek further info □				
Hester et al. 2015	Response options			
1. Were the criteria for inclusion in the sample clearly defined?	***Yes***	No	Unclear	Not applicable
2. Were the study subjects and the setting described in detail?	Yes	***No***	Unclear	Not applicable
3. Was the exposure measured in a valid and reliable way?	***Yes***	No	Unclear	Not applicable
4. Were objective, standard criteria used for measurement of the condition?	Yes	***No***	Unclear	Not applicable

5. Were confounding factors identified?	***Yes***	No	Unclear	Not applicable
6. Were strategies to deal with confounding factors stated?	Yes	***No***	Unclear	Not applicable
7. Were the outcomes measured in a valid and reliable way?	Yes	No	***Unclear***	Not applicable
8. Was appropriate statistical analysis used?	***Yes***	No	Unclear	Not applicable
Overall appraisal: Include ☑ Exclude □ Seek further info □				
Ou et al. 2013	Response options			
1. Were the criteria for inclusion in the sample clearly defined?	***Yes***	No	Unclear	Not applicable
2. Were the study subjects and the setting described in detail?	***Yes***	No	Unclear	Not applicable
3. Was the exposure measured in a valid and reliable way?	***Yes***	No	Unclear	Not applicable
4. Were objective, standard criteria used for measurement of the condition?	***Yes***	No	Unclear	Not applicable
5. Were confounding factors identified?	***Yes***	No	Unclear	Not applicable
6. Were strategies to deal with confounding factors stated?	***Yes***[§]	No	Unclear	Not applicable
7. Were the outcomes measured in a valid and reliable way?	***Yes***	No	Unclear	Not applicable
8. Was appropriate statistical analysis used?	***Yes***	No	Unclear	Not applicable
Overall appraisal: Include ☑ Exclude □ Seek further info □				
Ou et al. 2012	Response options			
1. Were the criteria for inclusion in the sample clearly defined?	***Yes***	No	Unclear	Not applicable
2. Were the study subjects and the setting described in detail? (According to this reference of the paper: O'Keefe et al. 2007, PMID: 17182822.)	***Yes***	No	Unclear	Not applicable
3. Was the exposure measured in a valid and reliable way?	***Yes***	No	Unclear	Not applicable
4. Were objective, standard criteria used for measurement of the condition?	***Yes***	No	Unclear	Not applicable
5. Were confounding factors identified?	***Yes***[†]	No	Unclear	Not applicable
6. Were strategies to deal with confounding factors stated?	Yes	***No***	Unclear	Not applicable
7. Were the outcomes measured in a valid and reliable way?	***Yes***	No	Unclear	Not applicable
8. Was appropriate statistical analysis used?	***Yes***	No	Unclear	Not applicable
Overall appraisal: Include ☑ Exclude □ Seek further info □				

O'Keefe et al. 2009	Response options			
1. Were the criteria for inclusion in the sample clearly defined?	***Yes***	No	Unclear	Not applicable
2. Were the study subjects and the setting described in detail? (Only the time period of living in either area wasn't mentioned.)	***Yes***	No	Unclear	Not applicable
3. Was the exposure measured in a valid and reliable way? (According to this reference of the paper: O'Keefe et al. 2007, PMID: 17182822.)	***Yes***	No	Unclear	Not applicable
4. Were objective, standard criteria used for measurement of the condition?	***Yes***	No	Unclear	Not applicable
5. Were confounding factors identified?	***Yes***[†]	No	Unclear	Not applicable
6. Were strategies to deal with confounding factors stated?	Yes	***No***	Unclear	Not applicable
7. Were the outcomes measured in a valid and reliable way?	***Yes***	No	Unclear	Not applicable
8. Was appropriate statistical analysis used?	***Yes***	No	Unclear	Not applicable
Overall appraisal: Include ☑ Exclude □ Seek further info □				

Supplementary Table 4.3 Table showing which data is used for which analysis. Cross-sectional studies are highlighted in grey, and case-control studies are not highlighted. Total number of studies in each category is stated at the bottom of table. Note: for quantitative analyses in risk category, the meta-analyses in the **Figure 2** included 3, 3, 5, and 3 studies for measuring C2, C3, C4, and total SCFA, respectively. C2: acetic acid, C3: propionic acid, C4: butyric acid. ¶As explained in the text, the SCFA values in this study could not be converted to mean and SD, and thus were included in the qualitative analysis.

		CRC risk				CRC incidence			
		C2	C3	C4	Total SCFA	C2	C3	C4	Total SCFA
Qualitative data (values of SCFA levels not reported)	Sze et al. 2019	*	*	*		*	*	*	
	Lin et al. 2019					*	*	*	
	Lin et al. 2016					*	*	*	
	Weir et al. 2013					*	*	*	
	Ohigashi et al. 2013	*				*			
	Monleon et al. 2009					*	*	*	
	Weaver et al. 1988					*	*	*	*
	Ocvirk et al. 2020	*	*		*				
	Katsidzira et al. 2019	*	*	*					
	Hester et al. 2015	*	*	*	*				
	Ou et al. 2013	*	*	*					
	Ou et al. 2012	*	*	*					
	O'Keefe et al. 2009	*	*	*	*				
Quantitative data (reported SCFA values)	Nannini et al. 2021¶	*	*	*		*	*	*	
	Chen et al. 2021	*	*	*					
	Torii et al. 2019					*	*	*	
	Niccolai et al. 2019	*	*	*	*	*	*	*	*
	Yusuf et al. 2018					*	*	*	
	Song et al. 2018					*	*	*	
	Bridges et al. 2018	*	*	*	*				
	Chen et al. 2013	*	*	*					
	Boutron-Ruault et al. 2005	*	*	*	*				
	Kashtan et al. 1992			*	*				
	Ocvirk et al. 2020			*					
	Total number of studies, in **qualitative** data analysis	9	8	7	3	8	7	7	1
	Total number of studies, in **quantitative** data analysis	5	5	7	4	4	4	4	1
		C2	C3	C4	Total SCFA	C2	C3	C4	Total SCFA
		CRC risk				CRC incidence			

Supplementary Table 4.4 The summary of other systematic reviews related to fibre intake, SCFA and risk of colorectal cancer or adenoma.

Study	PMID	Title	Aim	Primary source	Meta-analysis	Outcome
Rao et al. 2021	32202158	Non-Digestible Carbohydrate and the Risk of Colorectal Neoplasia: A Systematic Review	To check whether RS and inulin should be offered to cancer/precancerous patients or healthy subjects to decrease their risk of CRC.	Interventional studies on resistant starch (RS) or inulin supplementation	Total SCFA and butyrate concentration and excretion	Total SCFAs and butyrate concentrations and excretions in feces did not increase significantly after RS/inulin supplementation.
Nucci et al. 2021	33920845	Association between Dietary Fibre Intake and Colorectal Adenoma: A Systematic Review and Meta-Analysis	Assessing the association between dietary fibre intake and the risk of colorectal adenoma in adults.	Interview and questionnaire-based case-control, cohort, and cross-sectional studies	Effect size is measured as odds ratio	There could be a protective effect of dietary fibre intake against colorectal adenoma.
Shuwen et al. 2019	31401674	Protective effect of the "food-microorganism-SCFAs" axis on colorectal cancer: from basic research to practical application	To elucidate the “food-microorganism-SCFAs” axis and to provide guidance for prevention and intervention in CRC.	Not specific. Human, animal, and cell-based studies	-	The concentrations of SCFAs in CRC patients and individuals with a high risk of CRC were higher than those in healthy individuals.
Oh et al. 2019	31495339	Different dietary fibre sources and risks of colorectal cancer and adenoma: a dose-response meta-analysis of prospective studies	To summarise the relationships of different fibre sources with colorectal cancer and adenoma risks.	Interview and questionnaire-based prospective cohort studies	Effect size is measured as relative risk	The evidence for colorectal cancer prevention is strongest for fibre from cereals/grains. Each 10 g/d increase in dietary intake of vegetable or fruit fibre was statistically significantly associated with a reduced risk of incident colorectal adenoma.
Gianfredi et al. 2018	29516760	Is dietary fibre truly protective against colon cancer? A systematic review and meta-analysis	To evaluate the association between dietary fibre intake and the risk of colon cancer.	Interview and questionnaire-based case-control, cohort, and cross-sectional studies	Effect size is measured as odds ratio	Results suggest a protective role of dietary fibre intake on colon cancer risk.

Chapter 5 - Identification of Factors Associated with Early-Onset Colorectal Cancer

This chapter is a finalised manuscript to be submitted to *MJA*, titled: **Increased risk of rectal cancer and aggressive disease in sporadic early-onset colorectal cancer: a single site study of 3609 consecutive cases in NSW from 1995 to 2020**. The current version is confirmed by all authors.

The format of the manuscript is modified to conform to the thesis.

5.1 Background

The incidence of CRC amongst developed nations is either decreasing or has remained static over the past 20 years [135,138,196,324]. One of the main reasons for this trend is the roll out of CRC screening programs from the age of 50 years[135,138,196-199]. However, in contrast to the overall success in reducing CRC, the worldwide data point to a rising incidence of CRC diagnosed before the age of 50 years, conventionally termed early-onset CRC (EOCRC)[135,138,196-199,324]. Significantly, EOCRC more frequently involves the rectum or distal colon and usually presents at a more advanced stage, which may be attributable to delayed diagnosis and lack of screening for this cohort[198,199].

The increasing incidence of EOCRC is observed in the national and regional cancer registries collected between 1990 and 2016 from 20 European countries[203]. In the US, evaluation of the

population-based database [Surveillance, Epidemiology, and End Results (SEER)] between 1980–2016 also revealed a rising trend of EOCRC and increased rate of rectal cancer in this group[206]. In addition, the rising incidence rate of EOCRC in this age group was identified in the US Cancer Statistics for the period 2001-2017[202]. Interestingly, investigating the EOCRC incidence between US and Europe indicated a significantly higher rate in the US compared to that in Europe[207]. The authors suggested a higher incidence of obesity in the US population was a potential contributing factor[207]. All these studies point to a concerning increase in EOCRC within the US and Europe, which was also found in other countries such as the UK[324], Canada[208], New Zealand[325], and South Korea[326].

Colorectal cancer incidence in Australia is one of the highest among developed nations[209] and although the rate has been declining from the mid-1990s for people aged >50 years, an increased rate of both colon and rectal cancer in individuals <50 years of age has been reported[210]. Interestingly, two population-based studies from New South Wales (2001-2008) and Victoria (2000-2010) reported stable incidence rates of EOCRC[211,212]. Further, several other single-centre cohort studies have been conducted, which found no increase in the proportion of EOCRC (UK[213] and Japanese[214] cohorts). However, an increased incidence of EOCRC incidence was observed in a single centre study in Brazil[327].

The variance in the Australian state-wide population-based studies and the perceived gap in understanding the impact of EOCRC in the local context prompted us to perform a cohort analysis on a large consecutive series of CRC patients from a speciality colorectal surgical unit in Sydney. The focus of this study was to investigate the trend in incidence of sporadic EOCRC and to examine EOCRC related risk factors, clinicopathologic features and survival of patients

over a 26-year period. For clarity we have use the term sEOCRC to define sporadic EOCRC, which does not include EOCRC resulting from familiar syndromes.

5.2 Methods

5.2.1 Patient Cohort and Data Source

A consecutive series of 3609 patients who underwent surgical resection for colorectal cancer from January 1995 to December 2020 at the Concord Hospital Colorectal Surgical Unit (Sydney, Australia) were included in this study. Clinicopathological and follow up data for all patients were extracted from a dedicated surgical resection registry established in 1971 and approved by the Sydney Local Health District Ethics Committee (CH62/62011-136-P Chapuis HREC/11/CRGH206)[328]. Patients gave written informed consent for use of their clinical and tumour data for research. The standard follow-up procedure included six-monthly examinations for the first two years after resection and then on a yearly basis for five to six years[329].

On detailed review of the cohort, 150 patients were identified as undergone a subsequent surgery and all secondary records were excluded for these patients. No familial adenomatous polyposis (FAP) cases were identified in the cohort and three confirmed Lynch Syndrome cases were excluded to ensure only sporadic patients were examined. Age at diagnosis was inferred from the recorded age at date of resection, based on the short interval between diagnosis and surgery[330,331]. The age cut-off for sEOCRC was set at <50 years and for late-onset CRC (LOCRC) at ≥50 years in line with the globally accepted threshold[138,204]. The colon cancer sidedness was defined as proximal/right (cecum, ascending colon, hepatic flexure, and

transverse colon) and distal/left (splenic flexure, descending colon, and sigmoid colon) colon and the peritoneal reflection was considered as the borderline for the location of rectal tumours (above and below the peritoneal reflection was defined as upper and lower rectal tumour, respectively). All staging data were harmonised with the universally accepted American Joint Committee on Cancer (AJCC) staging system 8th edition[153].

5.2.2 Statistical analysis

All data were assessed for normality of distribution using Kolmogorov-Smirnov test with Lilliefors significance correction. For all continuous variables the Mann Whitney U test or chi-square (χ^2) test was utilised to compare categorical variables. Pearson correlation coefficient was used to assess temporal changes in trend over time. Where possible, median and interquartile range (IQR) are presented (defined as the 25th to 75th percentile) and odds ratios (OR) together with their 95% confidence intervals (CIs) were included. The time frame in years between the date of diagnosis and the date of death from any cause or CRC was measured to calculate overall survival (OS) and cancer-specific survival (CSS), respectively. The Kaplan-Meier (KM) method was used to assess survival trends. Competing risk analysis was assessed, which included the KM-based risk estimate and the proportion of all events that were a competing event to predict CSS event risk. This equates to any event risk that was not biased by the presence of competing events. A cut-off of 10% relative increase was not obtained as noted to be relevant by van Walraven and colleagues[332] for competing risk analysis to be performed. All tests were two tailed and statistical significance was defined at $P<0.05$. Analyses were performed using both SPSS version 28 (IBM Inc., Chicago, Illinois, USA) and GraphPad version 9 (GraphPad Software, San Diego, CA, USA) software.

5.3 Results

5.3.1 Patient characteristics and risk factors

The characteristics of the 3609 patients included in this study were categorised into sporadic EOCRC and LOCRC groups and shown in **Table 5.1**. The median age at diagnosis for the EOCRC and LOCRC groups was 44.7 (IQR: 22-49.9) and 72.2 (IQR: 50-98.1) years, respectively. Two hundred sixty-three (7.3%) patients were in the EOCRC group. More men than women developed sEOCRC (53-2% vs 46.8%). The proportion of men and women was similar in both EOCRC and LOCRC groups (53.2% vs 57.1% for men and 46.8% vs 42.9% for women, P=0.228).

Table 5.1 Characteristics of 3609 patients who have undergone colorectal tumour resection

	sEOCRC (<50 y)		LOCRC (≥50 y)			
	n	%	n	%	Total	P-value
All cases	263	7.3	3346	92.7	3609	
Sex						0.228
Male	140	53.2	1909	57.1	2049	
Female	123	46.8	1437	42.9	1560	
Primary tumour location						<.001
Cecum	15	5.7	384	11.5	399	
Ascending colon	22	8.4	410	12.3	432	
Hepatic flexure	4	1.5	122	3.6	126	
Transverse colon	11	4.2	267	8	278	
Splenic flexure	7	2.7	72	2.2	79	
Descending colon	10	3.8	106	3.2	116	
Sigmoid colon	70	26.6	898	26.8	968	
Rectum	124	47.1	1086	32.5	1210	
Colon or rectal cancer						<.001
Colon cancer	139	52.9	2260	67.5	2399	
Rectal cancer	124	47.1	1086	32.5	1210	
Colon cancer						<.001
Proximal/Right colon	52	37.4	1183	52.4	1235	
Distal/Left colon	87	62.6	1076	47.6	1163	
Rectal cancer						0.965
Upper rectum	27	52.9	211	52.6	238	
Lower rectum	24	47.1	190	47.4	214	
CRC Stage						<.001
0	2	0.8	33	1	35	
I	47	17.9	691	20.7	738	

	sEOCRC (<50 y)		LOCRC (≥50 y)			
	n	%	n	%	Total	P-value
II	61	23.2	1162	34.7	1258	
III	98	37.3	974	29.1	1072	
IV	49	18.6	457	13.7	506	
Tumour Histology						0.161
Adenocarcinoma	233	88.6	3010	90	3243	
Mucinous	25	9.5	310	9.3	335	
Signet ring	5	1.9	26	0.8	31	
Metastasis						<.001
Present	147	55.9	1431	42.8	1578	
Absent	116	44.1	1915	57.2	2031	

In our cohort analysis we specifically excluded patients with hereditary cancer syndromes, yet found sEOCRC patients had a significantly higher number of first-degree relatives diagnosed with CRC than their older counterparts (16.7% vs 12.1%, P=0.037). As expected, patients with sEOCRC had significantly fewer comorbidities at the time of diagnosis compared to LOCRC (summarised in **Supplementary Table 5.1**). One notable exception was inflammatory bowel disease (IBD) which was significantly associated with sEOCRC patients (3.4% vs 0.7%, P<0.001).

5.3.2 Sporadic EOCRC trend analysis over 26 years

The 5-year interval trends for case number, age at diagnosis, tumour location, and metastasis are summarised in **Table 5.2**. A small but non-significant decrease in the percentage of sEOCRC was observed over time. The details of these temporal changes are shown in **Supplementary Table 5.2**.

Table 5.2 Characteristics of the study population in 5-year intervals. The last period entails the data of six years.

	1995-1999	2000-2004	2005-2009	2010-2014	2015-2020	Total (1995-2020)
All cases, n	660	798	844	620	687	3609
Men, n (%)	382 (57.9)	472 (59.1)	481 (57)	327 (52.7)	387 (56.3)	2049 (56.8)
Age at Diagnosis						
All cases, median (Y)	69.3	71.2	70.3	71.8	72.3	70.8

	1995-1999	2000-2004	2005-2009	2010-2014	2015-2020	Total (1995-2020)
Men, median (Y)	68.8	70.8	69.2	71.5	72.4	70.4
Women, median (Y)	70.7	71.6	71.1	72.7	72.4	71.7
sEOCRC, n (%)	51 (7.7)	62 (7.8)	66 (7.8)	42 (6.8)	42 (6.1)	263 (7.3)
Men, n (%)	32 (62.7)	33 (53.2)	31 (47)	21 (50)	23 (54.8)	140 (53.2)
Late-onset CRC, n (%)	609 (92.3)	736 (92.2)	778 (92.2)	578 (93.2)	645 (93.9)	3346 (92.7)
Men, n (%)	350 (57.5)	439 (59.6)	450 (57.8)	306 (52.9)	364 (56.4)	1909 (57.1)
Primary tumour location						
Colon, n (%)	410 (62.1)	478 (59.9)	573 (67.9)	434 (70)	504 (73.3)	2399 (66.5)
Men, n (%)	223 (54.4)	259 (54.2)	313 (54.6)	215 (49.5)	274 (54.3)	1283 (53.5)
Right colon, n (%)	202 (49.3)	241 (50.4)	295 (51.5)	224 (51.6)	273 (54.3)	1235 (51.5)
Men, n (%)	108 (48.4)	112 (43.2)	139 (44.4)	111 (51.6)	143 (52.4)	613 (47.8)
Rectum, n (%)	250 (37.9)	320 (40.1)	271 (32.1)	186 (30)	183 (26.7)	1210 (33.5)
Men, n (%)	159 (63.6)	213 (66.6)	168 (62)	112 (60.2)	113 (61.7)	765 (63.2)
Lower rectum, n (%)*			36 (43.9)	92 (49.5)	86 (46.7)	214 (47.3)
Men, n (%)*			22 (43.1)	54 (48.2)	57 (50)	133 (48)
Metastasis						
All cases, n (%)	304 (46.1)	327 (41)	369 (43.7)	256 (41.3)	320 (46.6)	1576 (43.7)
Men, n (%)	170 (55.9)	188 (57.5)	223 (60.4)	139 (54.3)	187 (58.4)	907 (57.6)
Colon cancer, n (%)	176 (42.9)	185 (38.7)	238 (41.5)	187 (43.1)	236 (46.9)	1022 (42.6)
Proximal colon, n (%)	82 (40.6)	85 (35.3)	117 (39.7)	95 (42.4)	121 (44.3)	500 (40.5)
Distal colon, n (%)	94 (45.2)	100 (42.2)	121 (43.5)	93 (44.3)	116 (50.4)	524 (45.1)
Rectal cancer, n (%)	128 (51.2)	142 (44.4)	131 (48.3)	69 (37.1)	83 (45.4)	553 (45.7)
Upper rectum, n (%)*			22 (47.8)	42 (44.7)	45 (45.9)	109 (45.8)
Lower rectum, n (%)*			17 (47.2)	27 (29.3)	39 (45.3)	83 (38.8)
sEOCRC, n (%)	30 (58.8)	33 (53.2)	39 (59.1)	19 (45.2)	26 (61.9)	147 (55.9)

There was an overall decreasing trend for rectal cancer which was mainly associated with LOCRC, but for sEOCRC the trend was stable (r=-0.08, P=0.7). Further, for sEOCRC patients there was a statistically significant increased risk of metastases associated with colon cancer over time (r=0.39, P=0.05), but not with rectal cancer (r=-0.13, P=0.51).

5.3.3 Tumour location and histological subtype in sEOCRC

In sEOCRC patients the majority of tumours were distal to the splenic flexure (77.6%), with more than half being rectal tumours. The primary tumour location data is shown in **Table 5.1**. When comparing tumour location between sEOCRC and LOCRC there was a greater percentage of rectal cancers (47.1 vs 32.5%, P<0.001), and 15% more distal colon cancer (P<0.001) associated with sEOCRC patients. Our analysis showed there was no significant difference in combined mucinous and signet-ring cell tumour histology between sEOCRC and

LOCRC [30/263 (11.4%) vs 336/3346 (10%), P=0.48], but a significant difference was revealed when assessing these subtypes in the rectum and sigmoid colon only [sEOCRC: 23/194 (11.9%) vs LOCRC: 127/1985 (6.4%), P=0.004]. Further, when analysing signet-ring cell histology, 1.9% of sEOCRC tumours exhibited this histology which was statistically no different for LOCRC [5/263 (1.9%) vs 26/3346 (0.8%), P=0.057]. However, when considering rectum and sigmoid colon tumours alone signet-ring cell histology was found to be significantly enriched [sEOCRC: 5/194 (2.6%) vs LOCRC: 7/1985 (0.35%), P<0.001].

5.3.4 Tumour stage and pattern of metastasis of sEOCRC

The number of CRC patients with metastasis (based on the AJCC staging system) is shown in **Table 5.1**. Patients with LOCRC had more early-stage cancer (stage 0, I, or II), whilst the number of patients with advanced stage cancer (stage III or IV) was significantly more associated with EOCRC patients (P=0.005, or P=0.025, respectively) (**Table 5.3**). The percentage of lymph node metastasis was significantly more associated with sEOCRC compared to LOCRC (P<0.001), which was also reflected in the involved-to-examined node ratio and lymphatic vessel permeation in sEOCRC vs LOCRC (P=0.004). Additionally, distant metastases to the liver and lung were significantly greater in sEOCRC than LOCRC (P=0.021, and 0.023, respectively). Overall, the percentage of patients presenting with metastatic disease at the time of diagnosis was significantly higher for sEOCRC than LOCRC (P<0.001) (**Table 5.3**). High grade tumours were more prevalent in sEOCRC compared to LOCRC (27.4% vs 19.2%, P=0.004), as well as the percentage of poorly differentiated tumours (22.8% vs 16.8%, P=0.026).

