Defining the molecular, genetic and transcriptomic mechanisms underlying the variation in glycation gap between individuals

Fakhra Naseem. FHEA.

A thesis submitted in partial fulfilment of the requirements of the University of Wolverhampton for the degree of Doctor of Philosophy.

This work or any part thereof has not previously been presented in any form to the University or to any other body whether for the purposes of assessment, publication or for any other purpose (unless otherwise indicated). Save for any express acknowledgements, references and/or bibliographies cited in the work, I confirm that the intellectual content of the work is the result of my own efforts and of no other person.

Signature: Fakhra Naseem

Date: 23/10/2019

Abstract

The discrepancy between HbA_{1c} and fructosamine estimations in the assessment of glycaemia has frequently been observed and is referred to as the glycation gap (G-gap). This could be explained by the higher activity of the fructosamine-3-kinase (FN3K) deglycating enzyme in the negative G-gap group (patients with lower than predicted HbA_{1c} for their mean glycaemia) as compared to the positive G-gap group. This G-gap is linked with differences in complications in patients with diabetes and this potentially happens because of dissimilarities in deglycation. The difference in deglycation rate in turn leads to altered production of advanced glycation end products (AGEs). These AGEs are both receptor dependent and receptor independent. It was hypothesised that variations in the level of the deglycating enzyme fructosamine-3-kinase (FN3K) might be as a result of known Single Nucleotide Polymorphisms (SNPs): rs1056534, rs3848403 and rs1046896 in *FN3K* gene, SNP in *ferroportin1/SLC40A1* gene (rs11568350 linked with FN3K activity), differentially expressed genes (DEGs), differentially expressed transcripts or alternatively spliced transcript variants. Previous studies reported accelerated telomere length shortening in patients with diabetes.

In this study, 184 patients with diabetes were included as dichotomised groups with either a strongly negative or positive G-gap. This study was conducted to analyse the differences in genotype frequency of specific SNPs via real time qPCR,determine soluble receptors for AGE (sRAGE) concentration via ELISA, finding association of sRAGE concentration with SNPs genotype, and evaluate relative average telomere length ratio via real time qPCR. This study also aimed at the investigation of underlying mechanisms of G-gap via transcriptome study for the identification of the DEGs and differentially expressed transcripts and to

consequently identify pathways, biological processes and diseases linked to situations in which DEGs were enriched. The relative length of the telomere was normalised to the expression of a single copy gene (S). Chi-squared test was used for estimating the expected genotype frequencies in diabetic patients with negative and positive G-gap.

Genotype frequencies of FN3K SNPs (rs1056534, rs3848403 and rs1046896) and SLC40A1/ferroportin1 SNP (rs11568350) polymorphisms within the studied groups were non-significant. With respect to genotypes, the rs1046896 genotype (CT) and rs11568350 genotype (AC) were only found in heterozygous state in all the investigated cohorts. No association between sRAGE concentration and FN3K SNPs (rs3848403 and rs1056534) was observed as the sRAGE concentration was also found not to be different between the groups. Similarly, the relative average telomere length was not different in both groups. Plasma sRAGE levels were not different in the cohort studied even though the Wolverhampton Diabetes Research Group (WDRG) previously reported that AGE is higher in positive G-gap. The latter is a more likely consequence of lower FN3K activities. In this study, it was found that SNPs in the *FN3K/ferroportin1* gene are not responsible for the discrepancy in average glycaemia. The transcriptomic study via RNA-Seq mapped a total of 64451 gene transcripts to the human transcriptome. The DEGs and differentially expressed transcripts were 103 and 342 respectively ($p < 0.05$, fold change > 1.5). Of 103 DEGs, 61 were downregulated in G-gap positive and 42 were upregulated in positive G-gap individuals while 14 genes produced alternatively spliced transcript variants. Four pathways (Viral carcinogenesis, Ribosome, Phagosome and Dorso-ventral axis) were identified in the bioinformatics analysis of samples in which DEGs were enriched. These DEGs were also found to be associated with raised blood

pressure and glycated haemoglobin (conditions that coexist with diabetes). Future analysis based on these results will be necessary to elucidate the significant drivers of gene expression leading to the G-gap in these patients.

Acknowledgement and Dedications

I want to acknowledge my Director of Studies, Dr Simon Dunmore and my supervisor, Professor Paul Kirkham. Without their support, guidance and help it would not have been possible for me to continue my research work. They continuously provided valuable suggestions and showed keen interest that was a source of encouragement to me throughout this study.

I also wish to thank Professor Baldev Singh (Diabetes Centre, Newcross Hospital, Wolverhampton, UK), Dr Adam Watkins (Nottingham University, UK) and Dr James Brown (Aston University, UK) who provided the facilities and support to complete my research.

I would like to express my thanks to Dr Riffat Naseem, Dr Niraj Shah, Dr Amr Alderawi and all the staff and colleagues for all their help.

Words are confined and inefficacious to express my immense indebtedness for my gracious parents Muhammad Sawal (late) and Hanifan Bibi, my beloved husband Muhammad Tahir Mahmood, my son Haris Mahmood, my brother-in-law Nasir Mahmood and his wife Shazia Mahmood for their moral guidance, encouragement, kindness, and dedication throughout my student life.

Table of Contents

List of figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of tables

Chapter 2

Chapter 3

Chapter 5

Abbreviations

3DG	3-deoxyglucosone
ADA	American Diabetes Association
AGEs	Advanced glycation end products
CAGE	Cap analysis of gene expression
cDNA	Complementary DNA
CML	Carboxymethyl-lysine
C-peptide	Connecting-peptide
Ct	Cycle Threshold
CVD	Cardiovascular disease
DCCT	Diabetes Control and Complications Trial
DEGs	Differentially Expressed Genes
DMF	Deoxymorpholinofructose
dNTP	Deoxyribonucleotide triphosphate
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
ENCODE	Encyclopaedia of DNA elements
esRAGE	Endogenously secreted isoform
F3P	Fructose 3-phosphate
FL	Fructoslysine
FL3P	Fructoselysine-3-phosphate
FM Index	Ferragina-Manzini index
FN3K	Fructosamine-3-kinase
FN3K-RP	FN3K-related protein
FPG	Fasting Plasma Glucose

FPKM	Fragments per kilobase of transcripts per million mapped reads
GAD65	Glutamic acid decarboxylase
G-gap	Glycation Gap
GHb	Glycohaemoglobins
GTF	Gene Transfer Format
GLO1	Glyoxalase 1
GWAS	Genome-wide association studies
HbA_{1c}	Haemoglobin A_{1C}
HGI	Haemoglobin Glycation Index
HISAT	Hierarchical indexing for spliced alignment of transcripts
HPLC	High-performance liquid chromatography
IA-2	Islet antigen-2
IFCC	International Federation of Clinical Chemistry
IFG	Impaired fasting glucose
IGT	Impaired glucose tolerance
KEGG	Kyoto encyclopedia of genes and genomes
lncRNA	long non-coding RNA
MBG	Mean blood glucose
MMSEQ	Minimum mean square error quantizer
MCP-1	Monocyte chemoattractant peptide 1
MISO	Mixture-of-Isoforms
mRNA	Messenger RNA
NADPH	Nicotinamide adenine dinucleotide phosphate
ncRNA	Non-coding RNA
NFκB	Nuclear factor kappa B
NGS	Next Generation Sequencing

NO	Nitric oxide
PCR	Polymerase chain reaction
Pi	Inorganic phosphate
Poly-A	Polyadenylated
PAI-1	Plasminogen activator inhibitor 1
PTDM	Post transplantation diabetes mellitus
RAGE	Receptor of advanced glycation end products
RNA	Ribonucleic acid
RNA-Seq	RNA-sequencing
ROS	Reactive oxygen species
RP-HPLC	Reverse-phase high-performance liquid chromatography
RPKM	Reads per kilobase of transcripts per million mapped reads
rRNA	Ribosomal RNA
S	Single Copy Gene
SAGE	Serial analysis of gene expression
SAM	Sequence Alignment Map
SNPs	Single nucleotide polymorphisms
SPSS	Statistical Package for Social Sciences
sRAGE	Soluble receptor for AGEs
T/S	Relative average telomere length
T1D	Type 1 diabetes
T1DM	Type 1 Diabetes Mellitus
T2D	Type 2 diabetes
T2DM	Type 2 Diabetes Mellitus
TPM	Transcripts Per Million
TRF1	Telomere repeat-binding factor 1

tRNA	Transfer RNA
TSS	Transcription start site
UTRs	Untranslated regions
WBCs	White blood cells
WDRG	Wolverhampton diabetes research group

Chapter 1

Introduction

1.1 Diabetes mellitus

Diabetes mellitus is a group of metabolic disorders resulting in increased blood glucose level over a prolonged period because of defects in insulin secretion, insulin action, or both. Higher morbidity and mortality risk are present in diabetic individuals suffering from diabetes as compare to the general population. International Diabetes Federation study found that the number of adults with diabetes will be expected to rise from 415 million in 2015 to a projected 642 million by 2040 (Ogurtsova *et al.*, 2017).

In developed countries, diabetes mellitus is more common and is, associated with increase in obesity and changes in the lifestyle. Two broad categories of diabetes are: type 1 diabetes (T1DM), also known as insulin-dependent diabetes or juvenile-onset diabetes) which constitutes only 5-10% of diabetes cases (American Diabetes Association, 2014 and Atkinson and Eisenbarth, 2014) and type 2 diabetes (T2DM) which is more prevalent and accounts for 90-95% of diabetes cases worldwide (Wild *et al*., 2004). Insulin secretion stops in T1DM which leads to hyperglycaemia. This is a result of autoimmune destruction of the insulin-producing β-cells leading to an absolute insulin deficiency in the system (American Diabetes Association, 2014). Hereditary characteristics and the impact of environmental factors on the autoimmune destruction of pancreatic β-cells have been implicated in changes leading to the onset of the disease (Soleimanpour and Stoffers, 2013). Islet cell antibodies (ICA) and ketoacidosis are apparent in T1DM. Islet cell auto-antibodies (antibodies produced against proteins of islet cell cytoplasm), insulin auto-antibodies, GAD65, and tyrosine phosphatases IA-2 and IA-2b are markers of immune destruction (Wesche *et al*., 2001). T1DM can be diagnosed with a

decreased level of circulating C-peptide with reasonable accuracy (American Diabetes Association, 2014 and Arvan *et al*., 2012).

On the contrary, insulin resistance and inadequate insulin secretory response appear in T2DM, leading to raised circulating glucose levels (American Diabetes Association, 2014). Specific aetiologies of T2DM are not known but islet cell antibodies are absent in individuals with T2DM (Thomas and Philipson, 2015). Glucose provides carbon atoms for biosynthesis and fuel for the provision of energy in living organisms. However, long-term exposure to increased glucose concentrations can produce damaging effects. Type 2 diabetes for many years may remain undetected. During this period when patients show no appearance of symptoms, it is sufficient to cause pathological changes (American Diabetes Association, 2014).

1.2. The economic impact of diabetes

In the UK, diabetes is one of the chronic illnesses that has a significant economic impact. There is not only direct cost associated with the treatment of illness and its associated complications, but indirect social and productivity costs involved with diabetes. Hex *et al*. (2012) reported that the total overall cost of diabetes was £23.7 billion in 2010-2011. The direct cost was £9.8 billion, of which £1 billion was for Type 1 and the remaining £8.8 billion for Type 2 diabetes. The indirect cost was £13.9 billion, of which £0.9 billion was for T1D and the remaining £13.0 billion was for T2D. Hex *et al*. (2012) estimated that the cost is expected to rise to £39.8 billion by 2035-2036 for this disease. Comprising direct cost of £16.9 billion: of this £1.8 billion will be for Type 1 and the remaining £15.1 billion will be for Type 2 diabetes. Whilst the

£22.9 billion will be indirect cost: of this £2.4 billion will be for T1D and the remaining £20.5 billion will be for T2D.

1.3. Diagnostic criteria for diabetes

After several years of debate and modification, the diagnostic criteria for diabetes and pre-diabetes (intermediate hyperglycaemia such as impaired fasting glucose (IFG) and impaired glucose tolerance (IGT)) was set. The cut-off level for fasting glucose was 7.0 mmol/l (Kumar *et al*., 2016 and World Health Organization, 2006) and the threshold for IFG was changed from 6.1 to 5.6 mmol/l by the American Diabetes Association (ADA) (Expert Committee on the Diagnosis and Classification of Diabetes Mellitus, 2003). In addition, for the diagnosis of diabetes and to identify the increased risk for future diabetes, ADA introduced the use of glycated haemoglobin (HbA_{1c}) (American Diabetes Association, 2010) with a diagnostic cut-off point ≥ 6.5% (48mmol/mol) for diabetes (Farmer, 2012). HbA_{1c} levels are thought to better predict the development of long-term complications in type 1 and type 2 diabetes.

1.4. Glycation of protein/Maillard reaction

Glucose or other reducing sugars react with proteins and this process is called glycation of proteins. Protein glycation/Maillard reaction provides an association between chronic hyperglycaemia and a series of pathophysiological alterations that are critical in the development of the chronic complications of diabetes (Fournet, Bonte and Desmouliere, 2018). Enhanced development of dicarbonyl species (such as glyoxal, methylglyoxal, and 3-deoxyglucosone) leads to the production of AGEs during protein glycation. During hyperglycemia and oxidative stress, the carbonyl group of glucose non-enzymatically react with amino groups of proteins leading to

the formation of a Schiff base. This thermodynamically unstable base after rearrangement converts into a stable ketoamine, the Amadori product. During the intermediate stage of the Maillard reaction oxidation, fragmentation and dehydration reactions occur that result in the degradation of the Amadori product into more reactive carbonyl species than the sugars from which they are derived (Amore *et al*., 2004). The carbonyl species are formed not only through the non-oxidative rearrangement and hydrolysis of the Amadori product, but also through the fructose-3-phosphate which is an intermediate of the polyol pathway.

Auto-oxidation of carbohydrates, decomposition of triose phosphates and lipid peroxidation result in methylglyoxal production, whilst glyoxal formation occurs via fructose-phosphate fragmentation, lipid peroxidation and oxidative fragmentation of the Schiff base. In addition, these carbonyl compounds can also be formed by glucose degradation in vitro (Thornalley, Langborg and Minhas,1999). Methylglyoxal is a very reactive alpha-oxoaldehyde (Nemet, Varga and Turk, 2006). Oxidation, dehydration, fragmentation and cyclisation reactions take place in the late stage of the Maillard reaction and carbonyl species react with free amino groups of proteins very rapidly and form insoluble, irreversible compounds, usually called Advanced Glycation End-Products (AGEs) (Figure 1.1).

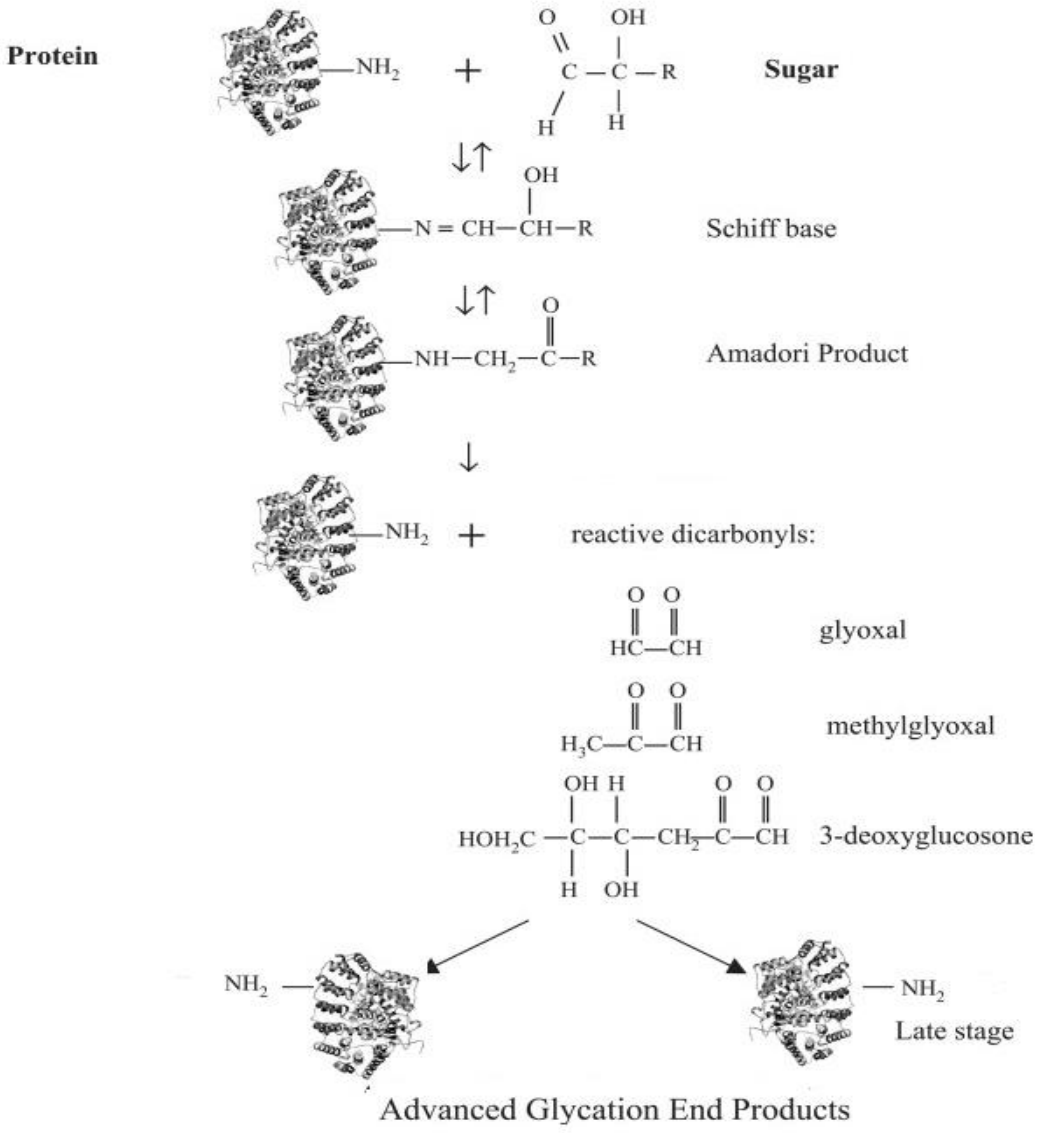


Figure 1. 1 Three stages of non-enzymatic glycation reaction/Maillard reaction. (Adapted and modified from Lapolla, Trald and Fedele, 2005).

1.4.1. Glycation of haemoglobin

The non-enzymatic reaction between glucose and amino groups in lysine and arginine residues or in the amino-terminal position leads to the glycation of proteins. A minor fraction of adult haemoglobin (HbA_{1c}) is glycated slowly. Erythrocytes are considered to be freely permeable to glucose because of the presence of glucose

transporter protein 1 (GLUT1) (Hajjawi, 2013). The non-enzymatic glycation reaction takes place within erythrocytes over the average life of red blood cells. The lifespan of red blood cells is approximately 120 days. During non-enzymatic glycation glucose covalently binds to the haemoglobin molecule (Goldstein *et al*., 2004). This is denoted as HbA_{1c} and its rate of formation is directly proportional to the glucose concentration inside the cell. Following the formation of Schiff base and ketoamine, this glycation adduct produces advanced glycation end products (AGEs) (Boteva and Mironova, 2019) which are associated with diabetic complications. This provides a link between the deleterious effects of chronic hyperglycaemia and several complications (Stitt, Jenkins and Cooper, 2002 & Cohen *et al*., 1999). Glycation of proteins is an important process in terms of biomarkers for the diagnosis of diabetes. The estimation of glycated haemoglobin or other glycated blood proteins (indicators of average glycaemia) helps in monitoring the efficacy of diabetes treatment. The American Diabetes Association recommended the use of glycated haemoglobin in the diagnosis of diabetes because of the established association between glycated haemoglobin and microvascular disease (American Diabetes Association, 2010).

Various molecular species of haemoglobin are present in human blood arising from the many potential glycation sites on haemoglobin which indicates that glycated haemoglobin is not a single molecular entity. Glycation of α- and β-chains in various species of HbA are collectively called the glycohaemoglobins (GHb). Total GHb includes all glycated subfractions of haemoglobin. Glycation reaction of haemoglobin takes place not only at the valine residues of N-terminal beta chain but this also occurs non-specifically at the epsilon amino groups on lysine residues along the alpha and beta chains and at the N-terminal valine of the alpha chains

(Little and Rohlfing, 2013). Consideration of all glycation sites makes total GHb value approximately 50% higher as compared to HbA_{1c} alone. HbA_{1c} makes about 5% of the total haemoglobin in normoglycaemic subjects. HbA_0, HbA_2, HbF are different molecular forms of human haemoglobin. Haemoglobin variants and “minor components” HbA_{1a}, HbA_{1b}, and HbA_{1c} in human blood were detected based on their elution from chromatography column (Makris and Spanou, 2011 & Weykamp, John and Mosca, 2009). HbA_{1c} was used for estimating the risk of diabetes complications after the publication of Diabetes Control and Complications Trial (DCCT) results and the prospective diabetes study performed inside the UK (Kilpatrick, Rigby and Atkin, 2008) as well as International Committee and the ADA also recommended HbA_{1c} as a diagnostic test for diabetes (International Expert Committee, 2009).

1.4.2. Glycation gap and consistency of the glycation gap in diabetes

Blood proteins are glycated in diabetes and in non-diabetes states. Protein glycation is important with respect to diabetes because the rate of reaction (covalent binding of glucose with protein) is slow. Moreover, and the level of protein glycation is directly proportional to glucose concentration (Ott *et al*., 2014 & Magalhaes, Appell, and Duarte, 2008). Protein glycation is an integrated measure of the blood glucose concentration over time. These are represented by the levels of glycated haemoglobin and serum fructosamines (glycated albumin) and they are commonly assayed to assess the efficacy of treatment (Vos, Schollum, and Walker, 2011 & Smaldone, 2008). Secondly, fructosamines may go through further spontaneous reactions, producing AGEs (advanced glycation end products). Both fructosamines and AGEs are thought to be involved in the development of diabetic complications

because they cross link with collagen and reduce vessel elasticity (Singh *et al.,* 2014). Vascular stiffness leads to vascular comlications.

Fructosamine, a ketoamine, is a product of non-enzymatic mechanism in blood plasma derived from glycation between a sugar molecule and a protein (usually albumin). This process involves the labile Schiff base intermediate and the Amadori rearrangement. Different kinetics of interaction of glucose with valine of haemoglobin versus lysine of fructosamine are observed in research studies.

In diabetic patients, the differences in HbA_{1c} contrasted with fructosamine were compared by Cohen *et al*. (2003). The possible causes for such discrepancy have been described, and various erythrocytic factors may affect HbA_{1c} independently of glycaemia (Beltran *et al.,* 2016 & Jeffcoate, 2004 & Hare, Shaw and Zimmet, 2012). These include impacts on red cell life expectancy and glucose concentration across the red cell membrane (Cohen *et al*.,2008 and Khera *et al*., 2008).

The glycation gap (G-gap), which is a variation between the actual HbA_{1c} concentration and that predicted by the fructosamine concentration, could provide an explanation of inter-individual differences in HbA_{1c} levels (Cohen *et al*., 2006). Within the intra-erythrocyte space, glycaemic control is measured by the haemoglobin A_{1C} (HbA_{1c}) level whereas serum proteins are glycated in the extracellular compartment. Thus, the G-gap might be linked with the complications associated with intracellular glucose metabolism. This might alter an individual's risk of vascular complications for any given level of long-term glycaemia by modification in one of the key pathological processes. These included protein glycations and the formation of advanced glycation end products (Cohen *et al*., 2003, McCarter *et al*.,

2004 & Gkogkolou, and Bohm, 2012 & and Goh & Cooper, 2008). The co-use of glycated haemoglobin and serum fructosamine is recommended as glycated HbA_{1c} alone may not be suitable for calculating prevailing glycaemia because of the factors described above (life expectancy of red cell and glucose gradient). In a similar way serum fructosamine may itself be affected by variability in protein turnover and serum albumin concentration (Speeckaert *et al*., 2014 & Duran *et al*., 2015). This highlights the importance of determining the G-gap in any individual. The above-mentioned factors affecting HbA_{1c} and serum fructosamine might have an overall impact on G-gap but Nayak *et al*. (2011) provided evidence for the consistent G-gap in patients with diabetes. This suggested that there is a consistent underlying mechanism that could be a genetic or biochemical process. Other studies also demonstrated such consistency in people with diabetes.

Considerable discrepancy in HbA_{1c} levels between-individual diabetic patients despite similar preceding levels of mean blood glucose was observed (Monnier, Lapinski and Colette, 2003). This variation is not random, as evidence has been provided by several investigators for between-individual biological variation in HbA_{1c} levels from populations of patients with diabetes (Hempe *et al*., 2002; Hudson *et al*., 1999; Madsen, Kjasrgaard and Ditzel, 1982). Hempe *et al*. (2002) used the Haemoglobin Glycation Index (HGI) for measuring HbA_{1c} variation between individuals with respect to mean blood glucose. HGI enabled the quantification of predictable and patterned differences in glycation level of an individual's intracellular haemoglobin relative to the rest of the population in response to a given level of MBG. However, the use of fructosamine instead of MBG for the comparison of the differences of HbA_{1c} in diabetic patients leads to the computation of glycation gap (G- gap) (Cohen *et al*., 2003). Individuals show consistency in G-gap and HGI over

time that represents a constant underlying mechanism for variation in intracellular glycation compared to extracellular glycation or glycaemia that is estimated by serum fructosamine or MBG (Nayak *et al.,* 2011, Cohen *et al.,* 2003 and Hempe *et al.,* 2002).

Haemoglobin glycation index (HGI) and G-gap may be important in any individual as identification of the G-gap might direct safer therapeutics. HGI and G-gap are two indices of between-individual differences in glycated haemoglobin (HbA_{1c}), adjusted for glycaemia that are based on a mean blood glucose level (MBG) and fructosamine measurement respectively.

1.4.2.1. Mechanisms of glycation gap (G-gap)

As described in the previous section (1.1.2 paragraph 3), glycated HbA_{1c} concentration is affected by various factors. Haemolysis, sickle cell disease, transfusion, cystic fibrosis, iron deficiency (anaemia) are potential underlying mechanisms that affect the erythrocytes lifespan independent of glycaemia. Therefore, HbA_{1c} can underestimate average glucose levels in some patients. Individuals may have a lower or higher HbA_{1c} concentration than predicted from prevalent glycaemia (Macdonald *et al.,* 2008, Hempe *et al.,* 2002, Chan *et al*., 2018 & Wei, Zheng, and Nathan, 2014). Dunmore *et al.* (2018) excluded mechanisms above which do not adequately account for the G-gap in many cases and focused on the factor of deglycation. Oxidases, isomerases, and kinases are three different types of enzymes involved in the metabolism of ketoamines (Wiame *et al*., 2002; Monnier and Wu, 2003). However, protein deglycation under physiological conditions is carried out by only the kinases. Fructosamine-3-kinase (FN3K) (Szwergold, Howell and Beisswenger, 2001 and Van Schaftingen *et al.,* 2012) is

found to be involved in the deglycation pathway. Deglycation is suggested to be responsible for the systematic variation of the glycation phenotype within the human population with diabetes measured using the G-gap as a metric. Therefore, the determination of G-gap is important in an individual with diabetes.

1.4.2.2. Clinical importance of G-gap

There are potential limitations of the use of HbA_{1c} in the monitoring of glycaemic control. There are different factors amongst others that affecting HbA_{1c} level. Haemolytic anaemia and chronic kidney disease decrease the HbA_{1c} level whilst splenectomy and reticulocytes increase the HbA_{1c} level (Sodi *et al.*, 2018). These and previously described factors are the reasons for potentially discrepant HbA_{1c} values when compared with other measures of average glycaemia including plasma fructosamine or glycated albumin. Glycation gap is a biochemical phenomenon affecting glycation independent of the factors described above (Dunmore *et al.,* 2018). The misclassification of glycaemic control based on HbA_{1c} alone (standard classifications of diabetes) could be avoided by considering another measure of glycaemic control, such as fructosamine alongside HbA_{1c}. In addition, continuous glucose measurement or glycated albumin should at least be used alongside HbA_{1c} in both diagnosis and monitoring treatment of diabetes. Different lifespans of haemoglobin and albumin do not affect the glycation gap. Individuals diagnosed with diabetes show fairly constant mean blood glucose (MBG) even if fluctuating over short periods, so fructosamine and $HbA1_c$ should give similar estimates of MBG if there is no G-gap (Nayak *et al*., 2011).

1.5. Fructosamine 3-kinase (FN3K)

1.5.1. Properties

FN3K is expressed at higher levels in the kidneys, heart, nervous tissue and erythrocytes (Mohas *et al*., 2010). This 35 kDa protein encoded by the *FN3K* gene which is located on chromosome 17q25.3 and ordered in six exons (Monnier, 2006). *FN3K* gene may have emerged during the duplication process of the gene, FN3K-related protein (FN3K-RP), as the gene encoding FN3K-RP, is located proximal to the gene encoding FN3K and has 65% sequence homology with FN3K and an identical genome organisation (Collard *et al*., 2003 and Delplanque *et al*., 2004). FN3K and FN3K-RP both have 6 exons where starting ATG codon is present in exon 1 and a termination codon is present in exon 6. They have intron-exon junctions in homologous position and lack TATA or CCAAT sequence but with high GC content in promoter. Their promoter sequences are similar to the housekeeping genes which indicate that they are likely to be expressed in various tissues (Conner, Beisswenger and Szwergold, 2005). The level of expression of FN3K in different tissues remains unchanged regardless of starvation and diabetes. No regulation of FN3K expression in human fibroblasts treated with the similar condition as in the diabetic state. This provides the stronger evidence that FN3K is a housekeeping gene (Conner, Beisswenger and Szwergold, 2005). Both FN3K and FN3K-RP catalyse the phosphorylation of psicosamines and ribulosamines. However, only FN3K acts on fructosamines (Collard *et al*., 2003) and deglycates proteins under physiological conditions. FN3K is more active in those tissues that contain proteins with long half-lives, such as erythrocytes (Collard *et al*., 2003). It was discovered initially with the identification of F3P (fructose 3-phosphate) in human and animal tissues (Szwergold *et al*., 2007), where 3-phosphokinase enzyme phosphorylates

fructose into fructose-3-phosphate (F3P) in the lens and erythrocyte, leading to the development of 3-deoxyglucosone (Lal *et al*., 2005) in diabetes.

FN3K is important in cellular defence and/or repair system where it deglycates proteins including HbA_{1c} by a catalytic pathway (Szwergold, Howell and Beisswenger, 2001; Van Schaftingen *et al.,* 2012). This involves ketoamines removal and prevention of AGE production. In addition to fructosamines, this enzyme also phosphorylates C3-epimers psicosamines, as well as ribulosamines and erythrulosamines. However, FN3K has a much lower affinity for psicosamines than the other ketoamines (Szwergold, Howell and Beisswenger, 2001).

1.5.2. FN3K involvement in deglycation

Delpierre *et al.* (2000) discovered, cloned and characterised FN3K enzyme in erythrocytes based on earlier NMR studies. They found that this enzyme deglycates both small molecular weight products as well as protein bound fructosamines. Erythrocytes of FN3K-knock out mouse were incubated with 200 mmol/L glucose leading to the increased formation of glycated haemoglobin alongside an increased level of inorganic phosphate. However, when this experiment was repeated with FN3K competitive inhibitor deoxymorpholinofructose (DMF), the concentration of DMF increased about two-fold the rate of accumulation of glycated haemoglobin with the decreased concentration of inorganic phosphate. This indicated the involvement of FN3K in deglycation (Delpierre *et al*., 2002).

FN3K is involved in the deglycation of glycated haemoglobin in erythrocytes as well as other glycated proteins that are present in other tissues. The difference in FN3K activity could result in an apparent difference in glycated haemoglobin levels. In

some individuals FN3K activity is high: then carry out deglycation at a higher rate and this results in a lower level of glycation, whilst those with low FN3K activity have a low rate of deglycation and results in high level of glycation (Mohas *et al*., 2010). The FN3K enzyme is involved in the control of intrinsic toxicity of glucose due to intracellular non-enzymatic glycation, this is a complex process that involves many stages as it proceeds and leads to the formation of irreversible end products known as advanced glycation end products (AGEs) (Figure 1.1).

The FN3K enzyme catalyses the phosphorylation of fructosamines with high affinity and leads to the development of F3P (fructosamine-3-phosphate) at the expense of ATP (Szwergold and Beisswenger, 2003). This phosphorylation destabilises the fructosamine linkage, which causes the breakdown of F3P (fructosamine-3-phosphate) to lysine, 3-deoxyglucosone, and inorganic phosphate (Conner, Beisswenger and Szwergold, 2004), thereby reversing the non- enzymatic glycation process at an early stage (Figure 1.2). The reduction or oxidation of 3-deoxyglucosone to 3-deoxyfructose (Sakiyama *et al.*, 2006) or 2-keto-3-deoxygluconic acid respectively neutralises its potential toxic effects (Ruckriemen *et al*., 2018). The discovery of the FN3K deglycating enzyme increased the understanding of cellular metabolism and its role in protein deglycation in the aetiology of diabetes.

Research study demonstrated the phosphorylation of glycated haemoglobin by FN3K in intact cells to prove partial deglycation. The aim was the identification of fructosamines removal from haemoglobin in intact erythrocytes by action of the FN3K (Delpierrre *et al.,* 2004). Study indicated that lysines bound fructosamines are relevant substrates for FN3K in addition to other low-molecular-mass glycated

substrates due to the high affinity of FN3K for FL (Fructoslysine, Amadori product) that was detected in vivo (Avemaria *et al*., 2015). Krause *et al.* (2006) used reverse-phase high-performance liquid chromatography (RP-HPLC) to estimate FN3K activity within human erythrocytes. Krause *et al.* (2006) used erythrocyte lysate and measured the product Nα hippuryl-Nε-(phosphofructosyl) lysine (BzGpFruK) that was formed by the transformation of the synthetic substrate Nα-hippuryl-Nε-(1-deoxy-D-fructosyl) lysine (BzGFruK) by FN3K. FN3K is constitutively expressed in human cells, previous studies found a correlation between FN3K activity and the levels of HbA_{1c} in patients with diabetes (Delpierre *et al*., 2002) where certain single nucleotide polymorphisms were found in *FN3K* gene. FN3K`s involvement in deglycation of intracellular proteins such as haemoglobin may provide a logical reason behind the variation in the glycation (measured as G-gap) because haemoglobin is intracellular as is FN3K which therefore does not affect fructosamine (albumin) that reflects extracellular circulating protein. The International Federation of Clinical Chemistry working group (IFCC) defined haemoglobin A_{1c} (HbA_{1c}) as haemoglobin that undergoes irreversible glycation at one or both N-terminal valines of the beta chains (Nasir, Thevarajah and Yean, 2010) as well as epsilon amino group of lysine residue on both α and β chains and form fructosylvaline and fructosylysine respectively. Fructosvaline is a low-affinity substrate for FN3K as compare to fructoslysine that is an excellent substrate.

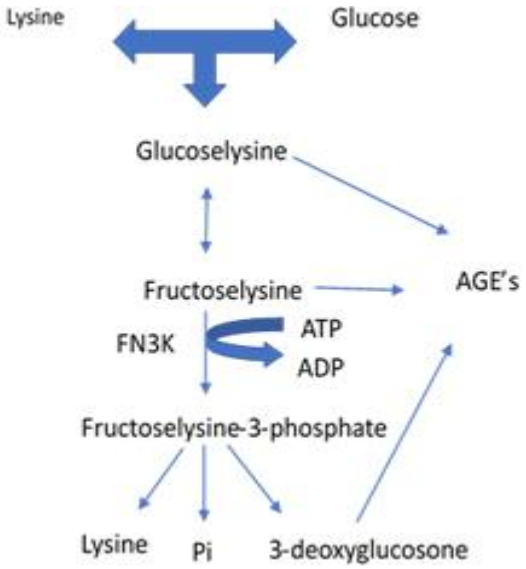


Figure 1. 2 FN3K role as a catalyst in deglycation of Fructosyl amino-acid Fructoselysine-3-phosphate (FL3P) is the unstable product formed from the phosphorylation of fructose-lysine. This phosphorylation destabilises fructose-lysine linkage and FL3P breaks down to unmodified lysine by β-elimination with the accompanied formation of inorganic phosphate (Pi) and 3-deoxyglucosone (3DG) (Szwergold *et al*., 2001).

