University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Microbiome Ireland
Interfacing Food & Medicine

Single-Cell Transcriptomics of Intestinal Epithelial Cells: Insights into the Prediabetic condition

Glenn Ross-Dolan
B.A. (Mod) Microbiology - Trinity College Dublin

Project Thesis in partial fulfilment for the degree of Masters in Bioinformatics and Computational Biology

Supervised by Dr. Silvia Melgar, APC Microbiome Ireland

Table of Contents

List of abbreviations

Acknowledgements

Abstract

Single-Cell Transcriptomics of Intestinal Epithelial Cells: Insights into the Prediabetic condition
Supervisor: Silvia Melgar, APC Microbiome Ireland, S.Melgar@ucc.ie

Background: Obesity is a chronic condition characterised by excess adiposity that may be accompanied by different structural and functional abnormalities and with increased comorbidity and premature mortality risk thus reducing quality of life (Jastreboff et al., 2019; Lustig et al., 2022).

The main cause of obesity is the energetic imbalance due to increased caloric intake and little expenditure. This induces metabolic and hormonal changes e.g. increase in blood sugar levels that induces a prediabetic status with increased risk of developing type 2 diabetes mellitus (T2DM), heart disease, and stroke.

T2DM is a major non-communicable disease and one of the world's fastest growing health problems, with a projected increase in the number of diabetic patients to 700 million by 2045 (Saeedi et al., 2019). T2DM is associated with significant morbidity, including increased risk of cardiovascular diseases (CVD) and stroke, hypertension, etc. These place an enormous burden on individuals, society, and the healthcare system (Brorsson & Pociot, 2015). T2DM is a non-reversible but preventable condition with overweight and obesity being major risk factors. The onset of T2DM is gradual, with most individuals progressing from normoglycemia through a pre-diabetic state.

There is substantial evidence for the role of gut microbiota and impaired barrier in metabolic diseases including T2DM (Brunkwall & Orho-Melander, 2017). Recent clinical trials using Glucagon-like

peptide-1 (GLP-1) receptor agonists (an incretin hormone produced by the enteroendocrine L cells in the distal intestine) have shown benefits to patients with these conditions. However, a role for other intestinal epithelial cell subsets in obesity and diabetes is yet to be determined. Thus, the aim of this project is to:

1) Identify alterations in intestinal epithelial cell subsets
2) Identify alterations in pathways contributing to obesity and/or diabetes pathophysiology.

For this purpose, the student will be performing a bioinformatics analysis on in house and publicly available bulk and single cells RNASeq data sets. Findings from this project, will provide a better understanding on how gut epithelial cells, which are in close contact to the microbiota, might support and/or sustain the development of these chronic conditions.

Aim

1) Identify alterations in intestinal epithelial cell subsets
2) Identify alterations in pathways contributing obesity and/or diabetes pathophysiology.

Introduction

1. Type II Diabetes Mellitus

Type II Diabetes Mellitus (T2DM), is a global health concern growing to pandemic proportions. The number of people with T2DM has increased four fold over the last three decades and is the ninth leading cause of death and approximately x in 100 people above the age of 18 have T2DM in the United States of America. With the rapid developments in technology and urbanisation the lifestyle of the average person across the globe has changed to one with increased obesity rates, sedentary lifestyles and increased ageing, the driving forces behind the disease. Particularly over the last two decades, obesity has become a pressing issue in every country across the globe leaving no exceptions and leading to a vast array of health complications. For these reasons, there is an urgent demand to take actions in slowing down the rate of obesity together with the complications associated with obesity, one of which being T2DM. T2DM is currently defined as a chronic metabolic disorder characterised by insulin resistance and an inadequate insulin secretory response ultimately leading to hyperglycemia. This definition follows the scheme set by the American Diabetes Association which classifies diabetes into three categories: type 1, type 2 and other specific forms such as genetic defects or secondary to other conditions. Characterised by insulin resistance and impaired insulin secretion, T2DM aetiology involves a intricate mix of genetic, environmental and lifestyle factors.

Risk Factors

Overweight and obesity in adults is defined as a body mass index greater than 25 kg/m2 and x kg/m2 respectively. Although high BMI is not a direct measure of health but rather size, a large portion of deaths from non communicable diseases are driven by high BMI.