Table 5.3 Comparing EOCRC and LOCRC with respect to metastasis and primary tumour location

	sEOCRC (<50 y)		LOCRC (≥50 y)		
	n	%	n	%	p-value
CRC					
Metastasis					<0.001
Absent	116	44.1	1915	57.2	
Present	147	55.9	1431	42.8	
Lymph node metastasis					0.005
Absent	165	62.7	2372	70.9	
Present	98	37.3	974	29.1	
Distant metastasis					0.025
Absent	214	81.4	2889	86.3	
Present	49	18.6	457	13.7	
Colon cancer					
Metastasis					0.026
Absent	67	48.2	1307	57.8	
Present	72	51.8	953	42.2	
Lymph node metastasis					0.67
Absent	98	70.5	16.31	72.2	
Present	41	29.5	629	27.8	
Distant metastasis					0.01
Absent	108	77.7	1936	85.7	
Present	31	22.3	324	14.3	
Rectal cancer					
Metastasis					<0.001
Absent	49	39.5	608	56	
Present	75	60.5	478	44	
Lymph node metastasis					0.001
Absent	67	54	741	68.2	
Present	57	46	345	31.8	
Distant metastasis					0.47
Absent	106	85.5	953	87.8	
Present	18	14.5	133	12.2	

By separating the rectal and colon cancer data, we found that the percentage of metastasis was significantly higher for sEOCRC than LOCRC, and this was even more pronounced for rectal cancer (60.5% vs 44%; P< 0.001) (**Table 5.3**). Sporadic EOCRC patients had a 1.7-fold increased risk of metastasis (OR=1.7; 95% CI: 1.32-2.18; P<0.001) compared to LOCRC

patients. Additionally, the risk of developing metastases from the colon was almost 1.5-fold for sEOCRC than for LOCRC (OR=1.47; 95% CI: 1.05-2.08; P=0.01) and this risk increased to almost two-fold for metastases from the rectum (OR=1.95; 95% CI: 1.33-2.84; P=0.013).

Subgroup analyses of the study population revealed the percentage of lymph node metastasis was significantly higher for rectal cancer, compared to colon cancer for both sEOCRC and LOCRC patients (P=0.006 and 0.019, respectively). Interestingly, there was a significantly higher percentage of distant metastasis in the distal compared to the proximal colon (P=0.003), and from tumours of the upper rectum rather than the lower rectum (P=0.005) in patients with LOCRC. However, for sEOCRC the same comparison was not statistically significant. The details of these comparisons are shown in **Supplementary Table 5.3 and 5.4**.

5.3.5 Surgical outcomes of sEOCRC

No difference was observed in the surgical techniques used for sEOCRC or LOCRC (P=0.826). However, stoma construction was performed more often during surgery for EOCRC patients (P<0.001). Also, patients with sEOCRC were administered more neoadjuvant radiotherapy (P<0.001) and adjuvant chemotherapy (P=0.02) compared with LOCRC patients. Thirty (30)-day post-operative complications were similar for sEOCRC patients when compared with their older counterparts, except for some age-related complications. **Supplementary Table 5.5** provides a summary of these comparisons.

5.3.6 Survival outcomes of sEOCRC

Despite the generally more advanced stage of sEOCRC tumours, the five-year overall survival (OS) and cancer-specific survival was greater for sEOCRC patients than their older counterparts (OS: 71% vs 61%, CSS: 72% vs 62%). The five-year survival data is detailed in **Supplementary Table 5.6**. The Kaplan–Meier survival curves comparing sEOCRC and LOCRC

are shown in **Figure 5.1**. Estimated mean overall survival time (EMOST) in years was significantly greater for EOCRC (16.6 ± 0.78 [95% CI, 15-18.1]) than for LOCRC (10.65±0.18 [95% CI, 10.3-11]) (P<0.001). Likewise, the estimated mean cancer-specific survival time (EMCSST) in years was also significantly greater for sEOCRC [18.1±0.75 (95% CI, 16.6-19.5)] than for LOCRC [13.9±0.33 (95% CI, 13.2-14.5)] (P<0.001). **Table 5.4** provides the details of all estimated mean survival times analysed in this study. We found no difference in survival between men and women with either sEOCRC or LOCRC. Regarding the primary tumour location, no significant difference in survival was observed between colon and rectal cancer for sEOCRC patients (**Table 5.4**).

Figure 5.1 Kaplan–Meier curves comparing the survival of EOCRC and LOCRC group

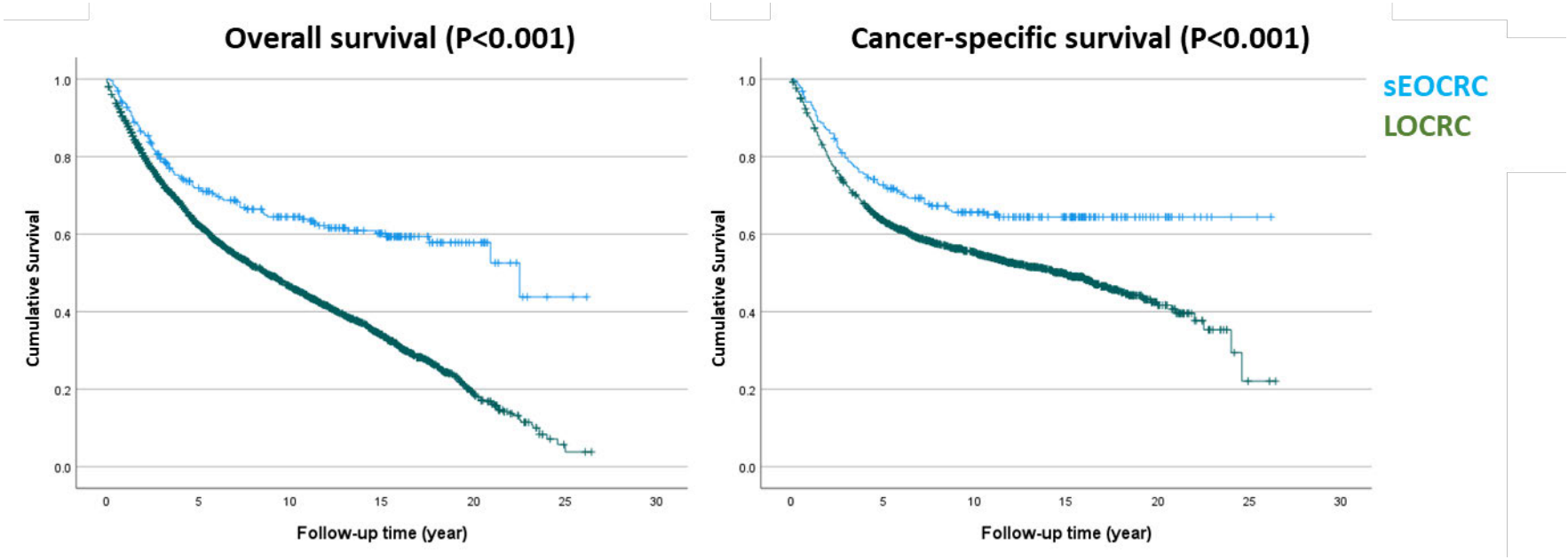


Table 5.4 Estimated mean of survival time in sEOCRC and LOCRC and groups

	Overall (EMOST)					Cancer-specific (EMCSST)				
	Mean	SE	95% CI: Lower band	Upper band	p-value	Mean	SE	95% CI: Lower band	Upper band	p-value
All CRC cases					<0.001					<0.001
Early	16.58	0.78	15.04	18.11		18.06	0.75	16.59	19.54	
Late	10.65	0.18	10.30	11.00		13.88	0.33	13.24	14.52	
sEOCRC										
Sex					0.768					0.777
Male	15.90	1.01	13.92	17.89		17.75	1.01	15.77	19.73	
Female	16.76	1.20	14.40	19.11		17.88	1.09	15.75	20.01	
Colon vs rectal cancer					0.663					0.657
Colon cancer	15.74	1.02	13.74	17.74		17.91	1.03	15.89	19.92	
Rectal cancer	17.40	1.03	15.37	19.42		17.72	1.07	15.62	19.82	
Disease spread					<0.001					<0.001
Localised tumour	22.29	0.82	20.68	23.90		24.09	0.53	23.07	25.12	
Lymph node metastasis	15.94	1.18	13.62	18.26		17.38	1.26	14.91	19.84	
Distant metastasis	3.56	0.78	2.04	5.08		3.79	0.84	2.16	5.43	
LOCRC										
Sex					0.21					0.068
Male	10.27	0.22	9.83	10.71		13.71	0.40	12.92	14.50	
Female	11.16	0.29	10.60	11.72		14.07	0.46	13.17	14.97	
Colon vs rectal cancer					0.096					0.004
Colon cancer	10.85	0.22	10.42	11.28		14.16	0.38	13.42	14.91	
Rectal cancer	10.20	0.30	9.62	10.78		13.24	0.50	12.26	14.21	
Disease spread					<0.001					<0.001
Localised tumour	12.60	0.23	12.15	13.06		18.42	0.45	17.54	19.30	

	Overall (EMOST)					Cancer-specific (EMCSST)				
	Mean	SE	95% CI: Lower band	95% CI: Upper band	p-value	Mean	SE	95% CI: Lower band	95% CI: Upper band	p-value
Lymph node metastasis	10.37	0.33	9.73	11.01		12.63	0.47	11.71	13.54	
Distant metastasis	2.87	0.25	2.38	3.36		2.66	0.23	2.22	3.10	

EMOST: estimated mean overall survival time, EMCSST: estimated mean cancer-specific survival time

Kaplan-Meier curves with respect to tumour spread is shown in **Figure 5.2**. Sporadic EOCRC patients with localised tumour, lymph node, or distant metastasis survived longer than their older counterparts [(P<0.001); **Table 5.4**]. However, in the case of distant metastasis, the difference in cancer-specific survival, despite being statistically significant (P<0.001), was only 1.1 years [sEOCRC (3.79±0.84; 95% CI, 2.16-5.43) vs LOCRC (2.66±0.23; 95% CI, 2.22-3.1)] (**Table 5.4**). This indicates similar survival irrespective of age at diagnosis, and gender (**Figure 5.2A** and **5.2B**). Regarding primary tumour location, significantly greater survival was observed for patients with localised colon cancer than those with rectal cancer (**Figure 5.2C**); but the mean survival difference for patients with either colon or rectal cancer was less than a year (**Supplementary Table 5.7**).

We also looked at the estimated mean survival time between sEOCRC vs LOCRC with respect to primary tumour location and metastasis (summarised in **Supplementary Table 5.8**). The largest difference in estimated mean survival was found between localised tumour and distant metastasis for sEOCRC patients with colonic cancer only (21.7 years, P<0.001).

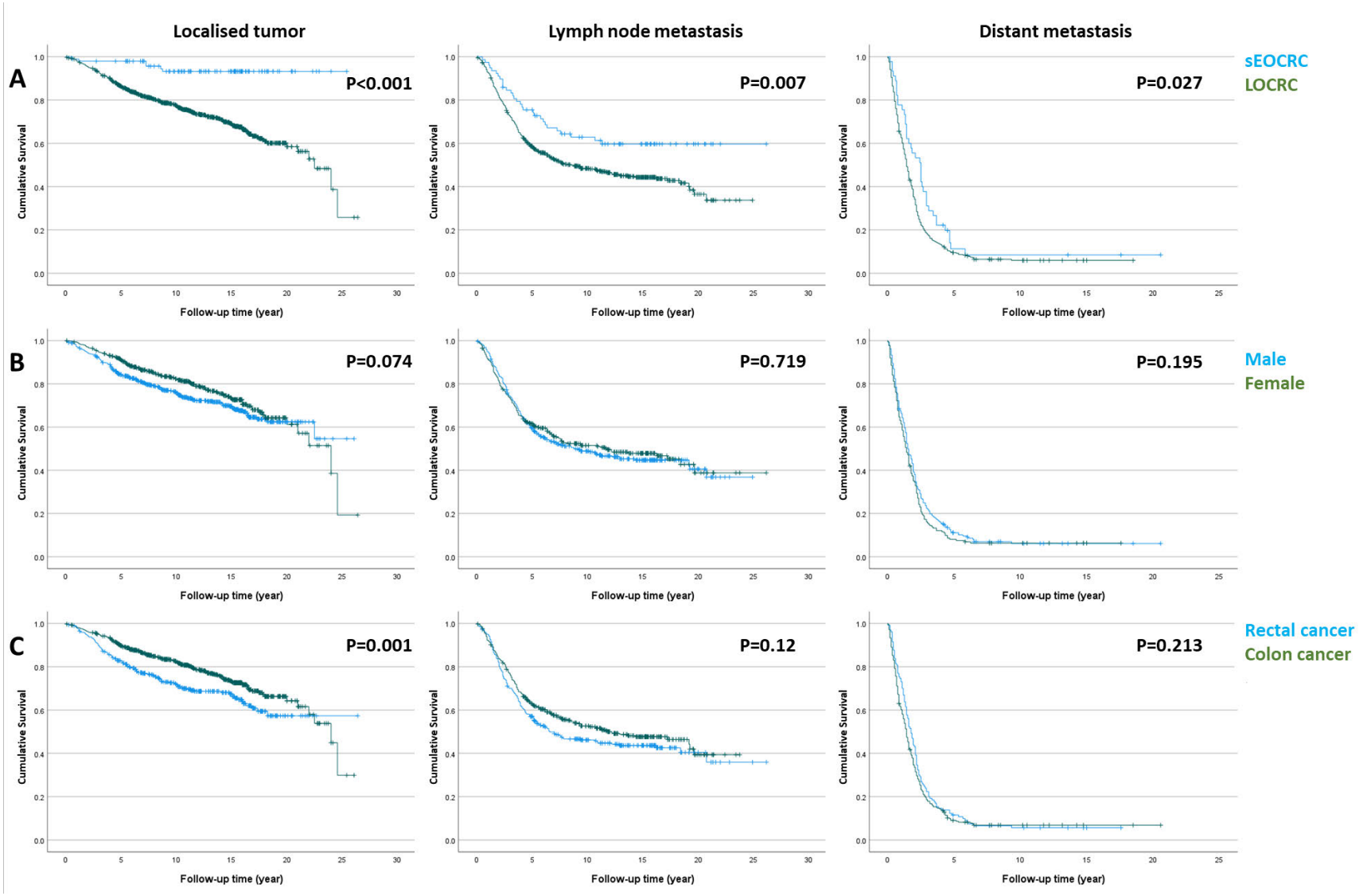


Figure 5.2 Kaplan–Meier curves comparing the cancer-specific survival with respect to CRC spread in three categories: A) CRC onset, B) sex, and C) primary tumour location

5.4 Discussion

The Concord Colorectal Surgical Registry was analysed in this study to investigate if the reported rising incidence of EOCRC was reflected in our large patient cohort between 1995-2020 and if any clinicopathologic/surgical and outcome features were associated with sEOCRC compared to their older counterparts.

Over the last decade a number of international population-based studies have reported on the rising incidence of EOCRC in high-income countries[202,203,206,208,210,324,325,333], yet in Australia, two large state-wide population-based studies have reported a stable incidence[211,212]. This lack of concordance may be due to the shorter time period assessed in the Australian studies (NSW: 2001-2008, VIC: 2000-2010)[211,212]. Two other studies performed in Western Australia (1982-2007)[237], and South Australia (1990-2010)[139] used a threshold of 40 years of age for early-onset of disease; with the former reporting a significantly increased age-standardized incidence, whilst the latter found a non-significant increasing trend[139,237]. We did not observe a rising trend of sEOCRC in our cohort over the 26-year (1995-2020) period reviewed. However, compared with a falling overall percentage of rectal cancer reported in a previous analysis performed on this cohort[328], we found a significantly higher percentage of rectal cancer associated with sEOCRC patients, consistent with other studies[206,211-214,325,333-335]. In addition, left-sided colon cancer was significantly associated with sEOCRC compared with LOCRC ($P < 0.001$) which was similarly consistent with other studies[137,211,212,327].

Our data confirm the recent findings and known links concerning IBD and family history as EOCRC risk factors[219,336] . Males were more associated with sEOCRC (Male; 53.2% vs Female; 46%) which was similar to that determined in other studies[211,214,337]. However, this was no

greater a risk factor than for CRC in general[217]. We found 11.4% mucinous and signet-ring cell tumour histology in sEOCRC which was similar to the 16.6% reported by Willauer et al.[338] and was no difference for LOCRC. However, we could not confirm the 13% percentage signet-ring cell histology reported in the study of Chang et al. (nor the 87.3% value in their Table 1 which appears erroneous)[339]. Instead, for sEOCRC we determined 1.9% signet-ring cell histology which is consistent with the 2% reported by Willauer et al.[338]. Significantly, we reveal an important difference between sEOCRC and LOCRC when considering only rectal and sigmoid colon location, which shows an enrichment of signet-ring cell histology (2.6% vs 0.35%). This is a potentially important observation as signet ring cell histology is highly linked to lymph node metastasis[340].

By subgrouping the cohort population into sEOCRC and LOCRC, we found a significantly higher proportion of patients in the sEOCRC group presented with metastasis at the time of diagnosis (42.8%), including metastasis to lymph nodes, liver, and lung. The higher incidence of metastatic tumours in this age group was also observed in other studies[211,224,327,333,337], including the NSW state-wide study[211]. We also found when analysed separately, a significantly higher percentage of metastasis in both rectal and colon cancers in sEOCRC patients. The odds ratio of having an advanced rectal or colon EOCRC cancer was 1.95- and 1.47-fold higher than LOCRC, respectively. Additionally, the presence of lymph node metastasis was associated with tumours of the rectum and distal colon in sEOCRC, indicative of an increased risk of metastasis potentially due to delayed diagnosis and/or tumour biology (histological subtype) associated with sEOCRC.

In our study, the higher incidence of lymph node and distant metastasis associated with sEOCRC and importantly the low median age at diagnosis of sEOCRC patients (44.7 years)

supports the need for greater awareness of the potential of CRC in young adults as well as the call for lowering the CRC screening age to 45 years in asymptomatic individuals.

The CRC screening program in Australia currently recommends screening from 50 years of age[341]. However, it has been recently lowered to 45 years in the US and this was based on an evaluation of the benefits and risks of lowering the age, which concluded an overall moderate net benefit[322]. Evaluation studies have been recommended in Australia and other countries to determine the cost and benefits of such updates to current screening programs.

We found better survival for sEOCRC patients in our cohort, which is consistent with other Australian and international studies[211,213,229,334]. In addition to the previous report of improved 5-year survival for all CRC patients in the Concord cohort[329], we found that in the case of patients presenting with distant metastasis, the estimated survival time of sEOCRC patients was only 1.1 years more than LOCRC patients indicative of the effect of distant metastasis on survival irrespective of age. Thus, when considering the significantly higher incidence of metastasis in sEOCRC patients, lowering the age of CRC screening to 45 years could potentially catch a proportion of patients prior to developing metastatic disease and improve survival. Further, based on a recent report on projected mortality rates for CRC to 2040 in Australia, there would be an increase in the mortality rate of both rectal and colon cancer in patients <50 years of age despite decreasing rates for their older counterparts[342]. This adds to the importance of adopting new policies for commencing CRC screening at a younger age and for also increasing CRC awareness in young adults and health professionals.

Our study has limitations to be considered when interpreting the results. First, this is a single-centre study at a speciality CRC surgical referral service in one region of Sydney. Secondly, there is potential patient selection bias with respect to age, socioeconomic status, and ethnic

background. Finally, since CRC patients attend the Concord Hospital colorectal surgical department for successful restorative surgery, there is a bias towards earlier tumour stage in the registry database. Although the prevalence of Lynch syndrome is only 1-3% of all CRC patients[343], their exclusion is important to focus on sEOCRC cases. This may be another limitation as this information was only collected from 2018 onwards. However, from the data it is clear there were very few potential Lynch syndrome patients in this registry. In addition, the registry only included pathological data for the upper and lower rectum from 2008, limiting the data available for this category.

Despite these limitations, accessing the data of 3609 patients over 26 years (1995-2020) provided us with significantly greater patient data set than other single-centre studies[137,213,214,334,337]. It is important to note that our cohort size allowed us to perform several detailed analyses on histological subtype, lymph node or distant metastasis with respect to tumour location in patients with sEOCRC and add further insights and understanding of this patient group.

We confirm the significantly higher percentage of lymph node and distant metastasis as well as poorly differentiated tumours irrespective of tumour location, and the significantly higher incidence of rectal cancer and distal (left-sided) colon cancer in EOCRC patients. Importantly, we identified a linkage between signet-ring cell histology, lymph node metastasis and survival in rectal and sigmoid colon sEOCRC. We also identified an almost two-fold increased risk of developing metastasis from rectal tumour in patients with sEOCRC than their older counterparts. Despite better survival, the estimated mean survival time of sEOCRC patients presenting with distant metastasis was only approximately one year greater than their LOCRC counterparts. Our results provide additional supportive evidence for reducing the current

commencing age of the National Bowel Cancer Screening Program to 45 or even 40 years of age to better identify younger individuals at risk of advanced sEOCRC.

5.5 Supplementary Materials

Supplementary Table 5.1 Comparison of medical history between sEOCRC and LOCRC. IBD: Inflammatory Bowel Disease; T2D: Type 2 Diabetes; CVD: Cerebrovascular Disease; PVD: Peripheral Vascular Disease

	sEOCRC (<50 y)		LOCRC (≥50 y)			
	n	%	n	%	Total	p-value
IBD						<.001
Present	9	3.4	25	0.7	34	
Absent	254	96.6	3321	99.3	3575	
T2D						<.001
Present	14	5.4	578	17.5	592	
Absent	246	94.6	2731	82.5	2977	
CVD						<.001
Present	1	0.4	215	6.4	216	
Absent	262	99.6	3131	93.6	3393	
PVD						<.001
Present	1	0.4	233	7	234	
Absent	262	99.6	3113	93	3375	
Hypertension						<.001
Present	26	9.9	1624	48.5	1650	
Absent	237	90.1	1722	51.5	1959	

Supplementary Table 5.2 Temporal changes of several factors in different subgroups of the study population. The trend analyses with significant changes are shown in bold. sEO: sporadic early-onset.