1.5.3. Genetic variants of FN3K

The consistent FN3K activity in human erythrocytes within an individual is highly variable between individuals regardless of the blood glucose levels or the type of diabetes (Delpierre *et al*., 2006). Previous study demonstrated the contribution of the FN3K enzyme in the development of diabetic complications by the comparison of the patients with complications with low and high FN3K activities (Delpierre *et al*., 2006). Dunmore *et al.* (2018) reported a remarkable difference in the activity between individuals despite similar blood glucose levels after the application of HPLC method (Krause *et al*., 2006) and introduced the term high deglycator group

(with high FN3K activity) and low deglycator group (with low FN3K activity). An alternative approach to understanding the possible role of FN3K in diabetic complications would be to analyse the presence of different SNPs in the FN3K gene that previously linked to FN3K enzymatic activity and HbA_{1c} level.

Genome-wide association studies (GWAS) of glucose and Insulin have identified variants at various genomic loci that affected HbA_{1c} through glycaemic, non-glycaemic (erythrocytic) and unclassified pathways. Glycaemic pathways involve factors that affect the blood glucose level whilst non-glycaemic, i.e. erythrocytic pathways include factors that affect the red blood cell (Leong and Wheeler, 2018). The FN3K locus identified through GWAS is an example of unclassified variants that affect the HbA_{1c} through the unclassified pathway (neither glycaemic nor erythrocytic) e.g. deglycation of glycated proteins (Wheeler *et al*., 2017). These unclassified genetic variants might act through mechanisms that are neither initiated through between-person variation in erythrocytes lifespan, erythrocyte biology or glycaemia.

In diabetic humans, certain single nucleotide polymorphisms (SNPs) of the *FN3K* gene i.e. C/G (rs1056534) alter its activity (Avemaria *et al.,* 2015*)*, C/G (rs1056534) influence sRAGE and glycated haemoglobin levels (Škrha *et al*., 2014 and Mohás *et al*., 2010), C/T (rs1046896) responsible for variability in glycated haemoglobin level (Soranzo *et al*., 2010), and C/T (rs3848403) affecting sRAGE concentration and linked with some diabetic complications (Škrha *et al*., 2014). Previous study also showed an association between FN3K enzymatic activity in red cells and some polymorphisms in the *FN3K* gene *(*Delpierre *et al*., 2006), where the SNP C/G (rs1056534) in exon 6 was connected with reduced enzymatic activity measured in

erythrocytes. However, no correlation has so far been detected between FN3K SNPs and HbA_{1c} levels (Delpierre *et al*., 2006). The study by Dunmore *et al.,* (2018) showed a notable correlation between erythrocyte FN3K and the G-gap. The study reported that FN3K enzyme activity and protein levels were both notably elevated (3-fold) in the negative G-gap group (lower HbA_{1c} levels than average glycaemia). These differences may arise through the single nucleotide polymorphisms in the *FN3K* gene (Avemaria *et al.,* 2015, Tanhauserova *et al.,* 2014 and Soranzo *et al.,* 2010), leading to the potential differences in enzyme activity or differentially expressed genes and transcript variants that are enriched in different biological processes, pathways and diseases that might give rise to the G-gap.

In this thesis, the correlation between FN3K SNPs and the glycation gap was investigated in patients with diabetes. Mohas. (2010) reported that the C allele of rs1056534 was linked with lower HbA_{1c} concentration and with later development of T2D. However, this variant was independent of diabetic complications, such as nephropathy, neuropathy or retinopathy during the investigation. In 2014, to verify the influence of genetic variability in FN3K genes on diabetic nephropathy progression, the correlation was reported between SNPs C/G (rs1056534) and diabetic nephropathy progression and cardiovascular morbidity and mortality in T2DM subjects (Tanhauserová *et al*., 2014). Skrha and co-workers also evaluated the genetic variability of FN3K and GLO1 with parameters of endothelial dysfunction and soluble receptor for AGEs (sRAGE) in patients with diabetes (Skrha *et al*., 2014). In diabetic patients, a significant relationship between the rs1056534 and rs3848403 of *FN3K* gene polymorphisms was proved with circulating sRAGE concentration.

1.6. Q248H ferroportin1 polymorphism

SLC40A1 gene encodes cellular ferroportin, a transmembrane protein that exports iron from the cells to the bloodstream (Girelli *et al*., 2008). Research evidence indicate an association between ferroportin polymorphism *SLC40A1* Q248H (localized in exon 6), and high levels of serum ferritin. It has been demonstrated that it may be responsible for primarily high iron levels in sub-Saharan African population (Beutler *et al*., 2003; Gordeuk *et al*., 2003; Kasvosve *et al*., 2005; McNamara *et al*., 2005; Barton *et al*., 2007 and Rivers *et al*., 2007). Ferroportin Q248H mutations (rs11568350) in gene leading to African iron overload disease (hemochromatosis type IV) or hyperferritinaemia as a result of increased duodenal iron absorption (Katchunga *el al.,* 2013, Kasvosve *et al.,* 2015 and Rivers *et al.,* 2007). Cikomola *et al.* (2017) found the correlation between FN3K activity and *Q248H* genotypes (rs11568350) in subjects. Research also indicate that serum ferritin is associated with the risk of developing diabetes mellitus (Sun *et al.,* 2008). Furthermore, ferroportin (encoded by*SLC40A1* gene) Q248H mutation is prevalent in Africa (Albuquerque *et al.,* 2011).

1.7. Single nucleotide polymorphisms (SNPs) genotyping

The single nucleotide polymorphisms (SNPs) are genetic signatures and caused by variation of a single nucleotide (A, T, C and G) in the DNA sequence. These are used for the study of predisposition to certain traits including disease (Chanock, 2001). Generally, SNPs (which do not always modify gene function) are important for the dynamics evaluation of the studied population. However, SNPs presence may lead to alternatively spliced transcript variants and may affect the transcription factor`s binding capacity (Uppu, Krishna and Gopalan, 2016). The SNPs may present in the coding/noncoding regions of genes. These are used for the

investigation of phenotypic characteristics based on their genotypes and they also play a role in diagnostic assays development (Karki *et al*., 2015). The TaqMan method is commonly used for SNP studies. This method is based on 5′ exonuclease chemistry. TaqMan assay use two fluorescently labelled probes to detect target amplicons which are accumulated during the PCR reaction. These target specific probes are highly sensitive (Martino, Mancuso and Rossi, 2010). The probes are non-fluorescent because of the close proximity between fluorophore and quencher. Once the Probe/target hybridisation occurs during each PCR cycle Taq DNA polymerase cleaves the probe which results in an increase in fluorescence (proportional to the accumulated target amplicon) because of conformational changes where the close proximity between fluorophore and quencher no longer exists. Fluorescent signals are detected using sequence detection software installed into the real-time PCR machine (Malkki and Petersdorf, 2012).

1.8. Diabetic complications and vascular diseases

Diabetic complications arise from chronic hyperglycaemia, including microvascular and macrovascular complications (Naess *et al.,* 2003 and Skyler & Oddo, 2002). Atherosclerotic cardiovascular, peripheral arterial and cerebrovascular diseases are examples of main macrovascular complications. Microvascular complications include retinopathy with potentially affecting the vision; nephropathy leading to renal damage; peripheral neuropathy with the potential risk of foot ulcers and amputations (Bailes, 2002). Macrovascular and microvascular complications in diabetes are caused by extended exposure to hyperglycaemia together with arterial hypertension, dyslipidaemia as well as genetic susceptibility, considered as risk factors (Krentz, Clough and Byrne, 2007). Obesity in T2D, insulin insensitivity or damage to insulin secretion leads to early glycaemia, responsible for variation in

structure and function of vessel wall reaching a climax with diabetic vascular complications. The disproportion between nitric oxide (NO) bioavailability and production of reactive oxygen species (ROS) is stimulated by high glucose concentrations leading to the alterations of vascular function (Creager *et al*., 2003).

To explain the association between the increased concentration of glucose and the development of complications “Non-enzymatic Glycation Hypothesis” was suggested. The “Non-enzymatic Glycation Hypothesis” hypothesises that immoderate non-enzymatic alteration of proteins and some phospholipids by glucose and its by-products are due to the detrimental effects of chronic hyperglycaemia (Stitt, Jenkins and Cooper, 2002). Diabetic complications dependent on several factors. The biochemical process of advanced glycation has been demonstrated to play a key role in the pathogenicity of long-term vascular complications in patients with diabetes (Brownlee, 2005, Genuth *et al*., 2005, Vlassara & Striker, 2013). Pentosidine and carboxymethyl-lysine (CML) are some important chemically characterised AGEs in humans, of which pentosidine has intrinsic fluorescence. It can also be used as a marker of AGE accumulation. AGEs are also originated from exogenous sources such as tobacco smoke and diet and they contribute to renal and vascular complications (Vlassara *et al*., 2002 and Zheng *et al*., 2002) and participate in insulin resistance and diabetes progression (Cai *et al*., 2012). Both fructosamines and AGEs can alter the biological and functional properties of proteins in living organisms (Ulrich and Cerami, 2001 & Chilelli, Burlina, and Lapolla, 2013). This highlights the importance of deglycation processes for retardation of vascular damage progression.

1.8.1. The function of advanced glycation end products (AGEs) and their receptors in diabetes

Spontaneous damage to proteins occurs due to the glycation. Many studies not only confirmed the traditional role of glycation in diabetes complications but also brought new insights. This is evident from AGEs that are involved in the pathogenesis of diabetes complications and in morbidity and mortality of diabetic patients. Nevertheless, it is critically important to identify proteins and their significant active sites vulnerable to glycation. These directed targets will help to develop inducers of enzymatic defence against the glycation of proteins (Schmidt and Stern, 2000; Ahmed and Thornalley, 2007) such as FN3K.

AGEs are generated from glucose-derived dicarbonyl precursors as well as glycation of proteins. These accumulate intracellularly and act as a stimulus for activating intracellular signalling pathways as well as altering the function of intracellular proteins (Brownlee, 2001 & Singh *et al*., 2014). AGEs are also present in the extracellular matrix and thus lead to the modification of matrix proteins and diminish matrix-matrix as well as matrix cell interactions. This may lead to cell death, cell differentiation or reduced cell resistance and relocation (Monnier *et al*., 2005). AGEs form cross-links with collagen and other proteins and leading to decreased vessel elasticity (Zieman *et al*., 2007). AGEs directly change protein structures and functions or cross-link which is potentially damaging (Singh *et al*., 2001). Chronic hyperglycaemia in a diabetic condition not only elevates the generation and accumulation of AGEs is also associated with impaired renal function because the kidney is the major site for clearing AGEs. AGE-modified proteins may be less susceptible to enzymatic degradation (Singh *et al*., 2014). AGE formation, AGE precursors and oxidative stress are accelerated with a high concentration of

circulating glucose in diabetic patients. High concentrations of AGEs have been found in serum and tissues of patients with diabetes, and accumulated AGE in diabetic tissue found to be correlated with diabetic complications (Monnier *et al*., 2005). Dunmore *et al*. (2018) identified differences in AGE levels and associated diabetic complications between the G-gap groups. In addition to direct impairing protein structures and functions, AGEs bind to the receptor of advanced glycation end products (RAGE) and have damaging effects (Sessa *et al*., 2014), which includes chronic inflammation, stimulating oxidative stress and playing an important role in cardiovascular disease and diabetic complications (Coughlan *et al*., 2009 and Yan, Ramasamy and Schmid., 2010). RAGEs are involved in the pathogenesis of diabetes angiopathy (Yan, Ramasamy, and Schmidt, 2007, Lindsey *et al*., 2009 & Skrha *et al*., 2012), atherosclerosis, cardiovascular disease, and other diabetes complications.

Total sRAGE is comprised of extracellular domain of wild-type full-length RAGE (cleaved from cell membrane) and an endogenously secreted isoform (esRAGE). The sRAGE is a potentially a useful biomarker for cardiovascular disease (CVD) in diabetes. It is possible to measure full-length RAGE and esRAGE separately (Zhang, Postina and Wang, 2009 & Hudson *et al*., 2008). RAGE antagonists are in the developmental phase as therapeutics for diabetes complication. Alongside this, there is a need for therapy focusing on total sRAGE and esRAGE as potential biomarkers for diabetes complications. RAGE is a multi-ligand receptor as it interacts with a wide range of AGEs and belongs to the immunoglobulin superfamily. *In vitro* studies showed that binding of AGEs with RAGE activates NAPDH oxidases and leads to the increased formation of intracellular ROS (Wautier *et al*., 2001 and Zhang *et al.,* 2006), which in turn leads to AGE formation. In addition to these

damaging mechanisms, AGEs also participate in the activation of the transcription factor, nuclear factor kappa B (NFκB) (Bierhaus *et al*., 2001). Expression of pro-inflammatory cytokines such as interleukin 6 (Sessa *et al*., 2014) and RAGE itself increases because of NFκB activation (Tobon, Cuevas and Torres, 2014), thus escalating the inflammatory response.

1.9. The nature of transcriptome

Examining specific SNPs may not provide a comprehensive view of the reasons for the differences in glycation in the G-gap whereas a transcriptomic approach has the potential to identify novel differences in transcription of a broader range of genes. Gene expression, or the transcription of DNA into RNA, can be used to determine the phenotype at the cellular level and the set of RNAs expressed in a given cell. In the process of transcription, DNA strands act as templates for the synthesis of single-stranded RNA molecules (i.e. transcripts) which are complementary to DNA. This process requires the presence of RNA polymerases. RNA polymerase II plays a key role in the transcription reaction because, in its presence, the binding of proteins (transcription factors) to a regulatory region located upstream of the gene, known as the promoter takes place (Fuda *et al*., 2009). After the assembly step and other conformational rearrangements (i.e. RNA releases from the transcription initiation complex and leads to promoter clearance), the elongation phase starts (Kwak and Lis, 2013). In the elongation phase, complementary RNA synthesis in 5' to 3' direction takes place from the transcription start site (TSS). The study of RNA produced in the cells, in terms of characterisation and quantification, is critical for the in-depth understanding of biological pathways. The term transcriptome is used for all RNA molecules, which is a complete set of transcripts present in a cell, tissue or in an organism at a particular time point. Transcriptomic studies provide a better

understanding of how different cell type exist in an organism from a genome and how genes are regulated during the development stages and in response to condition associated with diseases (Pertea, 2012). In contrast to DNA, RNA is single-stranded, backbone consists of ribose, and uracil (U) nucleotide is present instead of thymine (T). Diversity in human transcriptome has been revealed (Martin *et al*., 2014). The data published by Encyclopaedia of DNA elements (ENCODE) showed that more than 60% of the human genome consists of transcribed regions (Djebali *et al.,* 2012), which include protein-coding RNA and processed non-coding RNA. The majority of 55000 genes in the human genome do not encode proteins whilst only approximately 20000 are protein coding genes.

Several RNA variants exist including messenger RNA (mRNA) (encodes protein) and non-protein coding RNA (ncRNA), such as transfer RNA (tRNA), ribosomal RNA (rRNA) and regulatory RNAs including micro RNA (miRNA), small nuclear RNA (snRNA), short interfering RNA (siRNA) and long non-coding RNA (lncRNA). Cytoplasm contains protein-coding transcripts whilst the nucleus contains non-protein-coding transcripts (Djebali *et al.,* 2012). Total RNA in the cell contains a small part (<5%) of protein-coding mRNA whilst the majority of the total RNA is rRNA (International Human Genome Sequencing Consortium, 2004). mRNA contains all information that is necessary for protein synthesis. The mRNAs derived from protein-coding genes are translated into polypeptides. Expression of the polyadenylated (poly---A) mRNAs is of great interest in today's research. A precursor mRNA molecule is formed during transcription which undergoes several modifications, including 5' capping, the polyadenylation of the 3' end and the removal of introns by splicing (Darnell, 2013). In 5' capping, methylated guanine is ligated to its 5' end via an enzymatic reaction. Such capping makes mRNA distinct

from other RNA species, makes it more stable and helps in the initiation of the translation process by enabling ribosome binding to mRNA. During the polyadenylation process, poly---A---tail of about 200 nucleotides is ligated at the 3'end of the transcript which extends the half-life of mRNA. While in splicing event introns are removed from pre-mRNA and exons come together. These exons are stretches of sequence that contain the necessary information for protein synthesis. After various steps and the association of numerous RNA-binding and processing proteins this precursor mRNA is processed into mature mRNA (Maniatis and Reed, 2002). By changing the exon composition of the same mRNA and alternative inclusion or exclusion of exons a range of unique transcripts (isoforms) as well as proteins can be produced.

Alternative splicing is almost common for all human multiexon genes and on average around six transcripts per protein-coding gene were found and the majority of transcripts contained four exons, whilst 1.6 transcripts per gene were found for long non-coding RNA and the majority of them contain two exons (Consortium, 2012 & Harrow *et al.,* 2012). The nucleus is the prime location for alternative splicing, after that the molecule moves to the rough endoplasmic reticulum, where protein synthesis takes place. This process is called translation. The nucleus is also the place for another regulatory mechanism i.e. RNA editing, where small changes in nucleotide sequence occur by insertion, deletion or nucleotide replacement. These changes lead to amino acid substitutions, altering splice patterns and thus directly or indirectly can change the expression and function of many genes (Farajollahi and Maas, 2010; Gott and Emeson, 2000) as well as information stored in the RNA. In humans, A to I editing is common, where adenosine is replaced with inosine in RNAs (Valente and Nishikura, 2005). For the regulation of gene

expression, cells use different mechanisms, including binding of transcription factors to the promoter and enhancer regions. These transcription factors activate transcription. Other important gene expression regulators are epigenetic factors i.e. DNA methylation where a methyl group is added to the cytosine bases of the DNA sequence of the gene, resulting in cell suppressing the initiation of gene expression. Gene expression is also influenced by histone modifications (epigenetic factor), leading to a change in chromatin structure. Its importance appears in cell development (Smith and Meissner, 2013), cancer (Kulis and Esteller, 2010) and ageing (Horvath, 2013). In a human cell, RNA expression level is regulated at various levels and used as a surrogate to phenotypes such as disease (Vogel and Marcotte, 2012). Multiple technologies are available to measure RNA in a high-throughput manner (Marioni *et al*., 2008).

1.9.1. Transcriptomic analysis of gene expression

The transcriptome is a complete set of transcripts in a cell where transcription of the genetic code to RNA molecules takes place during a specific developmental stage or physiological condition. These polyadenylated products are formed as a result of RNA polymerase II activity (Tang *et al*., 2010 & Wang, Gerstein and Snyder, 2009). Despite the fact that the same genome is present within an organism or tissue, diverse phenotypes exist. This happens potentially because of mRNA transcripts variants that have varying types and amounts (Kalisky *et al*., 2011 & Tang, Lao and Surani, 2011) or divergence of cis-regulatory regions (Stern and Orgogozo, 2008). These cis-regulatory regions are the part of DNA sequence that regulates gene expression and divergence can also arise from epigenetic changes that alter chromatin structure (Wittkopp, 2005) and may alter transcription rate and/or transcript stability. Transcriptomic analysis (of mRNAs, non-coding RNAs and small

RNAs) in this sense provides an understanding of the transcriptional structure of genes with regards of their start sites, 5′ and 3′ ends, splicing patterns and other post-transcriptional modifications. This is therefore relevant in the quantification of the changing expression levels of each transcript during development and under different environmental conditions (Wang, Gerstein and Snyder, 2009). It also allows the identification of all genes with different expression patterns across various conditions. RNA-Seq allows the sequencing of a cell's entire transcriptome. The technique also allows the researcher to study this with unprecedented detail as only 1-2% of transcribed genes codes for and are translated to proteins. Often, 80-90% of the processed genes that did not translate to proteins are involved in the process of epigenetic and gene expression regulation (Elgar and Vavouri, 2008 & Shabalina and Spiridonov, 2004).

Regulation of a complex gene expression process occurred at transcription, post-transcriptional modifications, and translational level. Gene expression is a complex process. During transcription process multiple transcripts may be produced from a single gene as a result of multiple Transcription Start Sites (TSSs) (Davuluri *et al*., 2008). Alongside this, alternative splicing of the pre-RNA may produce several transcripts from the same gene (Chen and Manley, 2009 & Tian and Manley, 2013).

1.9.2. Methods for transcriptome analysis

For gene expression measurement many technologies have been in use over the years. These include Northern blots and qPCR, hybridisation-based microarray and more recently RNA-sequencing (RNA-Seq) technology.

1.9.2.1. Northern blots and qPCR

These techniques were initially used for measuring mRNA abundance and were suitable to investigate a limited number of genes at a time and provide insight into a few candidate genes. However, to measure the entire transcriptome these techniques were found to be impractical.

1.9.2.2. Microarray

Microarray (hybridisation-based) is a traditional method that was used to study gene expression and quantify RNA in a global manner (Lowe *et al*., 2017). It involves the incubation of fluorescently labelled complementary DNA (cDNA) fragments with microarray chips. The microarray chips are fixed with oligonucleotide probes containing a perfect match with target sequences. Laser scan for fluorescent detection of the probe-target pair quantifies gene expression levels. Useful information can be obtained from microarray results such as how gene expression changes during developmental states and how it varies between and among spaces (Zhang *et al*., 2007).

Although microarray is a relatively low-cost method, potential cross-hybridisation, reliability on existing knowledge of genome sequence and limited ability to fully detect and quantify the diverse RNA molecule that are expressed at various levels made it of limited use for RNA analysis (Okoniewski and Miller, 2006 & Pan *et al.,*

2004). Therefore, microarray has a limited interrogating capacity and is useful for only specific genes for which probes are designed.

1.9.2.3. RNA-Seq

RNA-Seq, a high- throughput sequencing approach is a relatively new technique. This technique is more sensitive for detecting genes at low expression level and it overcomes many weaknesses of hybridisation-based method.

In today's age, RNA sequencing has become a tool of choice for transcriptome analysis. It can detect all coding and non-coding transcripts in a cell as well as determine their sequence and abundance. It can also identify alternative splicing (Pan *et al*., 2008), novel transcript (Guttman *et al*., 2010), gene fusion events (Edgren *et al*., 2011), RNA editing products, mutations in the expressed coding sequence and allele-specific expression (Borodina, Adjaye and Sultan, 2011).

RNA-Seq does not rely on the prior knowledge of a gene to identify and quantify the novel transcripts. This shows that RNA-Seq can directly perform whole-genome transcriptome profiling in a high-throughput manner (Mortazavi *et al*., 2008 and Mutz *et al*., 2013). However, specialised algorithms and more powerful servers are needed for bioinformatic analysis of the big volume of data generated from this method (Pop and Salzberg, 2008).

1.9.3. Roadmap of an RNA sequencing experiment

Next generation or high throughput sequencing of RNA molecules is used to describe several sequencing technologies, including the one developed by Roche (454), Illumina (Hiseq 4000) and Applied Biosystems (ABI SOLiD). These

technologies require different experimental protocols and differ by sequencing chemistry, length of the read produced, running time for the experiment, throughput per run and reads produced per run (Di Bella *et al*., 2013). The most commonly used one, which is adopted with Illumina's machines because it is cost effective and it accommodates the samples with lower accuracy (Mardis, 2013), usually follow these steps (Figure 1.3):

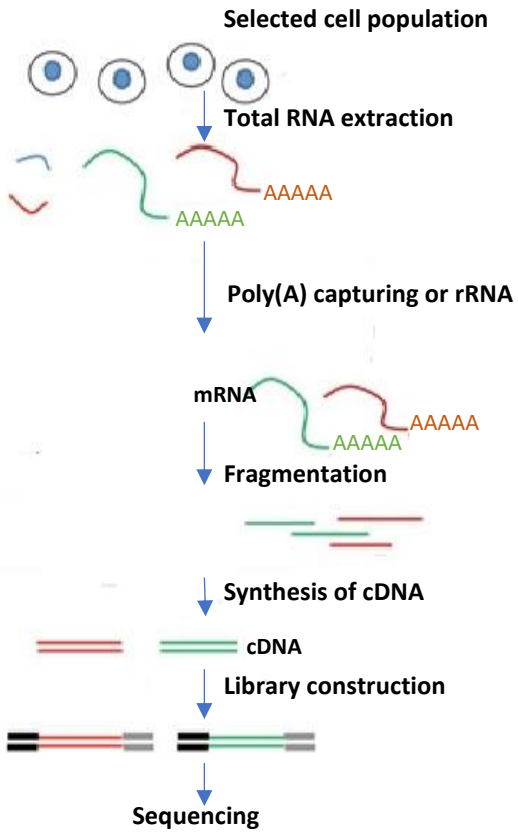


Figure 1. 3 Diagrammatic representation of RNA-Seq. Total RNA was isolated; enrichment and fragmentation of mRNA were performed. After that, the mRNA molecules were reverse transcribed into cDNA, which is then used to prepare a DNA library and for sequencing. (Zeng and Mortazavi, 2012).

1.9.3.1. Enrichment of protein-coding mRNAs from a larger pool of total RNA

More than 80% of the total RNA consists of rRNA transcripts. To avoid, the correspondence of sequencing reads with these transcripts, enrichment of protein-coding mRNAs of the transcriptome is performed for sequencing purposes via two methods which include poly-A enrichment of mRNA and depletion of rRNA. Poly-A enrichment process briefly involves, hybridisation of polyadenylated 3´-end of RNA transcripts to poly-d (T) probes, while ribosomal deletion involves hybridisation of abundant rRNA molecules to specific probes and subsequent deletion (Braasch and Corey, 2001).

1.9.3.2. Fragmentation of RNA

Next step involves the fragmentation of enriched RNA molecules (Illumina sequencing technology is only able to use DNA fragments of a specific length i.e. 150-200 bp) via RNA hydrolysis or nebulisation or uridine digestion to obtain a uniform sampling of sequences along the transcripts (Mortazavi *et al*., 2008). Alternatively, the cDNA (described below) may be fragmented instead of enriched RNA molecules via DNase I treatment or sonication (Wang, Gerstein and Snyder, 2009).

1.9.3.3. Synthesis of cDNA

DNA molecules are more stable in contrast to RNA. Reverse transcription process is used for the synthesis of cDNA molecules by using RNA molecules as templates. To start this process a random hexamers primer (sequence of the obtained fragments is unknown at this point) hybridises to the random positions of RNA sequence and synthesise single-stranded cDNA molecule. The hairpin loop at the

3’ end of single-stranded cDNA molecule act as a primer for the synthesis of a second DNA strand that is complementary to the first strand.

1.9.3.4. Adapter ligation and PCR amplification

Adaptors are ligated at both ends of cDNA molecules to satisfy two purposes. The first purpose is to enable the sequencing machine to recognise the fragments/hybridise the fragments into a flow cell where sequencing occurs and the second purpose to act as a primer for sequencing. Next step is the PCR amplification of cDNA molecules.

1.9.3.5. Size selection

To ensure that all molecules are of similar length, gel electrophoresis is performed to select the correct size of cDNA molecules, as a sequencing machine can sequence typically 150-200 bp fragments. After verification of cDNA library concentration and its fragment length samples are ready to load into a flow cell for sequencing (Mardis, 2013).

1.9.3.6. Sequencing-by-synthesis

In the initial step of sequencing, the double-stranded molecules are separated into single strands by denaturing and loaded into the flow cell. Adapters at both ends of cDNA molecules are complementary to the surface-bound oligonucleotide in the flow cell. Bridge amplification takes place after fragments hybridise to the surface-bound oligonucleotide in the flow cell. Synthesis of the new strand (complementary to the original strand) from the 3′ end of the flow cell oligonucleotide takes place in the elongation phase. The original strand is then removed by denaturation. Bridge formation and generation of a new site for the synthesis of a second strand occurs

by the hybridisation of the adaptor sequence at the 3′ end of the newly synthesised strand to a new surface-bound complementary oligonucleotide. In the end, after several hybridisations, extension and denaturation cycles many clusters originate from the same template with identical sequences. (Figure 1.4) (Bentley *et al*., 2008).

In each sequencing step, the Illumina system uses ‘cyclic reversible termination’ strategy. Under this strategy reversibly terminated fluorescently labelled nucleotides (specific for each nucleotide type) are incorporated into the growing complementary strand. Fluorescence image is captured via laser to identify which nucleotide was incorporated, and cleavage cycles are repeated (Bentley *et al*., 2008 and Metzker, 2005). The reaction begins with the use of a modified nucleotide containing a reversible terminator that terminates DNA synthesis after successful incorporation of a single base. The modified nucleotide also contains fluorescent dye. The fluorescent dye and terminal group are then removed from the new nucleotides and the process is repeated. Fluorescence imaging measurement helps to identify the incorporated single base. This process is repeated to produce the set of images. These images are interpreted with a base calling software and are converted into the set of sequences/reads (Das and Vikalo, 2013). These set of sequences/reads represent the set of molecules express in the initial samples. The length of that reads represents the number of cycles performed during the sequencing reaction. The sequence information and the quality of the identification of a base call at each given position of the read are stored as a plain text file in FASTQ format (Cock *et al*., 2010).

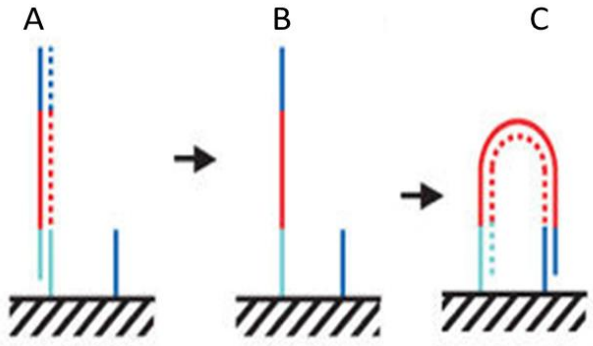


Figure 1.4 Hybridisation and bridge amplification. A. The single strand is hybridised to complementary adapter sequence on the flow cell surface. A new complementary strand (dotted) is synthesised in elongation reaction from the 3´end of the surface-bound adapter; **B.** Denaturation removes the master strand **C.** The adaptor sequence at the 3´end of the newly synthesized strand is hybridized to a new surface bound adapter, resulting a bridge formation and creating a site of synthesis of a new strand (dotted). Cluster formation takes place after multiple cycles of hybridisation, elongation, and denaturation. (Bentle *et al*., 2008).

1.9.4. Software for RNA-Seq data analysis

RNA-Seq studies have a number of objectives. These are commonly used for studies of specific genomic regions expression. These genomic regions include genes, isoforms, exons, splice junctions or novel transcribed regions. To study their expression, the identification of the features in the sequencing library is the prime requirement. Aligning short reads to these features is a challenge because of genome size (Trapnell and Salzberg, 2009). An efficient and accurate software is required for the analysis of data from RNA-Seq experiments. This software must have the ability to handle a large volume of sequencing reads produced from the experiment. Typically, analysis stage consists of the following stages: alignment of the reads to the reference, assembling the alignments into full-length transcripts, quantifying the expression level of each gene/transcript and calculating the differences in expression for all genes in different experimental groups/ conditions.

1.9.4.1. Alignment to the reference

Reads can be aligned to reference which can either be a genome or a transcriptome for quantification and differential expression of genes. The advantage of transcriptome alignment is simplification as no intronic sequences exist. On the other hand, it limits the downstream analysis because of its incompatibility with the identification of novel expressed regions, the study of intronic expression levels and splice junctions (Zhao, 2014). TopHat2 was a splice aware aligner package tool (Trapnell, Pachter and Salzberg, 2009) that generated spliced alignments and it was used at the beginning of RNA-Seq era, whilst Bowtie and BWA (Degner *et al*., 2009) were used for the alignment of short reads contiguously to a reference.

1.9.4.1.1. Alignment to the genome

RNA sequencing reads are aligned to a reference genome by a spliced alignment program HISAT2 (hierarchical indexing for spliced alignment of transcripts 2) that is a fast, sensitive and splice aligner program (Pertea *et al*., 2016). HISAT2 can identify splice sites by mapping RNA-Seq reads to a reference and it needs much less computer memory. HISAT2 is advanced version of the spliced alignment program HISAT (hierarchical indexing for spliced alignment of transcripts) (Kim, Langmead and Salzberg, 2015) and Bowtie 2 (Langmead and Salzberg, 2012). An independent algorithm Bowtie2 was able to align short reads and detect exon-exon junctions without the need for any advance information on the annotation. HISAT2 is a replacement of TopHat2. HISAT2 creates a global index (for the entire reference sequence) and lots of small local indexes (for regions of the reference). Small local indexes not only cover the entire reference but also cover the overlapping boundaries where overlapping boundaries simplify mapping of reads that span over exon-exon junction by covering two indexes. Kim *et al*. (2015) split exon spanning reads into three groups including long-anchored reads, where reads align across two exons with at least 16 bp aligning into one exon; intermediate-anchored reads, where reads align across two exons with 8–15bp aligning in one exon; and short-anchored reads, where reads align across two exons with 1–7bp in to one exon. Each group correlates to a different alignment strategy based on global and local indexes. For example a gene on human chromosome 22 consists of two exons and three introns (Figure 1.5) and after transcription and splicing of the genomic region there are three 100bp error-free reads including a reading mapped to exon, a read that spans over an exon-exon junction with an 8-bp sequence in one of these exons and a read that spans over two exons with 50bp anchored in each exon. Strategies (hierarchical indexing) for the alignment depend on two conditions that reads are at

least 28bp long and these align exactly onto one location. Firstly, global index is used Figure 1.5a: the 28bp long end of reading maps exactly, therefore, the global search is stopped and extended the partial alignment by using the mapped alignment as an index. The remaining 72bp sequence is aligned within one exon, the alignment of this read is completed. With the 2nd type of reading, the 28bp long end maps exactly whilst the read has an 8bp anchor as well, therefore, the alignment is extended until a mismatch occurs (Figure 1.5a), this mismatch appears when the extended alignment reaches an intron.

The local compressed full-text data structure FM index (Ferragina-Manzini index) is used for the alignment of 8bp. In the end, 8bp and 92bp are combined to generate a spliced alignment (Figure 1.5b). The process of alignment of 3rd (Figure 1.5c) read starts as described before where 28bp long end uniquely mapped, alignment extension continues which aligns further 22bp and stops when a mismatch appears. The local FM index is then selected to find 8bp anchor alignment. Once 8bp is found, the HISAT2 extension operation is used to align the rest of the read. If the 8bp prefix maps to too many locations within the local index than HISAT2 would use a longer prefix to reduce the number of potential locations. HISAT2 produce outputs in SAM (Sequence Alignment Map) format.

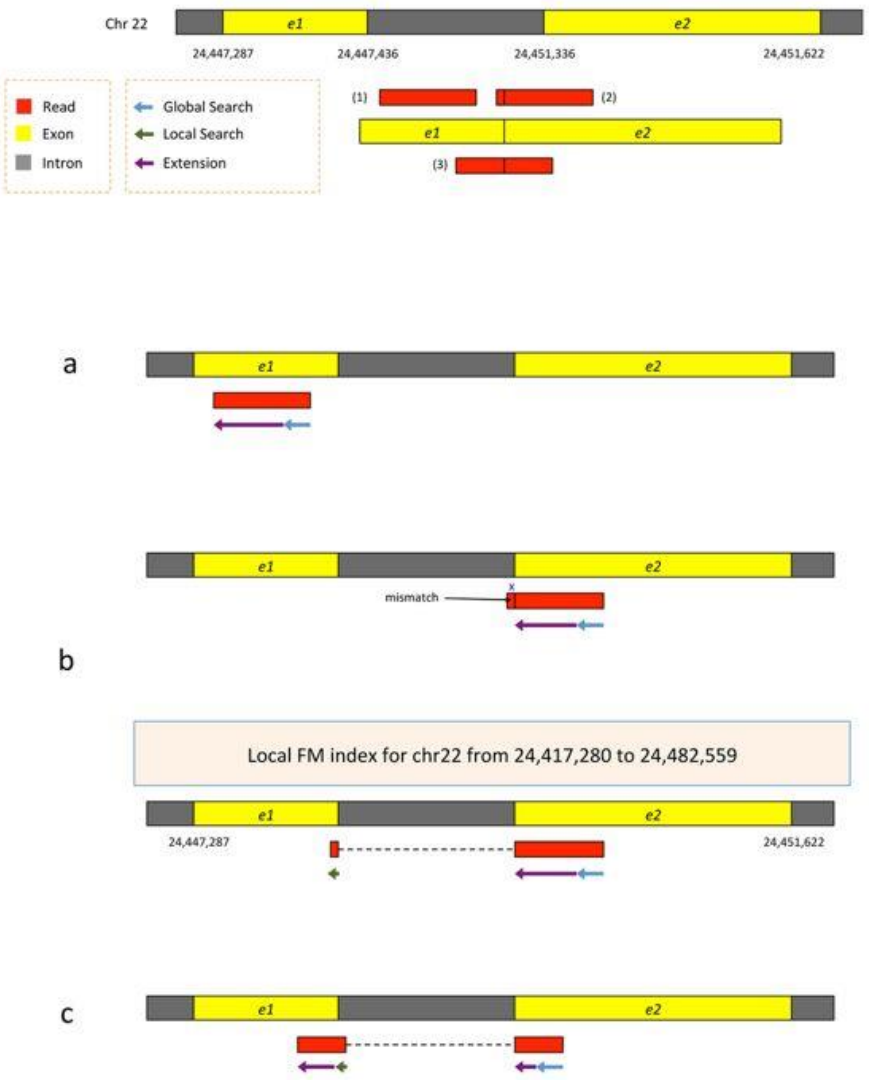


Figure 1. 5 HISAT2. The application of hierarchical indexing for fast and sensitive alignment with three different strategies. (Kim *et al*., 2015).