Approximately 80% of which are from diabetes, stroke, heart disease and cancer. Three quarters of this death and disease in adults is occurring in middle-income countries. In essence, the majority of people living and dying with non-communicable diseases have high BMI and are living in developing/lower resource regions of the globe. Current data from the world obesity atlas reports that data are showing that there is a positive correlation of increased GDP and increased levels of overweight/obesity. There also appears to be positive correlations between high BMI and GHG emissions, increased urbanisation, increased plastic waste, sedentary lifestyle, and consumption of animal proteins, sugars and sweeteners. Gradual weight gain serves as the trigger for later metabolic issues, with Type 2 diabetes mellitus (T2DM) being the most closely linked to obesity. In some cases however, T2DM may develop before obesity in those with innate insulin resistance, where increased glucose production and insulin levels ultimately trigger obesity.

Diet seems to be intimately related to the progression of various diseases, particularly metabolic disorders and with this, the search for high quality diets has been steadily increasing over the last number of decades. One group led by Toi et al. performed an umbrella review to unveil evidence in diet interventions and factors for preventing the progression of T2DM. In this study, sixty systematic reviews with meta analyses of randomised controlled trial/observational studies were analysed. The effect of each dietary intervention was analysed in its effect on the risk of T2DM. Results of the review indicate that dietary patterns such as Mediterranean (diets high in fruit, vegetables, legumes, beans, cereals, grains, fish, foods with high unsaturated fats like olive oil) and Dietary Approaches to Stop Hypertension diets, and diets with a high healthy eating index were beneficial in preventing T2DM. Specific food groups which exhibited beneficial effects in preventing T2DM included whole grain, olive oil, low fat dairy, fibre, magnesium and flavonoids. Contrastingly, diets exhibiting high glycemic index/glycemic load as well as red

meats, processed meats, sugar, and artificial sweeteners indicate an accelerating effect of progressing T2DM. Physical activity. Frequent physical activity plays a crucial role in managing type 2 diabetes mellitus. Numerous studies have shown that aerobic exercise, resistance training as well as a combination of the two can lead to significant improvements of the various systems in the bodies of people with T2DM. Often reported are the improvements in blood pressure, glucose regulation, insulin sensitivity, HbA1c levels as well as reduction in body fat/ visceral adipose tissue. Moreover, the extent of these improvements are proportional to the total energy consumed during the training rather than the intensity or the length of the training exclusively.

Smoking of tobacco has also been identified as a contributing factor in the progression of type 2 diabetes mellitus. Epidemiological research reveals distinct links between tobacco smoking and risk of developing T2DM. Similarly, medical research suggests that tobacco smoking and nicotine use directly affects insulin sensitivity, body composition and pancreatic beta cell mechanisms. Rimm et al. revealed that men smoking more than 25 cigarettes per day had nearly double the risk of developing T2DM compared to non-smokers. Another study investigated the impact of both active and passive smoking on diabetes risk using data from the HIPOP-OHP Study in Japan. Over a median follow-up of 3.4 years, the study found that individuals exposed to passive smoke at work and active smokers had significantly higher risks of developing diabetes compared to non-exposed individuals. Additionally, cross-sectional studies including EPIC-Norfolk and the National Health and Nutrition Examination Survey, found that smoking is linked to higher HbA1c levels indicating poorer glycemic control.
Although these epidemiological studies have shown that the incidence of T2DM is correlated with the incidence of smoking, this alone does not demonstrate that smoking drives the progression of

the disease. Numerous studies have also investigated physical mechanisms affirming these links. Although studies show that smoking results in a reduced body weight, it also has a negative effect on body composition. People that smoked more than 20 cigarettes per day exhibited an adjusted odds ratio of 1.93 for abdominal obesity compared to those who have never smoked. Similarly, other cross-sectional studies of men and women in the UK as well as Japanese men exhibit a greater waist-to-hip ratio compared to non-smokers. These changes in body composition appear to be caused primarily by nicotine signalling suggesting that non-tobacco alternatives may not mitigate these risk factors.

Smoking also exhibits negative effects on glucose tolerance and insulin sensitivity. A number of studies demonstrate that the glucose tolerance and insulin sensitivity index was reduced after smoking in a group of persistent smokers as well as total body glucose being reduced in smokers with diabetes versus non-smokers with diabetes. Furthermore other studies demonstrate that clamp analysis shows that the total glucose disposal in the body was lowered in smokers compared to those who do not. It is also apparent that use of nicotine, such as use through nicotine patches also decreases the efficacy of insulin.

Finally, clinical studies note that smoking disrupts B cell function. In studies involving Japanese men as well as studies on Swedish men, higher incidence of impaired insulin secretion as well as lower B cell function as measured by HOMA-B is exhibited compared to non-smokers. These effects are observed even after adjusting for other factors such as age, BMI, alcohol intake and physical activity.