Analysed population	Variable	Pearson r	p-value
All CRC cases	**Frequency**	**-0.45**	**0.022**
All CRC cases	**Frequency of men**	**-0.52**	**0.0065**
All CRC cases	Frequency of women	-0.21	0.31
sEOCRC	**Frequency**	**-0.39**	**0.049**
sEOCRC	**Frequency of men**	**-0.42**	**0.035**
sEOCRC	Frequency of women	-0.18	0.37
sEOCRC	Percentage	-0.21	0.29
Colon cancer	Frequency	-0.11	0.59
Colon cancer	Frequency of men	-0.17	0.41
Colon cancer	Frequency of women	-0.01	0.96
Rectal cancer	**Frequency**	**-0.68**	**0.0001**
Rectal cancer	**Frequency of men**	**-0.71**	**<0.0001**
Rectal cancer	**Frequency of women**	**-0.45**	**0.022**
Rectal cancer	**Percentage**	**-0.67**	**0.0002**
sEO colon cancer	Frequency	-0.15	0.46
sEO rectal cancer	**Frequency**	**-0.5**	**0.01**
sEO rectal cancer	Percentage	-0.08	0.7
Colon cancer (right colon)	Frequency	0.064	0.75
Colon cancer (right colon)	Frequency of men	0.071	0.73
Colon cancer (right colon)	Frequency of women	0.031	0.88
Colon cancer (left colon)	Frequency	-0.24	0.24
Colon cancer (left colon)	Frequency of men	-0.27	0.19
Colon cancer (left colon)	Frequency of women	-0.065	0.75
Rectal cancer (upper rectum)	Frequency	-0.37	0.21
Rectal cancer (upper rectum)	Frequency of men	-0.31	0.3
Rectal cancer (upper rectum)	Frequency of women	-0.24	0.43
Rectal cancer (lower rectum)	Frequency	-0.33	0.27
Rectal cancer (lower rectum)	Frequency of men	-0.24	0.43
Rectal cancer (lower rectum)	Frequency of women	-0.35	0.24
All CRC cases	Metastasis %	0.15	0.47
All CRC cases	Metastasis % in men	0.32	0.11
All CRC cases	Metastasis % in women	-0.28	0.17
Colon cancer	**Metastasis %**	**0.39**	**0.0477**
Rectal cancer	Metastasis %	-0.25	0.22
sEOCRC	Metastasis %	0.1	0.61
sEOCRC	Metastasis % in men	0.34	0.09
sEOCRC	Metastasis % in women	-0.33	0.095
sEO colon cancer	Metastasis %	0.39	0.0502
sEO rectal cancer	Metastasis %	-0.13	0.51
All CRC cases	**Age at resection, median (Y)**	**0.56**	**0.0029**

Analysed population	Variable	Pearson r	p-value
All CRC cases	**Age at resection in men, median (Y)**	**0.62**	**0.0008**
All CRC cases	Age at resection in women, median (Y)	0.18	0.38
Colon cancer	**Age at resection, median (Y)**	**0.55**	**0.0038**
Rectal cancer	Age at resection, median (Y)	0.015	0.94
sEOCRC	Age at resection, median (Y)	0.019	0.93

Supplementary Table 5.3 Analyses of metastasis rate with respect to CRC primary tumour location. *: The data related to upper/lower rectum was collected from 2008 in the registry.

	sEOCRC					*LOCRC*				
	Absent		Present			Absent		Present		
	n	%	n	%	p-value	n	%	n	%	p-value
Metastasis										
Colon vs rectal cancer					0.157					0.312
Colon cancer	67	57.8	72	49		1307	68.3	953	66.6	
Rectal cancer	49	42.2	75	51		608	31.7	478	33.4	
Right vs left colon					0.282					0.14
Proximal/Right colon	22	32.8	30	41.7		713	54.6	470	49.3	
Distal/Left colon	45	67.2	42	58.3		594	45.4	483	50.7	
Upper vs lower rectum*					0.615					0.74
Upper rectum	12	57.1	15	50		117	49	94	58	
Lower rectum	9	42.9	15	50		122	51	68	42	
Lymph node metastasis										
Colon vs rectal cancer					0.006					0.019
Colon cancer	98	59.4	41	41.8		1631	68.8	629	64.6	
Rectal cancer	67	40.6	57	58.2		741	31.2	345	34.5	
Right vs left colon					0.523					0.69
Proximal/Right colon	35	35.7	17	41.5		858	52.6	325	51.7	
Distal/Left colon	63	64.3	24	58.5		773	47.4	304	48.3	
Upper vs lower rectum*					0.137					0.942
Upper rectum	19	61.3	8	40		147	52.5	64	52.9	
Lower rectum	12	38.7	12	60		133	47.5	57	47.1	
Distant metastasis										
Colon vs rectal cancer					0.105					0.099
Colon cancer	108	50.5	31	63.3		1936	67	324	70.9	
Rectal cancer	106	49.5	18	36.7		953	33	133	29.1	
Right vs left colon					0.555					0.003
Proximal/Right colon	39	36.1	13	41.9		1038	53.6	145	44.8	
Distal/Left colon	69	63.9	18	58.1		898	46.4	179	55.2	

	sEOCRC Absent		Present			*LOCRC* Absent		Present		
	n	%	n	%	p-value	n	%	n	%	p-value
Upper vs lower rectum*					0.228					0.005
Upper rectum	20	48.8	7	70		181	50.3	30	73.2	
Lower rectum	21	51.2	3	30		179	49.7	11	26.8	

Supplementary Table 5.4 Lymph node vs distant metastasis rate with respect to CRC primary tumour location. *: The data related to upper/lower rectum was collected from 2008 in the registry.

	sEOCRC Lymph node		Distant			*LOCRC* Lymph node		Distant		
	n	%	n	%	p-value	n	%	n	%	p-value
Colon vs rectal cancer					0.014					0.023
Colon cancer	41	41.8	31	63.3		629	64.6	324	70.9	
Rectal cancer	57	58.2	18	36.7		345	34.5	133	29.1	
Right vs left colon					0.968					0.035
Proximal/Right colon	17	41.5	13	41.9		325	51.7	145	44.8	
Distal/Left colon	24	58.5	18	58.1		304	48.3	179	55.2	
Upper vs lower rectum*					0.121					0.023
Upper rectum	8	40	7	70		64	52.9	30	73.2	
Lower rectum	12	60	3	30		57	47.1	11	26.8	

Supplementary Table 5.5 Post-operational complications compared between sEOCRC and LOCRC group. The bold items were significantly higher in the LOCRC individuals.

	p-value
Predominant wound complications	0.703
Septicaemia present or not	0.619
Pelvic abscess	0.123
Intra-abdominal abscess	0.188
Abscess drained by interventional organ imaging	0.407
Infected vascular access line	0.193
Pelvic haematoma	0.206
Prolonged post-operative ileus	0.654
Fistula	0.491
Urinary complications	**0.003**
Bleeding necessitating reoperation	0.607
Bleeding necessitating transfusion	0.127
Renal failure necessitating dialysis	0.261
Respiratory problem needing consult	**0.007**
Cardiac complications	**<0.001**
Deep venous thrombosis	**0.031**
Pulmonary embolus	0.437
Cerebrovascular accident	0.247
Small bowel obstruction	0.312
Organic confusional state >24 hours	**<0.001**
Acute drug withdrawal	0.794
Multi-system failure	**0.033**
Ileostomy dysfunction	**0.007**
ICU admission	0.664
Anastomotic leak	0.952
Local drain	0.744
Perianastomotic abscess	0.684

Supplementary Table 5.6 Overall and cancer-specific five-year survival (OS and CSS) rate in all the study population, and the sEOCRC subgroup.

	Overall (OS)%	Cancer-specific (CSS)%
All CRC cases	62	63
Sex		
Male	59	61
Female	64	66
Onset		
Early	71	72
Late	61	62
Colon vs rectal cancer		
Colon cancer	63	65
Rectal cancer	60	59
Disease spread		
Localised tumour	76	86
Lymph node metastasis	57	58
Distant metastasis	10	9
sEOCRC	71	72
Sex		
Male	71	73
Female	71	70
Colon vs rectal cancer		
Colon cancer	70	71
Rectal cancer	73	72
Disease spread		
Localised tumour	98	98
Lymph node metastasis	70	73
Distant metastasis	11	12

Supplementary Table 5.7 Estimated mean of cancer-specific survival time with respect to CRC spread

		Mean	SE	95% CI:		p-value
Disease spread				Lower band	Upper band	
Localised tumour	sEOCRC	24.09	0.53	23.07	25.12	<.001
Localised tumour	LOCRC	18.42	0.45	17.54	19.30	
Lymph node metastasis	sEOCRC	17.38	1.26	14.91	19.84	0.007
Lymph node metastasis	LOCRC	12.63	0.47	11.71	13.54	
Distant metastasis	sEOCRC	3.79	0.84	2.16	5.43	0.027
Distant metastasis	LOCRC	2.66	0.23	2.22	3.10	
Localised tumour	Male	19.01	0.50	18.03	19.99	0.074
Localised tumour	Female	19.19	0.58	18.06	20.31	
Lymph node metastasis	Male	12.99	0.58	11.86	14.12	0.719
Lymph node metastasis	Female	13.81	0.71	12.42	15.19	
Distant metastasis	Male	3.01	0.32	2.37	3.64	0.195
Distant metastasis	Female	2.55	0.31	1.94	3.15	
Localised tumour	Rectal cancer	18.58	0.60	17.41	19.75	0.001
Localised tumour	Colon cancer	19.48	0.46	18.57	20.38	
Lymph node metastasis	Rectal cancer	12.906	0.714	11.506	14.307	0.120
Lymph node metastasis	Colon cancer	13.156	0.525	12.127	14.185	
Distant metastasis	Rectal cancer	2.852	0.348	2.17	3.533	0.213
Distant metastasis	Colon cancer	2.884	0.303	2.291	3.477	

Supplementary Table 5.8 Estimated mean of survival time in late- and early-onset colon and rectal cancer. EMOST: estimated mean overall survival time, EMCSST: estimated mean cancer-specific survival time.

	Overall (EMOST)					Cancer-specific (EMCSST)				
	Mean	SE	95% CI:		p-value	Mean	SE	95% CI:		p-value
			Lower band	Upper band				Lower band	Upper band	
sEOCRC										
Colon cancer					<.001					<.001
Localised tumour	22.53	0.96	20.64	24.41		25.05	0.36	24.35	25.76	
Lymph node metastasis	12.68	1.35	10.03	15.33		14.34	1.47	11.46	17.22	
Distant metastasis	3.01	0.96	1.12	4.89		3.32	1.09	1.18	5.46	
Rectal cancer					<.001					<.001
Localised tumour	20.21	0.93	18.38	22.04		20.47	0.93	18.65	22.29	
Lymph node metastasis	16.47	1.54	13.46	19.48		17.15	1.62	13.96	20.33	
Distant metastasis	4.21	1.16	1.93	6.49		4.21	1.16	1.93	6.49	
LOCRC										
Colon cancer					<.001					<.001
Localised tumour	12.90	0.28	12.36	13.45		18.82	0.48	17.88	19.76	
Lymph node metastasis	10.50	0.40	9.72	11.28		12.92	0.55	11.84	13.99	
Distant metastasis	2.91	0.31	2.30	3.52		2.68	0.28	2.13	3.23	
Rectal cancer					<.001					<.001
Localised tumour	11.93	0.41	11.12	12.73		17.88	0.65	16.61	19.15	
Lymph node metastasis	9.97	0.52	8.96	10.98		11.52	0.74	10.07	12.96	
Distant metastasis	2.51	0.30	1.92	3.09		2.50	0.30	1.91	3.10	

Chapter 6 - General Discussion and Future Directions

Type 2 diabetes and colorectal cancer are among the top health burden for patients and health systems worldwide[6,133]. Understanding the aetiology and identifying risk factors and biomarkers involved in their progression could help prevent or at least delay their onset. Chapters 2 and 3 of my thesis have investigated some of the molecular aspects of T2D progression. Chapter 4 introduced a potential biomarker for CRC progression, and Chapter 5 focused on features in early-onset CRC.

By investigating the telomere biology in the human islet-derived progenitor cell (hIPC) model, a significant decline in the telomere length during successive cell passages was observed. In addition, a significant increase in *TERC* level (lncRNA component of telomerase enzyme - the telomeric repeat template) was identified suggesting a compensatory mechanism against telomere shortening. Next, exposing the cells to three different glucose conditions (normal level, chronic high, and oscillating high and normal) did not result in different telomere shortening rates among them, however, the significant rise in *TERC* levels under normal glucose, but not in high and oscillating conditions, suggest an impaired mechanism in the latter two conditions. This was also corroborated by the significantly decreasing level of *TERF1*, as one of the main telomeres protecting factors, in the oscillating condition.

Furthermore, significantly lower plasma level of *TERC* in patients with T2D compared to healthy individuals was observed. This finding needs to be further investigated with a larger sample size accompanied by rLTL (relative leukocyte telomere length) measurement investigating the possibility of introducing these two telomere-related factors as potential biomarkers of T2D progression.

While there is evidence suggesting an association between telomere length and T2D progression[71,75], several factors need to be considered when evaluating its role as a biomarker. Telomere length has the potential to reflect cellular aging and cumulative oxidative stress[62,68]. Meta-analyses have found associations between shorter telomeres and an increased risk of developing T2D[71,75]. However, telomere length can be influenced by various factors including genetics, lifestyle, and environmental exposures[68,73], and thus it might not be specific enough to serve as a standalone marker for T2D progression. Additionally, the association between telomere length and T2D is still a subject of ongoing research, and causality has not been definitively established.

Another limitation for considering LTL as a definitive biomarker at the current assessment stage is the lack of a standardised methods to measure it in readily available plasma samples[68]. Currently, real-time PCR is a well-accepted technique due to its capacity to run hundreds of samples in parallel in a high-throughput format as well as easier interpretation of the results; compared to traditional time-consuming and low-throughput gel-based methods[68]. Nonetheless, the primer pair used for telomere amplification and the single-copy gene selected for normalisation of the data is not consistent among all the researchers[68]. Therefore, measuring telomere length in readily available plasma samples (rLTL) is still an emerging technique. As such, there remains significant scope for further exploration and deeper investigation to fully elucidate the complexities and nuances inherent in this area of research.

Investigating telomere biology presented in this thesis had two novel aspects with respect to T2D progression. First, the type of cell model used for the first time to observe the telomere length as there is no published report for utilising hIPCs for telomere biology. hIPC cells are a

primary cell model and considered representative of human pancreatic islets[93]. Second, in this thesis, *TERC* was studied as another component of telomere biology for the first time in hIPCs as well as human plasma samples. While considering telomere maintenance by the telomerase enzyme, it was hypothesized that *TERC* level could be measured in plasma samples in addition to the rLTL. What was examined in this thesis was the initial assessment steps in showing that plasma *TERC* levels in patients with T2D is significantly shorter than that of healthy controls. However, it is important to note that there is no claim to consider telomere length and *TERC* level as a definitive biomarker of T2D progression at this stage. This needs to be further investigated using larger patient cohorts together with the rLTL data while comparing patients with T2D, insulin resistance, and healthy individuals in matched age and sex settings. Another possibility is measuring rLTL and plasma *TERC* levels longitudinally in patients with insulin resistance and those at high-risk groups. A defined biomarker workflow suitable for clinical application is outside the scope of this thesis and additional studies are required to validate appropriately robust and accurate assays.

Nonetheless, it is important to note that different biomarkers provide different types of information. The current established biomarkers, such as HbA1c, OGTT, and FPG are effective at diagnosing T2D[54], while telomere length reflects cellular aging and potential oxidative stress[62,68]. Therefore, measuring telomere biology parameters is not aimed at diagnosing T2D but mostly for T2D progression prior to detecting the change in blood glucose. This could be integrated into comprehensive risk assessment models for T2D. Combining rLTL and *TERC* data with other clinical and genetic markers could yield more accurate predictions of an individual's risk for developing T2D.

Measuring plasma rLTL in the context of cardiovascular diseases is also new[76], but there is no study for plasma *TERC* measurement in those medical conditions. The same approach suggested for T2D can be applied to CVDs for monitoring their progression and risk prediction utilizing the changes in telomeres.

Analysing several cellular senescence biomarkers revealed a different pattern for the oscillating condition in successive cell passages compared to normal and high glucose, as well as increased transcript abundance of *GATA4* (as one of the master regulators of senescence-associated secretory phenotype) and *IL-8* (the main chemokine released by the senescent cells) in the oscillating condition.

Therefore, a different molecular profile was identified in the oscillating condition as the most similar environment in which islet cells are exposed to fluctuating concentrations of glucose during T2D progression. Similar analyses need to be performed on vascular endothelial cells to assess the telomere biology because of their direct exposure to high levels of blood glucose and their fluctuations during T2D progression.

In this thesis, the markers of cellular senescence were measured *in vitro* in a cell model for pancreatic islets that has not been used before for this purpose. This is a novel achievement; however, these primary findings cannot currently be recommended for application in the clinical setting. The next recommended step would be measuring the protein levels of these markers in a larger sample size *in vitro*. Also, diabetic animal models could be used to measure the plasma levels for these factors to define a signature of detectable senescence during T2D progression.

Other biomarkers were not within the scope of this thesis, however, ongoing work on SCFA and other studies focusing on the identification of microRNA signature are continuing in our

laboratory. Further multi-omic analyses are also needed to identify the molecular criteria for prediction of T2D progression.

In another set of experiments presented in this thesis, the progression of T2D and the underlying insulin resistance was conducted *in vitro* and *in vivo* while focusing on weight cycling and the yoyo diet (dietary oscillation of healthy and unhealthy diets). Higher concentrations of SCFAs and chow diet were considered as "healthy", as opposed to lower concentrations of SCFAs or high-fat diet as "unhealthy".

Based on the *in vitro* experimental design, the gene expression levels of several beneficial gut hormones were measured longitudinally. The identified decreasing amplitude of change in the expression level of these hormones was a novel achievement that suggests yoyo dieting may not be a healthy strategy because eventually, the level of these hormones may not increase in proportion to the level of dietary SCFAs (or fibre) in the food. Epigenetic studies previously performed in our laboratory support these observations.

Future studies on other colonic epithelial cell lines and primary cells would enhance our understanding on the effect of dietary oscillation in T2D progression. These studies are planned for the future, in which human primary colonic epithelial cells will be utilised. Similar to what has been presented in this thesis, the first step would be a pilot study to identify the best sodium butyrate concentration and time period in which the cells can survive best. Next, in the main *in vitro* experiments, the cells will be exposed to sodium butyrate (C4) in three conditions: chronic low, chronic high, and oscillating conditions. In the latter, the time at which the cells will be exposed to intermittent low- and high-C4 is the one that already been identified in the pilot study.

At least three experimental replicates will be used per each condition, and the cell samples will be collected with each cell passage. The RNA will be isolated from the cells harvested in each passage, as well as baseline and endpoint. Following cDNA synthesis, the expression level of incretin genes and SCFA receptors will be measured to investigate longitudinal changes over time. Certain timepoints can be selected for epigenomic investigations to track the changes in gene expression profile with respect to the exposure conditions and time. These data, combined with the data generated in this thesis, will shape a better understanding of the potential detrimental effects of dietary oscillations on the genes of the colonic cells responsible for insulin secretion and glucose homeostasis.

In the mouse study, insulin resistance was identified mid-way through the dietary intervention in the oscillating group, which had a similar GTT profile at the end of the study compared to mice on a high-fat diet. Other evidence of T2D progression was identified in the behavioural tests, indicative of diabetic neuropathy in the form of weak neuromuscular function in the oscillating and high-fat diet group, and impaired olfactory sensing in the oscillating group as novel outcomes.

Future investigations in assessing 1) serum insulin levels, 2) analysing gene expression in the ileum, cecum, colon, pancreas, and liver tissue to study the factors involved in T2D and insulin resistance progression, 3) assessing gut microbiome composition and their generated metabolites in the faecal samples as well as ileum, cecum, and colon content, 4) analysing the expression level of genes related to neuromuscular function in several brain tissues, and identifying their potential correlation with the behavioural tests data.

While substantial progress has been made in understanding T2D pathophysiology[24,27], there remain gaps in the knowledge, particularly regarding the intricate interplay of factors

contributing to T2D progression. The animal work presented in this thesis has paved the way for several avenues to be explored in future. These include serum insulin levels, gene expression in various tissues, gut microbiome composition, and neuromuscular function to build upon the findings of the current thesis.

The investigation into serum insulin levels could provide additional valuable insights into the effect of dietary oscillation on T2D progression. This data would be complementary to FBG collected throughout the dietary intervention and GTT at the endpoint. These changes in glucose and insulin level may provide beneficial data on the effects of the diets, but not the underlying causes of these changes. Future studies could also delve into those aspects by analysing gene expression in various tissues to clarify the intricate underlying molecular mechanisms contributing to insulin resistance and T2D progression. Ileum, cecum, and colon gene expression profiling may reflect changes in the incretin hormone levels as well as SCFA receptors. The effect of dietary oscillation on insulin resistance could be identified by profiling gene expression in liver, muscle, and lipid tissue. Other gene expression analyses could include insulin and glucagon genes in pancreas and liver.

Overall, investigating epigenetic modifications and their impact on gene expression could provide further insights into the dynamic regulation of genes involved in metabolic pathways. It is important to note that all these gene expression profiles in the oscillating group need to be compared to those obtained from mice on chronic high fat diet as well as normal chow as the representative of unhealthy and healthy diet, respectively. Furthermore, males and females could be compared to identify a possible effect of gender on insulin resistance and T2D parameters while considering body weight gain between the two groups. Furthermore,

to build upon these findings, single-cell RNA sequencing could be employed to gain a deeper understanding of cellular heterogeneity within these tissues.

The exploration of gut microbiome composition and their generated metabolites has already generated a novel perspective on T2D progression[283,285]. In this regard, the faecal samples collected throughout the dietary intervention as well as the content of the ileum, cecum, and colon collected at the endpoint could potentially provide valuable to study the effect of dietary oscillation on insulin resistance and T2D progression.

Finally, the analysis of gene expression related to neuromuscular function in various brain tissues collected may shed light on potential neurological links to the potential detrimental effect of dietary oscillation on the nervous system. Exploring the correlation between gene expression and behavioural test data offers a promising avenue for understanding how central nervous system changes contribute to the senses and motor functions. This could pave the way for collaborations between neuroscientists and metabolic researchers to develop a holistic understanding of the intricate crosstalk between unhealthy diets and dysregulation in the nervous system and metabolic pathways.

In conclusion, future studies in combination of the data presented in this thesis could significantly advance our understanding of the effects of dietary oscillation on insulin resistance and T2D progression. The insights gained from assessing serum insulin levels, analysing gene expression, investigating the gut microbiome, and exploring neuromuscular function collectively contribute to a comprehensive view of the complex factors influencing T2D. Moving forward, integrating multi-omics approaches, such as transcriptomics, proteomics, and metabolomics, could help capture a more holistic picture of T2D progression.

Eventually, collaborative efforts between research disciplines could yield innovative interventions and personalized strategies for preventing and managing T2D.

In this thesis, the potential link between SCFA level and T2D was also studied in CRC progression by conducting a systematic review and meta-analysis. The combined faecal concentration of the three major SCFA molecules (acetate, propionate, and butyrate) was significantly lower not only in CRC patients compared to healthy controls, but also in individuals with high-risk of CRC. The novelty was consolidating the link between faecal SCFAs concentration and CRC risk and incidence. The latter was expected; however, the more interesting achievement was identifying lower faecal SCFA concentrations in apparently healthy individuals at risk of CRC. In this regard, gut SCFA level was inversely associated with CRC risk, suggesting faecal SCFA concentration as a potential biomarker to identify CRC progression. Therefore, faecal SCFA measurement can be utilised as a non-invasive and complementary method to the current CRC screening procedures such as gFOBT and FIT[192,193].