TopHat2

TopHat2 aligns reads contiguously to the genome with Bowtie (Langmead and Salzberg, 2012). However, to simplify this process, TopHat2 aligns the reads to the transcriptome first and the nonaligned reads will subsequently be mapped to the genome. Initially aligned reads are assembled into exons, which might become connected via spliced alignment. Nonaligned reads and aligned reads with low alignment scores in the initial step, are split into the smaller segments and independently aligned again to get information of possible splice junction when two consecutive segments from the same read do not align contiguously on a given genomic locus. In the next phase, identified splice sites and their flanking sequences are linked into a novel transcriptome to re-align the set of unmapped reads. Each read is processed separately where paired-end data is available and the alignments from genome and transcriptome are evaluated at the end phase by considering other sources of information such as fragment length and orientation of the reads. At the final stage, all the obtained information during the mapping process is reported in the SAM (sequence alignment map) format (Li *et al*., 2009).

1.9.4.2. Assemble the alignments into full-length transcripts

Around 30% of non-coding RNA genes (ncRNAs) and greater than 90% of transcripts produced from multi-exon protein-coding genes are subject to alternative splicing (Wang *et al*., 2008 & Cabili *et al*., 2011). Previous procedures for the assembly of genes de novo as well as coding regions have been replaced with highly accurate method such as StringTie (Pertea *et al*., 2016). These new methods use RNA-Seq data as a starting point. A research study for the reconstruction of transcript found that previous method such as Cufflinks (for the transcript

identification and expression quantification) though identified all exons of a transcript but failed to assemble the exons into complete isoforms (Pertea *et al*., 2015).

For the assembly of alignments into full and partial transcripts**,** producing multiple isoforms, and for the estimation of the expression levels of all genes and transcripts, StringTie was used (Pertea *et al*., 2015). StringTie was based on genome-guided transcriptome assembly and de novo genome assembly concept for the improvement of transcript assembly. Therefore, it used spliced read alignments as well as alignments of contigs that it pre-assembles from reading pairs.

1.9.4.3. Quantification of expression level

Once the alignment assembly and identification of the specific location of reads in the genome and transcriptome is completed, the next step in the analysis involves computing the level of expression for genes and transcripts. Expression level can be quantified based on existing information by counting how many reads align to each of the features i.e. genome and transcriptome.

1.9.4.3.1. Quantification of gene/transcript expression

At the gene level, quantification was achieved previously by the number of counts of the reads covering the locus. Many downstream analysis algorithms such as DESeq (Love, Huber, and Anders, 2014), DEXSeq (Anders, Reyes and Huber, 2012) and the count (Anders, Pyl and Huber, 2015) start the process by counting the number of reads aligned to the features. To avoid the double counts which results in an overestimation of expression level of the reads (multi-mapping reads) that map to the multiple locations in the genome htseq-count discards them, but this may result in a significant loss of information (Turro *et al*., 2011). Another problem

was that individual transcripts complicate the estimation of expression level because of the overlapping of reads with exons which are shared across multiple isoforms of the same gene. In this case, various algorithms are required. Some of them depend on various pools of information to randomly estimated transcript expression levels. But the reads aligned to one of the annotated transcripts within the loci are more important. However, reads (split reads) provide useful information when they span two different exons. This information could be about splice junctions that involve cassette exons such as during a splicing event an inclusion and exclusion of an exon between two other exons to form two distinct protein isoforms. In this case, sequencing from both sides of the initial cDNA fragment in paired-end form is important because it covers larger genomic regions where it is highly probable that a given read pair is aligned across different exons (i.e. spliced reads).

Various tools were available for the estimation of expression levels of the transcripts including MISO (Mixture-of-Isoforms) (Katz *et al*., 2010), Cufflinks (Trapnell *et al*., 2010) and MMSEQ (Minimum Mean Square Error Quantizer) (Turro *et al*., 2011). MISO and Cufflinks require that the reads need to be mapped to a reference genome whilst MMSEQ based on mapping reads to the transcriptome, which limits the scope of the downstream analysis. To overcome this, StringTie performs quantification. StringTie uses a NetFlow algorithm to perform both functions simultaneously such as assembly and quantification for the highly expressed transcripts and separate reads associated with that transcript and leads to the repetition of the process until all the reads are used. With the use of an algorithm, StringTie runs faster and use less memory. By using the reference annotation file StringTie can construct assembly for low abundance genes i.e. novel genes and transcripts and able to reflect the expression of rare or novel genes and alternative

splicing isoforms. StringTie-merger is able to merge the transcript assemblies obtained from G-gap negative and G-gap positive samples into a consolidated annotation set. This function was previously performed with the Cufflinks-merger. Full length of the transcript can be restored especially those transcripts assembled with low coverage with this merger step. StringTie store the transcript assembly as GTF (Gene Transfer format) file format.

1.9.4.3.2. Computing the differences in gene expression levels

Ballgown helps the visualisation of transcript assembly on a gene-by-gene basis, estimation of abundance for exons, introns, transcripts or genes and the performance of linear model-based differential expression analysis. Ballgown is based on the output of StringTie merger. It collects all the transcripts and abundances and converts them into groups based on experimental condition. Therefore, it is thus able to find which genes and transcripts were differentially expressed between conditions. Ballgown software consists of plotting tools as part of the R/Bioconductor package that helps in the visualisation of results (Pertea *et al*., 2016). Ballgown acts as a link between the assembly of the transcriptome and the fast, flexible differential expression analysis. Ballgown software links the HISAT2 and StringTie software to R/Bioconductor package.

1.9.5. Normalisation of expression

Regardless of the quantification approach used it is necessary to ensure that the results are comparable across different libraries/features within a sample. Normalisation is important because the number of reads/counts not only are proportional to the expression levels of the feature of interest, but this also depend on the length of the feature such as base pairs and the sequencing depth of the

experiment. Sequencing depth represents the total number of mapped reads in a sample that are mapped to any regions within a genome. Different length of the feature produces within-sample bias (Hansen, Irizarry and Wu, 2012; Risso *et al*., 2011 and Pickrell *et al*., 2010) because long features produce more reads (Garber *et al*., 2011 and Oshlack and Wakefield, 2009). Whilst sequencing depth is a source of between- sample bias (Mortazavi *et al*., 2008; Marioni *et al*., 2008). This normalisation concludes the meaningful expression values between the same or different features in different or the same samples respectively under different biological conditions or deals with biases within-sample and between-samples (Risso *et al*., 2014) as count data obtained from RNA-Seq are subject to certain biases (Garber *et al*., 2011; Mortazavi *et al*., 2008; Nagalakshmi *et al*., 2008 and Wang, Gerstein and Snyder, 2009).

Commonly used normalization strategies in RNA-Seq are the reads per kilobase of transcripts per million reads mapped (RPKM) (Mortazavi *et al*., 2008) and fragments per kilobase of transcripts per million reads mapped (FPKM) (Trapnell *et al*., 2010). RPKM is used when the mapped reads are single end whilst FPKM is used when the mapped reads are paired-end (Mortazavi *et al*., 2008). FPKM showed that if one million fragments are sequenced, the FPKM value for a certain feature represents the expected number of fragments identified per thousand bases in that feature. This method assumes that across all libraries total RNA levels are similar. Ballgown used FPKM for normalisation.

1.9.6. Challenges for RNA sequencing technology

RNA-Seq technology has some challenges such as during library preparation, the random hexamer priming step has some specificity to sequence composition

leading to the preferential conversion of some fragments to cDNA (Hansen, Brenner and Dudoit, 2010). Cufflinks (Trapnell *et al*., 2010), MMSEQ (Turro *et al*., 2011) are some examples from several algorithms that cope with the biases derived from the random hexamer amplification step. Another challenge is that fragments with higher or lower GC content undergo differential amplification during PCR (Benjamini and Speed, 2012), and during sequencing step wrong base calls is possibly due to the failure to stop the elongation phase or to remove the fluorescent dye (Metzker, 2010). To overcome amplification bias alternative library preparation methods are recommended where random barcodes i.e. molecular identifiers quantify the absolute number of molecules (Shiroguchi *et al*., 2012) whilst to account for wrong base call bias some downstream analysis algorithms are used that include information on the quality of the identification of a base call at each given position of the read, as reported by the Phred score (Kim *et al*., 2013).

1.10. The Telomere

1.10.1. Telomere biology

Telomeres are nucleoprotein positioned at the end of chromosomes. They themselves do not encode proteins but perform various biological functions, including maintenance of chromosomal structure and stability, and eventually the determination of the lifespan of cells (Lu *et al*., 2013). Telomere plays an important role in the process of the cell's ageing alongside the protection of chromosome ends from degradation or fusion (Blackburn, Greider, and Szostak, 2006). They contain stretches of repetitive DNA sequence of six nucleotide bases (TTAGGG) ending in a short single-stranded G-rich overhang, thus the 5′ to 3′ strand of double-stranded telomeric DNA is G rich. Figure 1.6 illustrates the structure and position of the telomeres.

In humans, the average telomere length is approximately 2-20 kilobase pairs. The double- and single-stranded telomeric DNA is associated with the shelterin complex. The shelterin complex consists of six proteins which help in the organisation and stabilisation of telomeric DNA into a loop structure (Diotti and Loayza, 2011). Telomeres are functionally active when these proteins interact with telomeric DNA sequences. Two telomeric sequence-specific binding proteins among others include repeat-binding factor 1 and 2 (TRF1 and TRF2). TRF2 protects the telomeric structure in vivo by the formation of a T-loop. Without T-loop formation telomeres are prone to non-homologous end-joining and fusion (Blackburn, 2001 and De Boeck *et al*., 2009).

TRF1 binds to double-stranded DNA and blocks replication by DNA polymerases (Smucker and Turchi, 2001). DNA polymerases will not replicate chromosomes completely because one primer presents on each daughter DNA strand. 5'-3' exonuclease activity removes the last primers, but gaps remain unfilled by DNA polymerases because of the absence of 3'-OH to which a nucleotide can be added. Because of this, the telomere, which is a small region at the end is left unreplicated due to the end-replication problem and absence of telomerase expression in somatic cells (Wai, 2004). The end of the replication process results in chromosome shortening with every round of cell division. After a period, length reduces below the Hayflick limit leading to the induction of replicative senescence (Sozou and Kirkwood, 2001) and elimination of the telomeres. This results in apoptosis (Buijsse *et al*., 2007) and triggers associated ageing disorders. For this reason, telomere length is considered as the life-clock of a cell.

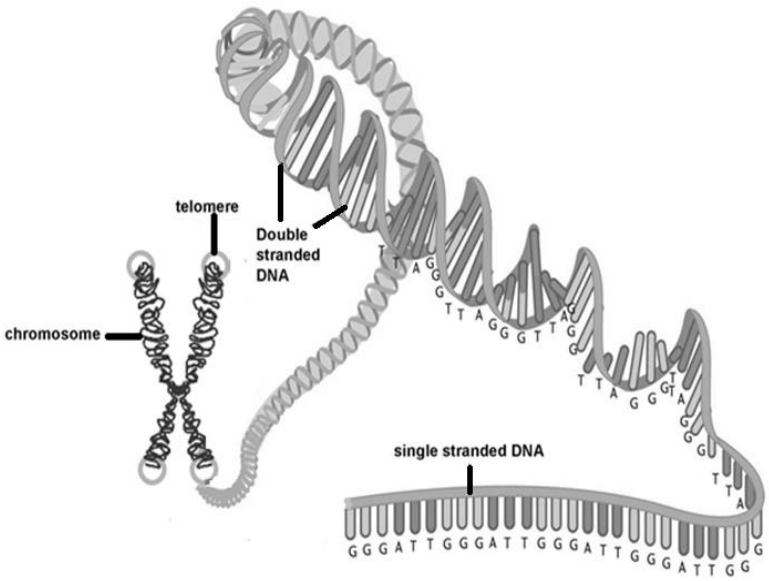


Figure 1. 6 Position and structure of telomere on the chromosome. (Zhu, Belcher and van, 2011).

1.10.1.1. End-replication problem

End replication problem leads to the shortening of telomere length with each round of cell division. Therefore, telomere length is proposed as a biomarker of cell ageing (Kaszubowska, 2008). All eukaryotic chromosomes consist of double-stranded DNA molecules: helicase enzymes unwind these double strands, where each strand of the double-stranded DNA is used as the template for the new strand synthesis (Nandakumar, Pandey and Patel, 2015). DNA polymerase synthesises continuously a new DNA strand in 5' to 3' direction. This strand is called the leading strand. To initiate the replication process, DNA polymerase requires primer to provide a 3' hydroxyl group. On the opposite side (3' to 5' direction) is the lagging strand. The DNA polymerase can only elongate from a 3' hydroxyl group. Because of this, the lagging strand is synthesised in short fragments termed as Okazaki fragments (Lei *et al*., 2005 and Wang *et al*., 2007). The leading strand can be synthesised 5'-3' completely and continuously to the end of the template. However, at the end of each round of DNA replication of the lagging strand, the last DNA primer is removed, leaving no 3' hydroxyl for DNA polymerase to initiate DNA synthesis. This leaves a gap and is unable to completely replicate the end of the lagging strand that is up to 500 nucleotides (mean value of 250 nucleotides) at the 5' end of the chromosome (Xin *et al*., 2007 and Yu *et al*., 2008). James Watson in 1972 identified this as the "end replication problem" (Herbert, 2011). The loss of terminal sequences continuously into the cell leads to cell death and cellular senescence. Therefore, the "end replication problem" provides the reason for the ability of the cultured cells to divide up to a certain number of generations (Palm and Lange, 2008) (also known as the Hayflick limit).

1.10.2. Relevance to disease

The attrition of telomeres is not only the distinctive feature of mammalian ageing but also involves numerous metabolic dysfunctions. A previous study suggests that shorter telomeres length is a risk factor for age-related pathologies such as insulin resistance, overt diabetes mellitus, cardiovascular disease or neurological disorders such as Alzheimer's disease (Kirchner *et al*., 2017). It is not clear whether the telomere shortening is a heritable feature, representing the cause of such chronic diseases or it is reflecting a disease consequence. But in either way, numerous cellular functions are compromised because of the loss of chromosomal DNA end-protection.

End replication problem (Maestroni, Matmati and Coulon, 2017) and oxidative stress (Von, 2002) have an impact on telomere length. The increased cellular concentration of free radicals can damage proteins, lipids and nucleic acids and their concentration increases because of cellular stress which is induced by chronic and autoimmune diseases as well as environmental factors (Rains and Jain, 2011). Telomere tandem repeats are guanine-rich which makes it more sensitive to react with reactive oxygen species (Fleming and Burrows, 2013) and leading to DNA breaks. In diabetes, oxidative stress is increased in leukocytes as well as in pancreatic β-cells. Oxidative stress in pancreatic β-cells could result in shortening of β-cell telomeres and subsequent dysfunction of insulin secretion (Tamura *et al*., 2016). Oxidative stress can also stimulate inflammation, which leads to an increased proliferation of immune cells. Increased proliferation results in an increased rate of telomere shortening (Shalev, 2012). Inflammation appears in chronic and autoimmune diseases, such as cardiovascular diseases, rheumatoid arthritis, type 1 and 2 diabetes (Shalev, 2012; Georgin *et al*, 2010). Studies have reported that a

longer telomere is associated with better health status, protection from age-related diseases and lower mortality risk (Shalev, 2012).

The proposed idea of the relationship between telomere length and the ageing process provided hindsight for the first association with telomeres and age-related diseases (Armanios, 2013). Cawthon *et al*., 2003 provided strong support that telomere length was linked to mortality. The study demonstrated that mortality rates from heart disease and infectious diseases were higher in elderly participants with telomere shortening as compared to participants with longer telomeres. This suggested that short telomeres could affect human health. Several studies showed that telomere length is apparently shorter in type 2 diabetes (T2D) (Adaikalakoteswari *et al*., 2007, Olivieri *et al*., 2009, Smart *et al*., 2009 and., Sampson *et al*., 2006) as T2D is considered an age-related disease in terms of its development and progression. Uziel *et al*. (2007) also reported shorter telomeres in type 1 diabetes patients. Although there is paucity of data, inflammation leads to shortened telomere length in cultured proliferative cells (Von, 2000). Thus, in theory, replications, oxidative stress and inflammation have an impact on telomere length. Previous studies demonstrated that patients with diabetes mellitus have shorter white blood cell telomeres than normal individuals (Sampson *et al*., 2006 and Zee *et al*., 2010).

1.10.2.1. Telomere length and type 2 diabetes

Type 2 diabetes (T2D) is characterised by high levels of blood glucose (hyperglycaemia) due to the absence or reduced production of insulin by pancreatic beta-cells and resistance to insulin. Beta cell failure is caused by increased levels of both glucose and free fatty acids (glucolipotoxicity) (Poitout *et al*., 2010), and is

often observed as an important factor in the initiation and progression of T2D. Increased levels of glucose increase mitochondrial metabolism which leads to increased production of reactive oxygen species (ROS) (Muoio and Newgard, 2008). Age is clearly a strong risk factor for T2D in terms of its onset and progression because the replication capacity of β-cells decreases with ageing (Meier *et al.,* 2008 and Mizukami *et al.,* 2014), potentially as a result of telomere shortening. Premature cell senescence due to chronic exposure to hyperglycaemia and free fatty acids over time has been proposed as an important cause and consequence of T2D and its complications (Sampson and Hughes, 2006). The β cell senescence and apoptosis leads to the development of T2D. Given that complications of the disease are consequent to senescent phenotypes in kidneys, the eyes, the peripheral nerves, and the vascular vessels, it can be postulated that telomere shortening may play a mediating role in inducing senescence of beta cells and the onset of the disease by affecting pancreatic β-cell metabolism directly (Guo *et al*., 2011), providing a link between telomere length and T2D.

1.10.2.2. Telomere length and type 1 diabetes

Telomeres in WBCs may be associated with the pathophysiology of T1D. T1D (IDDM) is mediated via autoimmune destruction of insulin-producing β cells by subsets of WBCs, primarily subsets of T-cells that express CD4 and CD8 surface markers (Graham *et al*., 2012). Wang *et al.* (2016) determined that the WBC telomere length in patients with Type 1 diabetes was shorter than in healthy individuals.

1.11. Research gaps in previous studies

HbA_{1c} estimation is widely recognised for the monitoring and control of glycaemic level. American Diabetes Association (2010) recommended the use of HbA_{1c} for diagnosing diabetes. Within-individuals variation in HbA_{1c} values are consistent in non-diabetics (Rohlfing *et al.,* 2002), in contrast to between-individuals (Cohen and Smith, 2008; Saaddine *et al.,* 2002). Macdonald *et al.* (2008) developed a methodology for measuring G-gap (a difference between glycated haemoglobin levels and other measures of average glycaemia such as fructosamine). Nayak *et al.* (2011) demonstrated that a glycation gap is found to be consistent in with-in diabetic individuals and also reported that G-gap is associated with the diabetic complications like retinopathy and nephropathy (Nayak *et al.,* 2013).

Among multiple factors that were involved in the variation of non-enzymatic haemoglobin (Hb) glycation (Leslie and Cohen, 2009; Cohen and Lindsell, 2012), this study focused on deglycation (fructosamine-3-kinase (FN3K)-catalyzed pathway). No previous study has explored the correlation between Single Nucleotide Polymorphisms (SNPs) of FN3K and Glycation gap status. Previous studies neglected to look at pathway/isoforms which are linked with the G-gap status at the transcriptome level. No study had been conducted to study the sRAGE level within consistent G-gap diabetic patients. There are no previous studies that have examined the telomere length in relation to the G-gap.

1.12. Research Questions, Aims, and Objectives

1.12.1. Research questions

Following are the questions that developed on the basis of research gaps in the previous studies;

1. Are SNPs in *FN3K* gene responsible for the underlying mechanism of the discrepancies in FN3K activity among individuals?

2. Are the sRAGE levels consistent with the AGE level in diabetic individuals according to their G-gap values?

3. Is FN3K the only enzyme involved in deglycation of glycated haemoglobin in erythrocytes or Are there other DEGs and differentially expressed transcripts and their associated pathways/biological processes involved in deglycation?

4. Do the different G-gap values influence the telomere length?

5. Are FN3K SNPs associated with the sRAGE in patients with different G-gap values?

6. Is Ferroportin1 SNP responsible for the variation of G-gap values in diabetic patients?

7. Can transcriptomic analysis provide novel information on the mechanisms causing the G-gap?

1.12.2. Research aims and objectives

● A previous study demonstrated shorter telomeres in T1D and T2D (Tamura *et al*., 2016). The first aim of this thesis was to test the hypothesis that the average telomere length ratio (T/S) is altered in G-gap negative and G-gap positive patients with diabetes.

● The second aim was to explore the relationship between specific FN3K polymorphisms (SNPs rs1056534, rs1046896, rs3848403) and G-gap in diabetic patients. The iron overload plays an important role in diabetes mellitus in sub-Saharan African (Katchunga *et al.,*2013) and this is linked with FN3K activity as well as with the glycation gap. In this thesis the Q248H ferroportin polymorphism A/C (rs11568350) (linked with FN3K activity) was studied in G-gap negative and G-gap positive patients.

● The third aim was to evaluate circulating levels of receptor for AGEs (RAGE) in G-gap groups. The association of rs3848403 SNPs of the *FN3K* gene was analysed with G-gap negative and G-gap positive patients. In this thesis, the correlation between FN3K SNPs and the glycation gap in patients with diabetes was examined.

● The final aim was to investigate DEGs and differentially expressed transcripts and alternatively spliced transcript variants in G-gap groups via RNA-Sequencing (RNA-Seq) technology.

Chapter 2

Material and Methods

2.1. Methods

2.1.1. Ethical committee approval

The UK National Health Service Research Ethical Committee (ref. no. 11/WM/0224) approved the genetic and transcriptomic studies. The study was also approved by the University Life Sciences Ethics Committee.

2.1.2. Selection of patients and data collection

Selection of appropriate patients for the study was critical after ethical approval. All simultaneous HbA_{1c} and fructosamine values measured over 10 years were identified from laboratory records at New Cross Hospital Wolverhampton, yielding 31,119 results amongst 12,253 people with diabetes.

From amongst those with at least 3 estimations, whose average G-gap was either <= -0.5 or >= +0.5, who were confirmed as having a long-term over 10 years consistent negative or positive G-gap status and who were alive (n=1663), 150 each of G-gap negative or positive subjects were invited by post to participate according to the ethical committee approved research protocol of whom in total 184 attended and had complete data. Pregnant women, participants with creatinine>200 µmol/L, with known haemoglobinopathy, or those with an unusual electrophoretic pattern on HbA_{1c} testing were excluded from this study. No other demographic or clinical selection criteria were used.

The 184 selected patients had 2 pairs of simultaneous estimations of HbA_{1c} and fructosamine and 3 calculated G-gaps that were consistently greater than +0.5% or lower than -0.5% in a positive or negative direction respectively.

2.1.2.1. Blood glucose analysis

All subjects were confirmed with diabetes using standard diagnostic tests as carried out in pathology lab of New Cross Hospital.

2.1.3. Blood collection

Venous blood was collected from all volunteers (refer to section 2.1.2) by venepuncture into heparin-based S-Monovette system (9 ml, Sarstedt), EDTA coated blood collection tubes (sterile 9 ml, Vacutainer, Becton Dickinson, UK) and Tempus Blood RNA tubes (Thermo Fisher Scientific). Plasma was separated by centrifugation (2000g for 10 minutes at 4°C), aliquoted into 200µl aliquots and stored at -80°C until needed.

2.1.3.1. Study participants

A total of 184 participants, distinguished according to their G-gap status (102 negatives and 82 positive groups), age 60.4±13 and 65±10 and with a body mass index (BMI) of 28.5±5.1 and 36±7.2 kg/m^2 respectively), were invited from the local community in Wolverhampton, UK (Table 2.1). Participants in this study were obese (BMI>30). Participants suffered from T1D and T2D. However, most of them had T2D (63% in negative G-gap and 91% in positive G-gap). After ethical approval, the written informed consent was given by all participants. Every patient was allocated a unique code, and volunteer specific data were recorded in the subject information sheet.

	Negative G-gap	Positive G-gap
Number	102	82
Age (years)	60±13	65±10
Gender (% male)	77%	61%
Ethnicity (% White, Asian, Black)	70%, 18%, 12%	73%, 23%, 4%
Body Mass Index (kg/m^2)	30.2±5.2	36.0±7.2

Table 2.1 The baseline characteristcs in the cohort in each glycation gap group

2.1.3.2. Analysis methods

Diabetes Control and Complications Trial (DCCT) presented HbA_{1c} aligned values. These were used in the analysis because HbA_{1c} values in mmol/mol were not available before 1 June 2009 as recommended by the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC). HbA_{1c} was measured by High-performance liquid chromatography (Tosoh G7 analyzer, Bioscience Ltd, U.K.). In the U.K. National External Quality Assurance Scheme (NEQAS), the performance scores as A (accuracy) score <100 and B (bias) score <2% reported by Nayak *et al*. (2013). These scores were within the acceptable range (maximum limits: A score <200 and B score less than 7.5%) for glycated haemoglobins. Between-batch the coefficient of variation was found to be 1.8 and 1.4 for an HbA_{1c} of 5.7% (39 mmol/mol) and 9.5% (80 mmol/mol), respectively. Fructosamine was estimated by using a Cobas kit (based on Nitrotetrazolium-blue reduction method) and coefficient of variation of between batches was 3.1 (at a level of 263 mmol/L) and 2.2 (at a level of 518 mmol/L) (Nayak *et al*., 2013).

2.1.3.3. Predicted HbA1c (FHbA1c) and the G-gap calculations

G-gap was determined by the method of Nayak *et al*. (2013), who calculated the predicted HbA_{1c} ($FHbA_{1c}$) from simultaneously measured $HbA_{1c,}$ and fructosamine values from fructosamine estimation, where fructosamine was converted to an HbA_{1c} equivalent measure (F_HbA_{1c}) using the Standardised Normal Deviate approach:$\mathrm{FHbA1c} = [\{(\text{actual Fructosamine value} - \text{mean Fructosamine}) / \text{SD Fructosamine}\} \times \text{SD HbA1c}] + \text{mean HbA1c}$ where SD = standard deviation**.** G-gap was calculated when the predicted HbA1c ($FHbA_{1c}$) was subtracted to the true HbA_{1c} $(G - gap = HbA1c - FHbA1c)$. Importantly correlation/regression methods were not used to derive the $FHbA_{1c}$ from HbA_{1c}. A negative G-gap is indicated when the true HbA_{1c} reading is lower than the $FHbA_{1c}$, and a positive G-gap denoted when the true HbA_{1c} reading exceeds the value predicted by fructosamine. The consistency of G-gap status was determined by the positivity of the product of the minimum and maximum G-gap value. To identify consistent G-gap among those with a second paired HbA_{1c}-fructosamine estimation, the product of two G-gaps was calculated. The G-gap would be consistent when the product of two G-gaps is positive (positive × positive = positive; negative × negative = positive) whilst any discrepancy in two paired readings of G-gaps would yield a negative G-gap product (negative × positive = negative) (Nayak *et al*., 2013).

2.1.3.4. Classification of the G-gap and statistical analysis

For simplification, the G-gap was categorised as negative when HbA_{1c} represented in % was less than or equal to 0.5 (i.e., <= -0.5), or positive when greater than or equal to 0.5 (i.e., >= +0.5).

In this study, the sRAGE level, SNPs genotyping, telomere length, and transcriptome level were compared between negative and positive G-gap patients. To analyse the data statistically t-tests and Chi-square tests were used, and charts drawn to represent these associations. All statistical tests were considered statistically significant if P value is < 0.05.

2.1.4. Study 1: Plasma sample analysis for sRAGE level

The plasma samples from G-gap negative and G-gap positive cohorts were prepared by collecting blood in plasma separator tubes and centrifuging at 2000g for 10 min at 4°C. To protect from repeated freeze-thaw cycles, the plasma samples were stored at -20°C after aliquoting and were then assayed for the levels of circulating soluble Receptor-AGE using the sRAGE immunoassay kit (R&D System, Minneapolis, MN 55413, USA) according to the instruction of the manufacturer. This assay is quantitative and is based on a sandwich enzyme-linked immunosorbent assay (ELISA) technique. The coefficient of variation (CV%) value is used to reflect the precision of an assay or how repeatable this assay is for producing these results (refer section 3.4.5). Full detail of assay parameters is provided in manufacturer`s leaflet.

2.1.4.1. Enzyme-linked immunosorbent assay (ELISA)

Plasma level of circulating soluble Receptors for AGE (sRAGE) for 146 patients was estimated by the use of a commercial ELISA kit (DuoSet® ELISA; R&D Systems, Minneapolis, MN 55413, USA) according to the manufacturer's instructions. In summary, the working concentration of the appropriate capture antibody was used to coat the 96-well microtiter plates overnight at room temperature and then next day washing was performed using wash buffer (0.05% Tween® 20 in PBS, pH 7.2 - 7.4. After that reagent diluent contains 1% BSA in PBS at pH 7.2 - 7.4 was used to block the plates for one hour at room temperature. Seven standard concentrations were prepared including 2000pg/ml, 1000pg/ml, 500pg/ml, 250pg/ml, 125pg/ml, 62.5pg/ml and 0pg/ml by using standards provided in the kit. Dilutions of plasma samples (1:4) in reagent diluent were carried out. Standards and diluted samples (100µl) were added to respective wells of the plates. After 2 hours incubation period at room temperature, the washing of microtiter plates was performed 3 times with the wash buffer (0.05% Tween® 20 in PBS) and incubated with the working concentration of 100µl biotinylated goat anti-human sRAGE detection antibody for 2 hours. The washing of plates was performed again, and streptavidin conjugated to horseradish peroxidase was added (100µl) to each well for 20 minutes. After washing, substrate solution carrying equal parts of reagent A (H_2O_2) and reagent B [tetramethylbenzidine (TMB)] was added into each well to quantify the HRP by the colourimetric reaction, and after 20 minutes the colour formation was terminated by adding 2N sulphuric acid stop solution (R&D Systems, # DY994), the colour that develops is directly proportional to the amount of antigen present in the initial step.

All ELISA reactions were in duplicate in neighbouring wells of the same 96 well plate for the assay of sRAGE within plasma samples. The mean was calculated for the

duplicate ELISA results. Each plate contained seven standard concentrations in duplicate. These standard concentrations were prepared by the serial dilutions method and represented the known concentration of the analyte that was present in the kit. The optical density of these wells was used to draw a standard curve. Optical density values of all wells were measured at 450nm on a microplate plate reader (Thermo Labsystems, Multiskan Ascent).

2.1.5. Study 2: SNPs Analysis

2.1.5.1. DNA extraction

Genomic DNA was isolated from aliquots (200µl) of whole blood of 125 patients with diabetes with consistent negative and positive G-gap using the QIAamp DNA Blood Mini Kit (QIAGEN, # 51104). The QIAamp based DNA extraction was performed according to the manufacturer`s instructions using QIAamp Mini spin columns in a standard microcentrifuge.

In summary, 20µl proteinase K (100mg/ml, Qiagen) was pipetted into the bottom of a 1.5ml microcentrifuge tube and then 200µl blood sample was added into it. In the next step 200µl Buffer AL (contains guanidine hydrochloride) was mixed to lyse the cells. The incubation time was 10 minutes at 56°C. The samples were briefly centrifuged to remove drops from inside the lid of microcentrifuge. Two hundred microlitres (200 µl) of 100% ethanol was added to the samples and mixed by pulse-vortexing for 15 seconds. The mixture was then carefully added to the QIAamp Mini Spin column. The caps were closed, and the samples centrifuged 8000 rpm for 1 min. Buffer AW1 (comprising of guanidine hydrochloride and chaotropic salts) (500 µl) was added to the QIAamp Mini spin columns, the caps were closed, and the

samples were centrifuged at 8000rpm for 1 minute. The QIAamp Mini spin column was then placed in a clean collection tube. Then 500 µl buffer AW2 (contains sodium azide) was added and centrifuged at 14000 rpm for 3 minutes. The samples were then centrifuged for an additional 1 minute at 14000 rpm to remove possible AW2 carryover. Buffer AW1 and AW2 were washing buffers that ensured complete removal of any residual contaminants without affecting DNA binding and significantly improving the purity of eluted DNA. The QIAamp Mini spin column was then placed in a clean 1.5 ml microcentrifuge tube and 100µl buffer AE (contains 10 mM Tris·Cl; 0.5 mM EDTA) was added and the sample was incubated for 5 minutes at room temperature before it was centrifuged at 8000 rpm for 1 minute. After centrifugation the concentration of DNA, eluted in Buffer AE was checked by Thermo Scientific 2000c/2000 nanodrop spectrophotometer; DNA purity was assessed through 260/280 ratio. The procedure involved was to first clean the UV spectrophotometer and then blanked the spectrophotometer by putting 1µl of elution fluid on the nanodrop window and then placing 1µl of samples to be quantified on the nanodrop window and each sample was quantified three times on the nanodrop. Nanodrop calculated the DNA concentration based on absorbance values obtained. Extracted genomic DNA was kept at -20°C until its use for the study of SNPs and telomere length.

2.1.5.2. DNA Amplification

2.1.5.2.1. Real-time qPCR

TaqMan (used for SNPs study) and SYBR-Green (used for telomere study) assays are two commonly used qPCR assays. They use TaqMan assay primers and a fluorescent probe that are complementary to template bind to the respective sequence while in SYBR-Green assay primers and a fluorescent dye sandwich between dsDNA (Figure 2.1). Despite the method used, targeting specific DNA sequences that are unique can provide a sensitive and specific way to test for the presence of SNPs and telomeric DNA. Primers (short synthetic oligo sequences) bind specifically to a DNA strand and act as an initiation site for DNA polymerase to synthesise a new DNA strand.

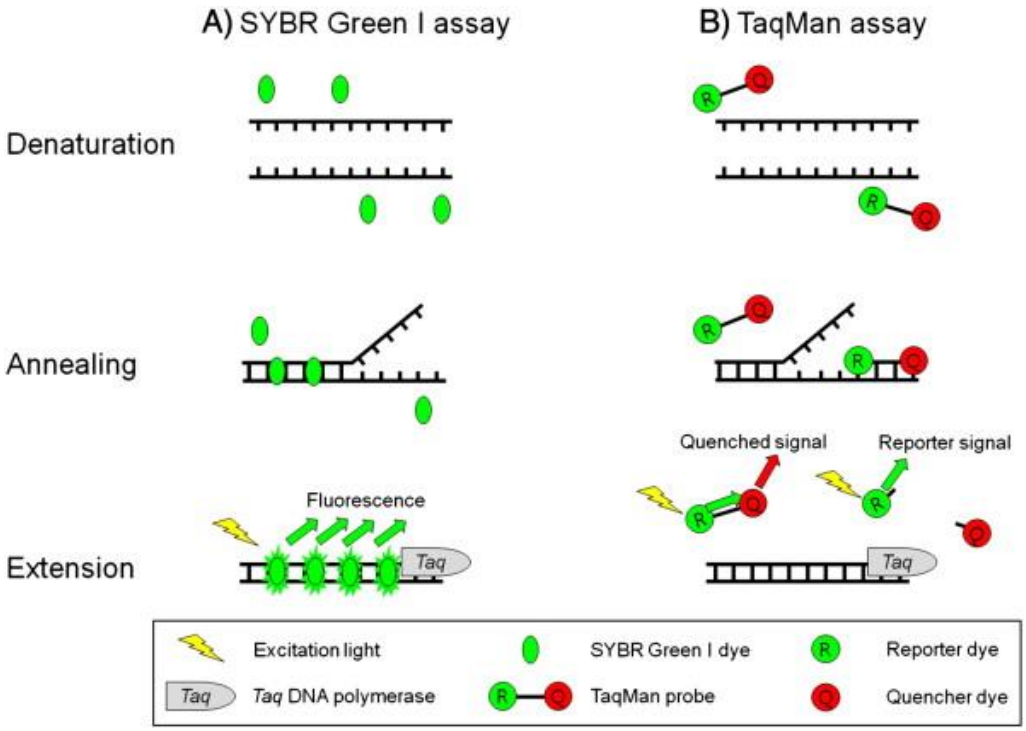


Figure 2. 1 SYBR-Green Vs TaqMan qPCR assays. (Kim *et al*., 2013).