These links may stretch past just the use of tobacco smoking into the realm of smokeless tobacco use, specifically in the use of tobacco pouches called ‘snus’. A number of studies are reporting an increased incidence of T2DM in people that use more than 5 boxes of snus per week at an odds ratio of 3.3. There has been mixed results however with other studies reporting no apparent correlation although, a recent meta analysis reported a hazard ratio of 1.15 in

snus users compared to those who have never used snus. This is a concerning figure given the lack of knowledge on tobacco free nicotine products, as well as the immense growth that the snus industry has seen in recent years with sales increasing from 126 million units in 2019 to 808 million units in march 2022 in the USA.

Alcohol intake has a complex relationship with risk of T2DM. Moderate consumption of alcohol is associated with a lower risk of developing T2DM compared to abstainers suggesting a potential protective effect although excessive alcohol consumption is linked to increased risk due to its contribution to obesity and insulin resistance. Several studies point to this ‘U-shaped’ relationship between alcohol consumption and T2DM risk. Reductions in risk in alcohol consumers seems to be somewhat confined to women and non-Asian populations.

As discussed previously, most risks associated with T2DM can be mitigated through lifestyle and dietary changes. There still however remain some risks which are not entirely in the individual's control, genomics. Genomics plays a significant role in T2DM susceptibility with genetic factors influencing a persons predisposition to overweight/obesity, insulin resistance and overall metabolic profile. The genetic susceptibility of Type 2 Diabetes mellitus is well documented and studied through twin and family based studies. Typically this involves mutations in genes regulating glucose levels and glucose-homeostasis hormones. However, most cases of T2DM involving genetic factors are driven by both the genetic and environmental factors together, otherwise known as epigenetics.

Microbiome is also linked to T2DM as a risk factor.

Diagnoses

Management of T2DM

Complications of T2DM

2. Prediabetes and Metabolic Syndrome

Defining Prediabetes

Prediabetes is a condition similar to diabetes where the metrics used to diagnose diabetes such as glucose homeostasis and blood glucose levels are elevated past normal levels but not elevated enough to classify someone as being diabetic. More specifically, prediabetes is characterised by impaired fasting glucose, impaired glucose tolerance, raised HbA1c levels or a combination of the three. The exact specifications of these three metrics are not agreed upon in the scientific literature leading to some discrepancies in prevalence estimates. The state of being prediabetic is associated with obesity, dyslipidemia, and hypertension and is considered a risk factor for developing T2DM as well as some other pathologies such as cardiovascular disease. Both prediabetes and T2DM share common risk factors including overweight, sedentary lifestyle, smoking, and genetics and both are associated with increased cardiovascular risks. However the cardiovascular risk is significantly higher in those with T2DM compared to those with prediabetes.

Prediabetes Pathologies

The state of being prediabetic is described as being physiologically stressful, having pathological changes linked to several systems in the body most notably the circulatory, nervous, digestive, urinary and endocrine systems.

The most common pathologies associated with prediabetes include microvascular complications, neurological disorders, cardiovascular disorders and metabolic disorders. Microvascular disorders such as nephropathy, retinopathy and neuropathy are prevalent in individuals with prediabetes. Research is indicating that prediabetes particularly with impaired glucose tolerance is independently linked to peripheral neuropathy and nerve dysfunction likely due to an increased oxidative stress and activation of neurotoxic pathways (Lee et al., 2015).

3. The Intestinal Epithelium in Metabolic Health and Disease

4. Key Signalling Pathways in Intestinal Epithelial Cells and Prediabetes

The proteasome

The proteasome is a complex of proteins involved in the homeostasis of redundant or damaged proteins. The mechanism of degradation employed is by process of proteolysis, breaking peptide-peptide bonds of the amino acid polymers with proteases. Proteins for degradation are tagged with a molecule called ubiquitin by enzymes called ubiquitin ligases. The ubiquitin-proteasome system plays a central role in the homeostasis of various cellular systems such as programmed cell death, cellular signalling and cell cycle. As a result, changes in this tightly-run system often lead to various disease pathologies. Three classes of enzymes exist which promote the ubiquitination process of target proteins: ubiquitin-activating enzymes (E1), ubiquitin-conjugating enzymes (E2) and ubiquitin-ligating enzymes (E3). The process begins with E1 binding to a molecule of ubiquitin via a thioester linkage in an ATP-dependent manner. (Here. E1 acts as the initial transporter (or pickup courier, for example from

the amazon warehouse) of the free floating ubiquitin molecules). The E1 molecule reaches an E2 molecule (you can think of this as the postal service) and transfers the ubiquitin molecule. The E2 molecule either transfers the ubiquitin molecule to an E3 molecule (final courier of the ubiquitin which has the final address) or else the E2 molecule binds the E3 molecule directly. The E3 molecule is known to have specificity for various substrates as well as specificity for the proteasome complex. Once a molecule has been tagged with 4 ubiquitin molecules, it is recognised and degraded by the 26S proteasome complex.