One major future step could focus on the standardisation of the method used to measure faecal SCFA concentration. Currently, gas chromatography, mass spectrometry, high-performance liquid chromatography, nuclear magnetic resonance spectroscopy, ion chromatography, and capillary electrophoresis are used for measuring metabolite concentrations in faeces samples[344]. The absence of a gold standard technique for measuring SCFAs in faecal samples poses a significant challenge. This lack of standardization can lead to inconsistencies and variations in reported SCFA concentrations across studies, making it difficult to compare and interpret results. The meta-analysis method used in this thesis helped reduce the effects of this diversity to a minimum, however, standardisation is needed in the laboratory and clinical setting.

Also, sample preparation can vary significantly among laboratories[344]. Factors such as sample extraction, derivatization and storage conditions can influence SCFA stability and recovery[345]. Inconsistencies in sample preparation can lead to variations in measured concentrations[346]. Therefore, developing a standardised protocols for sample collection, processing, and storage is crucial. These protocols should be validated to ensure that SCFAs are preserved in a consistent manner across different research settings.

Moreover, the absence of certified reference materials or standard reference compounds for SCFAs is another challenge for calibration and validation of the analytical assays. Without reliable standards, it is difficult to accurately quantify SCFA concentrations in faecal samples. Efforts should be made to produce and distribute certified reference materials for SCFAs. These materials can serve as benchmarks for calibration and quality control, ensuring the accuracy and comparability of SCFA measurements.

In summary, addressing the challenge of having a robust standardized approach for measuring SCFAs in faecal samples requires a coordinated effort from the scientific community. Researchers should collaborate to evaluate and validate existing methods, develop standardized protocols, and work towards the production of certified reference materials. This will ultimately enhance the reliability and comparability of SCFA measurements, facilitating more meaningful and consistent findings in SCFA research to introduce faecal SCFA measurement as a standard non-invasive method for CRC screening.

Apart from methodology standardisation, to comprehensively establish the link between SCFAs and CRC, researchers need to delve deeper into the molecular mechanisms at play. While current evidence suggests that SCFAs play a protective role by promoting a healthy gut environment and inhibiting inflammation[105,106,114], the precise signalling pathways and

molecular interactions are not yet fully understood[106]. Future studies should aim to uncover the mechanisms through which SCFAs influence CRC progression, including their effects on cell proliferation, apoptosis, and DNA damage repair. Additionally, future research should focus in more detail on characterizing the specific bacterial species responsible for SCFA production and how alterations in the microbiome, such as dysbiosis, may impact SCFA concentrations[113,114]. Moreover, investigating the interplay between host genetics and microbiome composition in SCFA metabolism is an avenue that holds great promise[113,114].

Despite the established inverse association between faecal SCFA concentration and CRC risk confirmed in this thesis, longitudinal studies are crucial to assess the temporal relationship between them. The meta-analysis presented in this thesis benefited from the data available from case-control and cross-sectional studies, however, longitudinal studies are required to follow at-risk individuals over extended periods to track changes in faecal SCFA levels. Such research would allow for a more precise determination of whether gradual changes in SCFA concentrations is aligned with CRC development to identify critical windows of intervention.

In addition, future potential work includes the use of SCFA in intervention studies aimed to slow down or prevent CRC progression. For example, future research could explore the potential of dietary interventions, including fibre-rich diets or probiotics, in manipulating SCFA production and mitigating CRC risk. These interventions should be tested in both animal models and human clinical trials to evaluate their effectiveness and safety. Further future research could explore how SCFA concentrations and their impact on CRC risk vary across populations. This could provide valuable insights into the interplay between genetics, lifestyle, and SCFA metabolism in different contexts.

The emerging field of research into the link between SCFA concentrations and CRC risk represents a promising avenue for reducing the burden of this disease. However, to harness the full potential of SCFAs in CRC prevention and management, future research must address the critical knowledge gaps outlined above. By advancing our understanding of SCFA biology and its implications for CRC, we can pave the way for innovative strategies to reduce the global impact of this devastating disease. In summary, this thesis has provided additional information about the beneficial aspect of SCFAs to create a healthy gut environment against CRC, T2D, and obesity progression, considering obesity and T2D as risk factors of CRC.

The work on CRC continued by focusing on patients with early-onset CRC. This was achieved by accessing the Concord colorectal surgery registry database to identify the clinicopathological and survival features that potentially could adversely impact individuals with EOCRC. Lynch Syndrome (LS) was excluded to allow the focused analysis of sporadic EOCRC, and thus a new abbreviation was suggested – sEOCRC as opposed to EOCRC in general, which includes LS. No increasing trend in sEOCRC incidence was observed, however, rectal cancer was significantly more prevalent in this age group, as well as the percentage of regional and distant metastasis, and poorly differentiated tumours. Also, sEOCRC patients had a 1.7-fold increased risk of metastasis compared to their older counterparts. Patients with sEOCRC had a better five-year survival rate, however, the estimated mean cancer-specific survival time was only about one year greater in this age group.

There are future research needs and avenues for better understanding sEOCRC. One fundamental aspect is to delve deeper into its genetic and molecular underpinnings. Future research could focus on conducting large-scale genomic studies to identify specific genetic alterations, such as mutations in oncogenes and tumour suppressor genes, that drive CRC at

earlier age. This could help in the development of targeted therapies and precision medicine approaches tailored to younger patients. Additionally, exploring the epigenetic modifications and gene expression profiles associated specifically with sEOCRC could provide valuable insights into the mechanisms underlying its aggressiveness and distinct clinical behaviour. The two key findings in this thesis, higher rate of metastasis and more advanced tumour at diagnosis, warrant investigating the interplay between genetic and environmental factors in sEOCRC susceptibility to unravel the complex aetiology of this disease.

Understanding the risk factors contributed specifically to sEOCRC, compared with LOCRC, was among the initial aims of this thesis. However, this was not fully achieved due to the nature of the database which was established with more of a focus on surgical parameters. Nonetheless, male sex, previous history of IBD, and family history of CRC was identified as sEOCRC risk factors. In general, some risk factors, such as family history and hereditary syndromes, are well-established, but there is a need to explore the role of lifestyle factors, diet, microbiome composition, and environmental exposures in CRC development in young patients[217,347]. Research should aim to identify modifiable risk factors that can be targeted through preventive interventions.

Additionally, investigating the potential role of the gut microbiota in sEOCRC development is an emerging area of interest[348,349]. Distinct microbial profile variations between sEOCRC and LOCRC could serve as a prospective diagnostic biomarker in the future[350]. This avenue of research holds promise for developing microbiome-based interventions to mitigate CRC risk[350].

Future research should also address disparities in sEOCRC incidence, diagnosis, and treatment[351]. There are socioeconomic, racial, or geographic factors contributing to

disparities in CRC outcomes among younger individuals[351]. Identifying these disparities and their root causes is essential for developing targeted interventions and improving access to care for all affected populations[352]. Furthermore, studying the economic burden of sEOCRC on patients and healthcare systems can inform healthcare policy decisions and resource allocation[351,352]. Research in this area can guide the development of cost-effective strategies for CRC prevention, screening, and treatment.

However, one of the most immediate and actionable steps that can be performed in parallel to the recommended future directions is the consideration of lowering the age at which individuals should begin CRC screening. Based on the what is reported in this thesis and previous findings[211,224], sEOCRC often presents at advanced stages, leading to poorer prognosis and treatment outcomes[199]. Therefore, lowering the age for CRC screening is crucial to lowering sEOCRC incidence[138,204]. Other future research can focus on developing non-invasive readily available biomarkers for detecting CRC at younger age by identifying new blood-based markers or novel imaging techniques. Investigating the feasibility of these markers in routine clinical practice and assessing their cost-effectiveness are essential steps in their implementation.

In summary, the significantly higher percentage of metastasis in patients with sEOCRC, higher chance of developing a metastatic tumour, and only a minor difference in survival of these patients in case of distant metastasis supports the reconsideration in the commencing age of the National Bowel Cancer Screening Program in Australia. The data presented on sEOCRC in this thesis provides further evidence about the complexities of sEOCRC and the need for detecting polyps or tumours in its preliminary stages before it turns into a full-blown disease. It is important to note that CRC should not be considered as a disease of the elderly and

screening programs should be extended to earlier age, at least to 45 years old. Future studies on larger cohort samples including the information on the metastasis status could help elucidate the adverse features that is associated with CRC at a younger age.

Globally, T2D and CRC are among the most prevalent metabolic disease, and cancers, respectively. This thesis has explored multiple aspects of these two medical conditions and has generated important research findings that contribute to understanding the molecular changes, biomarkers, and risk factors involved in T2D and CRC progression.

References

1. Vos, T.*, et al.* Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet* **396**, 1204-1222 (2020).
2. IDF Diabetes Atlas 2021 (https://diabetesatlas.org/atlas/tenth-edition/). (2021).
3. Laakso, M. Biomarkers for type 2 diabetes. *Mol Metab* **27s**, S139-s146 (2019).
4. Khan, M.A.B.*, et al.* Epidemiology of Type 2 Diabetes - Global Burden of Disease and Forecasted Trends. *J Epidemiol Glob Health* **10**, 107-111 (2020).
5. DeFronzo, R.A.*, et al.* Type 2 diabetes mellitus. *Nat Rev Dis Primers* **1**, 15019 (2015).
6. Chatterjee, S., Khunti, K. & Davies, M.J. Type 2 diabetes. *Lancet* **389**, 2239-2251 (2017).
7. Sung, H.*, et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **71**, 209-249 (2021).
8. Arnold, M.*, et al.* Global patterns and trends in colorectal cancer incidence and mortality. *Gut* **66**, 683-691 (2017).
9. Blüher, M. Obesity: global epidemiology and pathogenesis. *Nat Rev Endocrinol* **15**, 288-298 (2019).
10. Ma, Y.*, et al.* Obesity and risk of colorectal cancer: a systematic review of prospective studies. *PLoS One* **8**, e53916 (2013).
11. Dong, Y.*, et al.* Abdominal obesity and colorectal cancer risk: systematic review and meta-analysis of prospective studies. *Biosci Rep* **37**(2017).
12. Omata, F., Deshpande, G.A., Ohde, S., Mine, T. & Fukui, T. The association between obesity and colorectal adenoma: systematic review and meta-analysis. *Scand J Gastroenterol* **48**, 136-146 (2013).
13. Jayedi, A.*, et al.* Anthropometric and adiposity indicators and risk of type 2 diabetes: systematic review and dose-response meta-analysis of cohort studies. *Bmj* **376**, e067516 (2022).
14. Wojciechowska, J., Krajewski, W., Bolanowski, M., Kręcicki, T. & Zatoński, T. Diabetes and Cancer: a Review of Current Knowledge. *Exp Clin Endocrinol Diabetes* **124**, 263-275 (2016).
15. Gallagher, E.J. & LeRoith, D. Obesity and Diabetes: The Increased Risk of Cancer and Cancer-Related Mortality. *Physiol Rev* **95**, 727-748 (2015).
16. Jiang, Y.*, et al.* Diabetes mellitus and incidence and mortality of colorectal cancer: a systematic review and meta-analysis of cohort studies. *Eur J Epidemiol* **26**, 863-876 (2011).
17. Deng, L., Gui, Z., Zhao, L., Wang, J. & Shen, L. Diabetes mellitus and the incidence of colorectal cancer: an updated systematic review and meta-analysis. *Dig Dis Sci* **57**, 1576-1585 (2012).
18. Luo, W., Cao, Y., Liao, C. & Gao, F. Diabetes mellitus and the incidence and mortality of colorectal cancer: a meta-analysis of 24 cohort studies. *Colorectal Dis* **14**, 1307-1312 (2012).
19. Sun, L. & Yu, S. Diabetes mellitus is an independent risk factor for colorectal cancer. *Dig Dis Sci* **57**, 1586-1597 (2012).
20. De Bruijn, K.M.*, et al.* Systematic review and meta-analysis of the association between diabetes mellitus and incidence and mortality in breast and colorectal cancer. *Br J Surg* **100**, 1421-1429 (2013).
21. Guraya, S.Y. Association of type 2 diabetes mellitus and the risk of colorectal cancer: A meta-analysis and systematic review. *World J Gastroenterol* **21**, 6026-6031 (2015).
22. Peeters, P.J., Bazelier, M.T., Leufkens, H.G., de Vries, F. & De Bruin, M.L. The risk of colorectal cancer in patients with type 2 diabetes: associations with treatment stage and obesity. *Diabetes Care* **38**, 495-502 (2015).
23. Ahmad, E., Lim, S., Lamptey, R., Webb, D.R. & Davies, M.J. Type 2 diabetes. *Lancet* **400**, 1803-1820 (2022).
24. Galicia-Garcia, U.*, et al.* Pathophysiology of Type 2 Diabetes Mellitus. *Int J Mol Sci* **21**(2020).

25. Tinajero, M.G. & Malik, V.S. An Update on the Epidemiology of Type 2 Diabetes: A Global Perspective. *Endocrinol Metab Clin North Am* **50**, 337-355 (2021).
26. Davies, M.J.*, et al.* Management of Hyperglycemia in Type 2 Diabetes, 2022. A Consensus Report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetes Care* **45**, 2753-2786 (2022).
27. Pearson, E.R. Type 2 diabetes: a multifaceted disease. *Diabetologia* **62**, 1107-1112 (2019).
28. Martín-Peláez, S., Fito, M. & Castaner, O. Mediterranean Diet Effects on Type 2 Diabetes Prevention, Disease Progression, and Related Mechanisms. A Review. *Nutrients* **12**(2020).
29. Zheng, Y., Ley, S.H. & Hu, F.B. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. *Nat Rev Endocrinol* **14**, 88-98 (2018).
30. Toi, P.L.*, et al.* Preventive Role of Diet Interventions and Dietary Factors in Type 2 Diabetes Mellitus: An Umbrella Review. *Nutrients* **12**(2020).
31. Klein, S., Gastaldelli, A., Yki-Järvinen, H. & Scherer, P.E. Why does obesity cause diabetes? *Cell Metab* **34**, 11-20 (2022).
32. Sirdah, M.M. & Reading, N.S. Genetic predisposition in type 2 diabetes: A promising approach toward a personalized management of diabetes. *Clin Genet* **98**, 525-547 (2020).
33. Laakso, M. & Fernandes Silva, L. Genetics of Type 2 Diabetes: Past, Present, and Future. *Nutrients* **14**(2022).
34. Patterson, R.*, et al.* Sedentary behaviour and risk of all-cause, cardiovascular and cancer mortality, and incident type 2 diabetes: a systematic review and dose response meta-analysis. *Eur J Epidemiol* **33**, 811-829 (2018).
35. Carbone, S., Del Buono, M.G., Ozemek, C. & Lavie, C.J. Obesity, risk of diabetes and role of physical activity, exercise training and cardiorespiratory fitness. *Prog Cardiovasc Dis* **62**, 327-333 (2019).
36. Bellary, S., Kyrou, I., Brown, J.E. & Bailey, C.J. Type 2 diabetes mellitus in older adults: clinical considerations and management. *Nat Rev Endocrinol* **17**, 534-548 (2021).
37. Lingvay, I., Sumithran, P., Cohen, R.V. & le Roux, C.W. Obesity management as a primary treatment goal for type 2 diabetes: time to reframe the conversation. *Lancet* **399**, 394-405 (2022).
38. Czech, M.P. Insulin action and resistance in obesity and type 2 diabetes. *Nat Med* **23**, 804-814 (2017).
39. Christensen, A.A. & Gannon, M. The Beta Cell in Type 2 Diabetes. *Curr Diab Rep* **19**, 81 (2019).
40. Schwartz, S.S.*, et al.* The Time Is Right for a New Classification System for Diabetes: Rationale and Implications of the β-Cell-Centric Classification Schema. *Diabetes Care* **39**, 179-186 (2016).
41. Tomic, D., Shaw, J.E. & Magliano, D.J. The burden and risks of emerging complications of diabetes mellitus. *Nat Rev Endocrinol* **18**, 525-539 (2022).
42. Faselis, C.*, et al.* Microvascular Complications of Type 2 Diabetes Mellitus. *Curr Vasc Pharmacol* **18**, 117-124 (2020).
43. Viigimaa, M.*, et al.* Macrovascular Complications of Type 2 Diabetes Mellitus. *Curr Vasc Pharmacol* **18**, 110-116 (2020).
44. Patel, K., Horak, H. & Tiryaki, E. Diabetic neuropathies. *Muscle Nerve* **63**, 22-30 (2021).
45. Galiero, R.*, et al.* Peripheral Neuropathy in Diabetes Mellitus: Pathogenetic Mechanisms and Diagnostic Options. *Int J Mol Sci* **24**(2023).
46. Kaur, D., Tiwana, H., Stino, A. & Sandroni, P. Autonomic neuropathies. *Muscle Nerve* **63**, 10-21 (2021).
47. Selvarajah, D.*, et al.* Diabetic peripheral neuropathy: advances in diagnosis and strategies for screening and early intervention. *Lancet Diabetes Endocrinol* **7**, 938-948 (2019).
48. Qi, C., Mao, X., Zhang, Z. & Wu, H. Classification and Differential Diagnosis of Diabetic Nephropathy. *J Diabetes Res* **2017**, 8637138 (2017).
49. Thipsawat, S. Early detection of diabetic nephropathy in patient with type 2 diabetes mellitus: A review of the literature. *Diab Vasc Dis Res* **18**, 14791641211058856 (2021).

50. Wong, T.Y., Cheung, C.M., Larsen, M., Sharma, S. & Simó, R. Diabetic retinopathy. *Nat Rev Dis Primers* **2**, 16012 (2016).
51. Wang, W. & Lo, A.C.Y. Diabetic Retinopathy: Pathophysiology and Treatments. *Int J Mol Sci* **19**(2018).
52. van Sloten, T.T., Sedaghat, S., Carnethon, M.R., Launer, L.J. & Stehouwer, C.D.A. Cerebral microvascular complications of type 2 diabetes: stroke, cognitive dysfunction, and depression. *Lancet Diabetes Endocrinol* **8**, 325-336 (2020).
53. Biessels, G.J. & Despa, F. Cognitive decline and dementia in diabetes mellitus: mechanisms and clinical implications. *Nat Rev Endocrinol* **14**, 591-604 (2018).
54. Ortiz-Martínez, M.*, et al.* Recent Developments in Biomarkers for Diagnosis and Screening of Type 2 Diabetes Mellitus. *Curr Diab Rep* **22**, 95-115 (2022).
55. Califf, R.M. Biomarker definitions and their applications. *Exp Biol Med (Maywood)* **243**, 213-221 (2018).
56. Association, A.D. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2020. *Diabetes Care* **43**, S14-s31 (2020).
57. Ling, C. & Rönn, T. Epigenetics in Human Obesity and Type 2 Diabetes. *Cell Metab* **29**, 1028-1044 (2019).
58. Ling, C., Bacos, K. & Rönn, T. Epigenetics of type 2 diabetes mellitus and weight change - a tool for precision medicine? *Nat Rev Endocrinol* **18**, 433-448 (2022).
59. Arneth, B., Arneth, R. & Shams, M. Metabolomics of Type 1 and Type 2 Diabetes. *Int J Mol Sci* **20**(2019).
60. Chen, Z.Z. & Gerszten, R.E. Metabolomics and Proteomics in Type 2 Diabetes. *Circ Res* **126**, 1613-1627 (2020).
61. Smith, E.M., Pendlebury, D.F. & Nandakumar, J. Structural biology of telomeres and telomerase. *Cell Mol Life Sci* **77**, 61-79 (2020).
62. Hernandez-Segura, A., Nehme, J. & Demaria, M. Hallmarks of Cellular Senescence. *Trends Cell Biol* **28**, 436-453 (2018).
63. Vaiserman, A. & Krasnienkov, D. Telomere Length as a Marker of Biological Age: State-of-the-Art, Open Issues, and Future Perspectives. *Front Genet* **11**, 630186 (2020).
64. He, J., Tu, C. & Liu, Y. Role of lncRNAs in aging and age-related diseases. *Aging Med (Milton)* **1**, 158-175 (2018).
65. López-Otín, C., Blasco, M.A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194-1217 (2013).
66. Chakravarti, D., LaBella, K.A. & DePinho, R.A. Telomeres: history, health, and hallmarks of aging. *Cell* **184**, 306-322 (2021).
67. Ahmed, W. & Lingner, J. Impact of oxidative stress on telomere biology. *Differentiation* **99**, 21-27 (2018).
68. Cheng, F.*, et al.* Diabetes, metabolic disease, and telomere length. *Lancet Diabetes Endocrinol* **9**, 117-126 (2021).
69. Haycock, P.C.*, et al.* Leucocyte telomere length and risk of cardiovascular disease: systematic review and meta-analysis. *Bmj* **349**, g4227 (2014).
70. Roos, W.P., Thomas, A.D. & Kaina, B. DNA damage and the balance between survival and death in cancer biology. *Nat Rev Cancer* **16**, 20-33 (2016).
71. Wang, J.*, et al.* Association between telomere length and diabetes mellitus: A meta-analysis. *J Int Med Res* **44**, 1156-1173 (2016).
72. Turner, K.J., Vasu, V. & Griffin, D.K. Telomere Biology and Human Phenotype. *Cells* **8**(2019).
73. Shay, J.W. & Wright, W.E. Telomeres and telomerase: three decades of progress. *Nat Rev Genet* **20**, 299-309 (2019).
74. Okamoto, K. & Seimiya, H. Revisiting Telomere Shortening in Cancer. *Cells* **8**(2019).
75. Zhao, J., Miao, K., Wang, H., Ding, H. & Wang, D.W. Association between telomere length and type 2 diabetes mellitus: a meta-analysis. *PLoS One* **8**, e79993 (2013).