2.1.5.2.2. Single nucleotide polymorphism (SNPs) assay

In this thesis, SNPs were detected using the TaqMan SNP Genotyping Assays technology. The SNP genotyping was carried out using the Stratagene Mx3000PTM, PikoReal and ABI 7500 FAST Real-Time PCR System.

2.1.5.2.2.1. TaqMan SNP genotyping assay

The TaqMan SNP genotyping assay is a single-tube PCR assay that utilises the 5' nuclease chemistry for amplification and detection of specific polymorphisms in purified genomic DNA samples. Specific forward and reverse primers and two allele-specific oligonucleotide probes (designed to target the polymorphism) were used for the amplification of wild-type SNP Allele “A” and the mutated Allele “B”. Allele-specific oligonucleotide probes were labelled with a fluorescent reporter dye (VIC specific for allele “A” and FAM specific for allele “B”) attached to its 5′ end and a non-fluorescent quencher (NFQ) at its 3′ end (Applied Biosystems 2010a) (Figure 2.2). During real-time PCR amplification, each labelled probe binds specifically to one of the two alleles of the SNP of interest with different affinity. Once the amplification process starts the Taq polymerase enzyme cuts the bound probe and a fluorescent signal is generated, the strength of which is dependent on the accumulation of PCR product. Fluorescent signals were measured and interpreted automatically by using ABI 7500 FAST Real-Time PCR system. Homozygosity for Allele “A” and homozygosity for Allele “B” will be interpreted when only VIC dye fluorescence and only FAM dye fluorescence signals appeared respectively. Allele “A”/Allele “B” heterozygosity was interpreted by the presence of both fluorescence signals (VIC and FAM). This technique was suitable for the discrimination of alleles that differed by a single base change: this technique was robust, accurate and cost-effective.

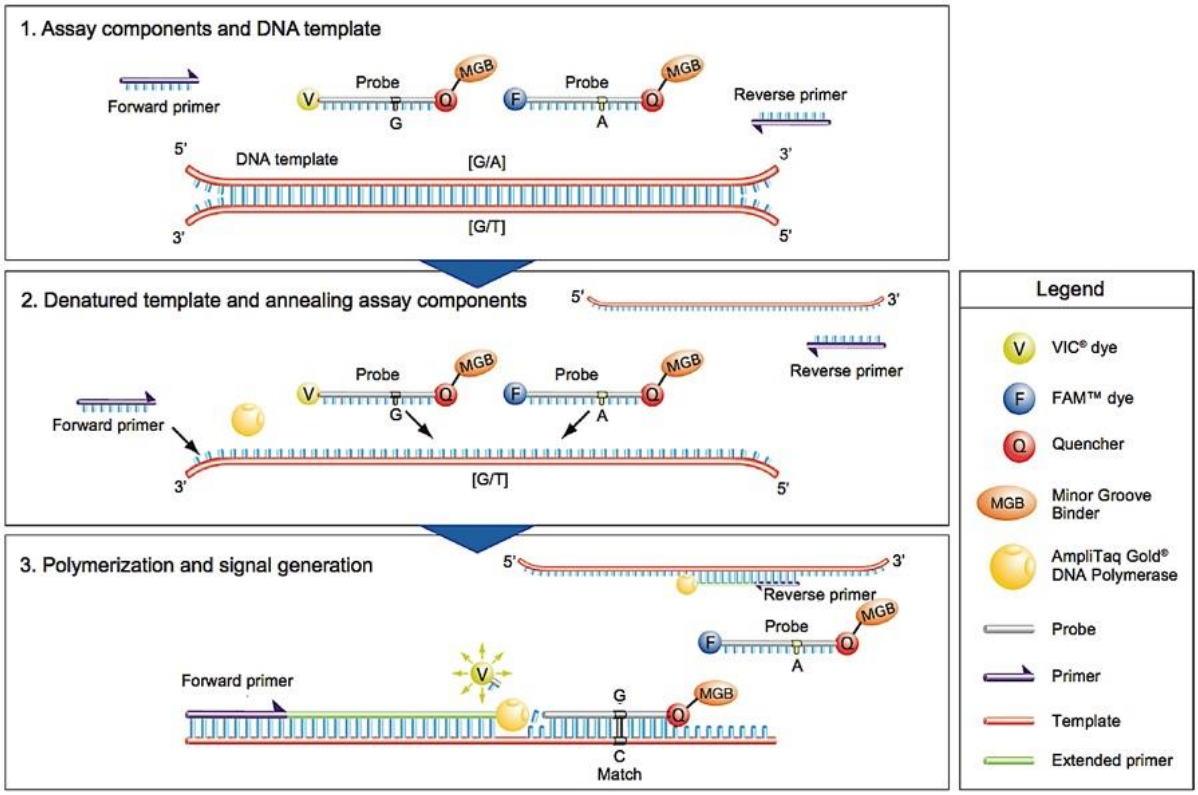


Figure 2. 2 Allelic discrimination is determined by the selective annealing of TaqMan MGB probes. (Applied Biosystems, 2010a).

2.1.5.2.3. Preparation of genomic DNA

DNA samples of 15 ng/µl concentrations were prepared in nuclease-free water. All samples should preferably have the same DNA concentration for genotyping so that a wide range of fluorescence activity within a genotype group would be avoided. The occurrence of SNPs C/G (rs1056534), C/T (rs1046896) and C/T (rs3848403) in FN3K gene and A/C (rs11568350) in the *SLC40A1* gene were analysed using the TaqMan Assay (ThermoFisher Scientific) with provided primers and probes by real time qPCR.

2.1.5.2.3.1. Sample preparation

TaqMan Genotyping Assay mix was supplied at 40x concentration. The assay was diluted to a 20x working concentration by adding one volume of PCR grade water (Thermofisher). For one well, a basic volume was 24µl containing 12.5µl 2× TaqMan universal PCR master mix (contains AmpliTaq Gold DNA Polymerase, deoxyribonucleotide triphosphate (dNTPs) and ROX Passive Reference), 1.25µl 20× SNP assay mix, and 10.25µL of nuclease-free water. The reaction mix was added into each well of a reaction plate. Then, 1µl of 15 ng/µl of the individual DNA sample was added into the bottom of wells, changing the tip between every DNA sample. Two wells in each plate were left free from template DNA as negative controls: for these wells, 1µl of water was added in place of DNA. The 8-strip cap seals were used to cover the plate and then centrifuged at 1,500 rpm for 1 min to collect reagents at the bottom of the wells and to remove any air bubbles. The plate was loaded into a Real-time PCR machine. For SNPs the thermal cycling conditions were 10 minutes at 95°C, followed by 40 cycles of 92°C for 15 seconds and 60°C for 60 seconds. The 92°C reads provided the cycle threshold (Ct) values for the amplification of the target sequence. The reagents used for the assay are summarised in Table 2.1. Results were transferred into Microsoft Excel spreadsheet for analysis. The 5' nuclease activity may produce false positive amplifications because of the repetitive nature and high throughput (Drosten, Seifried and Roth, 2001), but in general TaqMan® genotyping assays are considered to be highly accurate (Martino, Mancuso and Rossi, 2010).

A) Reagents used for the TaqMan genotyping assay for SNPs		
Reagents	**Concentration**	**Volume**
TaqMan genotyping assay	20x	1.25 µl
TaqMan Universal PCR MasterMix	2x	12.5 µl
Nuclease-free water	NA	10.25 µl
DNA template	15ng/µl	1.0µl
Total		25 µl

Table 2.1 Table showing the composition of the mixture for TaqMan genotyping assay.

2.1.6. Study 3: Telomere length (TL)

2.1.6.1. Extraction of DNA

DNA was extracted as described in section 2.1.5.1.

2.1.6.2. Real-time polymerase chain assay for telomere length (TL)

Average telomere length was estimated from total genomic DNA (15 ng/µl) samples of 125 diabetic patients with consistent negative and positive G-gap using real-time quantitative PCR using Cawthon's method (Telomere/Single Copy Gene ratio, T/S method) (Cawthon, 2002). The theory of this assay is to measure an average telomere length ratio by calculating telomeric DNA quantity with specially designed primer sequences and divide that quantity by the amount of a single-copy gene. The factor by which the ratio of telomere to single copy gene differs between G-gap negative samples and G-gap positive samples was measured. For each DNA sample 2 PCR reactions were designed using previously published telomere forward and reverse primers (Telomere forward: CGGTTTGTTTGGGTTTGGGTTTGGGTTTGGGTTTGGGTT and Telomere reverse: GGCTTGCCTTACCCTTACCCTTACCCTTACCCTTACCC; final concentrations 100nM each) (Cawthon, 2002 & Callicott and Womack, 2006) and *RPLP0* forward and reverse primers (RPLP0 forward: CAGCAAGTGGGAAGGTGTAATCC and RPLP0 reverse: CCCATTCTATCATCAACGGGTACAA; final concentrations 100nM each) (O'Callaghan *et al*., 2008). *RPLP0* represents a single copy gene (SCG) and encodes 60S acidic ribosomal protein P0. The 60S acidic ribosomal protein P0 is a human ribosomal protein and this is a component of the 60S ribosomal subunit. SCG was used as a control and to standardise the assay, as described in Cawthon

(2002). Each 20 µl reaction performed consists of 15 ng DNA, Precision FAST 2 × qPCR master mix (with Syber Green), 100 nM telomere primers or 100 nM Rplp0 primers on the 7500 FAST qPCR system (Applied Biosystem). The telomere assay uses SYBR Green to detect the PCR amplification product at 492nm-516nm wavelength indicating that only telomere/SCG is being amplified by the PCR. Both the telomere and SCG amplifications were performed in triplicates, whilst in each run duplicates of a non-template control included. Cawthon (2002) method was used to calculate the T/S ratio, based on the amplification of a telomere fragment (T) and an SCG (S) fragment. The thermal cycling conditions for both telomere and *RPLP0* consisted 20 seconds at 95°C. For telomere 40 cycles of 5 seconds at 95°C, 30 seconds at 65°C. For *RPLP0* 40 cycles of 30 seconds at 95°C, 30 seconds at 60°C. Followed by 30 seconds at 72°C for both telomere and *RPLP0*. The 95°C reads provided the cycle threshold value (Ct) values for the amplification of the target sequence. After completion, the data were transferred from PCR software to a Microsoft Excel spreadsheet for analysis. “T” represented the experimental DNA sample that was calculated by taking the average of the cycle threshold value for the telomere product whilst “S” represented the average of the cycle threshold value for the SCG product and the T/S ratio was calculated in the negative and positive G-gap groups by dividing the relative average input amount of the telomere PCR by the relative average input amount of the *RPLP0* PCR of the same sample. The cycle threshold (Ct) value represented the number of cycles required for the fluorescent signal to cross background level to exceed the threshold level. The results were reported as the relative average telomere length ratio.

A) Telomere length assay				
Steps	**Temperature**	**Time**	**Cycles**	**PCR**
Initial denature	95	20 seconds	1	Telomere and *RPLP0*
Denature	95	5 seconds	40	Telomere
	95	30 seconds		*RPLP0*
Annealing	65	30 seconds		Telomere
	60	30 seconds		*RPLP0*
Extension	72	30 seconds		Telomere and *RPLP0*

B) Reagents used for the Telomere Length Assay		
Reagents	**Concentration**	**Volume**
TaqMan genotyping assay	20x	2.0 µl
FAST qPCR Master Mix	2x	10.0 µl
Nuclease-free water	NA	7.0 µl
DNA template	15ng/µl	1.0 µl
Total		20 µl

Table 2. 2 Telomere length assay. (A): Thermal cycling program. (B): Reagents used for the telomere length assay on 7500 FAST qPCR System (Applied Biosystem).

2.1.7. Study 4: RNA extraction

Tempus Blood RNA tubes (ThermoFisher) contains 6ml of stabilising reagents. These tubes were used to collect 3ml peripheral venous blood samples from 32 diabetic patients with G-gap negative and G-gap positive, immediately mixed vigorously after collection, incubated at room temperature for 2 hours, and kept at -80˚C until processing.

A commercial kit from Norgen Biotek RNA purification Kit I (Norgen, Thorold, ON, Canada) was utilised to extract RNA. Samples were thawed and left for 2 hours at room temperature before extracting RNA. RNA extraction was performed according to the manufacturers' instructions briefly described here. Blood samples were transferred to 50ml conical tubes marked respectively with sample numbers and filled with Tempus Blood RNA tube diluent (contains phosphate-buffered saline) to make up a final volume of 12 ml. Tubes were mixed by vortexing vigorously for 30 seconds. Mixed samples were centrifuged at 4°C for 30 min at 4500rpm. After discarding supernatant, the pellet was dried by inverting the tube on a paper towel for 2 minutes. The dried pellet was re-suspended with 600µl lysis solution (comprising of Tris and EDTA) by vortexing for a few seconds. Pure ethanol 100% (300µl) was added to the samples and mixed. Afterwards, the solutions were transferred to the RNA isolation spin columns where a column was assembled with one of the provided collection tubes. Centrifugation for 1 min at 14,000 rpm was performed to make sure that the entire lysate volume had passed through the column. After that samples were washed with washing buffer (contains chaotropic salts). DNA may affect the sensitive downstream application. Therefore, RNase-Free DNase I Kit (Norgen, Thorold, ON, Canada) was used at this point to remove maximum residual DNA and 100µl RNase-free DNase 1 was applied to the sample

and centrifuged for 60 seconds at 14,000 rpm. Flow through solution was pipetted to the column again for maximum DNase activity and to get maximum RNA yields and centrifuged at 14000 rpm. DNA digestion was carried out by incubating the column assembly at room temperature for 15 min. Thereafter, 400µl of the washing solution was applied to the column and centrifuged for 1 min at 14,000rpm and then for 2 min at 14000 rpm to dry the column. RNA was eluted by adding 50µl nuclease-free water to the samples and incubated for 1 min followed by centrifugation for 2 min at 2000 rpm and 1 min at 14,000 rpm. The RNA elution step was repeated for maximum RNA recovery. NanoDrop spectrophotometer (Thermo Fisher Scientific) was used to determine the yield of isolated RNA. The samples were stored at -80˚C until RNA sequencing and library preparation by Oxford Genomic Centre.

2.1.7.1. RNA quantification

RNA quantification was performed using a RiboGreen assay (Invitrogen) according to manufacturer`s instruction. The RiboGreen assay utilises a fluorescent RNA stain that is 200-fold more sensitive than ethidium bromide. RiboGreen reagent binds RNA and fluoresces with the maximum emission wavelength of 525nm. Fluorescence can be detected at a wavelength with detection as low as 1 ng/ml RNA (McKiernan *et al*., 2018). Briefly, 200-fold RiboGreen reagent was prepared into TE buffer and kept away from light. Six standard concentrations were prepared including 50ng/ml, 25ng/ml, 12.5ng/ml, 6.25ng/ml, 3.12ng/ml and 1.5ng/ml by using standard provided in the kit (ribosomal RNA, 100 µg/ml). Dilution of samples (1:10) in TE buffer were carried out. Diluted standards and samples (100 µl) in TE buffer were added to respective wells of 96-well plate in triplicate. Finally, 100 µl of the prepared RiboGreen reagent was added to each well. Plates were incubated for 5–30 min at room temperature, protected from light until reading at fluorescein

emission settings on a FLUOstar OPTIMA plate reader (BMG Labtech) at wavelength 525nm.

2.1.7.2. Quality control checks

Quality control checks were conducted pre- and post-alignment to ensure sufficient quantity and quality of RNA samples. Pre-alignment quality control checks (RNA abundance and quality) were performed by using 2200 or 4200 Tape Station software using the High Sensitivity RNA ScreenTape kit (Agilent Technologies) according to the manufacturer's protocol. Briefly, each sample was prepared for RNA screen tape assay by mixing 5µl of Agilent RNA Buffer with 1µl of RNA sample. Similarly, the Ladder was prepared by mixing 5µl of Agilent RNA Buffer with 1µl of RNA ladder (a molecular-weight size marker). Samples and ladder were vortexed using IKA vortexer followed by centrifugation at 2000 rpm for 1 minute. After this brief spin the samples/ladder were denatured by heating at 72ºC for 3 minutes and placed on ice for 2 minutes and then centrifuged again and then loaded into Tape station. The Tape station calculates an RNA integrity number (RIN) which on a scale of 1 to 10 provides an assessment of RNA quality with 1 being highly degraded RNA and 10 being the heavily intact RNA (Mueller, Lightfoot and Schroeder, 2004). If the sequence base quality of RNA is not satisfactory then it can affect the downstream analysis. Therefore, it is better to perform quality trimming of sequence reads before alignment (Kallio, *et al*., 2011).

2.1.7.3. RNA-Seq library preparation

2.1.7.3.1. Ribosomal RNA removal

Samples for RNA-Seq from 32 negative and positive G-gap patients with diabetes were sequenced as recommended by the manufacturer (Illumina). Input material (total RNA) was normalised to 1µg prior to library preparation. For the removal of ribosomal RNA (rRNA), the Ribo-Zero rRNA Removal Kit (Illumina, Human/Mouse/Rat) was used according to the manufacturer's instructions. Ribosomal RNA (rRNA) removal is based on the hybridisation/bead capture procedure that selectively hybridises to rRNA species using biotinylated capture probes. The hybridised compound (probe-rRNA) is then arrested by magnetic beads and then removed. RNA 1µg was used to deplete ribosomal RNA. Ribosomal RNA (rRNA) depleted samples were purified using an ethanol precipitation method. Ribo-Zero-treated RNA was quantified by using a fluorometric method and the size profile was assessed on 2200 Agilent TapeStation system.

2.1.7.3.2. TruSeq stranded total RNA (Illumina) library preparation

After rRNA depletion, whole RNA libraries were prepared using TruSeq Stranded Total RNA kit (Illumina) following the manufacturer's instructions. Elute, Prime and Fragment High Mix were added to fragment RNA. ERCC (Synthetic Spiking Standards) (1:1000) (LifeTechnologies) was added before incubation at 94˚C for 8min. First and second cDNA strands were synthesised by two consecutive steps.For the first step (first strand synthesis), Act D (Actinomycin D) and SuperScript II, Reverse Transcriptase and random hexamer primer were added to fragment RNA before the thermocycler incubation at 25˚C for 10 minutes, at 42˚C for 15 minutes and at 70˚C for 15 minutes, while in the second step second strand

marking master mix was added and incubated at 16˚C for sixty minutes. The purified double-stranded cDNA was obtained by AMPure XP beads (Beckman Coulter), according to manufacturer's instructions. The indexing adaptors were ligated to the end of double-stranded cDNA strands by incubation for 10 mins at 30°C. Additional adaptors were removed with the help of AMPure XP beads (Beckman Coulter), following manufacturer's instructions. The amplification of cDNA libraries was performed on a Tetrad (Bio-Rad) using in-house unique dual indexing primers (Lamble *et al*., 2013) with this thermal cycler programme: 98°C for 30 seconds,15 cycles of 98°C for 10 seconds, 60°C for 30 seconds, 72°C for 30 seconds and finally 72°C for 5 minutes. The purification of amplified libraries was carried out by AMPure XP Beads (Beckman Coulter), according to manufacturer's instructions, before analysing quality and size profile on Agilent 2200 TapeStation and quantified using Qubit 2.0 Fluorometer, under standard operating procedures.

2.1.7.3.3. Libraries quantification and normalisation

The quantification of libraries was conducted by Qubit Kit Assay (Invitrogen, Life Technologies). This quantification method is fluorometric-based and uses DNA binding dyes. Briefly, samples and standards were kept on ice whereas the Qubit buffer and reagent were kept at room temperature. The working concentration of Qubit reagent 1:200 was prepared in Qubit buffer. The standard tube contained 190 µl of Qubit working solution and 10µl of Qubit standard whilst the sample tube included 199 µl of Qubit working solution (contains Quant-iT reagent and Quant-iT buffer) and 1µl of the sample. Each tube had a final volume of 200 µl. The tubes were vortexed briefly and incubated for 2 minutes at room temperature. Finally, tubes were inserted in the Qubit 2.0 Fluorometer and quantification was conducted. The Qubit readings were used for library normalisation. Individual libraries were

normalised to 10 nM before being pooled for processing. Each cDNA library was denatured and further diluted in resuspension buffer (10 mM Tris-HCl, 10% Tween-20) to 2 nM prior to loading on the sequencer. Paired-end sequencing of the libraries was performed at Oxford Genomic Centre using a HiSeq4000 75bp platform after preparation with Illumina, HiSeq 3000/4000 PE Cluster kit and 150 cycles SBS kit, generating a raw read count of >45 million reads per sample.

2.1.7.4. Roadmap from data to results

Raw data were obtained for analysis in FASTQ format from the sequencing machine. In this format, four lines represents one sequence read; where the first line consists of the name or ID of the sequence and starts with the sign “@", the second line contains actual sequence, the third line starts with + sign and has no specific function whilst the fourth line indicates the quality of each base present in the sequence. FastQC (v 0.11.5) software acts as a quality control tool for data sequence (Andrews, 2010), check GC content for each base as well as adapter content (Cock *et al*., 2009) by using the FASTQ files as a starting point and produce a report about the statistics of the data. These reports also contain a Phred quality score per base. Base Callers set the Phred quality scores [Q = −10log10(P)], where P represents the likelihood of the base call being incorrect (Cock *et al*., 2009). Filtration of garbage reads i.e. reads with low-quality base call/reads with low-quality scores (Andrews, 2010) or trimming of the sequence to remove low-quality bases can be performed before reading alignment. Trim Galore tool can be easily applied to paired-end reads to remove low-quality base calls and adapters from raw sequences (Krueger, 2015). Trimming not only improves the mapping yield but also improves mapping speed. In paired-end reads experiments, two ordered files are

prepared, one with the first end and the other with the second. FastQC reports help in qualitative judgment about the quality of the data.

Next step of the analysis includes the mapping of reads to the reference genome for the identification of their origin for the transcription process. Reference genomes are commonly available as FASTA/GTF file format and can be downloaded from the Ensembl FTP site (ftp://ftp.ensembl.org/pub/).

2.1.7.5. RNA-Seq read mapping

The reference genome, as well as gene annotations, exist for human. Therefore, sequenced reads were directly mapped to the reference genome. Data on human genome (GRCh38 human assembly) was downloaded from Ensembl FTP site that was used for all the steps of analysis. This assembly was used as a base to produce other files and results. HISAT2 was used to map the results with the reference genome and for the discovery of transcript splice sites. StringTie was used for the creation of multiple isoforms through the assembly of the alignments into full and partial transcripts, and for estimating the expression levels of all genes and transcripts. Ballgown takes the output (transcripts and expression levels) from StringTie and uses rigorous statistical methods to identify which transcripts were differentially expressed between G-gap negative and G-gap positive groups.

2.1.7.6. Bioinformatics tools/ RNA sequencing mapper

Bioinformatics tools were used to perform various activities including sequence mapping, identification of gene expression, identification of differential expression and aligning RNA-Seq reads (Luscombe, Greenbaum and Gerstein, 2001). High throughput sequencing of mRNA (RNA-Seq) was used for the measurement and

comparison of gene expression in negative and positive G-gap groups. RNA-Seq experiments generated large and complex data sets. HISAT2, String Tie, and Ballgown which provide fast, accurate and flexible software, were used to reduce the raw read data to comprehensive results and provide a comprehensive analysis of the RNA-Seq experiments. Together these were used for the alignment of reads to the genome, assembly of transcripts including novel splice variant and computing the abundance of these transcripts in each sample. These software packages were used to compare G-gap negative and G-gap positive groups to identify differentially expressed genes and transcripts.

2.1.7.6.1. HISAT2

For the downstream analysis of RNA sequencing, reads were mapped to a reference genome. Previously used TopHat2 programme was replaced with HISAT2 (hierarchical indexing for spliced alignment of transcripts) that is a faster, accurate and which uses a sensitive mapping programme for the alignment of RNA sequencing reads (Kim, Langmead and Salzberg, 2015). It is different in the sense that it maps the spliced reads that span over exon-exon junctions by using the indexing approach. The indexing approach is based on the FM index. In this tool, the algorithms use two types of indexes: one global FM index (denoting the whole genome) and various small around 48000 local FM indexes (collectively cover the entire genome) where every local FM index shows a genomic region consist of 64,000bp. HISAT2 aligns the reads to the reference genome first on the basis of the global index where it applies exonic alignments with selected segments obtained from reading sequences and identified the candidate mapping region. After this, alignment was finalised with the help of the local indexes of the candidate regions. The small indexes in combinations with other alignment approach made it possible

to effectively align the RNA-Seq reads particularly those that span exon-exon junctions.

2.1.7.6.2. String Tie assembler

Various approaches after sequencing of the transcriptome produce more than 200 million short sequence reads consisting of 100-150bp/reads. Assembly of numerous transcripts with variable sequence coverage and alternative splicing transcripts from the same locus that share the exons is hard. String Tie ensured that a read was assembled correctly across the entire region where the introns have been retained. Therefore, it was able to assemble the alignments into transcripts, producing multiple isoforms and computing the expression level of all genes and transcripts. String Tie was used because it was accurate and faster to align the reads into potential transcripts (Pertea *et al*., 2015) as compared to Cufflinks and can evaluate expression level effectively and reconstruct the genes precisely and comprehensively. String Tie can assemble transcripts regardless of whether gene annotation was available and can assemble a larger number of splice variants per gene locus. Specialised software i.e. Ballgown can process the output of String Tie. To eliminate the gene length and sample biases, normalisation was performed by Ballgown software.

2.1.7.6.3. Ballgown

Ballgown takes the output i.e. gene and transcripts expression levels from String Tie to determine differentially expressed genes between the studied cohort. Ballgown software helps in RNA-Seq data analysis as it is the part of the R/Bioconductor package. It helps in the visualisation of expression level for transcriptome assembly. The Ballgown R package is dependent on pre-processing activities (carried out by

HISAT2 and String Tie) including alignment of RNA sequencing reads to the reference and assembly of aligned reads. The obtained expression level of transcript, intron, and exons for the evaluation purpose should be in Ballgown readable format. FPKM is used for normalisation. Normalisation is done using a *stattest* function in a Ballgown.

2.1.7.7. Gene set enrichment and pathway analysis of differentially expressed genes

Gene set enrichment (GSE) testing, also known as functional analysis of genes, was used to identify enriched functions of differentially expressed genes between G-gap groups. All expressed genes were ranked according to their fold changes with $p<0.05$. This analysis provides an understanding of which pathways are important targets and more affected in G-gap groups. This study has identified which diseases the G-gap may play a role. The computational tool Database for Annotation, Visualization and Integrated Discovery (DAVID) (www.david.ncifcrf.gov/) v 6.8, which provides meaningful insight on biological meanings of large gene lists derived from genomic studies, was used for gene set enrichment and pathway analysis. Disease and gene ontology for functional annotation and KEGG for pathway analyses were applied using DAVID.

Chapter 3

“Do SNPs in FN3K and Ferroportin1 explain G-gap variation and potentially associate with sRAGE?”

3.1. Introduction

Research has demonstrated that *FN3K* and *SLC40A1/Ferroportin1* SNPs were linked with FN3K activity, HbA_{1c} levels and diabetes complications (Mohas *et al.*, 2010; Tanhauserova *et al.,* 2014). The discrepant HbA_{1c} values in comparison with other measures of average glycaemia including plasma fructosamine or glycated albumin could shift the glycation gap (G-gap) in the negative or positive direction. The study reported that a positive G-gap is associated with increased diabetes complications in patients compared with those with a negative G-gap as well as being associated with differences in mortality (Nayak *et al.,* 2013). Dunmore *et al.* (2018) reported that the individually consistent glycation gap may be the result of differences in activity of a deglycating enzyme FN3K, which is highly expressed intracellularly in the erythrocyte. FN3K can only affect the HbA_{1c} level without affecting glycated circulating protein. Glycation of circulating proteins is estimated by fructosamines.

The study of *FN3K* gene is an approach in the dissection of the genetic basis of FN3K 3-fold higher activity in the negative G-gap group in patients with diabetes. Dunmore *et al*. (2018) also reported a low level of AGEs in negative G-gap group whilst in the positive G-gap group the high AGEs level was reported. Advanced glycation end products (AGEs) plays a key role in the development of diabetic vascular disease (Peppa and Vlassara, 2005, Genuth *et al*., 2005 and Monnier, Sell and Genuth, 2005) and are used as biomarkers because they are supposed to provide a link between hyperglycaemia and diabetes complications (Monnier, Sell and Genuth, 2005). AGEs are involved in pro-oxidant activities. Through these activities, AGEs might be involved in the progression of vascular disease in people with diabetes (Brownlee, 2005 and Peppa, Uribarri and Vlassara, 2008). AGEs-

RAGEs interaction induces inflammation and oxidative stress and is thought to nurture disease progression. RAGE as multi ligands is involved in several physiological processes (neural growth, acute inflammation) and disease pathologies (diabetes, tumour growth/metastasis, Alzheimer's disease, chronic inflammation), implicating RAGE as a promising lead for therapeutic intervention (Schmidt *et al*., 2000 & Stern, 2002). N-terminal truncated RAGE, produced from alternative splicing of RAGE mRNA, and C-terminal truncated RAGE, formed by proteolytic cleavage of membrane-bound RAGE protein, are two other forms of RAGE, expressed in human endothelial cells of diabetic patients (Yonekura *et al*., 2003). Endogenous factors like glucose, inflammation and exogenous factors such as smoking, and diet affect the circulating levels of AGE receptors (Selvin *et al.*, 2013).

The extracellular surface of the RAGE consists of COOH terminus whilst the amine (NH) terminus of RAGE is important for the activation of the proinflammatory nuclear factor (NF)-KB-mediated signalling. RAGE induces inflammation after stimulation by AGEs and contributes to tissue injury. RAGE is also involved in the stimulus of progression of the chronic disease through NFkB-mediated signalling (Yan *et al*., 2008, Yan, Ramasamy and Schmidt, 2009, Yan *et al*., 2007 and Ramasamy *et al*., 2005). Previous studies reported the role of RAGE and soluble RAGE (sRAGE) in the pathophysiology of diabetes angiopathy (Ramasamy *et al*., 2007, Lindsey *et al*., 2009, Skrha *et al*., 2012). COOH terminal and transmembrane domains are absent in plasma sRAGE. sRAGE competes with RAGE to bind with the ligands. In type 2 diabetes, the sRAGE level was directly proportional to urinary albumin excretion (Humpert *et al.*, 2006). Considering the role of sRAGE in the pathogenesis of diabetes complication, this study examined the relative expression of soluble RAGE

(sRAGE) protein in G-gap negative and G-gap positive groups. This study was conducted to evaluate the sRAGE levels in 146 plasma samples collected from diabetic patients with negative and positive G-gaps and single nucleotide polymorphisms of genes encode FN3K/Ferroportin1 in the studied cohort. The relationship of FN3K polymorphisms (rs3848403 and rs1056534) with the sRAGE concentration was also explored in those cohorts.

3.2. The rationale of the study

The aim of the present study was investigating the genetic variation affecting FN3K activity in diabetic patients with negative and positive G-gap. This study, focused on the variant of FN3K gene, whose product is involved in non-enzymatic glycation reaction, in mixed populations with diabetes. The purpose was to identify the relationship between SNPs within the *FN3K* gene (with their known functional effects) and sRAGE levels in negative and positive G-gap. This may offer a better comprehension of the glycaemic control and provide a predictive role in the middle and long period of the disease.

3.3. Materials and methods

3.3.1. SNPs study for FN3K/Ferroportin1

3.3.1.1. DNA extraction

DNA was extracted according to the method described in Chapter 2 from whole blood that had been kept frozen at -20° C.

3.3.1.2. SNP selection and genotyping

The SNPs in *FN3K* and *SLC40A1/Ferroportin1* genes, according to their known functional effects were selected. The features of analysed SNPs are provided in Table 3.1. PCR with the fluorescent-based chemistry was used to determine *FN3K* and *SLC40A1/Ferroportin1* SNPs (TaqMan® SNP Genotyping Assay, Applied Biosystems). Preparation of genomic DNA and PCR mixture can be found in Chapter 2 (2.1.6.2.3.1. Sample preparation).

	Chromosome	**rs number**	**Nucleotide Substitution**
FN3K	17q25.3	1056534	C/G
		1046896	C/T
		3848403	C/T
SLC40A1/Ferroportin1	2q32.2	11568350	A/C

Table 3. 1 Characteristics of studied SNPs for FN3K/Ferroportin1 gene. The table shows FN3K, fructosamine 3-kinase; SLC40A1/Ferroportin1; SNPs, Single Nucleotide Polymorphism.

3.3.1.3. TaqMan assay

The TaqMan assay was used for genotyping polymorphisms of the *FN3K/Ferroportin1* gene with a view to discriminate between two alleles of a specific SNP of *FN3K/Ferroportin1 gene* (refer to Chapter 2, TaqMan genotyping assay for details).

3.3.2. sRAGE study

3.3.2.1 Materials

Details of materials including equipment (microplate reader), reagents (phosphate buffer saline, wash buffer, reagent diluent, substrate solution and stop solution) and kit (Duoset ELISA kit for Human sRAGE) used for the (soluble Receptor-AGE) sRAGE assay can be found in Appendix 1.

3.3.2.2. sRAGE assay

Plasma sRAGE expression was measured by immunoassay (R&D Systems, Minneapolis, MN 55413, USA) with a detection range (1:4 diluted samples) of 0–2000 pg/ml. Details can be found in chapter 2 (2.1.5.1. Enzyme-linked immunosorbent assay (ELISA)).

3.3.3. Statistical analysis

3.3.3.1. SNPs for *FN3K/Ferroportin1*

Data were analysed on SPSS v 24. The comparison of means between the groups was done by Student`s t-tests for 2 independent measures or otherwise one-way analysis of variance. The comparison of proportions was by Chi-Square. Binary regression was used to compare the association of dichotomised variables with other factors and univariate analysis for continuous variables. Results of statistical testing were considered significant at $p \leq 0.05$.

3.3.3.2. sRAGE

Data were analysed using SPSS/ Excel. Values were expressed in pg/ml. Student`s t-test (Two-Sample Assuming Unequal Variances) on Excel was then used to find the statistical significance ($p \leq 0.05$) an association between the negative/positiveve G-gap and sRAGE values. sRAGE plasma concentration in relation to the genotype of FN3K was expressed as the mean ± standard deviation (SD). Differences between the genotype of FN3K SNPs rs3848403 and rs1056534 in relation to sRAGE concentration were analysed by one-way analysis of variance (ANOVA). P-values ≤0.05 show statistically significant results.

3.4. Results

3.4.1. Patients

A total of 184 diabetic subjects were recruited at different times at New Cross Hospital in Wolverhampton and divided into two different glycation gap categories i.e. negative and positive (see Chapter 2 for details) This cohort was previously analysed for the FN3K activity (Dunmore *et al*., 2018). The baseline characteristics of patients are reported in Table 3.2. Negative G-gap and positive G-gap subjects differed significantly ($p<0.05$) in age (60±13 and 65±10 respectively), BMI (30.2±5.2 and 36.0±7.2 respectively) and type of diabetes (63% T2D and 91% T2D respectively) but not gender, ethnicity, duration of diabetes ($p>0.05$). As expected from the selection criteria, negative G-gap and positive G-gap subjects differed for HbA_{1c}, fructosamine and the G-gap (all $p<0.001$). All clinical and biochemical data for the negative G-gap and positive G-gap groups are shown in Table 3.2. Glycated HbA_{1c} level was 7.6% ±1.3 and 9.4%±1.3 in G-gap negative and G-gap positive groups respectively. While the fructosamine level was 347±62 and 293±43 µmol/l in

G-gap negative and G-gap positive groups respectively. Glycation gap in negative and positive groups was -1.5±0.6 and +1.6±0.6 respectively.