The structure of the proteasome complex. The proteasome is a barrel-shaped complex containing a 20S core particle, and a 19S regulatory particle. Two 19S particles bind bilaterally (sandwichly) to the 20S particle forming a 30S complex. The 20S inner unit is responsible for the ATP dependent catalytic activity and is made up of two external alpha rings and two internal beta rings. Both alpha rings and beta rings each containin seven distinct subunits (alpha 1-7 and beta 1-7). Another variation of this is the 11S-20S-11S or 19S-20S-11S hybrid. These variations allow for ubiquitin-independant degradation. Some of the 11S sandwich variations replace the 1,2, and 5 beta subunits with beta 1i, 2i and 5i counterparts encoded by (LMP2, encoded by PSMB9, MECL-1, encoded by PSMB10 and LMP7, encoded by PSMB8 respectively). This variation is triggered by interferon gamma and is known as the immunoproteasome.

A number of studies report the role of the ubiquitin proteasome system in the progression of type-2 diabetes related pathologies including cardiovascular diseases. Poor protein quality homeostasis can result in the accumulation of protein aggregates and increased levels of pro-hypertrophic and pro-apoptotic factors promoting disease pathologies. The UPS is also involved impacts cardiac, beta-adrenergic signalling, cell excitability, conductance and insulin gene transcription regulation. It is also responsible for the degradation of

the insulin receptor and the insulin receptor substrate contributing to insulin resistance.

Ferroptosis

Tight Junction

Oxidative phosphorylation

PI3/Akt

5. Multi-omics Approaches in Prediabetes & T2DM Research

6. Mouse Models in Prediabetes Research

7. Current Pharmacological Interventions for Prediabetes and T2DM

8. Emerging Therapies: Focus on Brown Seaweed Extract

Materials and Methods

Mouse models

The dataset used throughout this study consists of scRNA-seq (10X) data from the small intestinal crypts of control diet and HFHSD obese FVF-enriched mice (Foxa2-FVF/FVF) as well as villus samples from the small intestine of control diet and HFHSD C57BL/6N mice.

All mice subjects in the dataset are male. Mice living conditions consisted of 2 to 4 mice per group, 23C 45-65% humidity with a 12 hr light/dark cycle.

Mice diets are reported to have begun at 10-12 weeks old and were randomised into different test groups matched for body weight with similar variance and given ad libitum access to either the control diet or HFHS diet for 11-13 weeks.

Glucose tolerance and insulin secretion tests.

To assess glucose tolerance, an oral glucose tolerance test (oGTT) was performed on FVF mice after 12 weeks on either a control diet (CD) or a high-fat high-sugar diet (HFHSD). Following a 6-hour fast, each mouse received an oral dose of glucose (1.5 mg/g body weight of 20% (wt/v) d-(+)-glucose solution in PBS). Blood glucose levels were measured at 0, 15, 30, 60, and 120 minutes post-glucose administration using a handheld glucometer (Abbott). For insulin secretion analysis, blood samples were collected from the tail vein at 0, 15, and 30 minutes during the oGTT. Plasma was isolated by centrifuging the blood samples (3,500 rpm, 15 minutes, 4°C), and insulin levels were quantified using the Ultra-Sensitive Mouse Insulin ELISA Kit (Crystal Chem, 90080) following the manufacturer's protocol.

Insulin resistance and beta-cell function were evaluated using the homeostasis model assessment of insulin resistance (HOMA-IR) and HOMA-β, respectively. These indices were calculated 12 weeks after diet initiation, using fasting blood glucose and plasma insulin levels obtained after a 6-hour fast. The conventional formulas applied were HOMA-IR = fasting blood glucose (mg/dL) × fasting insulin (μU/mL) / 405 and HOMA-β = fasting insulin (μU/mL) × 360 / (fasting glucose (mg/dL) - 63).