76. D'Mello, M.J*., et al.* Association between shortened leukocyte telomere length and cardiometabolic outcomes: systematic review and meta-analysis. *Circ Cardiovasc Genet* **8**, 82-90 (2015).
77. Blazer, S*., et al.* High glucose-induced replicative senescence: point of no return and effect of telomerase. *Biochem Biophys Res Commun* **296**, 93-101 (2002).
78. Matthews, C*., et al.* Vascular smooth muscle cells undergo telomere-based senescence in human atherosclerosis: effects of telomerase and oxidative stress. *Circ Res* **99**, 156-164 (2006).
79. Kuhlow, D*., et al.* Telomerase deficiency impairs glucose metabolism and insulin secretion. *Aging (Albany NY)* **2**, 650-658 (2010).
80. Gutmajster, E*., et al.* Possible association of the TERT promoter polymorphisms rs2735940, rs7712562 and rs2853669 with diabetes mellitus in obese elderly Polish population: results from the national PolSenior study. *J Appl Genet* **59**, 291-299 (2018).
81. Wang, G*., et al.* C1q/TNF-Related Protein 9 Attenuates Atherosclerosis by Inhibiting Hyperglycemia-Induced Endothelial Cell Senescence Through the AMPKα/KLF4 Signaling Pathway. *Front Pharmacol* **12**, 758792 (2021).
82. Opstad, T.B*., et al.* TERT and TET2 Genetic Variants Affect Leukocyte Telomere Length and Clinical Outcome in Coronary Artery Disease Patients-A Possible Link to Clonal Hematopoiesis. *Biomedicines* **10**(2022).
83. Perl, S*., et al.* Significant human beta-cell turnover is limited to the first three decades of life as determined by in vivo thymidine analog incorporation and radiocarbon dating. *J Clin Endocrinol Metab* **95**, E234-239 (2010).
84. Jacovetti, C. & Regazzi, R. Mechanisms Underlying the Expansion and Functional Maturation of β-Cells in Newborns: Impact of the Nutritional Environment. *Int J Mol Sci* **23**(2022).
85. Butler, A.E*., et al.* Adaptive changes in pancreatic beta cell fractional area and beta cell turnover in human pregnancy. *Diabetologia* **53**, 2167-2176 (2010).
86. Granger, A. & Kushner, J.A. Cellular origins of beta-cell regeneration: a legacy view of historical controversies. *J Intern Med* **266**, 325-338 (2009).
87. De Tata, V. Age-related impairment of pancreatic Beta-cell function: pathophysiological and cellular mechanisms. *Front Endocrinol (Lausanne)* **5**, 138 (2014).
88. Reers, C*., et al.* Impaired islet turnover in human donor pancreata with aging. *Eur J Endocrinol* **160**, 185-191 (2009).
89. Zhong, F. & Jiang, Y. Endogenous Pancreatic β Cell Regeneration: A Potential Strategy for the Recovery of β Cell Deficiency in Diabetes. *Front Endocrinol (Lausanne)* **10**, 101 (2019).
90. Tamura, Y*., et al.* Telomere attrition in beta and alpha cells with age. *Age (Dordr)* **38**, 61 (2016).
91. Tamura, Y*., et al.* β-cell telomere attrition in diabetes: inverse correlation between HbA1c and telomere length. *J Clin Endocrinol Metab* **99**, 2771-2777 (2014).
92. Victorelli, S. & Passos, J.F. Telomeres and Cell Senescence - Size Matters Not. *EBioMedicine* **21**, 14-20 (2017).
93. Gershengorn, M.C*., et al.* Epithelial-to-mesenchymal transition generates proliferative human islet precursor cells. *Science* **306**, 2261-2264 (2004).
94. Bray, G.A. & Popkin, B.M. Dietary fat intake does affect obesity! *Am J Clin Nutr* **68**, 1157-1173 (1998).
95. Paternoster, S. & Falasca, M. Dissecting the Physiology and Pathophysiology of Glucagon-Like Peptide-1. *Front Endocrinol (Lausanne)* **9**, 584 (2018).
96. Martin, A.M., Sun, E.W. & Keating, D.J. Mechanisms controlling hormone secretion in human gut and its relevance to metabolism. *J Endocrinol* **244**, R1-r15 (2019).
97. Efeyan, A., Comb, W.C. & Sabatini, D.M. Nutrient-sensing mechanisms and pathways. *Nature* **517**, 302-310 (2015).
98. Nauck, M.A. & Meier, J.J. Incretin hormones: Their role in health and disease. *Diabetes Obes Metab* **20 Suppl 1**, 5-21 (2018).

99. Boer, G.A. & Holst, J.J. Incretin Hormones and Type 2 Diabetes-Mechanistic Insights and Therapeutic Approaches. *Biology (Basel)* **9**(2020).
100. Müller, T.D*., et al.* Glucagon-like peptide 1 (GLP-1). *Mol Metab* **30**, 72-130 (2019).
101. Andersen, A., Lund, A., Knop, F.K. & Vilsbøll, T. Glucagon-like peptide 1 in health and disease. *Nat Rev Endocrinol* **14**, 390-403 (2018).
102. van der Beek, C.M., Dejong, C.H.C., Troost, F.J., Masclee, A.A.M. & Lenaerts, K. Role of short-chain fatty acids in colonic inflammation, carcinogenesis, and mucosal protection and healing. *Nutr Rev* **75**, 286-305 (2017).
103. Alexander, C., Swanson, K.S., Fahey, G.C., Jr & Garleb, K.A. Perspective: Physiologic Importance of Short-Chain Fatty Acids from Nondigestible Carbohydrate Fermentation. *Advances in Nutrition* **10**, 576-589 (2019).
104. Gill, P.A., van Zelm, M.C., Muir, J.G. & Gibson, P.R. Review article: short chain fatty acids as potential therapeutic agents in human gastrointestinal and inflammatory disorders. *Aliment Pharmacol Ther* **48**, 15-34 (2018).
105. Parada Venegas, D*., et al.* Short Chain Fatty Acids (SCFAs)-Mediated Gut Epithelial and Immune Regulation and Its Relevance for Inflammatory Bowel Diseases. *Frontiers in Immunology* **10**(2019).
106. Liu, P*., et al.* The role of short-chain fatty acids in intestinal barrier function, inflammation, oxidative stress, and colonic carcinogenesis. *Pharmacol Res* **165**, 105420 (2021).
107. O'Keefe, S.J. Diet, microorganisms and their metabolites, and colon cancer. *Nat Rev Gastroenterol Hepatol* **13**, 691-706 (2016).
108. Martin, A.M., Sun, E.W., Rogers, G.B. & Keating, D.J. The Influence of the Gut Microbiome on Host Metabolism Through the Regulation of Gut Hormone Release. *Front Physiol* **10**, 428 (2019).
109. Sender, R., Fuchs, S. & Milo, R. Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans. *Cell* **164**, 337-340 (2016).
110. Glowacki, R.W.P. & Martens, E.C. In sickness and health: Effects of gut microbial metabolites on human physiology. *PLoS Pathog* **16**, e1008370 (2020).
111. Bultman, S.J. Interplay between diet, gut microbiota, epigenetic events, and colorectal cancer. *Mol Nutr Food Res* **61**(2017).
112. Yang, J. & Yu, J. The association of diet, gut microbiota and colorectal cancer: what we eat may imply what we get. *Protein Cell* **9**, 474-487 (2018).
113. Wong, S.H. & Yu, J. Gut microbiota in colorectal cancer: mechanisms of action and clinical applications. *Nat Rev Gastroenterol Hepatol* **16**, 690-704 (2019).
114. Song, M., Chan, A.T. & Sun, J. Influence of the Gut Microbiome, Diet, and Environment on Risk of Colorectal Cancer. *Gastroenterology* **158**, 322-340 (2020).
115. Mehta, T., Smith, D.L., Jr., Muhammad, J. & Casazza, K. Impact of weight cycling on risk of morbidity and mortality. *Obes Rev* **15**, 870-881 (2014).
116. Mackie, G.M., Samocha-Bonet, D. & Tam, C.S. Does weight cycling promote obesity and metabolic risk factors? *Obes Res Clin Pract* **11**, 131-139 (2017).
117. Rhee, E.J. Weight Cycling and Its Cardiometabolic Impact. *J Obes Metab Syndr* **26**, 237-242 (2017).
118. Contreras, R.E., Schriever, S.C. & Pfluger, P.T. Physiological and Epigenetic Features of Yoyo Dieting and Weight Control. *Front Genet* **10**, 1015 (2019).
119. Montani, J.P., Schutz, Y. & Dulloo, A.G. Dieting and weight cycling as risk factors for cardiometabolic diseases: who is really at risk? *Obes Rev* **16 Suppl 1**, 7-18 (2015).
120. Kasubuchi, M., Hasegawa, S., Hiramatsu, T., Ichimura, A. & Kimura, I. Dietary gut microbial metabolites, short-chain fatty acids, and host metabolic regulation. *Nutrients* **7**, 2839-2849 (2015).
121. Zhang, L., Liu, C., Jiang, Q. & Yin, Y. Butyrate in Energy Metabolism: There Is Still More to Learn. *Trends Endocrinol Metab* **32**, 159-169 (2021).

122. Cencic, A. & Langerholc, T. Functional cell models of the gut and their applications in food microbiology--a review. *Int J Food Microbiol* **141 Suppl 1**, S4-14 (2010).
123. Jochems, P.G.M., Garssen, J., van Keulen, A.M., Masereeuw, R. & Jeurink, P.V. Evaluating Human Intestinal Cell Lines for Studying Dietary Protein Absorption. *Nutrients* **10**(2018).
124. Beilharz, J.E., Maniam, J. & Morris, M.J. Diet-Induced Cognitive Deficits: The Role of Fat and Sugar, Potential Mechanisms and Nutritional Interventions. *Nutrients* **7**, 6719-6738 (2015).
125. Cordner, Z.A. & Tamashiro, K.L. Effects of high-fat diet exposure on learning & memory. *Physiol Behav* **152**, 363-371 (2015).
126. Holm-Hansen, S*., et al.* Behavioural effects of high fat diet in a mutant mouse model for the schizophrenia risk gene neuregulin 1. *Genes Brain Behav* **15**, 295-304 (2016).
127. Zieba, J., Morris, M.J. & Karl, T. Behavioural effects of high fat diet exposure starting in late adolescence in neuregulin 1 transmembrane domain mutant mice. *Behav Brain Res* **373**, 112074 (2019).
128. Kim, S.J., Windon, M.J. & Lin, S.Y. The association between diabetes and olfactory impairment in adults: A systematic review and meta-analysis. *Laryngoscope Investig Otolaryngol* **4**, 465-475 (2019).
129. Parasoglou, P., Rao, S. & Slade, J.M. Declining Skeletal Muscle Function in Diabetic Peripheral Neuropathy. *Clin Ther* **39**, 1085-1103 (2017).
130. Yagihashi, S., Mizukami, H. & Sugimoto, K. Mechanism of diabetic neuropathy: Where are we now and where to go? *J Diabetes Investig* **2**, 18-32 (2011).
131. Nguyen, L.H., Goel, A. & Chung, D.C. Pathways of Colorectal Carcinogenesis. *Gastroenterology* **158**, 291-302 (2020).
132. Mármol, I., Sánchez-De-Diego, C., Pradilla Dieste, A., Cerrada, E. & Rodriguez Yoldi, M. Colorectal Carcinoma: A General Overview and Future Perspectives in Colorectal Cancer. *International Journal of Molecular Sciences* **18**, 197 (2017).
133. Dekker, E., Tanis, P.J., Vleugels, J.L.A., Kasi, P.M. & Wallace, M.B. Colorectal cancer. *Lancet* **394**, 1467-1480 (2019).
134. Kuipers, E.J*., et al.* Colorectal cancer. *Nat Rev Dis Primers* **1**, 15065 (2015).
135. Akimoto, N*., et al.* Rising incidence of early-onset colorectal cancer - a call to action. *Nat Rev Clin Oncol* **18**, 230-243 (2021).
136. Connell, L.C., Mota, J.M., Braghiroli, M.I. & Hoff, P.M. The Rising Incidence of Younger Patients With Colorectal Cancer: Questions About Screening, Biology, and Treatment. *Curr Treat Options Oncol* **18**, 23 (2017).
137. Kasi, P.M*., et al.* Rising Proportion of Young Individuals With Rectal and Colon Cancer. *Clin Colorectal Cancer* **18**, e87-e95 (2019).
138. Patel, S.G., Karlitz, J.J., Yen, T., Lieu, C.H. & Boland, C.R. The rising tide of early-onset colorectal cancer: a comprehensive review of epidemiology, clinical features, biology, risk factors, prevention, and early detection. *Lancet Gastroenterol Hepatol* **7**, 262-274 (2022).
139. Young, J.P*., et al.* Rising incidence of early-onset colorectal cancer in Australia over two decades: report and review. *J Gastroenterol Hepatol* **30**, 6-13 (2015).
140. Keller, D.S., Berho, M., Perez, R.O., Wexner, S.D. & Chand, M. The multidisciplinary management of rectal cancer. *Nat Rev Gastroenterol Hepatol* **17**, 414-429 (2020).
141. Mármol, I., Sánchez-de-Diego, C., Pradilla Dieste, A., Cerrada, E. & Rodriguez Yoldi, M.J. Colorectal Carcinoma: A General Overview and Future Perspectives in Colorectal Cancer. *Int J Mol Sci* **18**(2017).
142. Harada, S. & Morlote, D. Molecular Pathology of Colorectal Cancer. *Adv Anat Pathol* **27**, 20-26 (2020).
143. Baran, B*., et al.* Difference Between Left-Sided and Right-Sided Colorectal Cancer: A Focused Review of Literature. *Gastroenterology Res* **11**, 264-273 (2018).
144. Yang, C.Y., Yen, M.H., Kiu, K.T., Chen, Y.T. & Chang, T.C. Outcomes of right-sided and left-sided colon cancer after curative resection. *Sci Rep* **12**, 11323 (2022).

145. Guinney, J.*, et al.* The consensus molecular subtypes of colorectal cancer. *Nat Med* **21**, 1350-1356 (2015).
146. Ashktorab, H. & Brim, H. Colorectal cancer subtyping. *Nat Rev Cancer* **22**, 68-69 (2022).
147. Sagaert, X., Vanstapel, A. & Verbeek, S. Tumor Heterogeneity in Colorectal Cancer: What Do We Know So Far? *Pathobiology* **85**, 72-84 (2018).
148. Hampel, H., Kalady, M.F., Pearlman, R. & Stanich, P.P. Hereditary Colorectal Cancer. *Hematol Oncol Clin North Am* **36**, 429-447 (2022).
149. Dinarvand, P.*, et al.* Familial Adenomatous Polyposis Syndrome: An Update and Review of Extraintestinal Manifestations. *Arch Pathol Lab Med* **143**, 1382-1398 (2019).
150. Cerretelli, G., Ager, A., Arends, M.J. & Frayling, I.M. Molecular pathology of Lynch syndrome. *J Pathol* **250**, 518-531 (2020).
151. Maratt, J.K. & Stoffel, E. Identification of Lynch Syndrome. *Gastrointest Endosc Clin N Am* **32**, 45-58 (2022).
152. Amin, M.B.*, et al.* The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA Cancer J Clin* **67**, 93-99 (2017).
153. Weiser, M.R. AJCC 8th Edition: Colorectal Cancer. *Ann Surg Oncol* **25**, 1454-1455 (2018).
154. Yu, G.H., Li, S.F., Wei, R. & Jiang, Z. Diabetes and Colorectal Cancer Risk: Clinical and Therapeutic Implications. *J Diabetes Res* **2022**, 1747326 (2022).
155. Chan, D.S.*, et al.* Red and processed meat and colorectal cancer incidence: meta-analysis of prospective studies. *PLoS One* **6**, e20456 (2011).
156. Zhao, Z.*, et al.* Red and processed meat consumption and colorectal cancer risk: a systematic review and meta-analysis. *Oncotarget* **8**, 83306-83314 (2017).
157. Gianfredi, V.*, et al.* Is dietary fibre truly protective against colon cancer? A systematic review and meta-analysis. *Int J Food Sci Nutr* **69**, 904-915 (2018).
158. Nucci, D.*, et al.* Association between Dietary Fibre Intake and Colorectal Adenoma: A Systematic Review and Meta-Analysis. *Int J Environ Res Public Health* **18**(2021).
159. Oh, H.*, et al.* Different dietary fibre sources and risks of colorectal cancer and adenoma: a dose-response meta-analysis of prospective studies. *Br J Nutr* **122**, 605-615 (2019).
160. Amitay, E.L.*, et al.* Smoking, alcohol consumption and colorectal cancer risk by molecular pathological subtypes and pathways. *Br J Cancer* **122**, 1604-1610 (2020).
161. Siegel, R.L.*, et al.* Colorectal cancer statistics, 2020. *CA Cancer J Clin* **70**, 145-164 (2020).
162. Monleon, D.*, et al.* Metabolite profiling of fecal water extracts from human colorectal cancer. *NMR in Biomedicine* **22**, 342-348 (2009).
163. Ohigashi, S.*, et al.* Changes of the intestinal microbiota, short chain fatty acids, and fecal pH in patients with colorectal cancer. *Digestive Diseases and Sciences* **58**, 1717-1726 (2013).
164. Amiot, A.*, et al.* H-1 NMR Spectroscopy of Fecal Extracts Enables Detection of Advanced Colorectal Neoplasia. *J. Proteome Res.* **14**, 3871-3881 (2015).
165. Lin, Y.*, et al.* NMR-based fecal metabolomics fingerprinting as predictors of earlier diagnosis in patients with colorectal cancer. *Oncotarget* **7**, 29454-29464 (2016).
166. Niccolai, E.*, et al.* Evaluation and comparison of short chain fatty acids composition in gut diseases. *World Journal of Gastroenterology* **25**, 5543-5558 (2019).
167. Torii, T., Kanemitsu, K. & Hagiwara, A. Simultaneous Assay of Fecal Short-Chain Fatty and Bile Acids and Ratio of Total Bile Acids to Butyrate in Colon Cancer. *Chromatography* **40**, 49-57 (2019).
168. Boutron-Ruault, M.C.*, et al.* Effects of a 3-mo consumption of short-chain fructo-oligosaccharides on parameters of colorectal carcinogenesis in patients with or without small or large colorectal adenomas. *Nutrition and Cancer* **53**, 160-168 (2005).
169. Bridges, K.M.*, et al.* Relating stool microbial metabolite levels, inflammatory markers and dietary behaviors to screening colonoscopy findings in a racially/ethnically diverse patient population. *Genes* **9**(2018).

170. Chen, C.Y.*, et al.* Bacteroides, butyric acid and t10,c12-CLA changes in colorectal adenomatous polyp patients. *Gut Pathogens* **13**, 9 (2021).
171. Chen, H.M.*, et al.* Decreased dietary fiber intake and structural alteration of gut microbiota in patients with advanced colorectal adenoma. *American Journal of Clinical Nutrition* **97**, 1044-1052 (2013).
172. Kashtan, H.*, et al.* Colonic fermentation and markers of colorectal-cancer risk. *American Journal of Clinical Nutrition* **55**, 723-728 (1992).
173. Lin, Y.*, et al.* H-1 NMR-based metabolomics reveal overlapping discriminatory metabolites and metabolic pathway disturbances between colorectal tumor tissues and fecal samples. *Int. J. Cancer* **145**, 1679-1689 (2019).
174. Song, E.M.*, et al.* Fecal Fatty Acid Profiling as a Potential New Screening Biomarker in Patients with Colorectal Cancer. *Digestive Diseases and Sciences* **63**, 1229-1236 (2018).
175. Sze, M.A., Topcuoglu, B.D., Lesniak, N.A., Ruffin, M.T. & Schloss, P.D. Fecal short-chain fatty acids are not predictive of colonic tumor status and cannot be predicted based on bacterial community structure. *mBio* **10**(2019).
176. Weaver, G.A., Krause, J.A., Miller, T.L. & Wolin, M.J. Short chain fatty acid distributions of enema samples from a sigmoidoscopy population: An association of high acetate and low butyrate ratios with adenomatous polyps and colon cancer. *Gut* **29**, 1539-1543 (1988).
177. Weir, T.L.*, et al.* Stool Microbiome and Metabolome Differences between Colorectal Cancer Patients and Healthy Adults. *PLoS ONE* **8**(2013).
178. Yusuf, F., Adewiah, S. & Fatchiyah, F. The level short chain fatty acids and HSP 70 in colorectal cancer and non-colorectal cancer. *Acta Informatica Medica* **26**, 160-163 (2018).
179. Nannini, G.*, et al.* Fecal metabolomic profiles: A comparative study of patients with colorectal cancer vs adenomatous polyps. *World J Gastroenterol* **27**, 6430-6441 (2021).
180. Ocvirk, S.*, et al.* A prospective cohort analysis of gut microbial co-metabolism in Alaska Native and rural African people at high and low risk of colorectal cancer. *American Journal of Clinical Nutrition* **111**, 406-419 (2020).
181. Katsidzira, L.*, et al.* Differences in Fecal Gut Microbiota, Short-Chain Fatty Acids and Bile Acids Link Colorectal Cancer Risk to Dietary Changes Associated with Urbanization Among Zimbabweans. *Nutrition and Cancer* **71**, 1313-1324 (2019).
182. Hester, C.M.*, et al.* Fecal microbes, short chain fatty acids, and colorectal cancer across racial/ethnic groups. *World Journal of Gastroenterology* **21**, 2759-2769 (2015).
183. Ou, J.*, et al.* Diet, microbiota, and microbial metabolites in colon cancer risk in rural Africans and African Americans. *American Journal of Clinical Nutrition* **98**, 111-120 (2013).
184. Ou, J., DeLany, J.P., Zhang, M., Sharma, S. & O'Keefe, S.J.D. Association between low colonic short-chain fatty acids and high bile acids in high colon cancer risk populations. *Nutrition and Cancer* **64**, 34-40 (2012).
185. O'Keefe, S.J.D.*, et al.* Products of the colonic microbiota mediate the effects of diet on colon cancer risk. *Journal of Nutrition* **139**, 2044-2048 (2009).
186. Kanth, P. & Inadomi, J.M. Screening and prevention of colorectal cancer. *Bmj* **374**, n1855 (2021).
187. Zauber, A.G.*, et al.* Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. *N Engl J Med* **366**, 687-696 (2012).
188. Nishihara, R.*, et al.* Long-term colorectal-cancer incidence and mortality after lower endoscopy. *N Engl J Med* **369**, 1095-1105 (2013).
189. Ladabaum, U., Dominitz, J.A., Kahi, C. & Schoen, R.E. Strategies for Colorectal Cancer Screening. *Gastroenterology* **158**, 418-432 (2020).
190. Song, M., Garrett, W.S. & Chan, A.T. Nutrients, foods, and colorectal cancer prevention. *Gastroenterology* **148**, 1244-1260.e1216 (2015).
191. Thanikachalam, K. & Khan, G. Colorectal Cancer and Nutrition. *Nutrients* **11**(2019).