	Negative G-gap	Positive G-gap	P value
Number	102	82	
Age	60±13	65±10	p<0.001
Body mass index (kg/m^2)	30.2±5.2	36.0±7.2	p<0.001
Type of diabetes (% Type 2)	63%	91%	p<0.001
Gender (% male)	77%	61%	Ns
Ethnicity (% white, Asian, Black)	70%, 18%, 12%	73%, 23%, 4%	Ns
Duration of diabetes (years)	19±12	16±10	Ns
Fructosamine derived HbA1c (% glycated HbA_{1c})	9.1±1.9	7.8±1.3	p<0.05
Average HbA_{1c} (% glycated)	7.6±1.3	9.4±1.3	p<0.001
Average Fructosamine (µmol/l)	347±62	293±43	p<0.001
Average G-gap (% glycated HbA_{1c})	-1.5±0.6	+1.6±0.6	p<0.001

Table 3.2 Descriptive characteristics of patients involved in the study. Values are the mean± SD or otherwise percentages.

3.4.2. Nanodrop

Nanodrop was used for the analysis of DNA sample purification. The absorbance at 260nm provided a yield of DNA whilst proteins have an absorbance at 280 nm. The A_{260}/A_{280} between 1.8 and 2.0 are acceptable as these indicate that the sample does not have any contamination. The average concentration of DNA samples was approximately 53.44ng/µl and 57.61ng/µl in G-gap negative and G-gap positive groups respectively. In this study the A_{260}/A_{280} ratio for all studied DNA samples was ≥1.7 in both studied groups.

3.4.3. SNPs association with G-gap status

Three polymorphisms in the *FN3K* gene and one polymorphism in *SLC40A1/Ferroportin*1 were studied. These were single nucleotide polymorphisms (SNPs), a c.141+246C>T (rs3848403) in intron region, a c.900C>G (rs1056534) in exon 6 and a C/T (rs1046896) in promoter region of *FN3K* gene whilst c.744G>T (rs11568350) in exon 6 of the *SLC40A1* gene leading to Q248H amino acid substitution in the ferroportin-1 protein. In this study, A/C frequency that relates directly to G/T frequency was measured. G/T alleles were reported in reverse orientation to genome (rs11568350 in the NCBI database). The SNPs rs3848403, rs1056534, rs1046896, and rs11568350 were not significantly different between the negative G-gap and positive G-gap groups as the $p>0.05$.

The FN3K SNPs rs3848403 had 3 allele combinations. Forty-eight percent of these were C/T, twenty-seven percent were T/T and twenty-five percent were C/C. rs1056534 had two allele combination C/G (73%) and G/G (27%), whilst rs1046896 had only one allele combination C/T. The ferroportin1 SNP also had only one allele combination A/C. These all were reallocated into heterozygous (C/T, C/G, A/C) and

homozygous (C/C, T/T, G/G or A/A) between G-gap negative and G-gap positive groups. Genotypes distributions of genetic variants of *FN3K* and the *SLC40A1* gene in G-gap positive and G-gap negative patients and their relationship are shown in Table 3.3.

	Genotypes	Positive G-gap (%)	Negative G-gap (%)	χ^2 test
FN3K (3848403) N= 121	CT TT CC	21.5 11.6 9.9	26.4 15.7 14.9	ns
FN3K (1056534) N=124	CG GG	29.8 11.3	43.6 15.3	ns
FN3K (1046896) N=125	CT	41.6	58.4	ns
SLC40A1 (11568350) N=125	AC	41.6	58.4	ns

Table 3.3 Table showing genotypes present in g-gap positive and g-gap negative patients. The table shows that SNPs rs3848403 and rs1056534 were present in heterozygous and homozygous state whilst SNPs rs1046896 and rs11568350 were present only in the heterozygous state.

3.4.4. Characterisation of SNPs

3.4.4.1. c.141+246C>T

The polymorphism c.141+246C>T has been described in literature (rs3848403 in the NCBI database). In the present study (Table 3.4), the polymorphism was detected in a heterozygous state in 58 patients. Of these, 32 were negative G-gap and 26 positive G-gap. The homozygous state was detected in 63 patients. of these, 37 were negative G-gap and 26 positive G-gap.

3.4.4.2. c.900 G>C

In the present study (Table 3.4), this polymorphism was found in heterozygous state in 91 patients. These include 54 negative G-gap and 37 positive G-gap patients. The homozygous state was found in 33 patients, including 19 negative G-gap and 14 positive G-gap. This polymorphism was already described in the literature (rs1056534 in the NCBI SNPs database). In this study, significant differences in heterozygous and homozygous state were not observed between negative and positive G-gap groups.

3.4.4.3. c.486 C>T

This polymorphism (Table 3.4) was found only in the heterozygous state in 125 patients. These include 73 negative G-gap and 52 positive G-gap. This polymorphism was already described in the literature (rs1046896 in the NCBI SNPs database).

3.4.4.4. g.189565370 C>A/ c.744G → T mutation (G/T alleles in reverse orientation to the genome)

This polymorphism (Table 3.4) was present as a heterozygous state in all 125 (73 negative G-gap and 52 positive G-gap) patients in this study. Reference SNP of this polymorphism in NCBI SNPs database is rs11568350.

	Genotypes	Numbers	Negative G-gap (n)	Positive G-gap (n)
FN3K (3848403) N= 121	Heterozygous (CT) Homozygous (TT/CC)	58 63	32 37	26 26
FN3K (1056534) N=124	Heterozygous (CG) Homozygous (GG)	91 33	54 19	37 14
FN3K (1046896) N=125	Heterozygous (CT)	125	73	52
SLC40A1 (11568350) N=125	Heterozygous (AC)	125	73	52

Table 3.4 Table showing genotypes in heterozygous and homozygous state in G-gap negative and G-gap positive patients. The table shows that SNPs rs3848403 and rs1056534 were present in heterozygous and homozygous state whilst SNPs rs1046896 and rs11568350 were present only in the heterozygous state.

3.4.5. Plasma sRAGE in relation to G-gap

Plasma sRAGE was assayed from sample collected studied cohort with negative (<-0.5%, n=79) and positive (>+0.5%, n=67) G-gaps using sandwich ELISA (R&D Systems, Bio-Techne Ltd). Linear regression analysis was performed in excel to create a standard curve (Figure 3.1) and used a line of the best-fit equation to calculate the concentration of the sRAGE in the plasma samples. The measurement of sRAGE in negative and positive G-gap heparinised-plasma samples revealed similar results p=0.16 (independent sample t-test) and the corresponding mean±SD values were 344.64±271.58 and 285.38±230.33 respectively (Figure 3.2). These results showed that the plasma sRAGE level is not correlated with differences in G-gap values. Descriptive statistics indicated skewness of 1.3>1, indicating that data are very skewed. A normal distribution has a kurtosis of zero whilst in the present experiment, it is 1.61 indicating that data distribution has a heavier tail than a normal distribution (Figure 3.3). A non-parametric Mann–Whitney U test (Table 3.4) was performed using Excel, which showed a non-significant difference (P = 0.23), after adjustment for ties with 95% CI in sRAGE levels between subjects with the G-gap negative and G-gap positive. The reported intra-assay and inter-assay coefficient of the variant for the kit was 5.7% and 7.7% respectively.

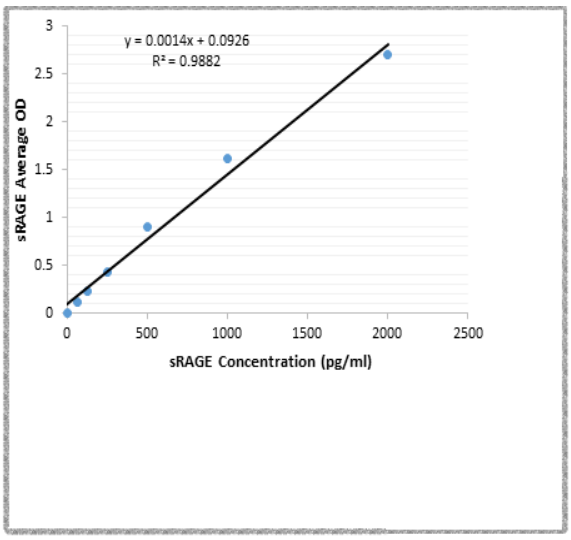


Figure 3. 1 The standard curve created from the standard concentrations of human sRAGE in ELISA assay. The R^2 value (0.9882) indicated data strong association between the variables used in constructing the standard curve and that unknown concentrations of sRAGE from plasma samples could be interpolated with accuracy from their optical density.

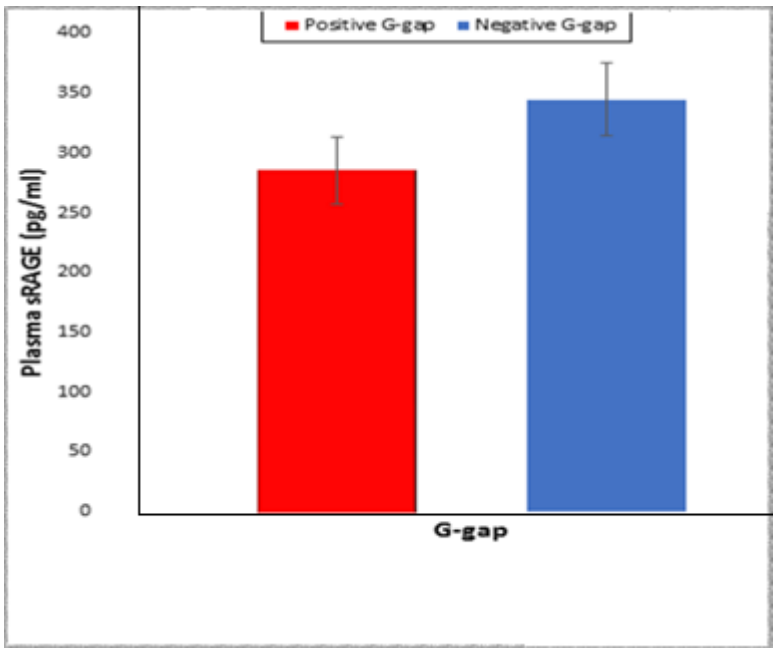
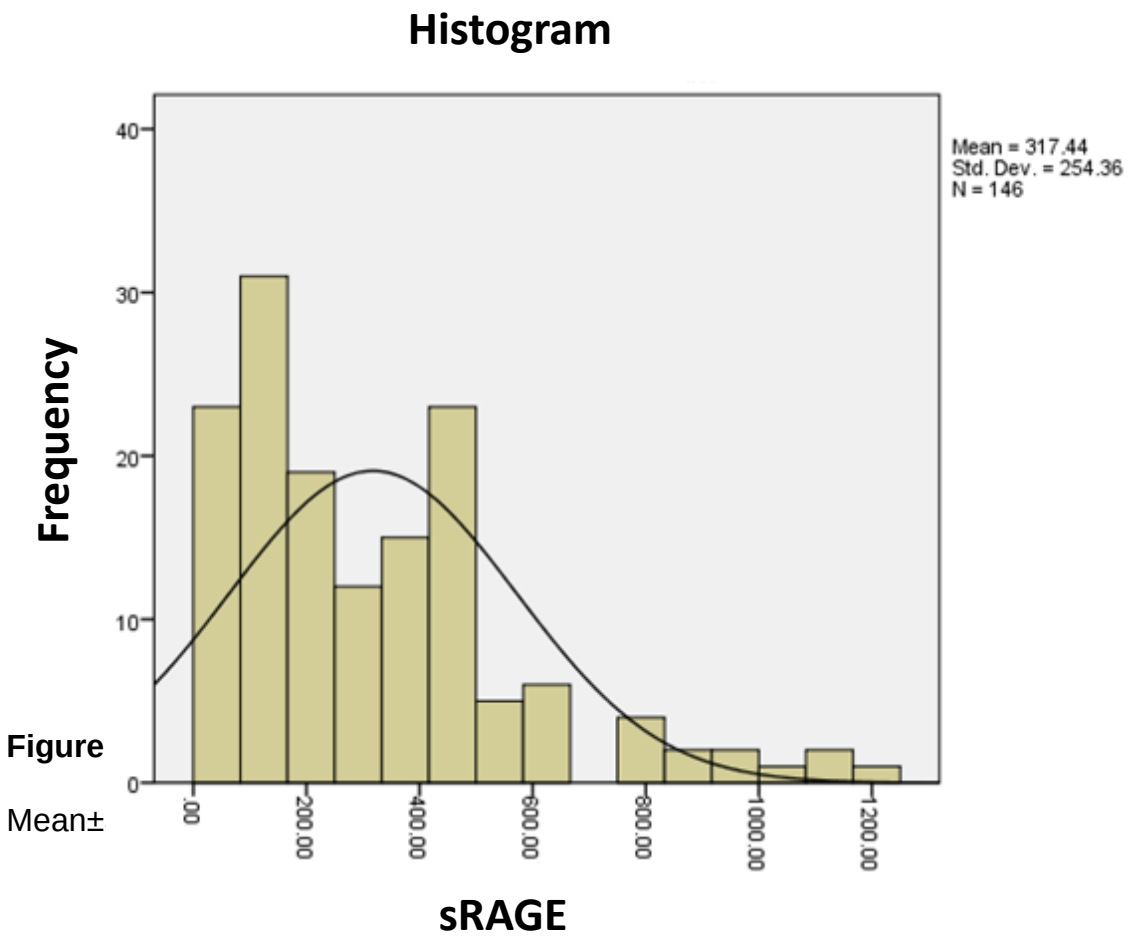


Figure 3. 2 Plasma sRAGE in relation to glycation gap. Values represent mean ± SE. P>0.05 using the t-Test, G-gap positive (n=67) vs G-gap negative (n=79).



Figure

Mean±

Figure 3.3 sRAGE data are positively skewed. Using the histogram. Mean±SD:317.44±254.36. The data distribution has a tail i.e. kertosis: 1.61>0.

G-gap (n)	Mean Ranks	Sum of Ranks	P Value (two-tailed)	Z Value	Mann–Whitney U Test
Negative (79)	77.34	6110	0.23/ns	1.19	2343
Positive (67)	68.97	4621			

Table 3. 4 Mann–Whitney U (non-parametric) test of normality for sRAGE analysis for negative and positive G-gap groups p>0.05.

3.4.6. Plasma sRAGE in relation to the genotype of FN3K SNPs rs3848403 and rs1056534.

In diabetic patients with G-gap negative and G-gap positive, statistically insignifcant association was observed in sRAGE concentrations according to their *FN3K* (rs3848403) and *FN3K* (rs1056534) genotype (Table 3.5). CT, TT and CC genotypes of *FN3K* (rs3848403) polymorphism were insignificantly associated with sRAGE concentration (C/T: 268.67±164.01, T/T: 312.69±332.46 and C/C 393.37±150.94 pg/ml; p>0.05 in G-gap positive and C/T: 290.98±211.26, TT: 383.02±359.84 and C/C: 180.9 pg/ml; p=0.5 in G-gap negative respectively). Moreover, statistically insignificant association was also observed between FN3K (rs1056534) genotype and sRAGE concentration in positive and negative G-gap groups (C/G: 284.19±248.20, G/G: 292.21±163.11 pg/ml; p=0.9 and C/G: 342.48±275.40, G/G: 247.92±170.97 pg/ml; p>0.05 respectively).

sRAGE (pg/ml)		Positive G-gap	Negative G-gap
FN3K **(rs3848403)**	CT	268.67±164.01	290.98±211.26
	TT	312.69±332.46	383.02±359.84
	CC	393.37±150.94	180.9
	ANOVA	ns	Ns
FN3K **(rs1056534)**	CG	284.19±248.20	342.48±275.40
	GG	292.21±163.11	247.92±170.97
	ANOVA	ns	Ns

Table 3.5 sRAGE concentration (pg/ml) in positive G-gap and negative G-gap diabetic patients with respect to their genotype. No significant association was observed between sRAGE concentration and FN3K genotypes ($p>0.05$). Data is shown as mean±SD.

3.5. Discussion

To understand the adverse effects of hyperglycaemia, previous studies have traditionally focused on the increased production of AGEs and activity of the receptor for AGEs (RAGE) produced because of its potentially harmful pathways (Wu, Xiao and Graves, 2015; Rhee and Kim, 2018). There was little focus on hypothetically protective mechanisms operating in parallel. These pathways have regulatory enzymes whose activities can lead to overall tolerance or intolerance of diabetes mellitus. For example, FN3K is involved in deglycation of proteins, control of AGE production and subsequently, the development of diabetes complications (Avemaria *et al*., 2015). The enzyme is highly expressed in erythrocytes (Delplanque *et al.,* 2004) and is involved in HbA_{1c} variation (Veiga *et al.,* 2006).

Dunmore *et al.* (2018) reported a significant relationship between erythrocyte FN3K and the G-gap and demonstrated that the negative G-gap group has significantly higher FN3K (3-fold difference) enzymatic activity and protein level. One of the factors affecting enzyme activity; contributing to the inter-individual differences in HbA_{1c}, could be genetic variability in the FN3K gene (Avemaria *et al.,* 2015, Tanhäuserová *et al.,* 2014 and Soranzo *et al.,* 2010). Single nucleotide polymorphisms (SNPs) may be possibly affecting FN3K enzyme activity among diabetic subjects with negative and positive G-gap. The combination of genes studied in this study encodes for enzymes that are linked with FN3K activity. Surprisingly no literature data exist on the effects of SNPs in these genes on the variability of the glycation gap.

SNPs (rs3848403, rs1056534, rs1046896, and rs11568350) in *FN3K* and *SLC40A1/Ferroportin1* genes did not reveal any significant associations with any of the studied endpoints (negative and positive G-gap groups) in this study. FN3K (rs1046896) and ferroportin1 (rs11568350) were both found only in the heterozygous state in the negative (58.4%) and positive (41.6%) G-gap groups and insignificantly distributed between the studied groups. Although FN3K SNPs (rs3848403 and rs1056534) were found both heterozygous and homozygous state in negative and positive G-gap groups, their distribution was still statistically insignificant in those groups. Previous studies found the association of SNPs with HbA_{1c} (Mohás *et al.,* 2010), sRAGE (Skrha *et al*., 2014), FN3K activity (Delpierre *et al*., 2006) and progression of diabetic nephropathy and cardiovascular morbidity and mortality (Avemaria *et al*., 2015).

It was hypothesised that the haemoglobin-level variant/iron homoeostasis may influence HbA_{1c}, which in turn could affect the rate of haemoglobin that is glycated and impacting on measured HbA_{1c} levels. This hypothesis requires further testing. The propensity of iron storage disease had a genetic basis (Gordeuk *et al.,* 1992). Because ferroportin/SLC40A1 mutations are linked with iron overload as well as with FN3K activity, ferroportin/*SLC40A1* gene may affect iron homeostasis which influencing HbA_{1c}. Previous study reported a *SLC40A1* (Ferroportin1) Q248H mutation (rs11568350) at rates ranging from 2.2 to 10% among sub-Saharan African natives (Albuquerque *et al.,*2011). A role for altered iron homeostasis influencing HbA_{1c} is not suggested for *SLC40A1* where no difference was observed at known functional variants in negative and positive G-gap groups ($p>0.05$). In this study, it was observed that the only the heterozygous state of this polymorphism (rs11568350) was found in both negative (58.4%) and positive (41.6%) G-gap groups. This present study does not establish a relationship between the ferroportin/SLC40A1 variant and G-gap groups. The possibility exists that the effect observed may be due to stratification with respect to the degree of Asian/European admixture without including the native people who resided in sub-Saharan Africa.

FN3K and ferroportin/SLC40A1 have not been studied in relation to G-gap status. Results of this study indicated that selected SNPs in these genes are neither associated with G-gap status, nor with the difference in FN3K activity in the studied G-gap groups. Previously, it was suggested that the level of FN3K enzymatic activity in the deglycation process defines the G-gap (Dunmore *et al*., 2018). It is known that FN3K deglycates lysine residues more efficiently as compare to the HbA_{1c}-defining N-terminal valine. This is suggesting that FN3K might not contribute to G-gap formation. Therefore, it might seem illogical to conclude that variation in FN3K

activity or the SNPs of FN3K gene accounts for the G-gap. The substrate (N-ε-fructosyl-lysine and N-terminal" N-α-fructosyl amino acids) specificity of FN3K falls in the ranges of 100 times to 10 times lower affinity (Szwergold, Howell and Beisswenger 2001; Delpierrre *et al*., 2004). However, lower affinity does not imply a zero-rate reaction but indicates a slower reaction because the low range is ten (10) indicating some deglycation at the N-terminal valine. Studies reported the inverse relationship between HbA_{1c} and FN3K activity when SNPs were found in the FN3K gene found (Mohas *et al*., 2010 and Delpierre *et al.,* 2006). Other factors contributing to inter-individual differences in HbA_{1c} could be transcript splice variants encoding products of differing activity.

The significantly low FN3K activity in the positive G-gap group was associated with an increase in the AGE level in the same group (Dunmore *et al*., 2018). This potentially links with a G-gap associated with mortality and morbidity that typically links with high HbA_{1c} levels (Nayak *et al*., 2013). Dunmore *et al.* (2018) found a fivefold change in PAI-1 (plasminogen activator inhibitor 1) (in positive G-gap patients, which would probably be a risk marker for cardiovascular disease (Grant, 2007). With higher expression in erythrocytes (Delplanque *et al.,* 2004), intracellular FN3K during the deglycation process provides a possible control mechanism for the production of AGEs and their associated complications (Avemaria *et al.,* 2015). Dunmore *et al*. (2018) showed that the plasma AGE level is associated with the glycation gap status and showed that positive G-gap group patients with low FN3K activity have a high level of AGE and negative G-gap group patients with high FN3K activity has a low level of AGE. To further investigate this, plasma sRAGE level was measured in negative and positive G-gap groups. Result obtained in this study showed that patients from negative G-gap group have slightly higher plasma sRAGE

concentrations (mean Rank 77.34). However, in comparison with the value obtained for positive G-gap group, this increase is not statistically significant (U = 2343, *p* = 0.23).

AGEs and other ligands bind to sRAGE. This plays an important role in the development of diabetic vascular complications (Bucciarelli *et al.,* 2002, Goldin *et al.,* 2006 and Selvin *et al.,* 2013). sRAGE is well expressed at vascular injury sites as in diabetic kidney disease (Goldin *et al.,* 2006). Interpreting the sRAGE level is controversial because previous studies reported an inverse relationship between sRAGE and poor health issues (Falcone *et al.,* 2005 and Lindsey *et al.,* 2009) like hyperglycemia and obesity. Other studies indicated that diabetic individuals, especially those with microvascular and/or macrovascular complications, have high circulating levels of sRAGE (Thomas *et al.,* 2011 and Colhoun *et al.,* 2011). Thomas *et al.* (2015) reported an independent association between sRAGE levels and new-onset/worsening renal disease as well as all-cause mortality in a type 2 diabetic cohort. Thomas *et al.* (2011) also found consistency in the results with their previous study in type 1 diabetic individuals (Thomas *et al.,* 2011). The study reported a positive correlation between sRAGE and mortality. The results of their study highlighted the AGE/RAGE axis is potentially important in the pathogenesis of diabetic complications and associated outcomes.

In this study the association between sRAGE levels and FN3K genotypes was also investigated. Skrha *et al.* (2014) previously reported a link between FN3K SNPs (rs 3848403, rs1056534) with soluble RAGE (sRAGE). However, the present study observed no association between sRAGE level, glycation status and genotypes of FN3K SNPs (rs3848403 and rs1056534) in negative and positive G-gap patients.

These SNPs do not appear to have any impact on sRAGE level. Results of this study indicated that sRAGE levels did not differ by G-gap status, suggesting that sRAGE may not be as specific to the G-gap as AGE level and sRAGE may be a generalised marker of ill health.

Overall, the glycation gap could be developed due to the production of transcript splice variants that encode products of differing activity. It might be possible that during alternative splicing or the presence of SNPs in exon/intron boundaries could lead to intron retention or exon skipping at the mRNA processing stage. These events would significantly change the structure of the resulting protein.

Chapter 4

Investigation of relative telomere length by quantitative real-time PCR in G-gap negative and G-gap positive patients.

4.1. Introduction

Many factors, including lack of telomerase activity, cell division, and oxidative stress, contribute to telomere shortening (Saretzki and Von, 2002). With each replication cycle, telomeric DNA gradually shortens and can lead to replicative senescence or apoptosis (Teixeira, 2013). Cell senescence contributes to T1D and T2D pathogenesis and plays an important role in the associated complications of the disease (Sampson *et al*., 2006). High levels of oxidative stress, and consequently increased oxidative DNA damage have been identified in prediabetes, metabolic syndrome (Hansel *et al*., 2004 & Su *et al*., 2008), and in T2D (Sampson *et al*., 2006). Reactive oxygen species (ROS) levels rise in hyperglycaemia because of high incorporation of reducing equivalents into the mitochondrial electron transport chain (Brownlee, 2001). Glucose is highly catabolised in hyperglycaemia, resulting in the high production of reducing equivalents and rise in NADPH-to-NADP ratio. This ratio stimulates the NADPH oxidases activity and results in increased ROS formation (Raddatz *et al*., 2011). The combination of above scenarios and reduced antioxidant cell defence mechanism (superoxide dismutase and/or glutathione peroxidase activity is low) could contribute to high ROS levels in prediabetes, metabolic syndrome or diabetes.

Special deoxyribonucleic acid (DNA) structures “telomeres” protects the chromosomes from end-to-end fusion in conjunction with specialised proteins called shelterin. The human telomere sequence consists of TTAGGG repeats (Lee *et al.*, 2018) which facilitates telomere recognition and quantification. In somatic cells, telomeres shorten with age. The telomeric DNA is susceptible to oxidative DNA damage, so the high ROS levels in diabetes presumably lead to accelerated telomere erosion (Petersen, Saretzki and von, 1998 & Serra *et al*., 2000). Various

cell types, including beta cells, may undergo vast and premature senescence because of the above processes and consequently, apoptosis occurs which if affecting the secretion or action of insulin could cause the onset of diabetes. Previously, shorter telomeres have been reported in patients with type 1 diabetes (Jeanclos *et al.*,1998) and type 2 diabetes compared with healthy individuals (Adaikalakoteswari, Balasubramanyam, and Mohan, 2005, Sampson *et al*., 2006 & Uziel *et al*., 2007). Although shortened telomeres were shown to be associated with diabetes, and a significant negative correlation has been reported between glycated haemoglobin (HbA_{1c}) and β-cell telomere length (Tamura *et al.,* 2016), there are no data available on such a relationship with the G-gap. Therefore, this study investigated the relationship between relative telomere to single copy gene ratio (T/S) and the G-gap in patients who are G-gap negative and G-gap positive.

4.2. Material and Methods

4.2.1. DNA source

DNA was extracted from whole blood from patients with diabetes using the QIAamp DNA Blood Mini Kit (QIAGEN, # 51104). (“DNA extraction” in Chapter 2).

4.2.2. Methods for quantitative analysis of mean telomere length in the studied cohort

4.2.2.1. Terminal restriction fragment (TRF) length analysis via Southern blotting

Southern blotting was used to estimate the mean terminal restriction fragment (TRF) length. This is the traditional method for measuring telomere length in total human

genomic DNA samples. Briefly, this method uses restriction enzymes that cleave DNA in the subtelomeric regions in order to produce TRF. Electrophoresis was used to separate the cleaved DNA before being transferred to a membrane where hybridisation takes place with a labelled telomere-specific probe. Different algorithms are used to convert the signals to calculate the actual telomere length, described in kilobase pairs.

This method is inappropriate where an available amount of DNA was restricted. TRF also gives false positive results in the mean telomere length estimation by measuring highly variable sub-telomeric regions. These regions include restriction site polymorphisms or length polymorphisms, that create confusion in the identification of primary factors responsible for inter-individual variation in the mean length of the true telomeric repeat sequence.

4.2.2.2. Standard PCR vs quantitative polymerase chain reaction (qPCR) for telomere

Recently, PCR based methods allow multiple samples to be compared for their relative content of pure telomeric repeats with low amounts of DNA. The technique also reduces the time required for analysis. Quantitative PCR (qPCR) estimates only average telomere length and does not give any information about the shortest telomere or the distribution of telomere lengths within the cell (Baird, 2005). Cawthon (2002), introduced the simplest quantitative PCR-based method that solves the amplification problem due to the repetitive nature of telomeric sequences by designing a set of primers which bind specifically to human telomere regions and lead to the amplification of only telomeric (TTAGGG) repeats. Prior to this development, the repetitive telomeric sequence used primers that were

complementary, leading to the formation of primer dimers, where two primers bind to one another, resulting in the amplification of the primer sequence, rather than amplification of the target DNA from the patient. In Cawthon's method (Cawthon, 2002) specifically designed telomere primers anneal to genomic DNA during hybridisation in a way that only four consecutive out of six base pairs are hybridised, followed by two mismatched bases and preventing DNA polymerase to extend from the 3'-end of each primer, thus eliminating the possibility for primer-dimer formation.

The principle for quantification of telomere repeats is that there will be an increase in fluorescence with more potential primer- annealing sites when there are longer telomeres, and hence decrease in the cycle numbers to reach a given threshold, and vice versa with shorter telomeres. Cawthon (2002) indicates that a factor that is used to determine the relative telomere length for each DNA sample, by which the sample differed from a reference DNA sample, is its ratio of telomere repeat number to single gene copy (SCG) number. Two quantitative PCRs for each sample, one for the telomere repeats (T) and one for an SCG (S), were performed (Figure 4.1) and Ct values were calculated respectively. The use of SCG (Rplp0) helps to normalise DNA loading and control variation in initial DNA quantity, the ratio of these two Ct values is used to calculate relative average telomere length. The T/S ratio of 125 of the samples was calculated. PCRs were performed in a 7500 FAST qPCR (Applied Biosystems) (refer Chapter 2 for detail). The melting curve produced for telomere assays suggested that only TTAGGG/SCG was amplified (Figure 4.2). On both telomere and SCG plates, all samples were run in triplicate. Each 96-well plate contained 15ng of DNA per well, from G-gap negative and G-gap positive patients. The Ct value represents, where a number of PCR cycles is required to amplify the telomeric DNA TTAGGG sequence in a given DNA sample to accumulate enough

product to pass a set threshold level. Small Ct values indicate that samples have more of the TTAGGG sequence and have longer telomeres as compared to samples with larger Ct values. This method provides the mean relative telomere content instead of actual telomere length.

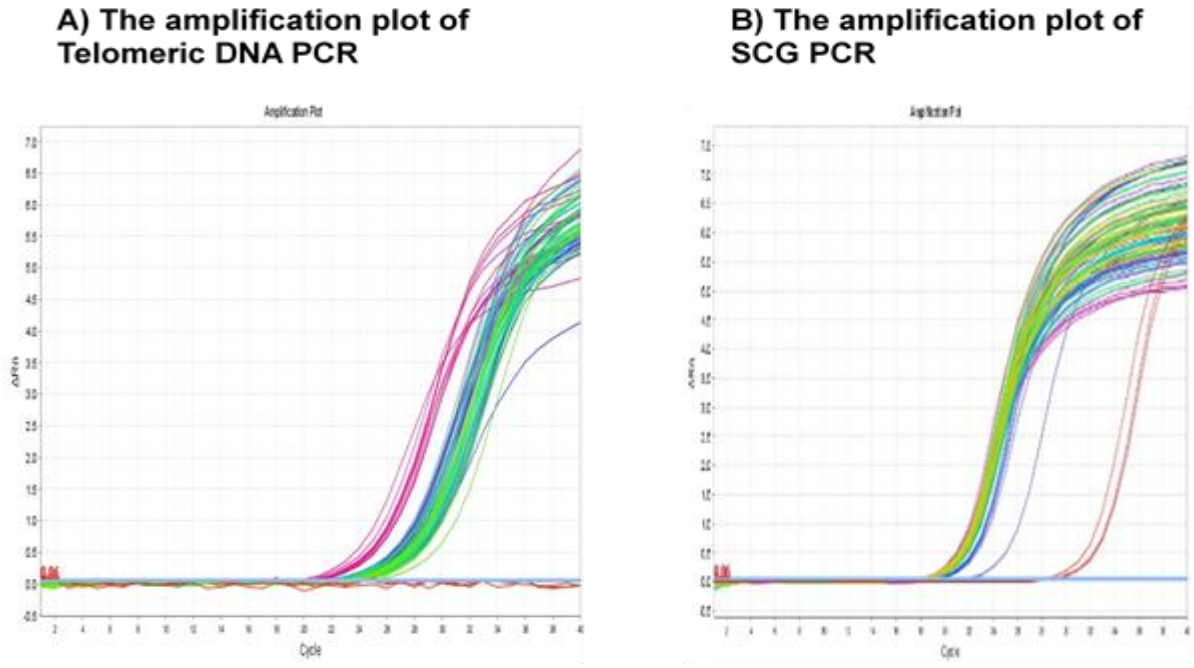


Figure 4. 1 Figure showing an amplification plot of telomere (A) and amplification plot of Single Copy Gene (SCG) (B)

A) Melting curve of telomeric DNA

B) Melting curve of SCG

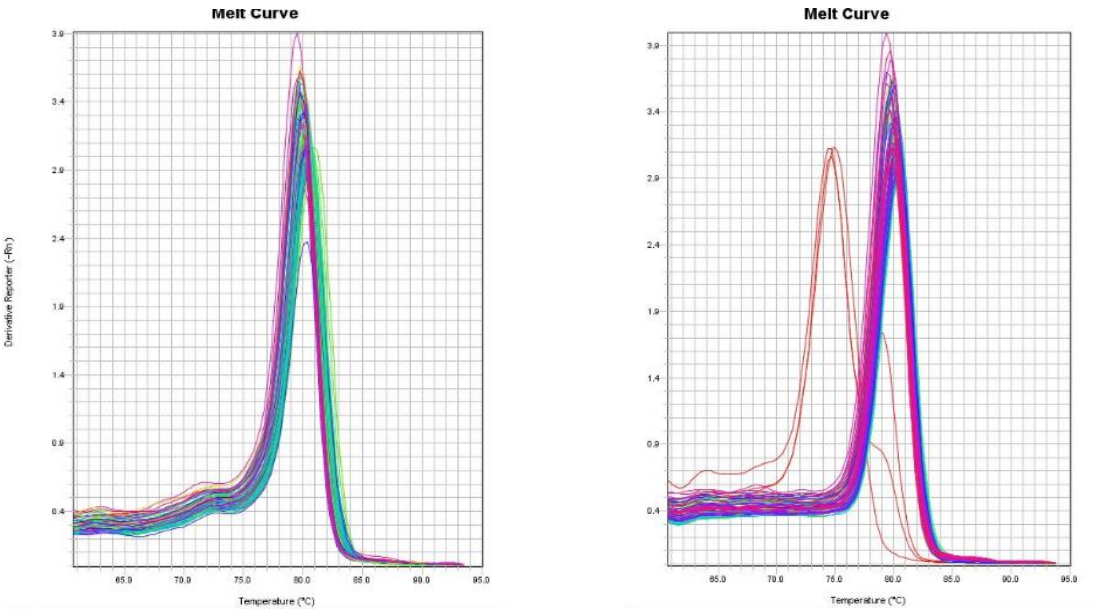


Figure 4. 2 Figure showing the melting curve of telomere (A) and melting curve of Single Copy Gene (SCG) (B)

4.2.3. Rationale behind the cycle threshold (Ct) value

In real-time PCR, at the crossing point (cycle threshold (Ct)), the fluorescence of a sample is detectable because it exceeds above the background fluorescence. This crossing point is very much dependent on the initial concentration of DNA. More amplification cycles are needed if there is a lower initial DNA concentration present in samples than with initial high DNA concentrations. Previous studies showed that with 35 ng/µl (Cawthon, 2002 & Gil and Coetzer, 2004) of the initial concentration of DNA the number of cycles used was 20 for telomere repeats and 30 for Rplp0 gene when quantitative real-time PCR was used for analysis of telomere length. The initial concentration of genomic DNA extracted from the whole blood of G-gap negative and G-gap positive patients was 15ng/µl and the number of cycles increased to 40 for both telomere length amplification and Rplp0 amplification. Rplp0 was a

“housekeeping” gene and this gene has constant expression levels amongst the samples cohort compared (Wang *et al.*, 2012). To monitor the amplification efficiency, a second primer pair was included that was specific for a “housekeeping” gene in the reaction. Acceptable quality of target nucleic acid and reaction components was verified by the amplification of the housekeeping gene.

4.3. Calculation and statistical analysis

Statistical analysis were performed with SPSS. The sample means (T/S) of the 2 groups of different G-gap groups were compared using an independent samples t-test with equal variances not assumed and the criteria for significance has been set ≤ 0.05 and the test is two-tailed. Descriptive statistics were calculated (the mean and standard deviation (SD)). Univariate analysis was used to compare the association of continuous variables with other factors i.e. age, gender, ethnicity, BMI, type of diabetes or G-gap status.