Crypt and Villus Isolation and Single-Cell Preparation

The isolation of small intestinal crypts was performed following established protocols. Briefly, small intestines (SIs) were excised and rinsed with cold PBS. The villi were carefully scraped away using a glass slide. The remaining tissue was then cut into 2-cm sections, repeatedly washed with cold PBS, and incubated in 2 mM EDTA/PBS for 35 minutes at 4°C with gentle agitation. Crypts were released by vigorous shaking and filtered through a 70-μm mesh to remove any villous debris. For the preparation of single cells, the isolated crypts were treated with TrypLE (Life Technologies, no. 12605) first on ice for 5 minutes, followed by 5 minutes at 37°C, and subsequently incubated with 10 μg/mL DNase in crypt complete medium (DMEM/F-12 with 10% FCS) for 5 minutes at 37°C. The resulting single-cell suspension was achieved by gentle, repeated pipetting. Cells were then washed twice with 2% FCS in PBS and pelleted by centrifugation at 300g for 5 minutes at 4°C. For flow cytometry analysis, cells were resuspended in 1-2 mL of FACS buffer (2% FCS, 2 mM EDTA in PBS; Sigma-Aldrich, no. Y0503) and passed through 40-μm cell strainer caps attached to FACS tubes.

For the isolation of villus cells, the villi were scraped and processed into a single-cell suspension using the same TrypLE treatment protocol as described for crypt cells.

Flow Cytometry

For gene expression analyses, including microarray, single-cell transcriptomics, and western blotting, small intestinal crypt cells were sorted using a FACS-Aria III (BD Bioscience) with FACSDiva software v.6.1.3 and a 100-µm nozzle. In all experiments, cells were gated based on their forward scatter area (FSC-A) and side scatter area (SSC-A). Singlets were identified using forward scatter width (FSC-W) and forward scatter height (FSC-H), and dead cells were excluded using 7-AAD (eBioscience, no. 00-6993-50). For quantitative PCR with reverse transcription (qRT-PCR), cells were directly sorted into Qiazol lysis reagent (QIAGEN, no. 79306). To enrich FVF-positive small intestinal crypt cells for scRNA-seq, 30,000 FVF+ (FVFlow and FVFhigh) cells were sorted along with 30,000 live crypt cells per sample. The sorted cells were collected in modified FACS buffer (2% FCS, 0.02 mM EDTA in PBS).

scRNA-seq

Crypt and villus samples were prepared as described above. Dead cells were excluded via flow cytometry after 7AAD labelling. Dead cell exclusion was controlled by trypan blue staining and sorted cells were counted. Single-cell libraries were generated using the Chromium Single cell 3′ library and gel bead kit v2 (10X Genomics, no. 120237) according to the manufacturer’s instructions. Libraries were sequenced on a HiSeq4000 (Illumina) with 150-bp paired-end sequencing of read 2.

Upstream scRNA-seq pipeline

Quality control was performed using the Scanpy package to ensure the integrity and accuracy of the single-cell RNA sequencing (scRNA-seq) data. The quality control steps focused on identifying and mitigating potential confounding factors, such as the presence of mitochondrial and ribosomal, which can indicate cellular stress or contamination. Identification of Mitochondrial and Ribosomal genes: Mitochondrial genes were identified using the prefix "mt-" for mouse genes. Ribosomal genes were identified using the prefixes "Rps" and "Rpl". The `calculate_qc_metrics()` function from Scanpy was used to compute common QC metrics, including the percentage of counts attributed to mitochondrial and ribosomal. The metrics were calculated and stored in the AnnData object for subsequent analysis.

Cells were removed with greater than 10% mitochondrial gene content, less than 200 genes by counts as well as cells with less than 500 total counts.

To identify potential doublets, the scrublet tool was used utilising a nearest neighbour classifier of observed transcriptomes and simulated doublets. Predicted doublets were subsequently removed from the dataset.

Counts were normalised per cell by total counts over all genes so that every cell has the same total count after normalisation. Each cell is normalised to a total count equal to the median of total counts for cells before normalisation. The data are then logarithmised using log1p where each count is transformed to the natural log of 1 plus the original count value. This method allows for the data to be transformed as well as zero values.

PCA Nearest neighbour graph and UMAP visualisation.

Filtering lymphocytes and ambient genes.

Mapping published annotations to the data

Downstream scRNA-seq Pipeline

KEGG pathway analysis

GO Term Analysis

Results

Discussion

References

Lee CC, Perkins BA, Kayaniyil S, Harris SB, Retnakaran R, Gerstein HC, Zinman B, Hanley AJ (2015) Peripheral Neuropathy and Nerve Dysfunction in Individuals at High Risk for Type 2 Diabetes: The PROMISE Cohort. Diabetes Care 38:793–800.

Appendix