192. Ferrari, A., Neefs, I., Hoeck, S., Peeters, M. & Van Hal, G. Towards Novel Non-Invasive Colorectal Cancer Screening Methods: A Comprehensive Review. *Cancers (Basel)* **13**(2021).
193. Jodal, H.C*., et al.* Colorectal cancer screening with faecal testing, sigmoidoscopy or colonoscopy: a systematic review and network meta-analysis. *BMJ Open* **9**, e032773 (2019).
194. Imperiale, T.F., Gruber, R.N., Stump, T.E., Emmett, T.W. & Monahan, P.O. Performance Characteristics of Fecal Immunochemical Tests for Colorectal Cancer and Advanced Adenomatous Polyps: A Systematic Review and Meta-analysis. *Ann Intern Med* **170**, 319-329 (2019).
195. Araghi, M*., et al.* Changes in colorectal cancer incidence in seven high-income countries: a population-based study. *The Lancet Gastroenterology & Hepatology* **4**, 511-518 (2019).
196. Sinicrope, F.A. Increasing Incidence of Early-Onset Colorectal Cancer. *N Engl J Med* **386**, 1547-1558 (2022).
197. Burnett-Hartman, A.N., Lee, J.K., Demb, J. & Gupta, S. An Update on the Epidemiology, Molecular Characterization, Diagnosis, and Screening Strategies for Early-Onset Colorectal Cancer. *Gastroenterology* **160**, 1041-1049 (2021).
198. Done, J.Z. & Fang, S.H. Young-onset colorectal cancer: A review. *World J Gastrointest Oncol* **13**, 856-866 (2021).
199. Garrett, C., Steffens, D., Solomon, M. & Koh, C. Early-onset colorectal cancer: why it should be high on our list of differentials. *ANZ J Surg* **92**, 1638-1643 (2022).
200. Mauri, G*., et al.* Early-onset colorectal cancer in young individuals. *Mol Oncol* **13**, 109-131 (2019).
201. Saad El Din, K*., et al.* Trends in the epidemiology of young-onset colorectal cancer: a worldwide systematic review. *BMC Cancer* **20**, 288 (2020).
202. Shah, R.R*., et al.* Trends in the incidence of early-onset colorectal cancer in all 50 United States from 2001 through 2017. *Cancer* **128**, 299-310 (2022).
203. Vuik, F.E*., et al.* Increasing incidence of colorectal cancer in young adults in Europe over the last 25 years. *Gut* **68**, 1820-1826 (2019).
204. Eng, C*., et al.* A comprehensive framework for early-onset colorectal cancer research. *Lancet Oncol* **23**, e116-e128 (2022).
205. Saraiva, M.R., Rosa, I. & Claro, I. Early-onset colorectal cancer: A review of current knowledge. *World J Gastroenterol* **29**, 1289-1303 (2023).
206. Loomans-Kropp, H.A. & Umar, A. Increasing Incidence of Colorectal Cancer in Young Adults. *J Cancer Epidemiol* **2019**, 9841295 (2019).
207. Wei, T., Luo, P., Zhang, X., Zhu, W. & Zhang, J. Comparison of the incidence of colorectal cancer in young adults between the USA and Europe. *Gut* **69**, 1540-1542 (2020).
208. Brenner, D.R*., et al.* Increasing colorectal cancer incidence trends among younger adults in Canada. *Prev Med* **105**, 345-349 (2017).
209. Australian Institute of Health and Welfare 2021. Cancer in Australia 2021. Cancer series no. 133. Cat. no. CAN 144. Canberra: AIHW.
210. Feletto, E*., et al.* Trends in Colon and Rectal Cancer Incidence in Australia from 1982 to 2014: Analysis of Data on Over 375,000 Cases. *Cancer Epidemiol Biomarkers Prev* **28**, 83-90 (2019).
211. Boyce, S*., et al.* Young-onset colorectal cancer in New South Wales: a population-based study. *Med J Aust* **205**, 465-470 (2016).
212. Sia, C.S*., et al.* No increase in colorectal cancer in patients under 50 years of age: a Victorian experience from the last decade. *Colorectal Dis* **16**, 690-695 (2014).
213. Jones, H.G*., et al.* Clinicopathological characteristics of colorectal cancer presenting under the age of 50. *Int J Colorectal Dis* **30**, 483-489 (2015).
214. Nagai, Y*., et al.* Clinicopathological Features of Colorectal Cancer Patients Under the Age of 50: Recent Experience and Case-Control Study of Prognosis in a Japanese Cohort. *Digestion* **93**, 272-279 (2016).

215. Wu, C.W. & Lui, R.N. Early-onset colorectal cancer: Current insights and future directions. *World J Gastrointest Oncol* **14**, 230-241 (2022).
216. Low, E.E.*, et al.* Risk Factors for Early-Onset Colorectal Cancer. *Gastroenterology* **159**, 492-501.e497 (2020).
217. O'Sullivan, D.E.*, et al.* Risk Factors for Early-Onset Colorectal Cancer: A Systematic Review and Meta-analysis. *Clin Gastroenterol Hepatol* **20**, 1229-1240.e1225 (2022).
218. Syed, A.R.*, et al.* Old vs new: Risk factors predicting early onset colorectal cancer. *World J Gastrointest Oncol* **11**, 1011-1020 (2019).
219. Gausman, V.*, et al.* Risk Factors Associated With Early-Onset Colorectal Cancer. *Clin Gastroenterol Hepatol* **18**, 2752-2759.e2752 (2020).
220. Álvaro, E.*, et al.* Clinical and Molecular Comparative Study of Colorectal Cancer Based on Age-of-onset and Tumor Location: Two Main Criteria for Subclassifying Colorectal Cancer. *Int J Mol Sci* **20**(2019).
221. Perea, J.*, et al.* Classifying early-onset colorectal cancer according to tumor location: new potential subcategories to explore. *Am J Cancer Res* **5**, 2308-2313 (2015).
222. Du, M.*, et al.* Integrated multi-omics approach to distinct molecular characterization and classification of early-onset colorectal cancer. *Cell Rep Med* **4**, 100974 (2023).
223. Ugai, T.*, et al.* Molecular Characteristics of Early-Onset Colorectal Cancer According to Detailed Anatomical Locations: Comparison With Later-Onset Cases. *Am J Gastroenterol* **118**, 712-726 (2023).
224. Sanford, N.N., Dharwadkar, P. & Murphy, C.C. Early-onset colorectal cancer: more than one side to the story. *Colorectal Cancer* **9**, CRC28 (2020).
225. AlZaabi, A.*, et al.* Early onset colorectal cancer: Challenges across the cancer care continuum. *Ann Med Surg (Lond)* **82**, 104453 (2022).
226. You, Y.N., Xing, Y., Feig, B.W., Chang, G.J. & Cormier, J.N. Young-onset colorectal cancer: is it time to pay attention? *Arch Intern Med* **172**, 287-289 (2012).
227. Lamprell, K.*, et al.* People with early-onset colorectal cancer describe primary care barriers to timely diagnosis: a mixed-methods study of web-based patient reports in the United Kingdom, Australia and New Zealand. *BMC Prim Care* **24**, 12 (2023).
228. Rosenberg, J., Angenete, E., Pinkney, T., Bhangu, A. & Haglind, E. Collaboration in colorectal surgical research. *Colorectal Dis* **23**, 2741-2749 (2021).
229. Cheng, E.*, et al.* Analysis of Survival Among Adults With Early-Onset Colorectal Cancer in the National Cancer Database. *JAMA Netw Open* **4**, e2112539 (2021).
230. Baars, J.E.*, et al.* Age at diagnosis of inflammatory bowel disease influences early development of colorectal cancer in inflammatory bowel disease patients: a nationwide, long-term survey. *J Gastroenterol* **47**, 1308-1322 (2012).
231. Jebeile, H., Kelly, A.S., O'Malley, G. & Baur, L.A. Obesity in children and adolescents: epidemiology, causes, assessment, and management. *Lancet Diabetes Endocrinol* **10**, 351-365 (2022).
232. Arroyo-Johnson, C. & Mincey, K.D. Obesity Epidemiology Worldwide. *Gastroenterol Clin North Am* **45**, 571-579 (2016).
233. Li, H., Boakye, D., Chen, X., Hoffmeister, M. & Brenner, H. Association of Body Mass Index With Risk of Early-Onset Colorectal Cancer: Systematic Review and Meta-Analysis. *Am J Gastroenterol* **116**, 2173-2183 (2021).
234. Spaander, M.C.W.*, et al.* Young-onset colorectal cancer. *Nat Rev Dis Primers* **9**, 21 (2023).
235. Rawla, P., Sunkara, T. & Barsouk, A. Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. *Prz Gastroenterol* **14**, 89-103 (2019).
236. Ireland, M.J.*, et al.* A systematic review of geographical differences in management and outcomes for colorectal cancer in Australia. *BMC Cancer* **17**, 95 (2017).

237. Troeung, L.*, et al.* Increasing Incidence of Colorectal Cancer in Adolescents and Young Adults Aged 15-39 Years in Western Australia 1982-2007: Examination of Colonoscopy History. *Front Public Health* **5**, 179 (2017).
238. Haggar, F.A., Preen, D.B., Pereira, G., Holman, C.D. & Einarsdottir, K. Cancer incidence and mortality trends in Australian adolescents and young adults, 1982-2007. *BMC Cancer* **12**, 151 (2012).
239. Khalangot, M., Krasnienkov, D. & Vaiserman, A. Telomere length in different metabolic categories: Clinical associations and modification potential. *Exp Biol Med (Maywood)* **245**, 1115-1121 (2020).
240. Di Micco, R., Krizhanovsky, V., Baker, D. & d'Adda di Fagagna, F. Cellular senescence in ageing: from mechanisms to therapeutic opportunities. *Nat Rev Mol Cell Biol* **22**, 75-95 (2021).
241. Kumari, R. & Jat, P. Mechanisms of Cellular Senescence: Cell Cycle Arrest and Senescence Associated Secretory Phenotype. *Front Cell Dev Biol* **9**, 645593 (2021).
242. Matthews, H.K., Bertoli, C. & de Bruin, R.A.M. Cell cycle control in cancer. *Nat Rev Mol Cell Biol* **23**, 74-88 (2022).
243. Sharpless, N.E. & Sherr, C.J. Forging a signature of in vivo senescence. *Nat Rev Cancer* **15**, 397-408 (2015).
244. Gorgoulis, V.*, et al.* Cellular Senescence: Defining a Path Forward. *Cell* **179**, 813-827 (2019).
245. González-Gualda, E., Baker, A.G., Fruk, L. & Muñoz-Espín, D. A guide to assessing cellular senescence in vitro and in vivo. *Febs j* **288**, 56-80 (2021).
246. Roger, L., Tomas, F. & Gire, V. Mechanisms and Regulation of Cellular Senescence. *Int J Mol Sci* **22**(2021).
247. Barnum, K.J. & O'Connell, M.J. Cell cycle regulation by checkpoints. *Methods Mol Biol* **1170**, 29-40 (2014).
248. Liu, J., Wang, L., Wang, Z. & Liu, J.P. Roles of Telomere Biology in Cell Senescence, Replicative and Chronological Ageing. *Cells* **8**(2019).
249. Kwon, S.M., Hong, S.M., Lee, Y.K., Min, S. & Yoon, G. Metabolic features and regulation in cell senescence. *BMB Rep* **52**, 5-12 (2019).
250. Kim, Y.M., Seo, Y.H., Park, C.B., Yoon, S.H. & Yoon, G. Roles of GSK3 in metabolic shift toward abnormal anabolism in cell senescence. *Ann N Y Acad Sci* **1201**, 65-71 (2010).
251. Lee, S.H., Lee, J.H., Lee, H.Y. & Min, K.J. Sirtuin signaling in cellular senescence and aging. *BMB Rep* **52**, 24-34 (2019).
252. Sikora, E., Bielak-Zmijewska, A. & Mosieniak, G. A common signature of cellular senescence; does it exist? *Ageing Res Rev* **71**, 101458 (2021).
253. Birch, J. & Gil, J. Senescence and the SASP: many therapeutic avenues. *Genes Dev* **34**, 1565-1576 (2020).
254. Saul, D.*, et al.* A new gene set identifies senescent cells and predicts senescence-associated pathways across tissues. *Nat Commun* **13**, 4827 (2022).
255. Lopes-Paciencia, S.*, et al.* The senescence-associated secretory phenotype and its regulation. *Cytokine* **117**, 15-22 (2019).
256. Koobotse, M.O., Schmidt, D., Holly, J.M.P. & Perks, C.M. Glucose Concentration in Cell Culture Medium Influences the BRCA1-Mediated Regulation of the Lipogenic Action of IGF-I in Breast Cancer Cells. *Int J Mol Sci* **21**(2020).
257. Joglekar, M.V.*, et al.* An Optimised Step-by-Step Protocol for Measuring Relative Telomere Length. *Methods Protoc* **3**(2020).
258. Cheng, F.*, et al.* Shortened Leukocyte Telomere Length Is Associated With Glycemic Progression in Type 2 Diabetes: A Prospective and Mendelian Randomization Analysis. *Diabetes Care* **45**, 701-709 (2022).
259. Cheng, F.*, et al.* Shortened Relative Leukocyte Telomere Length Is Associated With Prevalent and Incident Cardiovascular Complications in Type 2 Diabetes: Analysis From the Hong Kong Diabetes Register. *Diabetes Care* **43**, 2257-2265 (2020).

260. Murillo-Ortiz, B.*, et al.* Telomere length and type 2 diabetes in males, a premature aging syndrome. *Aging Male* **15**, 54-58 (2012).
261. Salpea, K.D.*, et al.* Association of telomere length with type 2 diabetes, oxidative stress and UCP2 gene variation. *Atherosclerosis* **209**, 42-50 (2010).
262. Shen, Q.*, et al.* Association of leukocyte telomere length with type 2 diabetes in mainland Chinese populations. *J Clin Endocrinol Metab* **97**, 1371-1374 (2012).
263. Zee, R.Y., Castonguay, A.J., Barton, N.S., Germer, S. & Martin, M. Mean leukocyte telomere length shortening and type 2 diabetes mellitus: a case-control study. *Transl Res* **155**, 166-169 (2010).
264. Menke, A., Casagrande, S. & Cowie, C.C. Leukocyte telomere length and diabetes status, duration, and control: the 1999-2002 National Health and Nutrition Examination Survey. *BMC Endocr Disord* **15**, 52 (2015).
265. Tudurí, E.*, et al.* The pancreatic β-cell in ageing: Implications in age-related diabetes. *Ageing Res Rev* **80**, 101674 (2022).
266. Aguayo-Mazzucato, C. Functional changes in beta cells during ageing and senescence. *Diabetologia* **63**, 2022-2029 (2020).
267. Da Silva Xavier, G. The Cells of the Islets of Langerhans. *J Clin Med* **7**(2018).
268. Tolhurst, G.*, et al.* Short-chain fatty acids stimulate glucagon-like peptide-1 secretion via the G-protein-coupled receptor FFAR2. *Diabetes* **61**, 364-371 (2012).
269. Le Poul, E.*, et al.* Functional characterization of human receptors for short chain fatty acids and their role in polymorphonuclear cell activation. *J Biol Chem* **278**, 25481-25489 (2003).
270. Muscogiuri, G.*, et al.* Gut microbiota: a new path to treat obesity. *Int J Obes Suppl* **9**, 10-19 (2019).
271. Lee, C.J., Sears, C.L. & Maruthur, N. Gut microbiome and its role in obesity and insulin resistance. *Ann N Y Acad Sci* **1461**, 37-52 (2020).
272. Pitocco, D.*, et al.* The role of gut microbiota in mediating obesity and diabetes mellitus. *Eur Rev Med Pharmacol Sci* **24**, 1548-1562 (2020).
273. Zhou, Z., Sun, B., Yu, D. & Zhu, C. Gut Microbiota: An Important Player in Type 2 Diabetes Mellitus. *Front Cell Infect Microbiol* **12**, 834485 (2022).
274. Jolivalt, C.G.*, et al.* Peripheral Neuropathy in Mouse Models of Diabetes. *Curr Protoc Mouse Biol* **6**, 223-255 (2016).
275. Zieba, J., Morris, M.J., Weickert, C.S. & Karl, T. Behavioural effects of high fat diet in adult Nrg1 type III transgenic mice. *Behav Brain Res* **377**, 112217 (2020).
276. Graber, T.G., Kim, J.H., Grange, R.W., McLoon, L.K. & Thompson, L.V. C57BL/6 life span study: age-related declines in muscle power production and contractile velocity. *Age (Dordr)* **37**, 9773 (2015).
277. Reeves, P.G., Nielsen, F.H. & Fahey, G.C., Jr. AIN-93 purified diets for laboratory rodents: final report of the American Institute of Nutrition ad hoc writing committee on the reformulation of the AIN-76A rodent diet. *J Nutr* **123**, 1939-1951 (1993).
278. Cândido, F.G.*, et al.* Impact of dietary fat on gut microbiota and low-grade systemic inflammation: mechanisms and clinical implications on obesity. *Int J Food Sci Nutr* **69**, 125-143 (2018).
279. Netto Candido, T.L., Bressan, J. & Alfenas, R.C.G. Dysbiosis and metabolic endotoxemia induced by high-fat diet. *Nutr Hosp* **35**, 1432-1440 (2018).
280. Martinez, K.B., Leone, V. & Chang, E.B. Western diets, gut dysbiosis, and metabolic diseases: Are they linked? *Gut Microbes* **8**, 130-142 (2017).
281. Zmora, N., Suez, J. & Elinav, E. You are what you eat: diet, health and the gut microbiota. *Nat Rev Gastroenterol Hepatol* **16**, 35-56 (2019).
282. Han, M.*, et al.* Dietary Fiber Gap and Host Gut Microbiota. *Protein Pept Lett* **24**, 388-396 (2017).

283. Canfora, E.E., Meex, R.C.R., Venema, K. & Blaak, E.E. Gut microbial metabolites in obesity, NAFLD and T2DM. *Nat Rev Endocrinol* **15**, 261-273 (2019).
284. Mandaliya, D.K. & Seshadri, S. Short Chain Fatty Acids, pancreatic dysfunction and type 2 diabetes. *Pancreatology* **19**, 280-284 (2019).
285. Agus, A., Clément, K. & Sokol, H. Gut microbiota-derived metabolites as central regulators in metabolic disorders. *Gut* **70**, 1174-1182 (2021).
286. Michałowska, J., Miller-Kasprzak, E. & Bogdański, P. Incretin Hormones in Obesity and Related Cardiometabolic Disorders: The Clinical Perspective. *Nutrients* **13**(2021).
287. Popkin, B.M., Adair, L.S. & Ng, S.W. Global nutrition transition and the pandemic of obesity in developing countries. *Nutr Rev* **70**, 3-21 (2012).
288. Pillon, N.J., Loos, R.J.F., Marshall, S.M. & Zierath, J.R. Metabolic consequences of obesity and type 2 diabetes: Balancing genes and environment for personalized care. *Cell* **184**, 1530-1544 (2021).
289. Yang, M., Liu, S. & Zhang, C. The Related Metabolic Diseases and Treatments of Obesity. *Healthcare (Basel)* **10**(2022).
290. Piché, M.E., Tchernof, A. & Després, J.P. Obesity Phenotypes, Diabetes, and Cardiovascular Diseases. *Circ Res* **126**, 1477-1500 (2020).
291. Dulloo, A.G. & Montani, J.P. Pathways from dieting to weight regain, to obesity and to the metabolic syndrome: an overview. *Obes Rev* **16 Suppl 1**, 1-6 (2015).
292. Yadav, H., Lee, J.H., Lloyd, J., Walter, P. & Rane, S.G. Beneficial metabolic effects of a probiotic via butyrate-induced GLP-1 hormone secretion. *J Biol Chem* **288**, 25088-25097 (2013).
293. Larraufie, P.*, et al.* SCFAs strongly stimulate PYY production in human enteroendocrine cells. *Sci Rep* **8**, 74 (2018).
294. Christiansen, C.B.*, et al.* The impact of short-chain fatty acids on GLP-1 and PYY secretion from the isolated perfused rat colon. *Am J Physiol Gastrointest Liver Physiol* **315**, G53-g65 (2018).
295. Tucker, K.R., Godbey, S.J., Thiebaud, N. & Fadool, D.A. Olfactory ability and object memory in three mouse models of varying body weight, metabolic hormones, and adiposity. *Physiol Behav* **107**, 424-432 (2012).
296. Bayley, J.S., Pedersen, T.H. & Nielsen, O.B. Skeletal muscle dysfunction in the db/db mouse model of type 2 diabetes. *Muscle Nerve* **54**, 460-468 (2016).
297. Page, M.J.*, et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Bmj* **372**, n71 (2021).
298. Muka, T.*, et al.* A 24-step guide on how to design, conduct, and successfully publish a systematic review and meta-analysis in medical research. *Eur J Epidemiol* **35**, 49-60 (2020).
299. Wells, G.A.*, et al.* The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Vol. 2021 (2011).
300. Moola, S.*, et al.* Systematic reviews of etiology and risk. in *JBI manual for evidence synthesis* (eds. Aromataris E & Munn Z) (JBI, 2020).
301. Ma, L.L.*, et al.* Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? *Mil Med Res* **7**, 7 (2020).
302. Higgins, J.*, et al.* Analysing data and undertaking meta-analyses. in *Cochrane Handbook for Systematic Reviews of Interventions* (2021).
303. Baldi, S.*, et al.* Free fatty acids signature in human intestinal disorders: Significant association between butyric acid and celiac disease. *Nutrients* **13**, 1-14 (2021).
304. Bartolucci, G.*, et al.* A method for assessing plasma free fatty acids from C2 to C18 and its application for the early detection of colorectal cancer. *J Pharm Biomed Anal* **215**, 114762 (2022).
305. Genua, F.*, et al.* Association of circulating short chain fatty acid levels with colorectal adenomas and colorectal cancer. *Clin Nutr ESPEN* **46**, 297-304 (2021).

306. Segal, I., Hassan, H., Walker, A.R.P., Becker, P. & Braganza, J. Fecal short chain fatty acids in South African urban Africans and whites. *Diseases of the Colon and Rectum* **38**, 732-734 (1995).
307. Wang, X., Wang, J., Rao, B. & Deng, L.I. Gut flora profiling and fecal metabolite composition of colorectal cancer patients and healthy individuals. *Experimental and Therapeutic Medicine* **13**, 2848-2854 (2017).
308. Borges-Canha, M., Portela-Cidade, J.P., Dinis-Ribeiro, M., Leite-Moreira, A.F. & Pimentel-Nunes, P. Role of colonic microbiota in colorectal carcinogenesis: a systematic review. *Rev Esp Enferm Dig* **107**, 659-671 (2015).
309. Rao, M.*, et al.* Non-Digestible Carbohydrate and the Risk of Colorectal Neoplasia: A Systematic Review. *Nutr Cancer* **73**, 31-44 (2021).
310. Ma, Y.*, et al.* Dietary fiber intake and risks of proximal and distal colon cancers: A meta-analysis. *Medicine (Baltimore)* **97**, e11678 (2018).
311. Sun, Q., Jia, Q., Song, L. & Duan, L. Alterations in fecal short-chain fatty acids in patients with irritable bowel syndrome: A systematic review and meta-analysis. *Medicine (Baltimore)* **98**, e14513 (2019).
312. Zhuang, X.*, et al.* Systematic Review and Meta-analysis: Short-Chain Fatty Acid Characterization in Patients With Inflammatory Bowel Disease. *Inflamm Bowel Dis* **25**, 1751-1763 (2019).
313. Rios-Covian, D.*, et al.* An Overview on Fecal Branched Short-Chain Fatty Acids Along Human Life and as Related With Body Mass Index: Associated Dietary and Anthropometric Factors. *Front Microbiol* **11**, 973 (2020).
314. Shuwen, H.*, et al.* Protective effect of the "food-microorganism-SCFAs" axis on colorectal cancer: from basic research to practical application. *J Cancer Res Clin Oncol* **145**, 2169-2197 (2019).
315. Kastrinos, F., Samadder, N.J. & Burt, R.W. Use of Family History and Genetic Testing to Determine Risk of Colorectal Cancer. *Gastroenterology* **158**, 389-403 (2020).
316. Baena, R. & Salinas, P. Diet and colorectal cancer. *Maturitas* **80**, 258-264 (2015).
317. Setia, M.S. Methodology Series Module 2: Case-control Studies. *Indian J Dermatol* **61**, 146-151 (2016).
318. Setia, M.S. Methodology Series Module 3: Cross-sectional Studies. *Indian J Dermatol* **61**, 261-264 (2016).
319. Ocvirk, S., Wilson, A.S., Appolonia, C.N., Thomas, T.K. & O'Keefe, S.J.D. Fiber, Fat, and Colorectal Cancer: New Insight into Modifiable Dietary Risk Factors. *Curr Gastroenterol Rep* **21**, 62 (2019).
320. Nguyen, T.T., Ung, T.T., Kim, N.H. & Jung, Y.D. Role of bile acids in colon carcinogenesis. *World J Clin Cases* **6**, 577-588 (2018).
321. Jia, W., Xie, G. & Jia, W. Bile acid-microbiota crosstalk in gastrointestinal inflammation and carcinogenesis. *Nat Rev Gastroenterol Hepatol* **15**, 111-128 (2018).
322. Davidson, K.W.*, et al.* Screening for Colorectal Cancer: US Preventive Services Task Force Recommendation Statement. *Jama* **325**, 1965-1977 (2021).
323. Anghel, S.A., Ioniță-Mîndrican, C.B., Luca, I. & Pop, A.L. Promising Epigenetic Biomarkers for the Early Detection of Colorectal Cancer: A Systematic Review. *Cancers (Basel)* **13**(2021).
324. Araghi, M.*, et al.* Changes in colorectal cancer incidence in seven high-income countries: a population-based study. *Lancet Gastroenterol Hepatol* **4**, 511-518 (2019).
325. Gandhi, J.*, et al.* Population-based study demonstrating an increase in colorectal cancer in young patients. *Br J Surg* **104**, 1063-1068 (2017).
326. Sung, J.J.Y.*, et al.* Increasing Trend in Young-Onset Colorectal Cancer in Asia: More Cancers in Men and More Rectal Cancers. *Am J Gastroenterol* **114**, 322-329 (2019).
327. Silva, A.C.B.*, et al.* Young-age onset colorectal cancer in Brazil: Analysis of incidence, clinical features, and outcomes in a tertiary cancer center. *Curr Probl Cancer* **43**, 477-486 (2019).