4.4. Results

After running the qPCR, for each sample with negative (<-0.5%, n=73) and positive (>+0.5%, n=52) G-gaps, mean Ct values were calculated by taking an average of triplet values. Subsequently, for each sample a mean T/S ratio (T stands for telomere, S stands for the single-copy gene) was calculated (Figure 4.3). Levene`s test for equality of variances showed no violations (p=0.116) whilst mean±SD in negative and positive G-gap groups were 1.097±0.009 and 1.075±0.010 respectively. Normality test was non-significant in the negative g-gap group (Kolmogorov-Smirnov; p =0.2) whilst it was found significant in the positive g-gap group (Kolmogorov-Smirnov; p =0.004). Non-parametric Mann Whitney U test was performed on SPSS, showed a non-significant difference (P = 0.17). These results

showed that the distribution of the T/S ratio is the same across negative and positive G-gap groups. In univariate analysis telomere length with duration of diabetes (p=0.029) was significantly associated but not with age, gender, ethnicity, BMI, type of diabetes or G-gap status (p>0.05).

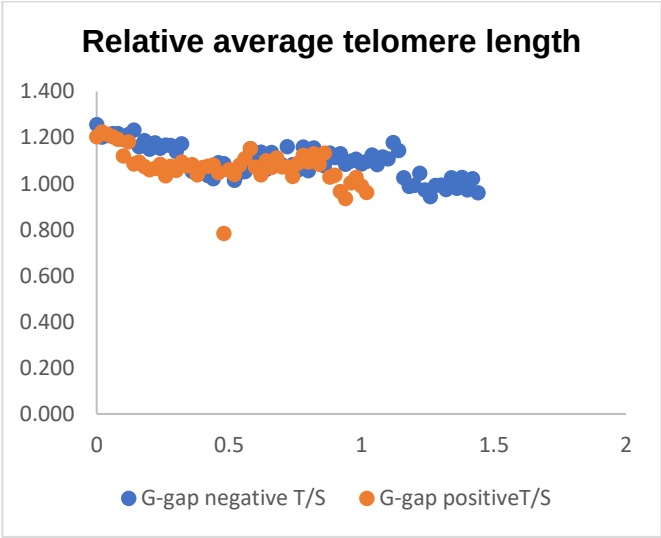


Figure 4. 3 Relative average telomere length. Data are shown as telomere to single-copy gene (T/S) ratio, G-gap negative (n=73) vs G-gap positive (n=52).

4.5. Discussion

Telomeres consist of TTAGGG tandem repeats (Stohr and Blackburn, 2010). The length of telomere of somatic cells is progressively shortened with each cycle of cell division as a function of the donor's age (Hornsby, 2007) because of the end-replication problem. In the end-replication problem, the DNA replication machinery is unable to synthesise the very end of the 3′ strand (Kim *et al*., 2003). *In vitro* studies indicated that the telomere length predicts the replicative history of cells (Hills *et al.*, 2009), because with a continuous replication process the telomere length progressively shortens, and cells undergo permanent cell cycle arrest and become senescent. This happens when telomere length reaches a critical length

(Mollano, Martin and Buckwalter, 2002). The "end replication problem", leads to the attrition in telomeres length.

The "end replication problem" involves the incomplete replication of the ends of linear chromosomes (Maestroni, Matmati and Coulon, 2017). Telomerase, a ribonucleoprotein complex counteracts telomeric attrition. This enzyme uses its RNA component as a template to add telomeric repeats onto the ends of chromosomes (Lindstrom *et al.*, 2002). Telomerase expression is very low in somatic cells. Therefore, the activity of the telomerase enzyme is insufficient to counterbalance the progressive attrition in the telomere length. Telomere length has therefore been considered an indicator of the replicative potential of cells *in vitro* as well as *in vivo*. Research reports demonstrated an association between telomere shortening in white blood cells (WBCs) of Type 1 (Jeanclos *et al*., 1998) and Asian Indian Type 2 diabetic patients (Adaikalakoteswari, Balasubramanyam, and Mohan, 2005). Telomere shortening might be a long-term risk marker of macro-vascular complications in negative and positive G-gap diabetic patients. This assumption was supported because Adaikalakoteswari *et al*. (2007) demonstrated that type 2 diabetic patients with atherosclerotic plaques had significantly shorter telomeres length and shorter telomere was associated with increased markers of systemic inflammation and oxidative stress (Adaikalakoteswari *et al*., 2007).

To understand normal cellular biology, telomere biology and function is of great significance. For understanding telomere biology, variation in telomere length and biological controls that control this length are of greater importance. Traditionally different methods were used to determine the telomere lengths, that were costly in terms of money and time. Usage of a standard PCR method for amplification was

limited because of the repetitive nature of telomeric DNA sequence, but this issue was resolved with the development of primers for the human telomere assay. The DNA used was extracted from whole human blood. Telomere length varies between various tissues (Campisi *et al.*, 2001), therefore measuring telomere length in blood equates with telomere length in other tissues (Saretzki and Von, 2002). Previous study found a strong correlation between telomere length in leukocytes and the telomere length in skeletal muscle, skin, and subcutaneous fat (Daniali *et al.*, 2013). Telomere length shorten at different rate in muscles and heart as their replication activities are low. In blood cells or skin, where the replication activities are higher, telomere length shortening rate is different. In addition to this, different pathophysiological condition associated with ageing also affect the rate of telomere length shorten in various tissues. Therefore, telomere length study in peripheral leukocytes is more common (Wang, 2018). Although short length telomeres have been reported in T1D and T2D, no difference between the studied groups G-gap negative and G-gap positive was observed and no relationship was found between telomere and glycation gap in this study.

The T/S data were normally distributed in negative G-gap (Kolmogorov-Smirnov; p = 0.2) but abnormal distribution was found in positive G-gap (Kolmogorov-Smirnov; p =0.004) (Figure 4.3). The observed abnormal distribution may be due to one anomaly/abnormality (Figure 4.3). When this anomaly was removed, still data were not normally distributed (Kolmogorov-Smirnov; p =0.001) in positive G-gap. Abnormal distribution of T/S data between studied groups could be due to the unequal sample sizes (G-gap negative; n=73 and G-gap positive; n=52). The data were not normally distributed between the groups. Therefore, the normality test is no longer valid. A non-parametric Mann–Whitney U test on SPSS was conducted,

which showed a non-significant difference (P = 0.17). These results showed that the distribution of the T/S ratio is the same across negative and positive G-gap groups.

The qPCR assay may produce different results if there are differences in the initial sample DNA concentrations. This was addressed by measuring the concentration of DNA in the samples shortly before qPCR analysis and preparing 125 samples with the same DNA concentrations (15ng/µl). Additional care was taken during pipetting to avoid any uncertainty for sample preparation before the qPCR run. For the present study, it is potentially impossible to completely exclude the impact of different sampling dates in individual telomere length but as the T/S data were normally distributed any such impact was minimal, if it exists. T/S data were not related to age as it might be linked with the biological age rather than chronological age.

Chapter 5

RNA-Seq from peripheral blood for the study of differential gene expression and transcript variants in G-gap negative and G-gap positive groups

5.1. Introduction

Nucleated, enucleated cells and cell-free ribonucleoproteins contain RNA within peripheral blood. Alongside this blood contains cell-free RNA that circulates in membrane vesicles i.e. apoptotic bodies, microvesicles, and exosomes. Physiological and pathological processes taking place in different cells and tissues of human cause specific changes in the RNA profile of whole peripheral blood (Yanez *et al*., 2015 & Mohr and Liew, 2007). Consistently, it was found by comparing the transcriptome of peripheral blood with the genes expressed in nine different human tissues such as brain, colon, heart, kidney, liver, lung, prostate, spleen, and stomach that more than 80% expression was shared with any given tissue (Liew *et al*., 2006). Peripheral blood is an ideal surrogate tissue because it continuously interacts with the whole body. This interaction may trigger specific changes in the gene expression in blood as a result of subtle changes within cell and tissue that occur in association with the disease.

Gene expression is critical in the understanding of cellular phenotypes and how it affects phenotypic demonstrations for the characterisation of disease risk genes (Altshuler, Daly and Lander, 2008 & Manolio *et al*., 2009). It is well understood that most part of the genome is transcribed into RNA and only a small proportion (1-2%) codes for a protein (ENCODE Project Consortium, 2012). Messenger RNA (mRNA) not only consists of the coding region but also contains untranslated sequences at both it's 5’ and 3’ ends that are important in regulatory function (Hinnebusch, Ivanov and Sonenberg, 2016). There are also many non-coding RNAs that play an important role in intracellular signalling and transcriptional modification (Cech and Steitz, 2014). It appears that the correlation of mRNA with its downstream protein-coding counterpart is weak which suggests that post-transcriptional regulation takes

place before translation occurs (Albert and Kruglyak, 2015). In the past, for gene expression studies, the widely recognised method was hybridisation-based microarray (DeRisi *et al*.,1996; Fodor *et al*.,1993). The technique was able to probe thousands of transcripts simultaneously, but its dependence on existing gene models and potential cross-hybridisation to probes with similar sequences could produce bias results.

Sequencing-based method i.e. RNA-sequencing (RNA-Seq) is another approach for accurately quantifying gene expression via transcriptome. This technique provides digital readouts for depicting and quantification of transcriptomes (Bentley *et al*., 2008; Lister *et al*., 2008; Mortazavi *et al*., 2008; Nagalakshmi *et al*., 2008; Wilhelm *et al*., 2008). RNA-Seq technology is based on deep sequencing that ensures each genomic region in the samples is sequenced repeatedly (Meyerson, Gabriel, and Getz, 2010). In this study, RNA-Seq involved the investigation of the complex aspects of transcriptomics, e.g. transcript variants, identification of differentially expressed genes and to investigate their association with disease, pathways and gene ontology and the putative mechanisms to explain observed gene expression changes between G-gap negative and G-gap positive groups.

RNA extraction, Selection/depletion of RNA (i.e. filtration of RNA with 3' poly(A) tails as these are mature, processed, coding sequences or depletion of ribosomal RNA (rRNA) because it is over 90% of the RNA in a cell, which if not removed would submerge other data in the transcriptome), cDNA synthesis with adaptors attached, and cDNA pair-end sequencing in a high-throughput manner to get short sequences typically 30 to 400 nucleotides in length. After sequencing the short sequences are then aligned to a reference genome and reference transcriptome. These sequences

help in the identification of exon regions from which the reads are generated or assembled de novo where genomic sequence does not exist. Counting the number of reads aligned to each locus in the transcriptome assembly step provides the quantification for the expression level for exons or gene. Nearly 95% of human's multi-exon genes are produced via alternatively spliced events and are reported in global surveys of alternative splicing (Pan *et al*., 2008). There is no genomic information regarding gene expression differences and transcript variants at the whole blood transcriptome level of negative and positive G-gap patients with diabetes. To solve this, the characterisation of specific gene expression related to the G-gap was performed in the whole blood of the studied cohort using RNA-Seq (RNA sequencing) technology. RNA was extracted from whole blood. The blood cells not only act as the first line of defence of the immune system and flow through the whole body (Chaussabel, Pascual and Banchereau, 2010) but the transcriptome extracted had a great resemblance with various tissues (Mohr and Liew, 2007). Thus, changes in gene expression in those tissues as well as in pre-erythrocytes (normoblasts) and other pathways constituting the G-gap might give a clear view via whole blood gene expression. RNA sequencing is one of the next generation sequencing (NGS) technology that made it possible to explore differentially expressed genes (Wilhelm and Landry, 2009) alongside assessing mRNA transcription patterns for all the genes in G-gap negative and G-gap positive group.

5.2. Materials and Methods

5.2.1. Sample collection and processing

RNA extraction and removal of chromosomal DNA were performed as reported in Chapter 2. RNA concentration and purity were assessed using a NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific). Samples were sent to Oxford Genomic Centre for library preparation (refer to chapter 2) and sequencing, where RNA was quantified using RiboGreen assay and the size profile and integrity analysed on the 2200 TapeStation (Agilent, RNA ScreenTape) (details are in Chapter 2). Cytoplasmic ribosomal RNA depleted using Ribo-Zero rRNA removal kit and library preparation was completed using TruSeq Stranded Total RNA Human kit (See Chapter 2). Sequencing was performed in the same facility (Oxford Genomic Centre).

5.2.2. Illumina sequencing Hi-Seq 4000 platform

Each sample was sequenced to a target depth of >45 million paired-ended 75 bp reads using a multiplex strategy on an Illumina Hi-Seq 4000 (Illumina, San Diego, CA, USA) by using PE Cluster kit and 150 cycle SBS kit. Illumina sequencing machine workflow consists of 4 stages: Preparing sample, cluster formation, sequencing of sample and data analysis. Sample preparation (Total RNA) comprises of series of stages to the RNA fragmentation, cDNA conversion, and formation of cDNA library (refer to Chap 2). cDNA library was loaded to six-lane flow cell where cDNA fragments hybridised to oligonucleotide primers that are complementary to adapters attached to cDNA. Clusters formed via bridge amplification. Illumina sequencing technology for RNA-Seq based on the standard dideoxy method (Lasken and McLean, 2014). During the assay for the sequencing

of each cluster flow cell inside the sequencing machine was flooded with DNA polymerase and nucleotides. These nucleotides were fluorescently labelled (the colour corresponding to the nucleotide base) and attached to a terminator at the 3'-end. The different fluorescent molecules emit four different wavelengths of light whilst the terminator inhibits elongation reaction. The sequencing cycle continues where only complementary nucleotides in the template strand added and the remaining i.e. terminator and fluorescent molecule washed away after fluorescent signals read correctly. This process is repeated i.e. adding one nucleotide at a time and imaging in between with high-resolution camera. The base-calls represented by the fluorescence intensity whilst the length of the reads indicated by cycle number and number of reads by clusters numbers (Bullard *et al*., 2010).

Sequencing billions of reactions are possible with the Illumina system at the same time. It produces approximately 200 nucleotides long individual sequencing read which is shorter though as compared to sequencing reads produced from other sequencing instruments but parallel sequencing in vast amount is possible by it. All sequencing data in this thesis are paired-end as pair end data provides accurate mapping and can quantify duplication rate with reasonable reliability.

5.2.3. Processing of RNAseq Data

Raw data from RNA sequencing experiment in FASTQ format was post-processed into Nottingham's Advanced Data Analysis Centre (ADAC). FastQC version 0.11.5 was used to display the quality of reads. After that raw reads were aligned to a reference using HISAT2 v 2.1.0. HISAT2 mapped the reads to a reference with the use of one global FM index and numerous small local FM indexes and identified transcript splice sites. HISAT2 stored the output as SAM files format. The reference

was downloaded from ENSEMBL (ftp://ftp.ensembl.org/pub/release95/fasta/homo_sapiens/dna/Homo_sapiens.GRCh38.dna.primary_assembly.fa.gz). StringTie v 1.2.3 program was used for the assembly of alignments into transcripts with human genome annotation. The human genome annotation was downloaded in the form of a GTF file from the web address (ftp://ftp.ensembl.org/pub/release-95/gtf/homo_sapiens/Homo_sapiens.GRCh38.95.gtf.gz). This assembly step generated an estimated transcript length that was stored in the form of GTF file for each of the 32 samples. Transcript assemblies were merged from all G-gap negative and G-gap positive samples by StringTie-merger in GTF format. Estimation of assembled transcripts abundance in each sample on the basis of gene expression values, the principal component analysis was done using 'cmdscale' (Classical multidimensional scaling) function. Differential expression analysis of gene, as well as a transcript, was performed using the Ballgown v 2.14.1 program. The statistical significance of fold changes was calculated by comparison of the read values of G-gap negative and G-gap positive groups. The normalisation of the data to correct distributional differences within the read counts, i.e. differences in total counts (sequencing depths), and within sample gene-specific effects i.e. gene length was performed using the FPKM function of Ballgown program.

5.2.4. Statistical analysis

Differential expression analysis was performed by using the stattest function of Ballgown. This is an open-source software based on FPKM values in a linear model (Pertea *et al*., 2016). The log in the model converts the count data and uses linear models for the testing of differential expression at the gene, transcript, exon or junction level. Contrasts between negative and positive G-gaps were performed.

The adjusted p-value <0.05 considered statistically significant. Differentially expressed genes (DEGs) and transcript variants with p-value <0.05 and fold change >1.5 were identified and submitted for gene enrichment to R Bioconductor package Ballgown. Downstream pathway analysis was performed using the DAVID software.

5.2.5. Database for Annotation, Visualisation and Integrated Discovery (DAVID)

DAVID is a bioinformatic tool that performs functional and pathway enrichment analysis of DEGs (refer to 2.1.8.7). DAVID identifies the molecular functions, biological processes and associated diseases amongst differentially expressed genes. The program not only calculates the enrichment of a gene associated with a specific pathway but also provide a statistical significance value of this enrichment. Enrichment will be considered statistical valuable when more genes in the list associated with a specific pathway. This statistical significance also depends on how many genes in total are in a pathway for e.g. having 5 differentially expressed genes in a pathway of 200 genes might have a lower significance than 5 genes in a pathway of only 50 genes.

5.3. Results

5.3.1. RNA quality analysis

5.3.1.1. Nanodrop

Downstream analysis requires intact and pure RNA. The purity of RNA control samples was analysed by NanoDrop. The absorbance at 260 nm (A260) provided yield and concentration of RNA. The average concentration of RNA samples was approximately 186.53ng/µl and 195.78ng/µl in G-gap negative and G-gap positive groups respectively. Absorbance ratio 260/280 was used to assess the purity of the samples. A range between 1.8 and 2.0 indicated adequate purity and is generally acceptable (Desjardins and Conklin, 2010). Contamination with proteins or phenol reduces this ratio because they absorb strongly near 280nm whilst alkaline conditions increase this ratio to over 2. The 260/280 ratio for all studied RNA samples in this study was in between 1.97-2.23 in both G-gaps groups. Another measure of nucleic acid purity is represented by the ratio 260/230 with a range of 2.0- 2.2. A lower 260/230 ratio could be due to the presence of ethylenediaminetetraacetic acid (EDTA), carbohydrates or phenol, which absorb near 230nm. In the present study, the RNA samples had desirable 260/280 ratios but 260/230 was lower (0.81-2.23) in both groups. An elution buffer and a phenol-based reagent were used to extract RNA from whole blood samples, the low 260/230 might be due to the contamination from salt or phenol reagent. To increase this ratio and to get the RNA of sufficient quality sodium acetate precipitation of small nucleic acids can be used.

5.3.1.2. Agilent 2200 TapeStation system

RNA samples need to be structurally intact to produce reliable results (Schroeder *et al*., 2006) in RNA sequencing. Otherwise, full-length cDNA could not be produced with the degraded RNA. Previous studies used 28S:18S ribosomal RNA ratio to determine the level of degradation, but due to the subjective nature of this method that produces discrepant results (Schroeder *et al*., 2006) a reliable method was introduced to detect the integrity of RNA samples by using the automated quality control instrumentation (Agilent 2200 Tape system). The tape station analyses the whole electrophoretic trace of the RNA and uses an algorithm to report on the integrity of each sample (Mueller, Lightfoot and Schroeder, 2004). The algorithm classifies RNA based on a scale of 1 to 10, where 1 indicates an extremely degraded sample whilst 10 represents perfectly intact RNA. RNA samples in this study had values ranging between 2.6- 6.0 RIN. RNA samples were then used in cDNA synthesis.

5.3.2. RNA sequencing (RNA-Seq) of G-gap patients' transcriptomes

RNA showed variable quality. Therefore, the proper library preparation kit was used to compensate for the low-input samples (see "TruSeq Stranded total RNA (Illumina) library preparation" section in chapter 2). After quality control checks, an average of over 45 million paired-end reads was obtained from each sample (n = 32).

5.3.3. FastQC report of RNA sequencing

ADAC bioinformatics unit of Nottingham University checked the quality of the RNA sequencing data using FastQC software. Per base, sequence quality is based on Phred quality scores. Sequencer uses Phred quality scores estimations for the quality of a base-call. Quality score (Q) of each given base was calculated on the basis of the probability of the base call being wrong. A sequencing quality score 10 (Q10) shows an error rate of 1 in 10 and inferred base call accuracy is 90%, Q20 represents an error rate of 1 in 100 and the base call accuracy is 99% whilst Q30 represents an error rate of 1 in 1000 and the corresponding base call accuracy 99.9%. The FastQC plots for forward (Figure 5.1) and reverse (Figure 5.2) reads represent the highest and lowest quality. Reads indicated consistent high-quality sequencing, with Q scores approaching 40. This indicated that the sequencing of cDNA libraries was highly accurate.

Overall quality of the reads was very good. Therefore, no trimming was done. Paired-end reads had a uniform length of 75bp and were not overlapping. The GC content was 45%. The range of average GC-content in human genomes is 35% to 60% across 100-Kb fragments, with a mean of 46.1 (Romiguier *et al.,* 2010).

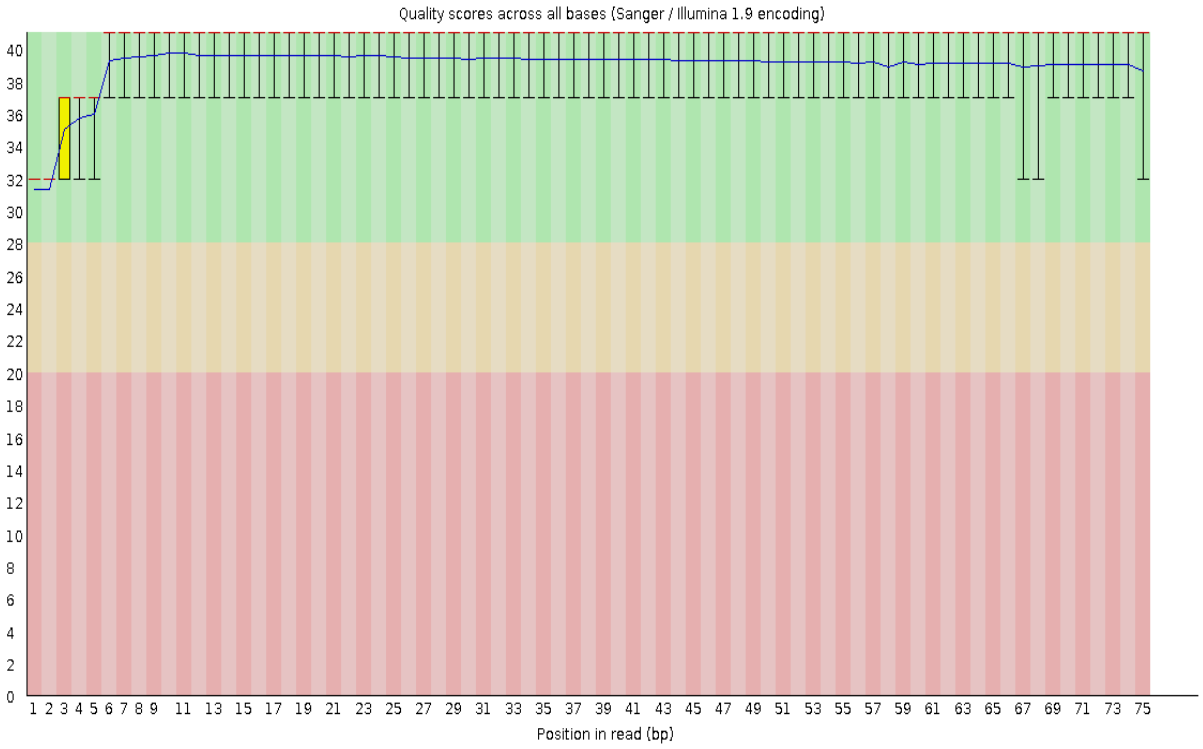


Figure 5. 1: Per base sequence quality of forward reads produced by FastQC (version 0.11.5). The quality scores are on the y-axis. The higher score indicates the better the base call. The y-axis has three parts on the basis of the background of the graph i.e. very good quality calls (green), calls of reasonable quality (orange), and calls of poor quality (red).

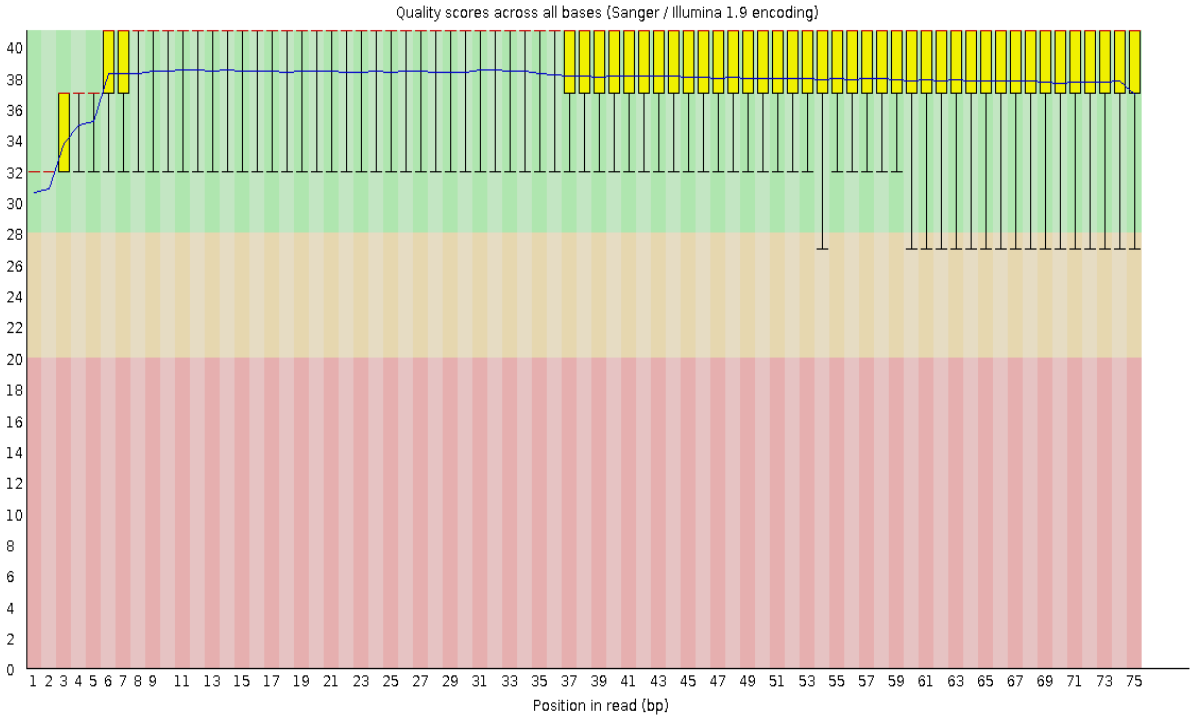


Figure 5. 2: Per base sequence quality of reverse reads produced by FastQC (version 0.11.5). The quality scores are on the y-axis. The higher score indicates the better the base call. The y-axis has three parts on the basis of the background of the graph i.e. very good quality calls (green), calls of reasonable quality (orange), and calls of poor quality (red).

5.3.4. Principal component analysis (PCA)

PCA was conducted to determine whether samples in each G-gap group clustered with each other or other groups. First, HISAT2 was used to count reads that uniquely mapped to one gene, and the data were then imported into StringTie to generate PCA plots. The PCA results demonstrated that most samples clustered together, regardless of the G-gap grouping Figure 5.3.

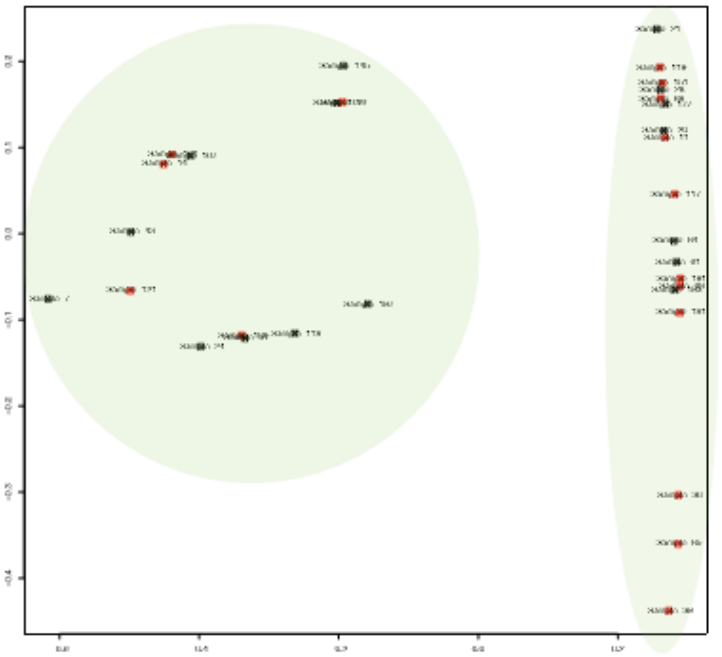


Figure 5.3: Principal component analysis (PCA) for all samples in G-gap groups. (Black: negative G-gap group; Red: positive G-gap group).

5.3.5. Transcript length and differential expression analysis of genes and transcripts

StringTie assembled the mapped reads into transcripts. These transcripts were of various lengths. The average length of the transcript was 2741.82 bp whilst the longest transcript consists of 205012 bp. The distribution of transcript length is shown in Figure 5.4. Overall, 64451 genes mapped to the human genome. Amongst them 103 genes and 342 transcripts were differentially expressed ($p < 0.05$, > 1.5-fold change). Of 103 DEGs 61 were upregulated in G-gap negative and 42 were upregulated in G-gap positive figure 5.5. Out of 103 DEGs 14 produced alternatively spliced transcript variants. On average 1 splice variant was encoded by a single gene whilst on higher side eight splice variants expressed from a single gene (*EPB41*). Out of 8 transcripts, 6 were upregulated in negative G-gap whilst 2 transcripts were upregulated in positive G-gap.

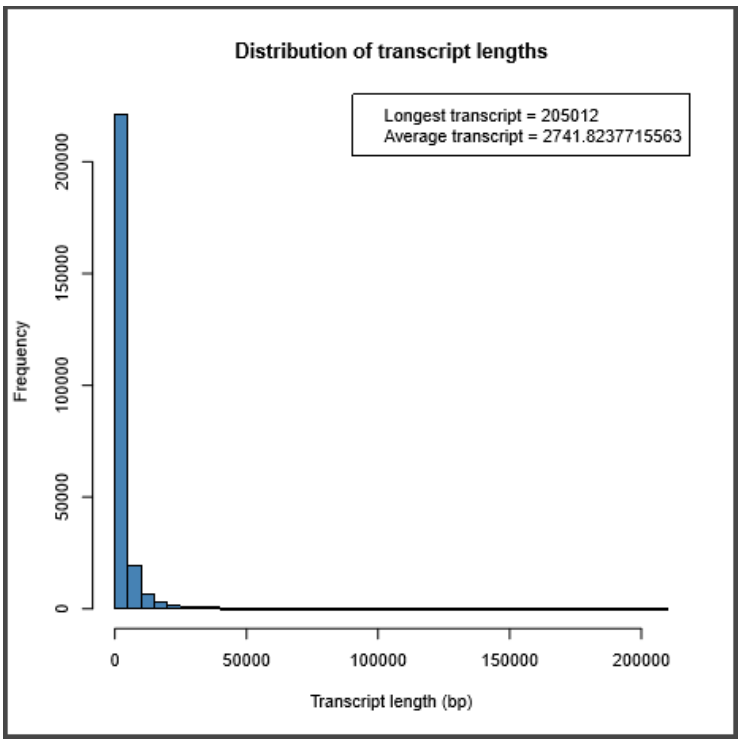


Figure 5. 4: Distribution of transcript length and number of sequences from G-gap negative and G-gap positive samples. The longest transcript consisted of 205012 bp and the average transcript length was 2741.82 bp. StringTie (v1.2.3) was used to assemble the mapped reads into transcripts.

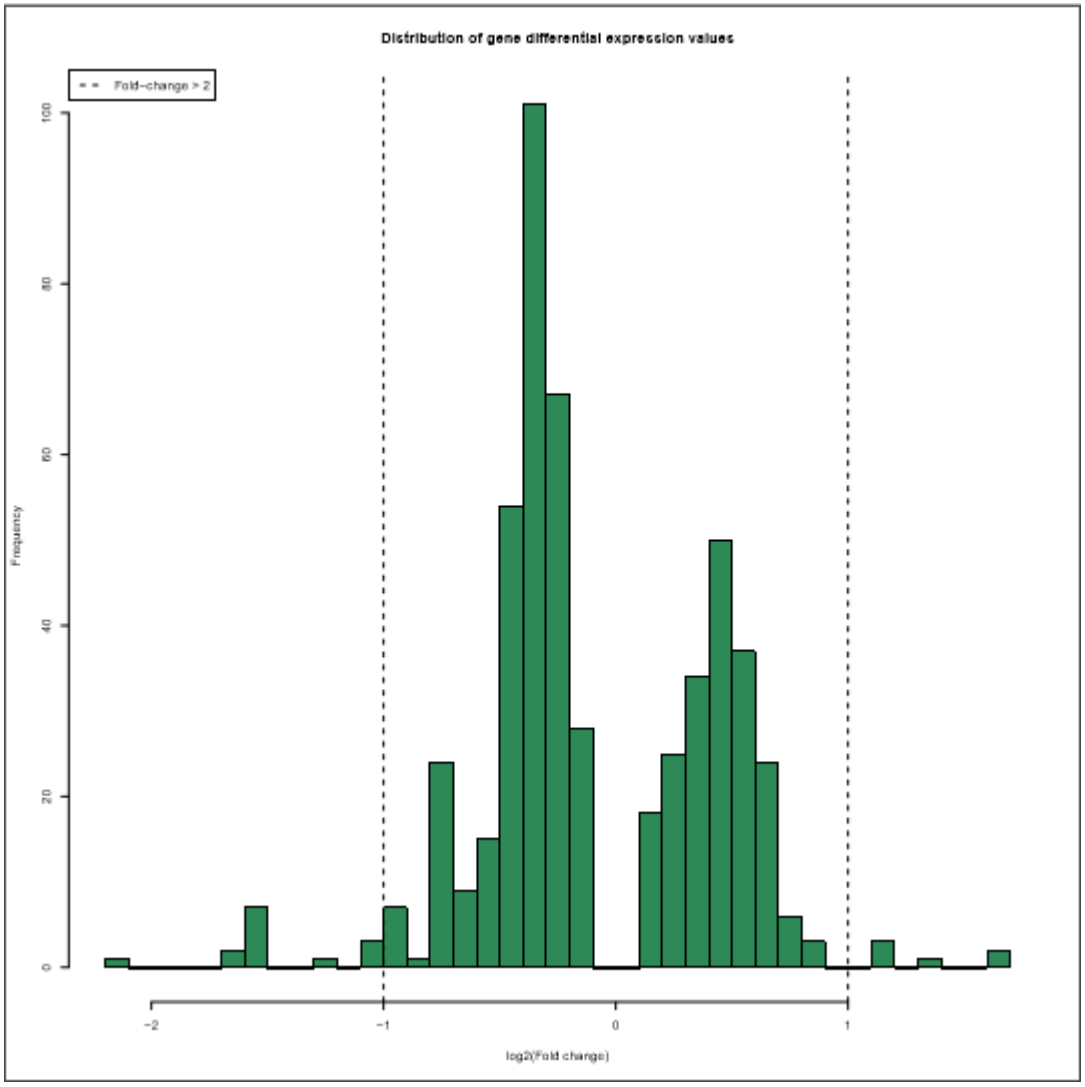


Figure 5.5: Distribution of differentially expressed genes between G-gap negative and G-gap positive groups. Overall, 103 genes were differentially expressed. Of these 61 were upregulated in negative G-gap whilst 42 were upregulated in G-gap positive samples.

5.3.6. Screening of differentially expressed genes using DAVID

Disease analysis demonstrated the DEGs from G-gap groups were enriched in seven diseases including blood pressure, glycated haemoglobin, erythrocytes indices, mean corpuscular volume, carotid stenosis, mean corpuscular haemoglobin and posttransplantation diabetes mellitus (PTDM) (Table 5.1), Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis demonstrated that the DEGs were enriched in four key pathways including viral carcinogenesis, ribosome, phagosome and dorso-ventral axis formation table (Table 5.2) while gene ontology showed biological processes that were enriched in the identified DEGs. In the biological process group, DEGs were mainly enriched in metabolic process, cellular component analysis and developmental process (Table 5.3).

Disease	Count	P value	Genes
Blood pressure	8	9.1E-1	*CITED2, PHF11, RAP1A, TBC1D4, CPVL, CCDC71L, EPB41L3 and GLRX5*
Glycated haemoglobin	5	9.2E-1	*CD44, CPVL, CDC123, SCAMP2* and *SYNE2*
Erythrocytes indices	4	8.9E-1	*CITED2, MAX, TYMP* and *TFRC*
Mean corpuscular volume	3	9.0E-1	*CITED2, TYMP* and *TFRC*
Carotid stenosis	3	7.8E1	*CX3CR1, TBC1D4* and *EPB41L3*
Mean corpuscular haemoglobin	2	9.8E-1	*CITED2* and *TFRC*
Post transplantation diabetes mellitus (PTDM)	2	9.9E-1	*CDC123* and *GCLC*

Table 5.1: Enriched disease association of differentially expressed genes (DEGs). Eight genes were associated with blood pressure whilst in posttransplantation diabetes mellitus, two genes were enriched (P≤0.05).

Pathway	Count	P value	Genes
Viral carcinogenesis	7	1.8E-2	*LYN*, *ACTN1*, *GRB2*, *HNRNPK*, *RHOA*, *ATAT5A* and *YWHAB*
Ribosome	5	4.9E-2	*RPL37*, *RPL41*, *RPS20*, *RPS26* and *RPSA*
Phagosome	5	6.6E-2	*CLEC7A*, *CTSS*, *CORO1A*, *NCF4* and *TFRC*
Dorso-ventral axis	3	3.0E-2	*ETS1*, *GRB2* and *NOTCH2*

Table 5.2: Enriched pathways of differentially expressed genes. In Viral carcinogenesis, seven genes were enriched whilst in Dorso-ventral axis, three genes were enriched (P≤ 0.05).

Biological processes	Gene count	P value
Metabolic process	88	9.7E-2
Cellular component analysis	60	6.4E-3
Developmental process	55	1.7E-2
Immune system process	32	2.9E-2
Multi organism process	28	8.1E-2
Biological adhesion	25	1.0E-3

Table 5.3: Gene ontology analysis of DEGs in different biological processes. A large number of DEGs genes were enriched in the metabolic process (88) whilst biological adhesion was enriched with 25 DEGs.

5.4. Results validation

Verifying the data from RNA sequencing analysis is of major concern (Zheng, Chung and Zhao, 2011). Possibilities that might influence the results include technical error during the blood collection phase or correct nucleotide quantification during RNA sequencing. The second possibility might arise during analysis whilst a third possibility could be the misinterpretation of the results.

To overcome these situations, the blood collection from patients was done by a professional and all the information and procedures were well documented. To overcome the second possibility obvious variations were eliminated whilst minor variations that could ± 0-10 % change the results (Rapaport *et al*., 2013) were ignored. However, no possible sources of error were noticed for the validation of results. RTqPCR could be employed by using the specific primers for the gene that were identified in RNA-sequencing in the remaining RNA extracted samples from patients and it would be possible to get the results that were nearly similar. However, RNA-Seq uses different algorithms and is more sensitive therefore this partly explain the difference between the results from two methods.

5.5. Discussion

The G-gap is common in patients with diabetes, However, the cause and underlying molecular events of G-gap are not clear. However, the study of transcriptome may contribute to understanding the underlying mechanism of G-gap occurrence and consistency. In this study, differentially expressed genes were enriched mainly in blood pressure and glycated haemoglobin A. These conditions coexist with diabetes. Studies showed that up to 75% of patients with diabetes suffer high blood pressure (Campbell *et al.,* 2009). Amongst DEGs linked to blood pressure and glycalated haemoglobin A, the *CPVL* gene was common to both diseases. This gene was upregulated in negative G-gap group. Hu *et al.* (2011) investigated that *CPVL* variant was associated with diabetic retinopathy.

According to GO and KEGG pathway enrichment analysis, various pathways were significantly downregulated in G-gap groups. Four pathways (Viral carcinogenesis, Ribosome, Phagosome and Dorso-ventral axis) came out top in the bioinformatic analysis in which DEGs were enriched. The identified pathways are important but the statistical significance of them appears quite small. These pathways are important because of their link with RBCs level (which could be affected by infection), ribosome (involves in the translation process) and Phagosome (involved in normal homeostasis of the immune system). These pathways are based purely on the difference between negative and positive G-gap groups. However, if the pathway analysis is performed based on gene expression profiles of the samples that cluster together under the PCA then the pathways that come out may be very different. Obviously, there is still the issue of what might be a more significant driver of gene expression leading to G-gap in these patients e.g. sex, age or ethnicity that is making them cluster other than G-gap basis. This requires further investigation.

5.4. Conclusion

In this study, RNA sequencing was performed for G-gap groups. Amongst a total of 64451 genes that were identified, 103 genes were differentially expressed in a significant manner whilst 342 differentially expressed transcripts were identified between negative and positive G-gap groups. It is important to understand the mechanism for the differences in diabetes complications by identifying the relevant pathway in which these genes play an important role. This would improve diabetes control and associated complications as well as enlightens the effective and specific ways for the prediction of diabetic complications in G-gap groups and to cure these.

Chapter 6

Final discussion, conclusion, limitation of the study and future study

6.1. Final Discussion

End-stage renal damage, blindness, neuropathies and cardiovascular disease (CVD), are vascular complications in diabetes. These complications result in disabilities and high mortality rates in patients with diabetes mellitus (Brownlee, 2001). Non-enzymatic glycation of proteins is an important process for patients with diabetes with regards of the development of diabetic complications through the generation of Advanced Glycated End-products (AGE) and in respect of the methodology for diagnosing and monitoring diabetes through measurement of glycated haemoglobin or other glycated blood proteins which are widely used as a sign of average glycaemia. Glycation is a non-enzymatic process mainly indicated by the glucose concentration, while deglycation is an enzyme-mediated pathway, one of the major enzymes of which is thought to be fructosamine-3-kinase (FN3K) (Van Schaftingen *et al*., 2012). FN3K gene knock-out mice studies demonstrated that glycation of proteins increases (Veiga *et al.,* 2006), highlighting the role of the enzyme in physiological deglycation. The gene that encodes FN3K is located on chromosome 17q25. Research studies demonstrated various polymorphisms of the FN3K gene, some SNPs from them alter FN3K activity in human erythrocytes (Delpierre *et al*., 2006) and some are responsible for variations in glycated haemoglobin levels (Mohás *et al.,* 2010) and some diabetic complications (Škrha *et al.,* 2014).

Non-enzymatic glycation and modification of proteins through Schiff base, Amadori, and Maillard reactions form AGEs that cross-link with extracellular matrix materials such as collagen and cause vascular stiffness (Goldin *et al*., 2006 and Semba *et al*., 2015) and vascular complications of diabetes (Genuth *et al*., 2005 and Nin *et al*., 2011). These AGEs accumulate in tissue and plasma during ageing and diabetes

(Meerwaldt *et al*., 2008) and contribute to age-related and diabetes-related pathologies, including atherosclerosis, cataract, nephropathy, retinopathy, and Alzheimer's disease (Roriz *et al*., 2009). Thus, FN3K provides a protective mechanism against non-enzymatic glycation by decomposing fructosamine 3-phosphates to 3-deoxyglucosone, inorganic phosphate and an amine (Collard *et al*., 2003). Studies in animal models proved that circulating sRAGE neutralises the adverse effects of cellular RAGE by binding with serum AGEs (Wautier *et al*., 1996 and Park *et al.,*1998) through the mechanisms that promote cleavage and shedding of full-length RAGE to develop sRAGE. But as previously discussed (section 3.5) that interpretation of sRAGE is controversial as different studies had different results depending on the disease type.

This study found a non- significant difference in sRAGE levels between two groups with negative and positive glycation gaps whereas Dunmore *et al*. (2018) demonstrated that the AGE level was higher. So, the finding of this study is consistent with the concept that AGEs act via receptor-dependent and receptor-independent ways (Goh and Cooper, 2008). When the AGEs act in a receptor-dependent manner this produces oxidative stress and increased inflammation as a of result immune cells proliferate. During this proliferation, telomere length shortens with each cycle of cell division. The present study found no difference in telomere length ratio between negative and positive G-gap indicating a potential low inflammatory response in those patients. This study did not observe any genotypic association between FN3K SNPs (rs3848403 and rs1056534) and sRAGE concentration among those groups so the AGE concentration is a more likely link between excessive glycation in subjects with a positive G-gap which is likely to be a consequence of lower FN3K activity. Since sRAGE levels could reflect the

expression of RAGE in a tissue but in reality, the sRAGE level is 1,000 times lower in humans than the level that is actually and efficiently needed to capture the circulating AGEs (Yamagishi and Matsui, 2010) and this supports the hypothesis that sRAGE is not acting as a decoy receptor for AGEs.

HbA_{1c} values vary among individuals, despite the similar preceding levels of blood glucose or fructosamine concentration. Dunmore *et al.* (2018) correlated the FN3K activity with the glycation gap and the development of diabetic complications. The non-enzymatic glycation links between an elevated level of blood glucose concentration and the development of diabetic complications, as evidenced from various research studies. Therefore, FN3K deglycating/protein repairing enzyme activity is of greater importance and various research studies proved the association of SNPs in FN3K gene with typical aspects of diabetes. FN3K SNPs and the ferroportin1 SNP were not responsible for functional alterations in the studied cohort indicating the potential involvement of DEGs and transcript variants. These could be enriched in various disease states, involving pathways and biological processes in the underlying mechanism of G-gap. The transcriptome study was conducted to check the hypothesis that DEGs and transcript variants in G-gap groups could collectively enhance or reduce the expression of other enzymes, the activity of enzymes and regulatory factors that take part in the deglycation reaction. This study identified 103 DEGs and 342 differentially expressed transcripts in G-gap groups. This study also found 14 DEGs that encoded various alternatively spliced transcript variants. These DEGs were mainly associated with blood pressure, glycated haemoglobin and erythrocyte indices, and they were enriched in various pathways (viral carcinogenesis, ribosome and phagosome) and biological processes (metabolic process, cellular component analysis, developmental process and

immune system process). This is the initial finding. Therefore, protein-protein interaction (PPI) analysis in STRING (Search tool for the Retrieval of Interacting Genes/Proteins) can be performed to find putative partners that affect the gene expression and provide an understanding for the mechanisms associated with G-gap.

6.2. Conclusion

The prime objective of the present thesis was to explore the findings via research related to the questions presented in Chapter 1. These questions were developed based on gaps from previous studies. In this study the role of SNPs was assessed for the differences of FN3K activity in G-gap negative and positive groups. This was explored by TaqMan genotyping assay on Real-time qPCR. Results demonstrated novel findings that studied SNPs were not involved in the mechanism related to higher and lower FN3K activity in the studied cohort. The sRAGE level was not different in both groups and this study confirmed direct (receptor-independent) and indirect (receptor-dependent) roles of AGEs in diabetic complications. The lack of differences in telomere length between group within the studied cohort, indicated a potentially low inflammatory response. The transcriptome study in this thesis identified various DEGs, differentially expressed transcripts and alternatively spliced variants and their association with various disease, pathways and biological processes which has opened up new ways of thinking about the underlying mechanism of the G-gap in patients with diabetes.

6.3. Limitations

Working with a small sample size of different ethnic groups reduces the possibility of accounting for genetic and lifestyle factors that are more important when

considering a wider population during multicentre research. The studied cohort had a limited ethnic background including mainly white population and without considering the relatively specific ethnic background and the country of origin. The limitation of this research including participant's involvement from only a single NHS trust hospital in Wolverhampton which might not reflect the country's overall ethnic diversity and thus makes it difficult to interpret the impact of genetic variability on the medical conditions of patients with diabetes. Ethnicity might have an impact on the variable expression of genes under different medical conditions alongside responding to antidiabetic drug management. It has been reported that Asian ethnicities are more vulnerable to develop diabetes and they are 60% of the global diabetic population (Guariguata *et al.,* 2014).

In this study, there is an unequal percentage of males and females (77% vs 23% respectively in negative G-gap group whilst 61% vs 39% respectively in positive G-gap group). Consequently, the analyses are not generally applicable to both genders because there might be gender specific variation. The lack of diversity in ethnicity and gender bias may affect variations in G-gap values. Research studies demonstrated that gender and ethnicity might be associated with potential variation in G-gap values (Cohen *et al*., 2006 and Nayak *et al*., 2011). This makes it questionable to represent the global population based on the studied cohort in this study. Another potential limitation is several freeze-thaw cycles that might affect the sRAGE level in relation to G-gap values over time.

6.4. Future study

Detailed studies are important to understand the difference in the glycation gap between individuals including epigenomics, involving reversible heritable changes

in actual DNA sequence. These epigenetic changes take place via DNA methylation and histone modification which affect gene expression via potential hindrance in the availability of DNA sequence for transcriptome formation.

Study of the glycation gap in a mouse with FN3K knock-out of the gene. In the absence of FN3K gene glycation of HbA1c is increased. Glycation gap study will help to identify and confirm whether there any other mechanism exists to control diabetic complications.

WDRG studied the glycation level in positive and negative G-gap groups in human with diabetes. In the future, this study could be conducted in normal individuals without diabetes as well as in other mammals such as a chimpanzee.

Validation study for differentially expressed genes identified by RNA-Seq can be conducted through RTqPCR. This will contribute to the analysis of the measured transcript abundance of selected genes that were involved in diabetes pathways and showed levels of differential expression.

Bioinformatic data from RNA-Seq study should be analysed further for the identification of the generic organisational principle of functional networks and for the investigation of novel protein function.

Chapter 7

References

Abreu, E., Aritonovska, E., Reichenbach, P., Cristofari, G., Culp, B., Terns, R.M., Lingner, J. and Terns, M.P. (2010) TIN2-tethered TPP1 recruits human telomerase to telomeres in vivo. *Molecular and cellular biology*, **30**(12), pp. 2971-2982.

Adaikalakoteswari, A., Balasubramanyam, M. and Mohan, V. (2005) Telomere shortening occurs in Asian Indian Type 2 diabetic patients. *Diabetic Medicine*, **22**(9), pp. 1151-1156.

Adaikalakoteswari, A., Balasubramanyam, M., Ravikumar, R., Deepa, R. and Mohan, V. (2007) Association of telomere shortening with impaired glucose tolerance and diabetic macroangiopathy. *Atherosclerosis*, **195**(1), pp. 83-89.

Ahmad, S., Khan, M.S., Akhter, F., Khan, M.S., Khan, A., Ashraf, J.M., Pandey, R.P. and Shahab, U. (2014). Glycoxidation of biological macromolecules: a critical approach to halt the menace of glycation. *Glycobiology*, **24**(11), pp.979-990.

Ahmed, N. and Thornalley, P. (2007). Advanced glycation endproducts: what is their relevance to diabetic complications? *Diabetes, Obesity and Metabolism*, **9**(3), pp.233-245.

Albert, F.W. and Kruglyak, L. (2015) The role of regulatory variation in complex traits and disease. *Nature Reviews Genetics*, **16**(4), pp. 197-212.

Albuquerque, D., Manco, L., Loua, K.M., Arez, A.P., Trovoada, M.d.J., Relvas, L., Millimono, T.S., Rath, S.L., Lopes, D. and Nogueira, F. (2011) SLC40A1 Q248H allele frequencies and associated SLC40A1 haplotypes in three West African population samples. *Annals of Human Biology***, 38**(3), pp. 378-381.

Altshuler, D., Daly, M.J. and Lander, E.S. (2008) Genetic mapping in human disease. *Science (New York, N.Y.)***, 322**(5903), pp. 881-888.

American Diabetes Association (2010) Standards of medical care in diabetes--2010. *Diabetes Care***, 33 Suppl 1**pp. S11-61.

American Diabetes Association (2010). Diagnosis and classification of diabetes mellitus. *Diabetes Care*, **33 suppl 1**pp. S62-69.

American Diabetes Association (2014) Diagnosis and classification of diabetes mellitus. *Diabetes care***, 37 Suppl 1**pp. S81-90.

Amore, A., Cirina, P., Conti, G., Cerutti, F., Bagheri, N., Emancipator, S.N. and Coppo, R. (2004) Amadori-configurated albumin induces nitric oxide-dependent apoptosis of endothelial cells: a possible mechanism of diabetic vasculopathy. *Nephrology Dialysis Transplantation***, 19**(1), pp. 53-60.

Anders, S., Pyl, P.T. and Huber, W. (2015) HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics***, 31**(2), pp. 166-169.

Anders, S., Reyes, A. and Huber, W. (2012) Detecting differential usage of exons from RNA-seq data. *Genome research*, **22**(10), pp. 2008-2017.

Andrews, S. (2010) FastQC: a quality control tool for high throughput sequence data.

Applied Biosystems., 2010a. Custom TaqMan® SNP Genotyping Assays: Protocol. Foster City, CA, USA: Applied Biosystems.

Armanios, M. (2013). Telomeres and age-related disease: how telomere biology informs clinical paradigms. *The Journal of clinical investigation*, **123**(3), pp.996-1002.

Arvan, P., Pietropaolo, M., Ostrov, D. and Rhodes, C.J. (2012) Islet autoantigens: structure, function, localization, and regulation. *Cold Spring Harbor perspectives in medicine*, **2**(8), pp. 10.1101.

Atkinson, M. and Eisenbarth, G. (2015). Type 1 diabetes. *Lancet*, pp.383:69-82.

Atkinson, M.A., Eisenbarth, G.S. and Michels, A.W. (2014) Type 1 diabetes. *The Lancet*, **383**(9911), pp. 69-82.

Avemaria, F., Carrera, P., Lapolla, A., Sartore, G., Chilelli, N.C., Paleari, R., Ambrosi, A., Ferrari, M. and Mosca, A. (2015) Possible role of fructosamine 3-kinase genotyping for the management of diabetic patients. *Clinical Chemistry and Laboratory Medicine (CCLM)*, **53**(9), pp. 1315-1320.

Bailes, B.K, (2002). Diabetes mellitus and its chronic complications. *AORN journal*, **76**(2), pp.265-282.

Baird, D.M. (2005) New developments in telomere length analysis. *Experimental Gerontology***, 40**(5), pp. 363-368.

Baker, D., Wijshake, T., Tchkonia, T., LeBrasseur, N., Childs, B., van de Sluis, B., Kirkland, J. and van Deursen, J. (2011). Clearance of p16Ink4a-positive senescent cells delays ageing-associated disorders. *Nature*, **479**(7372), pp.232-236.

Bandaria, J., Qin, P., Berk, V., Chu, S. and Yildiz, A. (2016). Shelterin Protects Chromosome Ends by Compacting Telomeric Chromatin. *Cell*, **164**(4), pp.735-746.

Barton, J.C., Acton, R.T., Lee, P.L. and West, C. (2007) SLC40A1 Q248H allele frequencies and Q248H-associated risk of non-HFE iron overload in persons of sub-Saharan African descent. *Blood Cells, Molecules, and Diseases***, 39**(2), pp. 206-211.

Beltran del Rio, M., Tiwari, M., Amodu, L.I., Cagliani, J. and Rodriguez Rilo, H.L, (2016). Glycated hemoglobin, plasma glucose, and erythrocyte aging. *Journal of diabetes science and technology*, **10**(6), pp.1303-1307.

Benjamini, Y. and Speed, T.P. (2012) Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic acids research***, 40**(10), pp. e72-e72.

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L. and Bignell, H.R. (2008) Accurate whole

human genome sequencing using reversible terminator chemistry. *Nature*, **456**(7218), pp. 53- 59.

Beutler, E., Barton, J.C., Felitti, V.J., Gelbart, T., West, C., Lee, P.L., Waalen, J. and Vulpe, C. (2003) Ferroportin 1 (SCL40A1) variant associated with iron overload in African-Americans. *Blood Cells, Molecules, and Diseases*, **31**(3), pp. 305-309.

Bierhaus, A., Schiekofer, S., Schwaninger, M., Andrassy, M., Humpert, P., Chen, J., Hong, M., Luther, T., Henle, T., Kloting, I., Morcos, M., Hofmann, M., Tritschler, H., Weigle, B., Kasper, M., Smith, M., Perry, G., Schmidt, A., Stern, D., Haring, H., Schleicher, E. and Nawroth, P. (2001). Diabetes-Associated Sustained Activation of the Transcription Factor Nuclear Factor- B. *Diabetes*, **50**(12), pp.2792-2808.

Blackburn, E. (2000). Telomere states and cell fates. *Nature*, pp.408:53–56.

Blackburn, E. (2001). Switching and Signaling at the Telomere. *Cell*, **106**(6), pp.661-673.

Blackburn, E., Greider, C. and Szostak, J. (2006). Telomeres and telomerase: the path from maize, Tetrahymena, and yeast to human cancer and ageing. *Nature Medicine*, **12**(10), pp.1133-1138.

Blencowe, B.J. (2006) Alternative splicing: new insights from global analyses. *Cell*, **126**(1), pp. 37-47.

Borodina, T., Adjaye, J. and Sultan, M. (2011) A strand-specific library preparation protocol for RNA sequencing. *in Methods in Enzymology*, **500**, pp. 79-98.

Boteva, E. and Mironova, R. (2019) Maillard reaction and aging: can bacteria shed light on the link? *Biotechnology & Biotechnological Equipment,* pp. 1-17.

Braasch, D.A. and Corey, D.R. (2001) Locked nucleic acid (LNA): fine-tuning the recognition of DNA and RNA. *Chemistry & Biology***, 8**(1), pp. 1-7.

Brownlee, M. (2001). Biochemistry and molecular cell biology of diabetic complications. *Nature*, **414**(6865), pp.813-820.

Brownlee, M. (2005) The pathobiology of diabetic complications: a unifying mechanism. *Diabetes***, 54**(6), pp. 1615-1625.

Bucciarelli, L.G., Wendt, T., Qu, W., Lu, Y., Lalla, E., Rong, L.L., Goova, M.T., Moser, B., Kislinger, T., Lee, D.C., Kashyap, Y., Stern, D.M. and Schmidt, A.M. (2002) RAGE blockade stabilizes established atherosclerosis in diabetic apolipoprotein E-null mice. *Circulation***, 106**(22), pp. 2827-2835.

Buijsse, B., Feskens, E.J., Moschandreas, J., Jansen, E.H., Jacobs Jr, D.R., Kafatos, A., Kok, F.J. and Kromhout, D., (2007). Oxidative stress, and iron and antioxidant status in elderly men: differences between the Mediterranean south (Crete) and northern Europe (Zutphen). *European Journal of Cardiovascular Prevention & Rehabilitation*, ***14***(4), pp.495-500.

Bullard, J.H., Purdom, E., Hansen, K.D. and Dudoit, S. (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics*, **11**(1), pp. 94.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A. and Rinn, J.L. (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes & development*, 25(18), pp. 1915-1927.

Cai, W., Ramdas, M., Zhu, L., Chen, X., Striker, G.E. and Vlassara, H. (2012) Oral advanced glycation endproducts (AGEs) promote insulin resistance and diabetes by depleting the antioxidant defences AGE receptor-1 and sirtuin 1. *Proceedings of the National Academy of Sciences*, **109**(39), pp. 15888-15893.

Callicott, R.J. and Womack, J.E. (2006) Real-time PCR assay for measurement of mouse telomeres. *Comparative medicine*, **56**(1), pp. 17-22.

Campbell, N.R., Leiter, L.A., Larochelle, P., Tobe, S., Chockalingam, A., Ward, R., Morris, D. and Tsuyuki, R. (2009) Hypertension in diabetes: a call to action. *Canadian Journal of Cardiology*, **25**(5), pp. 299-302.

Campisi, J., Kim, S., Lim, C. and Rubio, M. (2001) Cellular senescence, cancer, and ageing: the telomere connection. *Experimental Gerontology*, **36**(10), pp. 1619-1637.

Caselli, E., Rizzo, R., Ingianni, A., Contini, P., Pompei, R., and Di Luca, D. (2014) High prevalence of HHV8 infection and specific killer cell immunoglobulin-like receptors allotypes in Sardinian patients with type 2 diabetes mellitus. *Journal of medical virology*, **86**(10), pp. 1745-1751.

Cawthon, R.M. (2002) Telomere measurement by quantitative PCR. *Nucleic acids research*, **30**(10), pp. e47-e47.

Cawthon, R.M., Smith, K.R., O'Brien, E., Sivatchenko, A. and Kerber, R.A. (2003) Association between telomere length in blood and mortality in people aged 60 years or older. *The Lancet*, **361**(9355), pp. 393-395.

Cech, T.R. and Steitz, J.A. (2014) The noncoding RNA revolution—trashing old rules to forge new ones. *Cell*, **157**(1), pp. 77-94.

Chakhtoura, M. and Azar, S.T. (2013) The role of vitamin d deficiency in the incidence, progression, and complications of type 1 diabetes mellitus. *International journal of endocrinology*, **2013**pp. 148673.

Chan, C.L., Hope, E., Thurston, J., Vigers, T., Pyle, L., Zeitler, P.S. and Nadeau, K.J. (2018). Hemoglobin A1c Accurately Predicts Continuous Glucose Monitoring–Derived Average Glucose in Youth and Young Adults with Cystic Fibrosis. *Diabetes care*, **41**(7), pp.1406-1413.

Chanock, S. (2001). Candidate genes and single nucleotide polymorphisms (SNPs) in the study of human disease. *Disease markers*, **17**(2), pp.89-98.

Chaussabel, D., Pascual, V. and Banchereau, J. (2010) Assessing the human immune system through blood transcriptomics. *BMC biology*, **8**(1), pp. 80-84.

Chen, L., Redon, S. and Lingner, J. (2012) The human CST complex is a terminator of telomerase activity. *Nature*, **488**(7412), pp. 540-544.

Chen, M. and Manley, J.L. (2009) Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nature reviews Molecular cell biology*, **10**(11), pp. 741-754.

Chilelli, N., Burlina, S. and Lapolla, A. (2013) AGEs, rather than hyperglycemia, are responsible for microvascular complications in diabetes: a “glycoxidation-centric” point of view. *Nutrition, Metabolism and Cardiovascular Diseases*, **23**(10), pp. 913-919.

Cikomola, J.C., Kishabongo, A.S., Vandepoele, K., De Mulder, M., Katchunga, P.B., Laukens, B., Van Schie, L., Grootaert, H., Callewaert, N. and Speeckaert, M.M. (2017) A simple colorimetric assay for measuring fructosamine 3 kinase activity. *Clinical Chemistry and Laboratory Medicine (CCLM)*, **55**(1), pp. 154-159.

Classification and Diagnosis of Diabetes. (2014). *Diabetes Care*, **38**(Supplement_1), pp. S8-S16.

Cock, P.J., Fields, C.J., Goto, N., Heuer, M.L. and Rice, P.M. (2009) The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic acids research*, **38**(6), pp. 1767-1771.

Cohen, R.A., Hennekens, C.H., Christen, W.G., Krolewski, A., Nathan, D.M., Peterson, M.J., LaMotte, F., and Manson, J.E. (1999) Determinants of retinopathy progression in type 1 diabetes mellitus. *The American Journal of Medicine*, **107**(1), pp. 45-51.

Cohen, R.M. and Lindsell, C.J. (2012) When the blood glucose and the HbA(1c) don't match: turning uncertainty into opportunity. *Diabetes Care*, **35**(12), pp. 2421-2423.

Cohen, R.M. and Smith, E.P. (2008) Frequency of HbA1c discordance in estimating blood glucose control. *Current Opinion in Clinical Nutrition & Metabolic Care*, **11**(4), pp. 512-517.

Cohen, R.M., Franco, R.S., Khera, P.K., Smith, E.P., Lindsell, C.J., Ciraolo, P.J., Palascak, M.B. and Joiner, C.H. (2008) Red cell life span heterogeneity in hematologically normal people is sufficient to alter HbA1c. *Blood*, **112**(10), pp. 4284-4291.

Cohen, R.M., Holmes, Y.R., Chenier, T.C. and Joiner, C.H. (2003) Discordance between HbA1c and fructosamine: evidence for a glycosylation gap and its relation to diabetic nephropathy. *Diabetes Care*, **26**(1), pp. 163-167.

Cohen, R.M., Snieder, H., Lindsell, C.J., Beyan, H., Hawa, M.I., Blinko, S., Edwards, R., Spector, T.D. and Leslie, R.D. (2006) Evidence for independent heritability of the glycation gap (glycosylation gap) fraction of HbA1c in nondiabetic twins. *Diabetes Care*, **29**(8), pp. 1739-1743.

Colhoun, H.M., Betteridge, D.J., Durrington, P., Hitman, G., Neil, A., Livingstone, S., Charlton-Menys, V., Bao, W., Demicco, D.A., Preston, G.M., Deshmukh, H., Tan, K. and Fuller, J.H. (2011) Total soluble and endogenous secretory receptor for advanced glycation end products as predictive biomarkers of coronary heart disease risk in patients with type 2 diabetes: an analysis from the CARDS trial. *Diabetes***, 60**(9), pp. 2379-2385.

Collard, F., Delpierre, G., Stroobant, V., Matthijs, G., and Van Schaftingen, E. (2003) A mammalian protein homologous to fructosamine-3-kinase is a ketosamine-3-kinase acting on psicosamines and ribulosamines but not on fructosamines. *Diabetes***, 52**(12), pp. 2888-2895.

Conner, J.R., Beisswenger, P.J. and Szwergold, B.S. (2004) The expression of the genes for fructosamine-3-kinase and fructosamine-3-kinase-related protein appears to be constitutive and unaffected by environmental signals. *Biochemical and biophysical research communications***, 323**(3), pp. 932-936.

Conner, J.R., Beisswenger, P.J. and Szwergold, B.S. (2005). Some clues as to the regulation, expression, function, and distribution of fructosamine-3-kinase and fructosamine-3-kinase-related protein. *Annals of the New York Academy of Sciences*, **1043**(1), pp.824-836.

Consortium EP. (2012) an integrated encyclopedia of DNA elements in the human genome. *Nature,* **489** (7414), pp. 57-74.

Costes, S., Langen, R., Gurlo, T., Matveyenko, A.V. and Butler, P.C. (2013) beta-Cell failure in type 2 diabetes: a case of asking too much of too few? *Diabetes*, **62**(2), pp. 327-335.

Coughlan, M.T., Thorburn, D.R., Penfold, S.A., Laskowski, A., Harcourt, B.E., Sourris, K.C., Tan, A.L., Fukami, K., Thallas-Bonke, V., Nawroth, P.P., Brownlee, M., Bierhaus, A., Cooper, M.E. and Forbes, J.M. (2009) RAGE-induced cytosolic ROS promote mitochondrial superoxide generation in diabetes. *Journal of the American Society of Nephrology: JASN*, **20**(4), pp. 742-752.

Creager, M.A., Luscher, T.F., Cosentino, F. and Beckman, J.A. (2003) Diabetes and vascular disease: pathophysiology, clinical consequences, and medical therapy: Part I. *Circulation*, **108**(12), pp. 1527-1532.

Daniali, L., Benetos, A., Susser, E., Kark, J.D., Labat, C., Kimura, M., Desai, K.K., Granick, M. and Aviv, A. (2013). Telomeres shorten at equivalent rates in somatic tissues of adults. *Nature communications*, **4**, p.1597.

Darnell, J.E., Jr (2013) Reflections on the history of pre-mRNA processing and highlights of current knowledge: a unified picture. *RNA (New York, N.Y.)*, **19**(4), pp. 443-460.

Das, S. and Vikalo, H. (2013) Base-calling for high-throughput short-read sequencing: dynamic programming solutions. *BMC Bioinformatics*, **14**(1), pp. 129.

Davuluri, R.V., Suzuki, Y., Sugano, S., Plass, C. and Huang, T.H. (2008) The functional consequences of alternative promoter use in mammalian genomes. *Trends in Genetics*, **24**(4), pp. 167-177.

De Boeck, G., Forsyth, R.G., Praet, M., and Hogendoorn, P.C. (2009) Telomere-associated proteins: cross-talk between telomere maintenance and telomere-lengthening mechanisms. *The Journal of Pathology*, **217**(3), pp. 327-344.

De Lange, T. (2005) Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes & development*, **19**(18), pp. 2100-2110.

Degner, J.F., Marioni, J.C., Pai, A.A., Pickrell, J.K., Nkadori, E., Gilad, Y. and Pritchard, J.K. (2009) Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. *Bioinformatics*, **25**(24), pp. 3207-3212.

Delpierre, G., Collard, F., Fortpied, J., and Van Schaftingen, E. (2002) Fructosamine 3-kinase is involved in an intracellular deglycation pathway in human erythrocytes. *The Biochemical journal*, **365**(Pt 3), pp. 801-808.

Delpierre, G., Rider, M.H., Collard, F. and Stoobant, V. (2000) Identification, cloning, and heterologous expression of a mammalian fructosamine-3-kinase. *Diabetes*, **49**(10), pp. 1627-1634.

Delpierre, G., Veiga-da-Cunha, M., Vertommen, D., Buysschaert, M., and Van Schaftingen, E. (2006) Variability in erythrocyte fructosamine 3-kinase activity in

humans correlates with polymorphisms in the FN3K gene and impacts on haemoglobin glycation at specific sites. *Diabetes & Metabolism*, **32**(1), pp. 31-39.

Delpierrre, G., Vertommen, D., Communi, D., Rider, M.H. and Van Schaftingen, E. (2004) Identification of fructosamine residues deglycated by fructosamine-3-kinase in human hemoglobin. *The Journal of biological chemistry*, **279**(26), pp. 27613-27620.

Delplanque, J., Delpierre, G., Opperdoes, F.R., and Van Schaftingen, E. (2004) Tissue distribution and evolution of fructosamine 3-kinase and fructosamine 3-kinase-related protein. *The Journal of biological chemistry*, **279**(45), pp. 46606-46613.

Desjardins, P. and Conklin, D. (2010). NanoDrop microvolume quantitation of nucleic acids. *JoVE (Journal of Visualized Experiments)*, (45), p.e2565.

Di Bella, J.M., Bao, Y., Gloor, G.B., Burton, J.P. and Reid, G. (2013) High throughput sequencing methods and analysis for microbiome research. *Journal of microbiological methods*, **95**(3), pp. 401-414.

Di Giammartino, D.C., Nishida, K. and Manley, J.L. (2011) Mechanisms and consequences of alternative polyadenylation. *Molecular Cell*, **43**(6), pp. 853-866.

Diotti, R. and Loayza, D. (2011). Shelterin complex and associated factors at human telomeres. *Nucleus*, **2**(2), pp.119-135.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W. and Schlesinger, F. (2012) Landscape of transcription in human cells. *Nature*, **489**(7414), pp. 101-108.

Drosten, C., Seifried, E. and Roth, W.K. (2001). TaqMan 5′-nuclease human immunodeficiency virus type 1 PCR assay with phage-packaged competitive internal control for high-throughput blood donor screening. *Journal of clinical microbiology*, **39**(12), pp.4302-4308.

Dunmore, S.J., Al-Derawi, A.S., Nayak, A.U., Narshi, A., Nevill, A.M., Hellwig, A., Majebi, A., Kirkham, P., Brown, J.E. and Singh, B.M. (2018) Evidence That Differences in Fructosamine-3-Kinase Activity May Be Associated with the Glycation Gap in Human Diabetes. *Diabetes*, **67**(1), pp. 131-136.

Duran, L., Rodriguez, C., Drozd, D., Nance, R.M., Delaney, J.A., Burkholder, G., Mugavero, M.J., Willig, J.H., Warriner, A.H., Crane, P.K. and Atkinson, B.E. (2015). Fructosamine and hemoglobin A1c correlations in HIV-infected adults in routine clinical care: impact of Anemia and albumin levels. *AIDS research and treatment*, *2015*.

Edgren, H., Murumagi, A., Kangaspeska, S., Nicorici, D., Hongisto, V., Kleivi, K., Rye, I.H., Nyberg, S., Wolf, M. and Borresen-Dale, A. (2011) Identification of fusion genes in breast cancer by paired-end RNA-sequencing. *Genome biology*, **12**(1), pp. R6.

Elgar, G. and Vavouri, T. (2008) Tuning into the signals: noncoding sequence conservation in vertebrate genomes. *Trends in genetics*, **24**(7), pp. 344-352.

ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**(7414), pp. 57-74.

Expert Committee on the Diagnosis and Classification of Diabetes Mellitus (2003) Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care*, **26 Suppl 1**pp. S5-20.

Falcone, C., Emanuele, E., D'Angelo, A., Buzzi, M.P., Belvito, C., Cuccia, M. and Geroldi, D. (2005) Plasma levels of soluble receptor for advanced glycation end products and coronary artery disease in nondiabetic men. *Arteriosclerosis, Thrombosis, and Vascular Biology*, **25**(5), pp. 1032-1037.