328. Dent, O.F., Bokey, L., Chapuis, P.H., Chan, C. & Newland, R.C. Trends in short-term outcomes after resection of colorectal cancer: 1971-2013. *ANZ J Surg* **87**, 39-43 (2017).
329. Dent, O.F., Newland, R.C., Chan, C., Bokey, L. & Chapuis, P.H. Trends in pathology and long-term outcomes after resection of colorectal cancer: 1971-2013. *ANZ J Surg* **87**, 34-38 (2017).
330. Curtis, N.J*., et al.* Time from colorectal cancer diagnosis to laparoscopic curative surgery-is there a safe window for prehabilitation? *Int J Colorectal Dis* **33**, 979-983 (2018).
331. Hangaard Hansen, C., Gögenur, M., Tvilling Madsen, M. & Gögenur, I. The effect of time from diagnosis to surgery on oncological outcomes in patients undergoing surgery for colon cancer: A systematic review. *Eur J Surg Oncol* **44**, 1479-1485 (2018).
332. van Walraven, C. & McAlister, F.A. Competing risk bias was common in Kaplan-Meier risk estimates published in prominent medical journals. *J Clin Epidemiol* **69**, 170-173.e178 (2016).
333. Petersson, J*., et al.* Increasing incidence of colorectal cancer among the younger population in Sweden. *BJS Open* **4**, 645-658 (2020).
334. Ghodssi-Ghassemabadi, R., Hajizadeh, E., Kamian, S. & Mahmoudi, M. Clinicopathological features and survival of colorectal cancer patients younger than 50 years: a retrospective comparative study. *J Egypt Natl Canc Inst* **31**, 6 (2019).
335. Siegel, R.L*., et al.* Global patterns and trends in colorectal cancer incidence in young adults. *Gut* **68**, 2179-2185 (2019).
336. Gu, J*., et al.* A risk scoring system to predict the individual incidence of early-onset colorectal cancer. *BMC Cancer* **22**, 122 (2022).
337. Mueller, M*., et al.* Colorectal cancer of the young displays distinct features of aggressive tumor biology: A single-center cohort study. *World J Gastrointest Surg* **13**, 164-175 (2021).
338. Willauer, A.N*., et al.* Clinical and molecular characterization of early-onset colorectal cancer. *Cancer* **125**, 2002-2010 (2019).
339. Chang, D.T*., et al.* Clinicopathologic and molecular features of sporadic early-onset colorectal adenocarcinoma: an adenocarcinoma with frequent signet ring cell differentiation, rectal and sigmoid involvement, and adverse morphologic features. *Mod Pathol* **25**, 1128-1139 (2012).
340. Wu, X., Lin, H. & Li, S. Prognoses of different pathological subtypes of colorectal cancer at different stages: A population-based retrospective cohort study. *BMC Gastroenterol* **19**, 164 (2019).
341. Parkin, C.J., Bell, S.W. & Mirbagheri, N. Colorectal cancer screening in Australia: An update. *Aust J Gen Pract* **47**, 859-863 (2018).
342. Luo, Q*., et al.* Trends in colon and rectal cancer mortality in Australia from 1972 to 2015 and associated projections to 2040. *Sci Rep* **12**, 3994 (2022).
343. de la Chapelle, A. The incidence of Lynch syndrome. *Fam Cancer* **4**, 233-237 (2005).
344. Primec, M., Mičetić-Turk, D. & Langerholc, T. Analysis of short-chain fatty acids in human feces: A scoping review. *Anal Biochem* **526**, 9-21 (2017).
345. Pradhan, S., Gautam, K. & Pant, V. Variation in Laboratory Reports: Causes other than Laboratory Error. *JNMA J Nepal Med Assoc* **60**, 222-224 (2022).
346. White, G.H. & Farrance, I. Uncertainty of measurement in quantitative medical testing: a laboratory implementation guide. *Clin Biochem Rev* **25**, S1-24 (2004).
347. Gu, W.J*., et al.* The Burden of Early-Onset Colorectal Cancer and Its Risk Factors from 1990 to 2019: A Systematic Analysis for the Global Burden of Disease Study 2019. *Cancers (Basel)* **14**(2022).
348. Xiong, H*., et al.* Gut microbiota display alternative profiles in patients with early-onset colorectal cancer. *Front Cell Infect Microbiol* **12**, 1036946 (2022).
349. Yang, Y*., et al.* Dysbiosis of human gut microbiome in young-onset colorectal cancer. *Nat Commun* **12**, 6757 (2021).
350. Ahmad Kendong, S.M., Raja Ali, R.A., Nawawi, K.N.M., Ahmad, H.F. & Mokhtar, N.M. Gut Dysbiosis and Intestinal Barrier Dysfunction: Potential Explanation for Early-Onset Colorectal Cancer. *Front Cell Infect Microbiol* **11**, 744606 (2021).

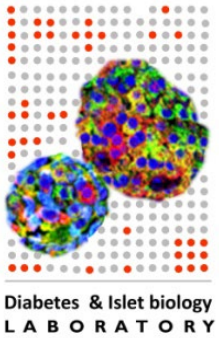
351. Muller, C., Ihionkhan, E., Stoffel, E.M. & Kupfer, S.S. Disparities in Early-Onset Colorectal Cancer. *Cells* **10**(2021).
352. Salem, M.E.*, et al.* Impact of Sociodemographic Disparities and Insurance Status on Survival of Patients with Early-Onset Colorectal Cancer. *Oncologist* **26**, e1730-e1741 (2021).

Appendices

All the protocols are established in Diabetes and Islet Biology laboratory, School of Medicine, Western Sydney University, NSW, Australia. Appendix C (relative telomere length measurement using qPCR) is written based on Joglekar et al. 2020 article (PMID: 32260112) and confirmed by Dr. Mugdha Joglekar.

Appendix A: Cell passaging and counting

Title:	Cell counting and cell distribution				
Protocol #:	16	Submitted:	21/04/2014	Approved:	21/04/2014
Category:	CC	Author(s):	WW, MVJ, AAH	Checked by:	AAH



Reagents:

1) Trypan Blue 0.4%, liquid, sterile-filtered, suitable for cell culture (Sigma Aldrich, catalog # T8154)
2) Serum Containing Media (SCM). Note- Media with vary depending on cell type
3) Trypsin + EDTA
4) Serum Free Media (SFM). Note- Media with vary depending on cell type
5) 70% Ethanol

All plastic ware should be nuclease free. Never stick your hands into containers of nuclease-free plastic ware. Always pour out whatever is required. Use of appropriate PPE is a must.

Equipment

1) Use appropriate PPE
2) Hemocytometer
3) Cover slip
4) Water bath
5) Biosafety cabinet
6) Media waste dispenser bottle
7) CO2 incubator
8) Microscope
9) Centrifuge
10) Tissue culture plate (Falcon well Clear Flat Bottom TC-Treated Cell Culture Plate- 24 wells : catalog # FAL353047; 12 wells : catalog # FAL353043; 6 well : catalog # FAL353046)
11) 1.7 ml Eppendorf tube

Reagent Setup

1) Set water bath to 37 °C
2) Place media bottles and Trypsin +EDTA into water bath

3) Turn UV on for 15 minutes before using biosafety cabinet
4) Switch UV light off after 15 minutes and turn on biosafety cabinet and open glass door of cabinet
5) Wipe biosafety cabinet working bench area with 70 % Ethanol when biosafety cabinet is “on”
6) Wipe media bottles and other reagent tubes with 70 % Ethanol before putting them into the biosafety cabinet
7) Take cell culture flask (ideally T75) out from CO2 incubator
8) Examine cells under microscope (ideally the cell culture should be confluent before proceeding cell count)

Procedure

Trypsinization

1) Wipe cell culture flask containing cells with 70 % Ethanol and place then flask into biosafety cabinet
2) Remove and discard the media in the flask
3) Add 4-5 ml of Trypsin +EDTA into the T75 flask to cover the surface of where the cells lie.
4) Incubate flask in CO_2 incubator for 2 minutes (to let the cell dislodge from the surface of the flask)
5) Examine the flask under the microscope to make sure the cells have dislodged (i.e. floating). Note- If necessary to increase the chances of dislodging all the cells- tap the sides of the flask to help the cell to dislodge
6) Add 1 ml of SCM into flask to stop the process of trypsinization
7) Resuspend media in flask and transfer media into a 15 ml tube
8) Centrifuge 15 ml tube at 1500 rpm for 3 minutes to pellet down the cells
9) Remove the supernatant
10) Resuspend cell pellet in 4 ml of SFM. Note- make sure you resuspend well to get a good distribution of the cells

Cell counting

11) Set up haemocytometer

12) Put on a dab of water on the side of the haemocytometer counting chambers (which is located in the centre). Note- make sure not to have any water on the counting chamber region.

13) Place carefully the cover slip on the haemocytometer to cover the counting chamber in the centre. Note- Make sure not to spread the water (which is only used to stick the cover slip on the haemocytometer) onto the counting chamber region

14) Resuspend the cells in 4 ml of SFM with a 1 ml pipette. Note- make sure you resuspend well to get a good distribution of the cells

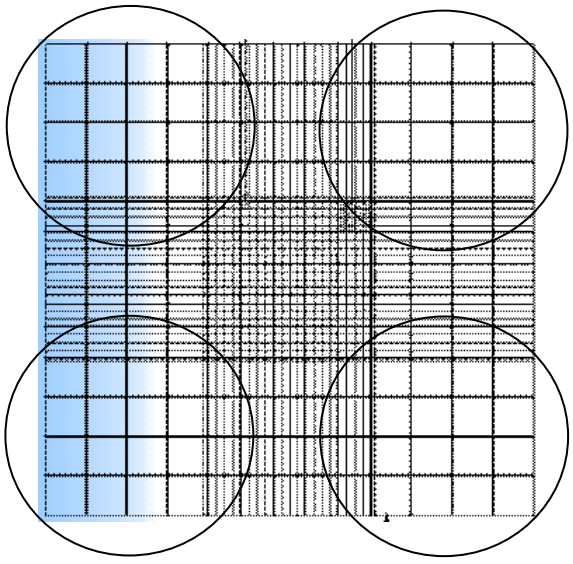
15) Transfer 10 µl of the cells into an Eppendorf tube

16) Add 10 µl of trypan blue into Eppendorf tube

17) Resuspend the cells with trypan blue mixture and transfer 10 µl of this mixture onto the haemocytometer. Note- 10 µl of mixture is loaded between the cover slip and haemocytometer. Make sure the 10 µl is spread through the rectangular cell counting region

18) Examine haemocytometer under the microscope at 10 x

19) Count each quadrant individually (the quadrants are circled in the image below. There are 16 squares per a quadrant. Include cells which sit on the edge of the line as well



20) Add the total cell count of all four quandrants together (i.e. all 64 squares)

21) Calculate the number of cells/ml

Example-

Quadrant 1 has 31 cells

Quadrant 2 has 33 cells

Quadrant 3 has 26 cells

Quadrant 4 has 24 cells

Total number of cells in all quadrants together = Q1 +Q2 +Q3 +Q4 = 31+ 33 +26 + 24

= 114

Number of Cells/ml = Total number of cells in all quandrants/4 x 10^4 x dilution factor

= **114/4 x 10^4 x 2**

= 5.7 x 10^5

Total number of cells = Number of cells per ml x total volume

= 5.7 x 10^5 x 4 ml

= 2.28 x 10^6

Cell distribution

22) Distribute a certain number of cells evenly into each well of a tissue cell culture plate

Example-

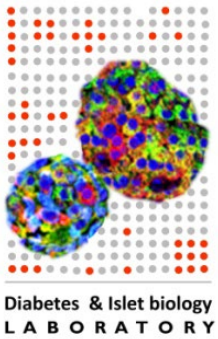
If you are intending to have 18 wells with each containing 100, 000 cells, however you have 2.28 x 10^6 cells in a total of 4 ml- it would be better to adjust the volume to make cell distribution easier

I.e. By adding 0.56 ml of SFM to the 4 ml of cell would make 500,000 cells per ml. Thus each 200 µl would contain 100 000 cells. Note- the 0.56 ml is obtained by doubling the decimal value of the million cells in total volume calculate (in this case is 0.28 ml)

23) Add additional amount of SFM media to make up volume to 500 µl in each well in a 24-tissue culture plate. Note- volume may vary in different size tissue culture plates
24) Examine cells under the microscope
25) Store plate in CO_2 incubator
26) Centrifuge any remaining cells at 1500 rpm for 3 minutes
27) Transfer supernatant into a Eppendorf tube
28) Label and store remaining dry pellet and supernatant separately in -80 °C.

Appendix B: DNA isolation from cells

Title:	**Cellular DNA extraction**				
Protocol #:	DNA1.1C/21	Submitted:	18/01/2021	Approved:	19/01/2021
Category:	MB	Author(s):	MVJ	Checked by:	AAH



Reagents

1) 1.7 ml microcentrifuge tubes (Axygen, MCT-175-C)
2) QIAmp DNA Blood Mini Kit (Qiagen, 50 rxns 51104, 250 rxns 51106)
3) Dulbecco's Phosphate Buffered Saline, DPBS (Gibco, 14190-250)
4) 100% Ethanol (Sigma, E7023-500ml)
5) Fresh aliquot for nuclease-free water
6) P10, P20, P200 and P1000 filtered tips

Equipment

1) Centrifuge for Eppendorf tubes.
2) Vortex mixer
3) Well calibrated micropipettes designated for (pre-PCR) nucleic acid isolation
4) Thermomixer (preferred) or convection oven set at 56°C.

Notes

1) All plastic wares should be nuclease free.
2) Never stick your hands into containers of nuclease-free plastic ware. Always pour out whatever is required.

Reagent Setup

1) Allow samples to thaw thoroughly on ice prior to commencing.
2) Set temperature on thermomixer or convection oven to 56°C.
3) Prepare buffers as per Qiagen manual.

Procedure

1. **Thaw the cell pellets and resuspend in 200 µl of DPBS.**
2. **Add 20 µl Qiagen protease.** Ensure proper mixing of enzyme and cell suspension by pipetting multiple times. DO NOT discard any part of the solution during mixing.
3. **Add 200 µl Buffer AL.**

4. **Vortex for 15 sec.** Use appropriate eye protection and other PPE always.
6. **Incubate at 56°C for 10 min.** Tubes may have bubbles and lids may be displaced from the gas. Ensure they are secure before proceeding.
7. **Briefly centrifuge for 10-15 sec to bring the liquid (if any) from the caps together at the bottom of the tube.**
8. **Add 200 µl of 96-100% Ethanol.** Use absolute ethanol, DO NOT dilute.
9. **Vortex for 15 sec and then briefly centrifuge.**
10. **Transfer the sample mixture (~600 ul) to a QIAamp Mini Spin column.** Make sure to add directly on the filter and not touch the rim of columns or the filter itself.
11. **Centrifuge at 6,000 xg for 1 min.** If lysate is still observed on the column, perform the spin again at the full speed.
12. **Place column in a clean 2 mL collection tube and discard the old collection tube with flow-through.**
13. **Add 500 µl Buffer AW1 to column.**
14. **Centrifuge at 6,000 xg for 1 min and discard flow-through but do not discard the collection tube at this time.**
15. **Add 500 µl Buffer AW2 to column.**
16. **Centrifuge at full speed (20,000 xg) for 3 min.**
17. **Put column in a new 2 ml collection tube and discard the old collection tube with the filtrate.**
18. **Centrifuge at full speed (20,000 xg) for 1 min. Discard the collection tube.**
19. **Put column in a 1.7 ml microcentrifuge tube.**
20. **Put 60 µl Buffer AE directly on the filter.** Switch tips between samples to avoid cross-contamination.
21. **Incubate at room temperature for 5 min.**
22. **Centrifuge at 6,000 xg for 1 min**. **Discard filters.**
23. **Measure concentration of DNA or store it at -20°C**. Nanodrop or Qubit or Bioanalyzer can be used for DNA quantitation. For long-term storage, place the DNA in -80°C. Several freeze-thaw cycles should not generally interfere with most downstream applications, however for long-range PCR/seq and high sensitivity applications store in aliquots and avoid multiple freeze-thaw. Mark with a dot on the side and cap for every freeze-thaw and record in the logbook.

Appendix C: Relative telomere length measurement using qPCR

This protocol is written based on Joglekar et al. 2020 article (PMID: 32260112) and is confirmed by Dr. Mugdha Joglekar.

Reagents

1) Nuclease-free water (Qiagen, catalog #: 129117)
2) TaqMan® 2X Fast SYBR green master mix (Applied Biosystems, catalog #: 4385612)
3) Telomere primers (F: CGGTTTGTTTGGGTTTGGGTTTGGGTTTGGGTTTGGGTT, R: GGCTTGCCTTACCCTTACCCTTACCCTTACCCTTACCCT)
4) hBG primers (F: GCTTCTGACACA ACTGTGTTCACTAGC, R: CACCAACTTCATCCACGTTCACC)

Consumables

1) MicroAmp® Fast Optical 96-Well Reaction Plate, 0.1 mL (with Barcode, catalog #: 4346906 (20 plates) or without Barcode catalog #: 4346907 (10 plates))
2) MicroAmp® Optical Adhesive Film (seal) (Applied Biosystems, catalog #: 4311971 (100 covers))
3) Adhesive Seal Applicator (Applied Biosystems, catalog #: AB1391 or 4333183)
4) MicroAmp™ 96-Well Support Base (Applied Biosystems, catalog #: 4379590)
5) Filter tips for pipettes (2ul, 20ul, 200ul, 1mL)
6) 1.7 mL centrifuge tubes

Equipment

1) Use appropriate PPE
2) NanoDrop / Qubit
3) ViiA™ 7 Real-Time PCR System
4) Refrigerated centrifuge for 96-well plate- set at 4°C

Procedure

1) Prepare DNA samples using commercially available kits such as from Qiagen or other extraction methods using phenol-chloroform. Dilute the DNA sample to obtain 200ng DNA in 24µl (8.33 ng/µl)

Rationale: There are two reactions (one for telomeres, and one for hBG gene (human beta globulin) as a single copy gene normalizer) per sample, and each is measured in triplicate: 6 reactions for each DNA sample.
The volume of DNA solution to be used for PCR must be calculated to provide 25ng DNA to each reaction → aiming for 6 reactions → top up to 8 reactions → 8*25ng = 200ng DNA is needed in 24µl

2) Mix the DNA samples well by vortex, and spin.

qPCR Run

Each run contains 96 reactions using a 96-well plate

Each plate contains 30 samples (triplicate), 1 positive control, and 1 no template control (NTC) sample

One plate for hBG gene, one plate for telomeres.

Reaction setup for **hBG** gene, per each reaction:

Reagent	Volume (µl) per reaction
2X Fast SYBR Green Master Mix	5
3µM Forward primer	1
7µM Reverse primer	1
8.33ng/µl DNA	3
Total	10

Reaction setup for **telomeres**, per each reaction:

Reagent	Volume (µl) per reaction
2X Fast SYBR Green Master Mix	5
1µM Forward primer	1
3µM Reverse primer	1
8.33ng/µl DNA	3
Total	10

1) Prepare the following mix on ice for hBG gene and telomeres separately:

Reagent	Volume for all 96 reactions(µl)
2X Fast SYBR Green Master Mix	515

telomere/hBG Forward primer	103
teloemere/hBG Reverse primer	103

2) Add 7µl of the mix to the wells
3) Add 3µl of DNA sample to each well following this pattern, avoiding any mixing or contamination:

	1	2	3	4	5	6	7	8	9	10	11	12
A	NTC	8	16	24	NTC	8	16	24	NTC	8	16	24
B	1	9	17	25	1	9	17	25	1	9	17	25
C	2	10	18	26	2	10	18	26	2	10	18	26
D	3	11	19	27	3	11	19	27	3	11	19	27
E	4	12	20	28	4	12	20	28	4	12	20	28
F	5	13	21	29	5	13	21	29	5	13	21	29
G	6	14	22	30	6	14	22	30	6	14	22	30
H	7	15	23	+ctrl	7	15	23	+ctrl	7	15	23	+ctrl

4) Seal the plate with adhesive film firmly, place it on a Support Base, and centrifuge at 3000rpm, 4°C, for 5 minutes. Make sure no bubble is in the wells after centrifuge.

5) Run the 96-well fast qPCR plate with the following program:

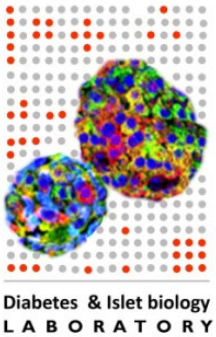
hBG gene	Temp	Time	Number of cycles
Initial Denaturation	95°C	3min	
Denaturation	95°C	15sec	40
Primer Annealing + extension	58°C	1min	
Melt curve initial temp	58°C	1min	

Telomeres	Temp	Time	Number of cycles
Initial Denaturation	95°C	3min	
Denaturation	95°C	15sec	40
Primer Annealing + extension	56°C	1min	
Melt curve initial temp	56°C	1min	

Data analysis is performed as detailed in Joglekar et al, 2020 (PMID: 32260112). Briefly, ΔΔCt values are calculated for each sample and then presented as relative telomere length.

Appendix D: RNA isolation from cells

Title:	**Cell RNA Isolation using Trizol**				
Protocol #:	MB_RNA12	Submitted:	15/06/2015	Approved:	15/06/2015
Category:	MB	Author(s):	WW, MVJ, RJF	Checked by:	MVJ, AAH



Reagents:

1) Nuclease-free 1.7 ml Eppendorf microcentrifuge tubes
2) Trizol (or TRI reagent)
3) Chloroform (molecular biology grade)
4) Isopropyl Alcohol / IPA (molecular biology grade)
5) Ethanol (96-100% pure) (Catalog #: 108543 from Merck or Catalog #: E7023500ml from Sigma Aldrich)
6) Fresh aliquot of nuclease-free water
7) P10, P20, P200 and P1000 filtered tips
8) RNase AWAY (highly preferred but not essential)

All plastic ware should be nuclease free. Never stick your hands into containers of nucleasefree plastic ware. Always pour out what is required. Use appropriate PPE. Know your hazards and question your colleague to confirm they have the knowledge of PPE. Report any shortcomings to your supervisor.

Equipment

1) Appropriate PPE;
2) Refrigerated centrifuge for Eppendorf tubes set at 4°C;
3) Vortex mixer;
4) Well calibrated micropipettes designated for (pre-PCR) nucleic acid isolation.

Reagent Setup

1) Allow samples to thaw thoroughly on ice prior to commencing.
    a. Due to the time involved in each stage of processing, a maximum of EIGHT samples should be processed at a time. DO NOT EXCEED this number for any reason.
2) Set temperature on centrifuge to 4°C
    a. Run “Fast Temp” program if changing rotor or a quick cool down is required.

b. Use a dedicated centrifuge with appropriate rotor

3) Clean gloves with RNase Away prior to commencing.

Procedure

1) **Lyse cells by adding 1 ml of Trizol to each sample in a 1.7 ml Eppendorf tube.** For this protocol use 1 ml of Trizol. If a different volume of Trizol is used (e.g. 0.5mL), the volume of all other reagents must be altered to maintain the correct ratios.
2) **Add 200 µl of chloroform to each sample and shake vigorously for 1 min.** This ensures correct mixing. Do not vortex at this stage.

3) **Immediately centrifuge the tubes at 12,000 x g for 15 mins at 4°C.**

4) **Carefully remove the upper aqueous (clear) layer (400-500 µl) into a fresh 1.7 ml microcentrifuge tube.** Use a 200 µl pipette and tips to remove this gradually. **Note:** There will be a large amount of white precipitate at the interface between the organic and aqueous layers. **Do not disturb this layer.** After the removal of 400-500 µl there will still be some of the aqueous layer left; try to remove as much as possible without contamination.

5) **Add 500 µl of IPA.**

6) **Mix gently by inverting 7-10 times.** Do not vortex or shake tubes.

7) **Incubate for 10 mins at RT.**

8) **Centrifuge at 12,000 x g for 15 mins at 4°C.** A pellet may not always be visible at this stage, so *always* orient the tubes with the hinge facing outwards to allow estimation of the pellet location – in the bottom towards the hinge.

9) **Prepare a fresh volume of 75% ethanol while the samples are spinning.** For 8 samples, add 6.75 ml of 100% ethanol to 2.25 ml of nuclease-free water (9 ml in total)

10) **Carefully aspirate and discard the supernatant by placing the pipette tip along the wall of the tube opposite to the hinge.** When removing the supernatant, reduce the size of the pipette as you go to minimise disturbance to the pellet. Ideally, use a P200 initially, then a P20 when getting closer to the pellet.

11) **Add [1] ml of freshly prepared 75% ethanol to each tube and briefly vortex.**

12) Centrifuge at 12,000 x g for 15 mins at 4°C. As in #8, orient tubes with the hinge facing outwards.

13) Carefully aspirate and discard the supernatant by placing the pipette tip along the wall of the tube opposite to the hinge. As before (#10), reduce the size of the pipette when removing the supernatant. Ideally, use a P200 initially then a P10 to remove as much ethanol as possible.

14) Allow the tubes to dry at RT for 5 minutes. Lay the tubes flat with their lids open. If there are large droplets, gently twist the tube to spread out the liquid, allowing it to dry faster. **Note:** Do not over dry the samples as this will reduce the solubility of the RNA.

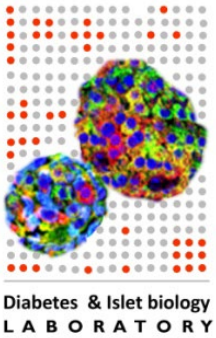
15) Add 15 µl of nuclease-free water to each tube and resuspend. Always store RNA on ice after this stage. The volume may vary depending on the pellet size.

16) Measure concentration of RNA using Nanodrop. If you are not going to proceed with downstream analysis immediately, skip this step and store your RNA. Always measure (Nanodrop, Qubit or Bioanalyzer) before downstream processing.

17) Store RNA at -80°C. Once you store samples at -80°C, log sample details in the freezer log on networked lab drive. IF SAMPLES ARE FOR OVERSEAS SHIPMENT then follow the vendor protocol for plating the samples on to RNA stable plates – DO NOT FREEZE.

Appendix E: cDNA Synthesis

Title:	cDNA synthesis for mRNA qPCR				
Protocol #:	28	Submitted:	03/03/2014	Approved:	03/03/2014
Category:	MB	Author(s):[2]	WW, MVJ, AAH	Checked by:	AAH



Reagents:

1) RT Random Primers (10x)
2) dNTPs with dTTP (100 mM)
3) RT Buffer (10x)
4) Enzyme (recombinant moloney murine leukemia virus (rMoMuLV) reverse transcriptase (50 U/µl)
5) Nuclease-free water (Qiagen, catalog #: 129117)
6) PCR (0.2 ml) tubes (Axygen, catalog #: PCR-02-C)

Note: Reagents 1 to 4 are components of the High-Capacity cDNA Reverse Transcription Kit (from Thermofisher, Catalog #: 4368813 or 4368814)

All plastic ware should be nuclease free. Never stick your hands into containers of nuclease-free plastic ware. Always pour out whatever is required. Use of appropriate PPE is a must.

Equipment

1) Use appropriate PPE
2) Thermocycler
3) Vortex mixer
4) Mini-centrifuge

Reagent Setup

Thaw the reagents on ice: Reverse Transcription (RT) Random Primers (10X), dNTPs with dTTP (100 mM) and RT Buffer (10x). **Do not** keep the enzyme (recombinant moloney murine leukemia virus (rMoMuLV) reverse transcriptase (50 U/µl) on ice. The enzyme is always stored at -20 °C and when using enzyme, make sure to keep it on a -20°C block and immediately store back in -20°C freezer after addition.

Procedure

1) Use this protocol for a total RNA input of 500 ng.

2) Gently vortex all reagents, **except** the enzyme (**N.B**. **Do not** vortex enzyme), and briefly centrifuge them maximum speed for 10 sec (using minifuge).

3) Combine reagents to make RT reagent mix, as described in table below. It is recommended to prepare at least 5% excess volume to compensate for pipetting error.

Components	**Volume per reaction (µl)**
10 X RT Buffer	0.5
25X dNTPs (100 mM)	0.2
10X Random Primer	0.5
Reverse transcriptase (50 U/µl)	0.25
Total	1.45

4) Vortex to mix and then centrifuge briefly at maximum speed for 10 sec (using minifuge).

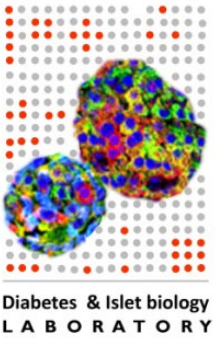
5) Add 500 ng of each RNA sample into a new PCR (0.2 ml) tube, and then add an appropriate volume of nuclease-free water to make a total of 3.55 µl.

6) Add 1.45 µl of the RT reagent mix into each PCR tube.

7) Vortex to mix, and then briefly centrifuge at maximum speed for 10 sec (using minifuge). Place samples into a thermocycler and start the RT program. Cycling conditions: 25 °C for 10 min, 37 °C for 2 h, 70 °C for 10 min and hold at 4 °C. Store cDNA at -15 to -25 °C or use immediately.

Appendix F: Gene expression analysis using TaqMan qPCR

Title:	TaqMan mRNA qPCR on a 96-well platform				
Protocol #:	29	Submitted:	30/09/2014	Approved:	30/09/2014
Category:	MB	Author(s):[3]	WW, MVJ, AAH	Checked by:	AAH



All plastic ware should be nuclease free. Never stick your hands into containers of nuclease-free plastic ware. Always pour out whatever is required. Use of appropriate PPE is a must.

Reagents:

1) TaqMan® Fast Universal PCR Master Mix (2X), No AmpErase® UNG (Applied Biosystems, catalog #: 4366072 (500 reactions))
2) TaqMan® primer/probe assay(s) (Applied Biosystems, catalog #: 4331182 (small))
3) Nuclease-free water (Qiagen, catalog #: 129117)
4) MicroAmp® Fast Optical 96-Well Reaction Plate, 0.1 mL (with Barcode, catalog #: 4346906 (20 plates) or without Barcode catalog #: 4346907 (10 plates))
5) MicroAmp® Optical Adhesive Film (seal) (Applied Biosystems, catalog #: 4311971 (100 covers))
6) Adhesive Seal Applicator (Applied Biosystems, catalog #: AB1391 or 4333183)

Equipment

1) Use appropriate PPE
2) ViiA™ 7 Real-Time PCR System
3) Refrigerated centrifuge for 96-well plate- set at 4°C

Reagent Setup

1) Make sure that the correct block, heated lid and sample carrier is installed in the and ViiA™ 7 Real-Time PCR Instrument (system).
2) Turn on computer, software program ViiA™ 7 Real-time PCR software (N.B. version or name of software may vary) and ViiA™ 7 Real-Time PCR System connected (N.B. Make sure system is connected to computer by selecting and checking from console tab of the software).
3) Set refrigerated centrifuge to 4°C
4) Thaw cDNA samples and selected primer/probe assays on ice.

Procedure

1) Dilute cDNA sample to 33.3 ng/µl if haven't already. For example- if stock cDNA sample was prepared in 500 ng total in a 5 µl reaction, take 3 µl of cDNA stock (i.e. 100 ng/µl) and add 6 µl of nuclease-free water to make a 1 in 3 dilution to obtain 33.3 ng/µl.

2) Combine reagents to make reagent mix for each primer/probe assay, as described in table below. It is recommended to prepare at least 5% excess volume to compensate for pipetting error.

Reagent	Volume (µl) per reaction
2X TaqMan® Fast Universal PCR Master Mix	2.5
20X TaqMan® primer/probe assay	0.25
Nuclease-free water	1.25
Total	4

3) Vortex to mix qPCR reagent mix and then centrifuge briefly at 10,000 x g for 10 sec (quickspin).

4) Add 4 µl of the appropriate qPCR reagent mix into the respective well(s) of 96-well fast qPCR plate. N.B**.** Additions on plate are performed while the plate is on a plate holder (provided along with ViiA™ 7 instrument) on ice.

5) Vortex to mix diluted cDNA sample and then centrifuge briefly at 10,000 x g for 5 sec (quick-spin).

6) Add 1 µl of 33.3 ng cDNA sample into the respective well(s).

7) Seal 96-well fast qPCR plate with adhesive film. N.B. use applicator to seal the adhesive film properly. Do not mark or write on the adhesive film, top and bottom of the in 96-well fast qPCR plate.

8) Centrifuge 96-well fast qPCR plate at 3000 rpm for 5 min, 4 °C. N.B. Make sure there are no visible air bubbles in 96-well fast qPCR plate.

9) Place 96-well fast qPCR plate into Real-time PCR system.

10) Set up experiment with fast 96-well block (0.1 mL), standard curve as experiment type, TaqMan ® Reagents and fast run.

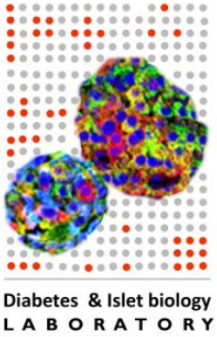
11) Enter the appropriate target (i.e. selected primer/probe assay and assay dye) and sample details and assign them on the 96-well plate layout.

12) Run 96-well fast qPCR plate with the following default run settings- Hold stage: 95 °C for 20 sec (ramp 2.63 °C/sec), PCR stage step: 50 cycles [95 °C for 1 sec (ramp 2.63 °C/sec) and 60 °C for 20 sec (ramp 2.42 °C/sec)]. Reaction volume per well is set at 10 µl.

13) Click “Run” and then “Start Run” with connected real-time PCR system.

Appendix G: Digital droplet PCR

Title:	**Digital droplet PCR for gene target(s) (automated droplet generator)**				
Protocol #:	32	Submitted:	08/04/2016	Approved:	08/04/2016
Category:	MB	Author(s):[4]	RJF,WW,AAH	Checked by:	AAH



All plastic ware should be nuclease free. Never stick your hands into containers of nuclease-free plastic ware. Always pour out whatever is required. Use of appropriate PPE is a must.

Reagents/Consumables:

1) TaqMan®Assay(s), 20X (primer/probe mix) (Applied Biosystems)
2) ddPCR Supermix for Probes, no dUTP (Bio-Rad #186-3024)
3) Semi-Skirted 96-well plate (Eppendorf #951020346)
4) 96-well aluminium foil plate seal (Beckman-Coulter #538619)
5) Automated droplet generator oil (Bio-Rad #186-4110)
6) Automated droplet generator tips (Bio-Rad #186-4121)
7) Automated droplet generator cartridges (Bio-Rad #186-4109)
8) Automated droplet generator cold block (stored at -20˚C)
9) Pierceable foil heat seal (Bio-Rad #1814040)
10) Filtered pipette tips

Equipment

1) Use appropriate PPE
2) Automated droplet generator
3) Heat sealer
4) QX200 Droplet Reader
5) Vortex mixer
6) Well calibrated micropipettes designated for pre- and post-PCR
7) C1000 thermocycler with deep 96-well block

Reagent Setup

1) Allow samples to thaw thoroughly on ice prior to commencing.
2) Thaw supermix for first use. Once thawed, store at 4°C and use within 2 weeks.
3) Turn on the QX200 droplet reader 30 mins before use.
4) Ensure the droplet generator oil is tightly screwed into the auto droplet generator. Prime the lines when a new bottle is installed.

Procedure

1) **Create the ddPCR mastermix.** Below is the volume (µl) of each reagent per reaction. It is recommended to add 5% extra to account for pipetting error.

a.	Supermix	12.5	
b.	Primer/Probe^{+}		1.25
c.	Water$^{\#}$		10
	Total	23.75	

2) **Briefly vortex and centrifuge the mastermix.**

3) **Add 23.75 µl of the mastermix to the respective wells in the Eppendorf 96-well plate.** Ensure that the samples are configured in groups of eight (8), positioned in the columns of the plate. If you do not have samples in multiples of eight, place the appropriate number of NTC controls to fill the remaining wells.

4) **Briefly vortex and spin the samples and then add 1.25 µl $^{\#}$ of the cDNA to the respective well. Mix by pipetting twenty (20) times.**

5) **Seal the droplet plate with the pierceable heat seal (180°C, 5 secs).** Make sure the red stripe on the seal is visible. Be careful when placing the block into the heat sealer; support the base to minimise damage to the rails. Remove the block after sealing. Centrifuge the plate at 1500rpm for 3 minutes.

6) **Allow the plate to sit at room temperature for 10 mins.** It is important to let the samples come to room temperature for optimal droplet generation.

7) **Set up the automatic droplet generator.** Load consumables from the back to the front of the machine to minimise contamination.

    a. Configure the sample plate.

    b. Insert the required droplet generation cartridges and tips (N.B. x2 amount of the tips for each sample is required) within the machine. Completely remove the lid of the tip boxes.

c. Check the tip waste.

d. Place the cool block into the machine.

e. Place empty (collection) Eppendorf 96-well plate on cool block (this is the droplet plate).

8) **Place the sample plate (i.e. the plate containing the mix) on the sample plate position of the machine. Note: <u>Do not remove</u> the seal (foil), this is to prevent chances of cross-contamination. In addition the tips can pierce through the seal.**

9) **Run the droplet generation program.** The cycle will take ~45 mins for a full 96-well plate. Remove the droplet plate within two (2) hours.

10) **Seal the droplet plate with the pierceable heat seal (180°C, 5 secs).** Make sure the red stripe on the seal is visible. Be careful when placing the block into the heat sealer; support the base to minimise damage to the rails. Remove the block after sealing.

N.B. Turn on the heat sealer a few minutes before use to allow heat sealer to heat up to 180°C.

11) **Thermo-cycle the sealed plate using the following cycling protocol.** Ensure that the ramp rate is set to 2°C/sec.

***Use a heated lid set to 105 °C and set the sample volume to 50 µl**

a. 95°C 10:00
b. 40 cycles:
    i. 94°C 00:30
    ii. 60°C 01:00
c. 98°C 10:00
d. 12°C Hold

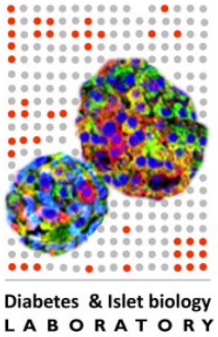
12) Enter the reaction information on the Quantsoft program (which is linked to the droplet reader).

13) Check to make sure there is enough droplet reader oil and also that the waste container is not full.

14) **Place the plate into the droplet reader, insert the sample details and run.**

Appendix H: OpenArray panel

Title:	**OpenArray Human mRNA custom (56) Panel**				
Protocol #:	31	Submitted:	25/01/2016	Approved:	25/01/2016
Category:	MB	Author(s):	WW, MVJ, AAH	Checked by:	MVJ, AAH



All plastic ware should be nuclease free. Never stick your hands into containers of nuclease-free plastic ware. Always pour out whatever is required. Use of appropriate PPE is a must.

Reagents:

1) OpenArray 384-well sample plate (Applied Biosystems, catalog #: 4406947)
2) Aluminium Seal (Beckman-Coulter, catalog #: 538619)
3) OpenArray Custom Slide (Applied Biosystem, catalog #: 4470201. Note- slide is custom-made)
4) OpenArray accessories kit (Applied Biosystems, catalog #: 4453975)
5) TaqMan OpenArray real-time PCR mastermix (Applied Biosystems, 1.5 ml catalog #: 4462159, 5 ml 4462164)
6) AccuFill tips (Applied Biosystems, 1 box, catalog #: 4457246, 10 boxes 4458107)
7) Nuclease-free water (Qiagen, catalog #: 129117)
8) Filtered pipette tips (Axygen)
9) PCR 0.2 ml tubes (Axygen, catalog #: PCR-02-C)

Equipment

1) Use appropriate PPE
2) QuantStudio™ 12K Flex Accufill System
3) QuantStudio™ 12K Flex Real-Time PCR System
4) Vortex mixer
5) Centrifuge

Reagent Setup

1) Download the relevant plate file (.tpf) from the website- https://www.thermofisher.com/au/en/home/products-and-services/product-types/download-openarray-tpf-and-spf-plate-files.html using the custom design *made OpenArray slide sales order, lot. and serial number. This contains the run information for the specific nanofluidics array slide.

2) Take the Openarray slide (before use) from the freezer and allow it to come to room temperature (~15 min).

Procedure

1) Thaw cDNA samples (100 ng/µl) and 2X TaqMan® OpenArray® Real-Time PCR Master Mix (if using for the first time) on ice. Mix qPCR reagent mix by swirling the bottle.

2) Mix the samples by vortexing, then centrifuge for 10 sec @ 10 000 rpm (quick spin).

3) Prepare the PCR mix-

   a) Mix the 2X TaqMan® OpenArray® Real-Time PCR Master Mix by gently inverting or swirling the tube/bottle a few times.

   b) Combine the following components:

Component	Volume (µl) for 1 area of the 384-well sample plate[1]	Stock concentration	Final concentration	Units
2X TaqMan® OpenArray® Real-Time PCR Master Mix	132.0	2	1	X
Nuclease-free water	68.6	-	-	-
Final volume of PCR mix	200.6	-	-	-

1 One area of a 384-well sample plate corresponds to a single TaqMan® OpenArray® slide.

   c) Mix diluted qPCR mix (1X) well by pipetting up and down.

4) Combine prepared qPCR (1X) with cDNA. In a new PCR (0.5 ml) tube-

   a) Load 3.8 µl of prepared PCR mix (1X) (see above) into PCR tube.

b) Add 1.2 µl of cDNA* (100 ng/µl) into allocated (labelled) PCR tube.

* Since 1.2 µl of 100 ng/µl cDNA is added the total amount of cDNA input (per this 5 µl reaction) is 120 ng.

5) Vortex and mix well the 5 µl reaction in PCR tube and load reaction sample into the allocated wells of the 384-well sample plate.

I. Each sample plate can contain up to eight slides worth of samples. However, the system for OpenArray can only process 4 OpenArrays in a single run. If more than four slides worth of samples are to be loaded on one sample plate, please ensure the remaining sections are sealed. Seal with OpenArray sample plate sealer.

NOTE: It is advisable to pre-cut the sealer into the required sections, so the sections may be sealed/unsealed individually to reduce evaporation. Alternatively, the plate may be sealed with an intact sealer, and then sections can be individually cut out when loading.

6) Seal the sample plate with aluminium seal

7) Centrifuge sample plate for 1 minute at 2000 rpm to eliminate bubbles. Sample plate can only be stored on ice for 1 hour. However it is recommended to process sample plate immediately.

Loading OpenArrays and performing qPCR

8) Download (if haven't already) the relevant plate file (.tpf) from the website- https://www.thermofisher.com/au/en/home/products-and-services/product-types/download-openarray-tpf-and-spf-plate-files.html using the custom nanofluidics array slide sales order, lot. and serial number. This contains the run information for the specific nanofluidics array slide.

9) Remove (if haven't already) the nanofluidics array slide from the freezer and allow it to come to room temperature (~15 min).

10) Ensure that the correct block, heated lid and sample carrier is installed in the nanofluidics array system. Turn on the computer, and real-time PCR system and the loading system. Access the respective software and ensure that the machines are

connected. Remove the loading system consumables (Array slide lid, plug and immersion fluid) from packaging.

11) Gently pull on the plunger of the immersion fluid syringe to loosen. Remove cap, place tip on and flush air from the tip. Place the loading system tips within the machine and remove lid. Place sample plate within PCR system.

12) Put gloves on. Ensure they are tightly fitting to minimize the risk of accidentally marking the slide lid. Carefully open slide packaging. Slowly tip slide into hand. Do not touch the top of the slide.

13) Place slide into the PCR system, with the barcode on the left. Remove sealer from the portion of the sample plate intended for loading. Use the loading system software to enter the slide barcode, slide position, sample position and tip configuration.

14) When all relevant checks are completed, press load slide. While the PCR system is loading the slide, remove the clear and red plastic from the bottom of the slide lid. When finished loading, carefully remove and seal the slide within 90 sec.

    I. Place the slide within the plate clamp. Place the slide lid onto the slide. Clamp for 30 sec. Ensure the lid is positioned so that barcode is correctly displayed. Remove the assembly from the plate clamp.

    II. Position immersion fluid syringe within the slide so that the tip is pressing against the lid. Slowly fill slide with immersion fluid, ensuring the fluid runs along the lid. Once full, seal the slide with the plug, turning the screw until the handle breaks off.

    III. Remove the plastic cover on the top of the slide lid, and then carefully place into the slide carrier of the real-time PCR system. Ensure there is support on the bottom of the slide as it is being lowered, so it does not drop suddenly, and do not touch the top of the slide. It is OK to touch the sides of the slide/cassette. Initialize the PCR system and start the program for qPCR within 1 hr.

15) Select “OpenArray” within the PCR-system software. Press “Find Slide IDs”. This will take a few mins. If the software cannot find the plate ID, it will ask for it to be entered manually.

16) Press “Confirm Plate Centres”. Again, this will take a few mins. Check that the red dot is within the centre and that there are no fingerprints/marks on the top of the slide. Load the respective .tpf file for each slide and specify a result file name and location. Press “Start Run”. The program will take approximately 2 hr to complete.