Farajollahi, S. and Maas, S. (2010) Molecular diversity through RNA editing: a balancing act. *Trends in Genetics*, **26**(5), pp. 221-230.

Farmer, A. (2012) Use of HbA1c in the diagnosis of diabetes. *BMJ (Clinical research ed.)*, **345**pp. e7293 .

Fleming, A.M. and Burrows, C.J. (2013) G-quadruplex folds of the human telomere sequence alter the site reactivity and reaction pathway of guanine oxidation compared to duplex DNA. *Chemical research in toxicology*, **26**(4), pp. 593-607.

Florez, J.C., and Barroso, I. (2014) Genome-Wide Association Studies of Glycaemic Traits: A Magical Journey. *in Genetics in Diabetes.* Karger Publishers, **23**, pp. 42-57.

Fournet, M., Bonte, F. and Desmouliere, A. (2018) Glycation damage: A possible hub for major pathophysiological disorders and aging. *Aging and disease***, 9**(5), pp. 880-900.

Francisco, M., Lazaruk, K.D., Rhodes, M.D. and Wenz, M.H. (2005) Assessment of two flexible and compatible SNP genotyping platforms: TaqMan® SNP Genotyping Assays and the SNPlex™ Genotyping System. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis***, 573**(1), pp. 111-135.

Fraser, C.G. (2001) Biological variation: from principles to practice. Amer. Assoc. for Clinical Chemistry.

Fuda, N.J., Ardehali, M.B. and Lis, J.T. (2009) Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature***, 461**(7261), pp. 186-192.

Garber, M., Grabherr, M.G., Guttman, M. and Trapnell, C. (2011) Computational methods for transcriptome annotation and quantification using RNA-seq. *Nature methods***, 8**(6), pp. 469-477.

Genuth, S., Sun, W., Cleary, P., Sell, D., Dahms, W., Malone, J., Sivitz, W. and Monnier, V. (2005) DCCT Skin Collagen Ancillary Study Group: Glycation and carboxymethyllysine levels in skin collagen predict the risk of future 10-year progression of diabetic retinopathy and nephropathy in the diabetes control and complications trial and epidemiology of diabetes interventions and complications participants with type 1 diabetes. *Diabetes*, **54**(11), pp. 3103-3111.

Georgin-Lavialle, S., Aouba, A., Mouthon, L., Londono-Vallejo, J.A., Lepelletier, Y., Gabet, A. and Hermine, O. (2010) The telomere/telomerase system in autoimmune and systemic immune-mediated diseases. *Autoimmunity reviews*, **9**(10), pp. 646-651.

Gil, M.E. and Coetzer, T.L. (2004) Real-time quantitative PCR of telomere length. *Molecular biotechnology*, **27**(2), pp. 169-172.

Girelli, D., De Domenico, I., Bozzini, C., Campostrini, N., Busti, F., Castagna, A., Soriani, N., Cremonesi, L., Ferrari, M. and Colombari, R. (2008) Clinical, pathological, and molecular correlates in ferroportin disease: a study of two novel mutations. *Journal of hepatology*, **49**(4), pp. 664-671.

Gkogkolou, P. and Bohm, M. (2012). Advanced glycation end products: key players in skin aging? *Dermato-endocrinology*, **4**(3), pp.259-270.

Goh, S. and Cooper, M.E. (2008) The role of advanced glycation end products in progression and complications of diabetes. *The Journal of Clinical Endocrinology & Metabolism*, **93**(4), pp. 1143-1152.

Goldin, A., Beckman, J.A., Schmidt, A.M. and Creager, M.A. (2006) Advanced glycation end products: sparking the development of diabetic vascular injury. *Circulation*, **114**(6), pp. 597-605.

Goldstein, D.E., Little, R.R., Lorenz, R.A., Malone, J.I., Nathan, D., Peterson, C.M. and Sacks, D.B. (2004) Tests of glycemia in diabetes. *Diabetes care*, **27**(7), pp. 1761-1773.

Gordeuk, V., Mukiibi, J., Hasstedt, S.J., Samowitz, W., Edwards, C.Q., West, G., Ndambire, S., Emmanual, J., Nkanza, N. and Chapanduka, Z. (1992) Iron overload in Africa. *New England Journal of Medicine*, **326**(2), pp. 95-100.

Gordeuk, V.R., Caleffi, A., Corradini, E., Ferrara, F., Jones, R.A., Castro, O., Onyekwere, O., Kittles, R., Pignatti, E. and Montosi, G. (2003) Iron overload in Africans and African-Americans and a common mutation in the SCL40A1 (ferroportin 1) gene☆. *Blood Cells, Molecules, and Diseases*, **31**(3), pp. 299-304.

Gott, J.M. and Emeson, R.B. (2000) Functions and mechanisms of RNA editing. *Annual Review of Genetics*, **34**(1), pp. 499-531.

Graham, K.L., Sutherland, R.M., Mannering, S.I., Zhao, Y., Chee, J., Krishnamurthy, B., Thomas, H.E., Lew, A.M. and Kay, T.W. (2012). Pathogenic mechanisms in type 1 diabetes: the islet is both target and driver of disease. *The review of diabetic studies: RDS*, **9**(4), p.148-168.

Grant, P. (2007) Diabetes mellitus as a prothrombotic condition. *Journal of internal medicine*, **262**(2), pp. 157-172.

Guariguata, L., Whiting, D.R., Hambleton, I., Beagley, J., Linnenkamp, U. and Shaw, J.E. (2014) Global estimates of diabetes prevalence for 2013 and projections for 2035. *Diabetes research and clinical practice*, **103**(2), pp. 137-149.

Guimarães, E.L., Empsen, C., Geerts, A., and van Grunsven, L.A. (2010) Advanced glycation end products induce production of reactive oxygen species via the activation of NADPH oxidase in murine hepatic stellate cells. *Journal of Hepatology*, **52**(3), pp. 389-397.

Guo, N., Parry, E.M., Li, L., Kembou, F., Lauder, N., Hussain, M.A., Berggren, P. and Armanios, M. (2011) Short telomeres compromise β-cell signaling and survival. *PloS one*, **6**(3), pp. e17858.

Guttman, M., Garber, M., Levin, J.Z., Donaghey, J., Robinson, J., Adiconis, X., Fan, L., Koziol, M.J., Gnirke, A. and Nusbaum, C. (2010) Ab initio reconstruction of cell type-specific transcriptomes in the mouse reveals the conserved multi-exonic structure of lincRNAs. *Nature Biotechnology*, **28**(5), pp. 503-510.

Hajjawi, O.S. (2013) Glucose transport in human red blood cells. *Am J Biomed Life Sci*, **1**(3), pp. 44-52.

Hansel, B., Giral, P., Nobecourt, E., Chantepie, S., Bruckert, E., Chapman, M.J. and Kontush, A. (2004) Metabolic syndrome is associated with elevated oxidative stress and dysfunctional dense high-density lipoprotein particles displaying impaired antioxidative activity. *The Journal of Clinical Endocrinology & Metabolism***, 89**(10), pp. 4963-4971.

Hansen, K.D., Brenner, S.E. and Dudoit, S. (2010) Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic acids research***, 38**(12), pp. e131-e131.

Hansen, K.D., Irizarry, R.A. and Wu, Z. (2012) Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics***, 13**(2), pp. 204-216.

Hare, M.J., Shaw, J.E. and Zimmet, P.Z. (2012). Current controversies in the use of haemoglobin A1c. *Journal of internal medicine*, **271**(3), pp.227-236.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., Barnes, I., Bignell, A., Boychenko, V., Hunt, T., Kay, M., Mukherjee, G., Rajan, J., Despacio-Reyes, G., Saunders, G., Steward, C., Harte, R., Lin, M., Howald, C., Tanzer, A., Derrien, T., Chrast, J., Walters, N., Balasubramanian, S., Pei, B., Tress, M., Rodriguez, J.M., Ezkurdia, I., van Baren, J., Brent, M., Haussler, D., Kellis, M., Valencia, A., Reymond, A., Gerstein, M., Guigo, R. and Hubbard, T.J. (2012) GENCODE: the reference human genome annotation for The ENCODE Project. *Genome research***, 22**(9), pp. 1760-1774.

Hempe, J.M., Gomez, R., McCarter, R.J. and Chalew, S.A. (2002) High and low hemoglobin glycation phenotypes in type 1 diabetes: a challenge for interpretation of glycemic control. *Journal of diabetes and its complications*, **16**(5), pp. 313-320.

Herbert, B.S. (2011). The impact of telomeres and telomerase in cellular biology and medicine: it's not the end of the story. *Journal of cellular and molecular medicine*, **15**(1), p.1-2.

Herman, W.H. and Cohen, R.M. (2012) Racial and ethnic differences in the relationship between HbA1c and blood glucose: implications for the diagnosis of diabetes. *The Journal of Clinical Endocrinology & Metabolism*, **97**(4), pp. 1067-1072.

Hex, N., Bartlett, C., Wright, D., Taylor, M. and Varley, D. (2012) Estimating the current and future costs of Type 1 and Type 2 diabetes in the UK, including direct health costs and indirect societal and productivity costs. *Diabetic Medicine*, **29**(7), pp. 855-862.

Hills, M., Lücke, K., Chavez, E.A., Eaves, C.J. and Lansdorp, P.M. (2009). Probing the mitotic history and developmental stage of hematopoietic cells using single telomere length analysis (STELA). *Blood*, **113**(23), pp.5765-5775.

Hinnebusch, A.G., Ivanov, I.P. and Sonenberg, N. (2016) Translational control by 5'-untranslated regions of eukaryotic mRNAs. *Science (New York, N.Y.)*, **352**(6292), pp. 1413-1416.

Hockemeyer, D., Daniels, J., Takai, H. and de Lange, T. (2006) Recent expansion of the telomeric complex in rodents: Two distinct POT1 proteins protect mouse telomeres. *Cell*, **126**(1), pp. 63-77.

Hornsby, P.J. (2007). Telomerase and the aging process. *Experimental gerontology*, **42**(7), pp.575-581.

Horvath, S. (2013) DNA methylation age of human tissues and cell types. *Genome biology*, **14**(10), pp. 3156.

Hu, C., Zhang, R., Yu, W., Wang, J., Wang, C., Pang, C., Ma, X., Bao, Y., Xiang, K. and Jia, W. (2011). CPVL/CHN2 genetic variant is associated with diabetic retinopathy in Chinese type 2 diabetic patients. *Diabetes*, **60**(11), pp.3085-3089.

Hudson, B.I., Carter, A.M., Harja, E., Kalea, A.Z., Arriero, M., Yang, H., Grant, P.J. and Schmidt, A.M. (2008) Identification, classification, and expression of RAGE gene splice variants. *The FASEB Journal*, **22**(5), pp. 1572-1580.

Hudson, P.R., Child, D.F., Jones, H. and Williams, C.P. (1999) Differences in rates of glycation (glycation index) may significantly affect individual HbA1c results in type 1 diabetes. *Annals of Clinical Biochemistry*, **36**(4), pp. 451-459.

Humpert, P.M., Kopf, S., Djuric, Z., Wendt, T., Morcos, M., Nawroth, P.P. and Bierhaus, A. (2006) Plasma sRAGE is independently associated with urinary albumin excretion in type 2 diabetes. *Diabetes Care*, **29**(5), pp. 1111-1113.

International Expert Committee (2009) International Expert Committee report on the role of the A1C assay in the diagnosis of diabetes. *Diabetes Care*, **32**(7), pp. 1327-1334.

International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature*, **431**(7011), pp. 931-945.

Jeffcoate, S. (2004) Diabetes control and complications: the role of glycated haemoglobin, 25 years on. *Diabetic Medicine*, **21**(7), pp. 657-665.

Kalisky, T., Blainey, P. and Quake, S.R. (2011) Genomic analysis at the single-cell level. *Annual Review of Genetics*, **45**pp. 431-445.

Kallio, M.A., Tuimala, J.T., Hupponen, T., Klemelä, P., Gentile, M., Scheinin, I., Koski, M., Käki, J. and Korpelainen, E.I. (2011) Chipster: user-friendly analysis software for microarray and other high-throughput data. *BMC Genomics*, **12**(1), pp. 507.

Karijolich, J. and Yu, Y. (2011) Converting nonsense codons into sense codons by targeted pseudouridylation. *Nature*, **474**(7351), pp. 395-398.

Karki, R., Pandya, D., Elston, R.C. and Ferlini, C. (2015). Defining “mutation” and “polymorphism” in the era of personal genomics. *BMC medical genomics*, **8**(1), p.37.

Karlseder, J. (2003) Telomere repeat binding factors: keeping the ends in check. *Cancer letters*, **194**(2), pp. 189-197.

Kasvosve, I., Gomo, Z.A., Nathoo, K.J., Matibe, P., Mudenge, B., Loyevsky, M. and Gordeuk, V.R. (2005) Effect of ferroportin Q248H polymorphism on iron status in African children–. *The American Journal of Clinical Nutrition***, 82**(5), pp. 1102-1106.

Kasvosve, I., Tshwenyego, U., Phuthego, T., Koto, G., Zachariah, M., Nyepetsi, N.G. and Motswaledi, M.S. (2015) Serum ferritin concentration is affected by ferroportin Q248H mutation in Africans. *Clinica Chimica Acta***, 444**pp. 257-259.

Kaszubowska, L. (2008) Telomere shortening and ageing of the immune system. *J Physiol Pharmacol***, 59**(Suppl 9), pp. 169-186.

Katchunga, P.B., Baguma, M., Jean-René, M., Philippé, J., Hermans, M.P. and Delanghe, J. (2013) Ferroportin Q248H mutation, hyperferritinemia and atypical type 2 diabetes mellitus in South Kivu. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews***, 7**(2), pp. 112-115.

Kato, H., van Chuyen, N., Shinoda, T., Sekiya, F. and Hayase, F. (1990) Metabolism of 3-deoxyglucosone, an intermediate compound in the Maillard reaction, administered orally or intravenously to rats. *Biochimica et Biophysica Acta (BBA)-General Subjects***, 1035**(1), pp. 71-76.

Katz, Y., Wang, E.T., Airoldi, E.M. and Burge, C.B. (2010) Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature methods***, 7**(12), pp. 1009.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, D. (2002) The human genome browser at UCSC. *Genome research*, **12**(6), pp. 996-1006.

Kilpatrick, E.S., Rigby, A.S. and Atkin, S.L. (2008) A1C variability and the risk of microvascular complications in type 1 diabetes: data from the Diabetes Control and Complications Trial. *Diabetes Care*, **31**(11), pp. 2198-2202.

Kim, D., Langmead, B. and Salzberg, S.L. (2015) HISAT: a fast-spliced aligner with low memory requirements. *Nature methods*, **12**(4), pp. 357-360.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S.L. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology*, **14**(4), pp. R36.

Kim, J., Lim, J. and Lee, C. (2013) Quantitative real-time PCR approaches for microbial community studies in wastewater treatment systems: applications and considerations. *Biotechnology Advances*, **31**(8), pp. 1358-1373.

Kim, S., Parrinello, S., Kim, J. and Campisi, J. (2003) Mus musculus and Mus spretus homologues of the human telomere-associated protein TIN2. *Genomics*, **81**(4), pp. 422-432.

Kirchner, H., Shaheen, F., Kalscheuer, H., Schmid, S., Oster, H. and Lehnert, H. (2017). The telomeric complex and metabolic disease. *Genes*, **8**(7), p.176.

Knoop, V. (2011) When you can't trust the DNA: RNA editing changes transcript sequences. *Cellular and Molecular Life Sciences*, **68**(4), pp. 567-586.

Krause, R., Oehme, A., Wolf, K., and Henle, T. (2006) A convenient HPLC assay for the determination of fructosamine-3-kinase activity in erythrocytes. *Analytical and bioanalytical chemistry*, **386**(7-8), pp. 2019-2025.

Krentz, A.J., Clough, G. and Byrne, C.D. (2007). Interactions between microvascular and macrovascular disease in diabetes: pathophysiology and therapeutic implications. *Diabetes, Obesity and Metabolism*, 9(6), pp.781-791.

Krueger, F. (2015) Trim galore. A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files.

Kulis, M. and Esteller, M. (2010) DNA methylation and cancer. In *Advances in genetics,* **70** pp. 27-56.

Kumar, R., Nandhini, L.P., Kamalanathan, S., Sahoo, J. and Vivekanadan, M. (2016) Evidence for current diagnostic criteria of diabetes mellitus. *World journal of diabetes*, **7**(17), pp. 396-405.

Kwak, H. and Lis, J.T. (2013) Control of transcriptional elongation. *Annual Review of Genetics*, **47**pp. 483-508.

Lal, S., Szwergold, B.S., Walker, M., Randall, W., Kappler, F., Brown, T. and Beisswenger, P.J. (2005). Production and metabolism of 3-deoxyglucosone in humans. In *the Maillard reaction in foods and medicine*, pp. 291-297.

Lamble, S., Batty, E., Attar, M., Buck, D., Bowden, R., Lunter, G., Crook, D., El-Fahmawi, B. and Piazza, P. (2013) Improved workflows for high throughput library preparation using the transposome-based Nextera system. *BMC Biotechnology***, 13**(1), pp. 104.

Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nature methods***, 9**(4), pp. 357-359.

Lapolla, A., Traldi, P. and Fedele, D. (2005) Importance of measuring products of non-enzymatic glycation of proteins. *Clinical biochemistry***, 38**(2), pp. 103-115.

Lasken, R.S. and McLean, J.S. (2014) Recent advances in genomic DNA sequencing of microbial species from single cells. *Nature Reviews Genetics***, 15**(9), pp. 577-584.

Lee, M., Teber, E.T., Holmes, O., Nones, K., Patch, A.M., Dagg, R.A., Lau, L.M.S., Lee, J.H., Napier, C.E., Arthur, J.W. and Grimmond, S.M. (2018). Telomere sequence content can be used to determine ALT activity in tumours. *Nucleic acids research*, **46**(10), pp.4903-4918.

Lei, M., Zaug, A.J., Podell, E.R. and Cech, T.R. (2005) Switching human telomerase on and off with hPOT1 protein in vitro. *The Journal of biological chemistry***, 280**(21), pp. 20449-20456.

Leong, A. and Wheeler, E. (2018) Genetics of HbA1c: a case study in clinical translation. *Current opinion in genetics & development***, 50**pp. 79-85.

Leslie, R.D.G. and Cohen, R.M. (2009) *Biologic variability in plasma glucose, hemoglobin A1c, and advanced glycation end products associated with diabetes complications, J Diabetes Sci Technol,* **3**, pp. 635– 643.

Li, B., Ruotti, V., Stewart, R.M., Thomson, J.A. and Dewey, C.N. (2009) RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics***, 26**(4), pp. 493-500.

Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *bioinformatics***, 25**(14), pp. 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics***, 25**(16), pp. 2078-2079.

Li, S., Tighe, S.W., Nicolet, C.M., Grove, D., Levy, S., Farmerie, W., Viale, A., Wright, C., Schweitzer, P.A. and Gao, Y. (2014) Multi-platform assessment of transcriptome profiling using RNA-seq in the ABRF next-generation sequencing study. *Nature Biotechnology***, 32**(9), pp. 915-925.

Liew, C.C., Ma, J., Tang, H.C., Zheng, R. and Dempsey, A.A. (2006). The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. *Journal of Laboratory and Clinical Medicine*, **147**(3), pp.126-132.

Lindsey, J.B., Cipollone, F., Abdullah, S.M. and Mcguire, D.K. (2009) Receptor for advanced glycation end-products (RAGE) and soluble RAGE (sRAGE): cardiovascular implications. *Diabetes and Vascular Disease Research***, 6**(1), pp. 7-14.

Lindsey, J.B., de Lemos, J.A., Cipollone, F., Ayers, C.R., Rohatgi, A., Morrow, D.A., Khera, A. and McGuire, D.K. (2009) Association between circulating soluble receptor for advanced glycation end products and atherosclerosis: observations from the Dallas Heart Study. *Diabetes Care***, 32**(7), pp. 1218-1220.

Lindström, U.M., Chandrasekaran, R.A., Orbai, L., Helquist, S.A., Miller, G.P., Oroudjev, E., Hansma, H.G. and Kool, E.T. (2002). Artificial human telomeres from DNA nanocircle templates. *Proceedings of the National Academy of Sciences*, **99**(25), pp.15953-15958.

Liton, P.B., Challa, P., Stinnett, S., Luna, C., Epstein, D.L. and Gonzalez, P. (2005) Cellular senescence in the glaucomatous outflow pathway. *Experimental Gerontology***, 40**(8), pp. 745-748.

Little, R.R. and Rohlfing, C.L. (2013) The long and winding road to optimal HbA1c measurement. *Clinica Chimica Acta***, 418**pp. 63-71.

Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, 15(12), pp. 550.

Lowe, R., Shirley, N., Bleackley, M., Dolan, S. and Shafee, T. (2017). Transcriptomics technologies. *PLoS computational biology*, **13**(5), p. e1005457.

Lu, W., Zhang, Y., Liu, D., Songyang, Z. and Wan, M. (2013). Telomeres—structure, function, and regulation. *Experimental cell research*, **319**(2), pp.133-141.

Luscombe, N.M., Greenbaum, D. and Gerstein, M. (2001) What is bioinformatics? A proposed definition and overview of the field. *Methods of information in medicine***,** **40**(04), pp. 346-358.

Lutz, C.S. (2008) Alternative polyadenylation: a twist on mRNA 3′ end formation. *ACS chemical biology***, 3**(10), pp. 609-617.

Macdonald, D.R., Hanson, A.M., Holland, M.R. and Singh, B.M. (2008) Clinical impact of variability in HbA1c as assessed by simultaneously measuring fructosamine and use of error grid analysis. *Annals of Clinical Biochemistry***, 45**(4), pp. 421-425.

Maestroni, L., Matmati, S. and Coulon, S. (2017). Solving the telomere replication problem. *Genes*, **8**(2), pp.55.

Magalhaes, P.M., Appell, H. and Duarte, J.A. (2008) Involvement of advanced glycation end products in the pathogenesis of diabetic complications: the protective role of regular physical activity. *European Review of Aging and Physical Activity***, 5**(1), pp. 17.

Makris, K. and Spanou, L. (2011) Is there a relationship between mean blood glucose and glycated hemoglobin? *Journal of diabetes science and technology***, 5**(6), pp. 1572-1583.

Malkki, M. and Petersdorf, E.W. (2012). Genotyping of single nucleotide polymorphisms by 5′ nuclease allelic discrimination. In *Immunogenetics* pp. 173-182.

Maniatis, T. and Reed, R. (2002) An extensive network of coupling among gene expression machines. *Nature***, 416**(6880), pp. 499-506.

Maniatis, T. and Tasic, B. (2002) Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature***, 418**(6894), pp. 236-243.

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R. and Chakravarti, A. (2009) Finding the missing heritability of complex diseases. *Nature***, 461**(7265), pp. 747-753.

Mardis, E.R. (2013) Next-generation sequencing platforms. *Annual review of analytical chemistry***, 6**pp. 287-303.

Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M. and Gilad, Y. (2008) RNA-Seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome research,* **18**(9), pp. 1509-1517.

Martin, A.R., Costa, H.A., Lappalainen, T., Henn, B.M., Kidd, J.M., Yee, M.C., Grubert, F., Cann, H.M., Snyder, M., Montgomery, S.B. and Bustamante, C.D. (2014). Transcriptome sequencing from diverse human populations reveals differentiated regulatory architecture. *PLoS genetics*, **10**(8), p.e1004549.

Martin, G. and Keller, W. (2007) RNA-specific ribonucleotidyl transferases. *RNA (New York, N.Y.)***, 13**(11), pp. 1834-1849.

Martin, J.A., Brown, T.D., Heiner, A.D. and Buckwalter, J.A. (2004) Chondrocyte senescence, joint loading, and osteoarthritis. *Clinical orthopaedics and related research***, 427**pp. S96-S103.

Martino, A., Mancuso, T. and Rossi, A.M. (2010). Application of high-resolution melting to large-scale, high-throughput SNP genotyping: a comparison with the TaqMan® method. *Journal of biomolecular screening*, **15**(6), pp.623-629.

McCarter, R.J., Hempe, J.M., Gomez, R. and Chalew, S.A. (2004) Biological variation in HbA1c predicts the risk of retinopathy and nephropathy in type 1 diabetes. *Diabetes Care***, 27**(6), pp. 1259-1264.

McKiernan, P., Lynch, P., Ramsey, J., Cryan, S. and Greene, C. (2018). Knockdown of Gene Expression in Macrophages by microRNA Mimic-Containing Poly (Lactic-co-glycolic Acid) Microparticles. *Medicines*, **5**(4), p.133.

Mcnamara, L., Gordeuk, V.R. and Macphail, A.P. (2005) Ferroportin (Q248H) mutations in African families with dietary iron overload. *Journal of gastroenterology and hepatology***, 20**(12), pp. 1855-1858.

Meerwaldt, R., Links, T., Zeebregts, C., Tio, R., Hillebrands, J. and Smit, A. (2008) The clinical relevance of assessing advanced glycation endproducts accumulation in diabetes. *Cardiovascular Diabetology***, 7**(1), pp. 29.

Meier, J.J., Butler, A.E., Saisho, Y., Monchamp, T., Galasso, R., Bhushan, A., Rizza, R.A. and Butler, P.C. (2008) Beta-cell replication is the primary mechanism subserving the postnatal expansion of beta-cell mass in humans. *Diabetes***, 57**(6), pp. 1584-1594.

Metzker, M.L. (2005) Emerging technologies in DNA sequencing. *Genome research***, 15**(12), pp. 1767-1776.

Metzker, M.L. (2010) Sequencing technologies—the next generation. *Nature reviews genetics***, 11**(1), pp. 31-46.

Meyerson, M., Gabriel, S. and Getz, G. (2010) Advances in understanding cancer genomes through second-generation sequencing. *Nature Reviews Genetics***, 11**(10), pp. 685-696.

Minamino, T., Orimo, M., Shimizu, I., Kunieda, T., Yokoyama, M., Ito, T., Nojima, A., Nabetani, A., Oike, Y. and Matsubara, H. (2009) A crucial role for adipose tissue p53 in the regulation of insulin resistance. *Nature medicine*, **15**(9), pp. 1082-1087.

Miyake, Y., Nakamura, M., Nabetani, A., Shimamura, S., Tamura, M., Yonehara, S., Saito, M. and Ishikawa, F. (2009). RPA-like Mammalian Ctc1-Stn1-Ten1 Complex Binds to Single-Stranded DNA and Protects Telomeres Independently of the Pot1 Pathway. *Molecular Cell*, **36**(2), pp.193-206.

Mizukami, H., Takahashi, K., Inaba, W., Osonoi, S., Kamata, K., Tsuboi, K. and Yagihashi, S. (2014) Age-associated changes of islet endocrine cells and the effects of body mass index in Japanese. *Journal of diabetes investigation*, **5**(1), pp. 38-47.

Mohás, M., Kisfali, P., Baricza, E., Mérei, A., Maász, A., Cseh, J., Mikolás, E., Szijártó, I., Melegh, B. and Wittmann, I. (2010) A polymorphism within the fructosamine-3-kinase gene is associated with HbA1c Levels and the onset of type 2 diabetes mellitus. *Experimental and clinical endocrinology & diabetes*, **118**(03), pp. 209-212.

Mohr, S. and Liew, C. (2007) The peripheral-blood transcriptome: new insights into disease and risk assessment. *Trends in molecular medicine*, **13**(10), pp. 422-432.

Mollano, A.V., Martin, J.A. and Buckwalter, J.A. (2002). Chondrocyte senescence and telomere regulation: implications in cartilage aging and cancer (a brief review). *The Iowa orthopaedic journal*, **22**, p.1-7.

Monnier, L., Lapinski, H. and Colette, C. (2003). Contributions of fasting and postprandial plasma glucose increments to the overall diurnal hyperglycemia of type 2 diabetic patients: variations with increasing levels of HbA1c. *Diabetes care*, **26**(3), pp.881-885.

Monnier, V.M. (2006) The fructosamine 3-kinase knockout mouse: a tool for testing the glycation hypothesis of intracellular protein damage in diabetes and aging. *The Biochemical journal***, 399**(2), pp. e11-3.

Monnier, V.M. and Wu, X. (2003) Enzymatic deglycation with amadoriase enzymes from Aspergillus sp. as a potential strategy against the complications of diabetes and aging. *Biochemical Society Transactions***, 31**(Pt 6), pp. 1349-1353.

Monnier, V.M., Mustata, G.T., Biemel, K.L., Reihl, O., Lederer, M.O., Zhenyu, D. and Sell, D.R. (2005) Cross-linking of the extracellular matrix by the Maillard reaction in aging and diabetes: an update on “a puzzle nearing resolution”. *Annals of the New York Academy of Sciences***, 1043**(1), pp. 533-544.

Monnier, V.M., Sell, D.R. and Genuth, S. (2005) Glycation products as markers and predictors of the progression of diabetic complications. *Annals of the New York Academy of Sciences***, 1043**(1), pp. 567-581.

Moore, M.J. and Proudfoot, N.J. (2009) Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell***, 136**(4), pp. 688-700.

Morgan, P.E., Dean, R.T. and Davies, M.J. (2002) Inactivation of cellular enzymes by carbonyls and protein-bound glycation/glycoxidation products. *Archives of Biochemistry and Biophysics*, **403**(2), pp. 259-269.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods*, **5**(7), pp. 621-628.

Mueller, O., Lightfoot, S., and Schroeder, A. (2004) RNA integrity number (RIN)–standardization of RNA quality control. *Agilent application note, publication* pp. 1-8.

Muoio, D.M., and Newgard, C.B. (2008) Mechanisms of disease: molecular and metabolic mechanisms of insulin resistance and [beta]-cell failure in type 2 diabetes. *Nature reviews. Molecular cell biology*, **9**(3), pp. 193-205.

Mutz, K., Heilkenbrinker, A., Lönne, M., Walter, J. and Stahl, F. (2013) Transcriptome analysis using next-generation sequencing. *Current opinion in biotechnology*, **24**(1), pp. 22-30.

Naess, S., Eriksen, J., Midthjell, K. and Tambs, K. (2003) Diabetes mellitus and comorbidity. Change between 1984–1986 and 1995–1997: Results of the Nord-Trøndelag Health Study. *Journal of diabetes and its complications*, **17**(6), pp. 323-330.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M. and Snyder, M. (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science (New York, N.Y.)***, 320**(5881), pp. 1344-1349.

Nandakumar, D., Pandey, M. and Patel, S.S. (2015). Cooperative base pair melting by helicase and polymerase positioned one nucleotide from each other. *Elife*, **4**, p. e06562.

Nasir, N.M., Thevarajah, M. and Yean, C.Y. (2010) Hemoglobin variants detected by hemoglobin A1c (HbA1c) analysis and the effects on HbA1c measurements. *International journal of diabetes in developing countries***, 30**(2), pp. 86-90.

Nayak, A.U., Holland, M.R., Macdonald, D.R., Nevill, A. and Singh, B.M. (2011) Evidence for consistency of the glycation gap in diabetes. *Diabetes Care***, 34**(8), pp. 1712-1716.

Nayak, A.U., Nevill, A.M., Bassett, P. and Singh, B.M. (2013) Association of glycation gap with mortality and vascular complications in diabetes. *Diabetes Care***, 36**(10), pp. 3247-3253.

Nemet, I., Varga-Defterdarovic, L. and Turk, Z. (2006) Methylglyoxal in food and living organisms. *Molecular nutrition & food research***, 50**(12), pp. 1105-1117.

Nin, J.W., Jorsal, A., Ferreira, I., Schalkwijk, C.G., Prins, M.H., Parving, H.H., Tarnow, L., Rossing, P. and Stehouwer, C.D. (2011) Higher plasma levels of advanced glycation end products are associated with incident cardiovascular disease and all-cause mortality in type 1 diabetes: a 12-year follow-up study. *Diabetes Care*, **34**(2), pp. 442-447.

O'Callaghan, N.J., Dhillon, V.S., Thomas, P. and Fenech, M. (2008) A quantitative real-time PCR method for absolute telomere length. *BioTechniques*, **44**(6), pp. 807-809.

Ogurtsova, K., da Rocha Fernandes, J., Huang, Y., Linnenkamp, U., Guariguata, L., Cho, N., Cavan, D., Shaw, J. and Makaroff, L. (2017) IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040. Diabetes research and clinical practice, 128pp. 40-50.

Okoniewski, M.J. and Miller, C.J. (2006) Hybridization interactions between probesets in short oligo microarrays lead to spurious correlations. *BMC Bioinformatics*, **7**(1), pp. 276.

Olivieri, F., Lorenzi, M., Antonicelli, R., Testa, R., Sirolla, C., Cardelli, M., Mariotti, S., Marchegiani, F., Marra, M. and Spazzafumo, L. (2009) Leukocyte telomere shortening in elderly Type2DM patients with previous myocardial infarction. *Atherosclerosis*, **206**(2), pp. 588-593.

Oshlack, A. and Wakefield, M.J. (2009) Transcript length bias in RNA-Seq data confounds systems biology. *Biology direct*, **4**(1), pp. 14.

Ott, C., Jacobs, K., Haucke, E., Santos, A.N., Grune, T. and Simm, A. (2014) Role of advanced glycation end products in cellular signaling. *Redox biology*, **2**pp. 411-429 .

Pachter, L. (2011) Models for transcript quantification from RNA-Seq. *arXiv preprint arXiv:1104.3889*.

Palm, W. and de Lange, T. (2008) How shelterin protects mammalian telomeres. *Annual Review of Genetics*, **42**pp. 301-334.

Pan, Q., Shai, O., Lee, L.J., Frey, B.J. and Blencowe, B.J. (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*, **40**(12), pp. 1413-1415.

Pan, Q., Shai, O., Misquitta, C., Zhang, W., Saltzman, A.L., Mohammad, N., Babak, T., Siu, H., Hughes, T.R. and Morris, Q.D. (2004) Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform. *Molecular Cell*, **16**(6), pp. 929-941.

Peppa, M. and Vlassara, H. (2005) Advanced glycation end products and diabetic complications: a general overview. *Hormones (Athens)*, **4**(1), pp. 28-37.

Peppa, M., Uribarri, J. and Vlassara, H. (2008) Aging and glycoxidant stress. *HORMONES-ATHENS-*, **7**(2), pp. 123-132.

Pertea, M. (2012) The human transcriptome: an unfinished story. *Genes*, **3**(3), pp. 344-360.

Pertea, M., Kim, D., Pertea, G.M., Leek, J.T. and Salzberg, S.L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature protocols*, **11**(9), p.1650.

Pertea, M., Kim, D., Pertea, G.M., Leek, J.T. and Salzberg, S.L. (2016) Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie, and Ballgown. *Nature protocols*, **11**(9), pp. 1650-1667.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T., Mendell, J.T. and Salzberg, S.L. (2015) StringTie enables improved reconstruction of transcriptome from RNA-Seq reads. *Nature Biotechnology*, 33(3), pp. 290-295.

Pickrell, J.K., Marioni, J.C., Pai, A.A., Degner, J.F., Engelhardt, B.E., Nkadori, E., Veyrieras, J., Stephens, M., Gilad, Y. and Pritchard, J.K. (2010) Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature*, **464**(7289), pp. 768-772.

Poitout, V., Amyot, J., Semache, M., Zarrouki, B., Hagman, D. and Fontés, G. (2010) Glucolipotoxicity of the pancreatic beta cell. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids*, **1801**(3), pp. 289-298.

Pop, M. and Salzberg, S.L. (2008) Bioinformatics challenges of new sequencing technology. *Trends in Genetics*, **24**(3), pp. 142-149.

Price, L.H., Kao, H., Burgers, D.E., Carpenter, L.L. and Tyrka, A.R. (2013) Telomeres and early-life stress: an overview. *Biological Psychiatry*, **73**(1), pp. 15-23.

Raddatz, E., Thomas, A.C., Sarre, A. and Benathan, M. (2011) Differential contribution of mitochondria, NADPH oxidases, and glycolysis to region-specific oxidant stress in the anoxic-reoxygenated embryonic heart. *American journal of physiology.Heart and circulatory physiology*, **300**(3), pp. H820-35.

Radobuljac, M.D., Bratina, N., Tomori, M. and Battelino, T. (2012) Type 1 diabetes and psychosocial risk factors in adolescence. *Zdravniski Vestnik*, **81**(9), pp.664-675.

Rains, J.L. and Jain, S.K. (2011) Oxidative stress, insulin signaling, and diabetes. *Free Radical Biology and Medicine*, **50**(5), pp. 567-575.

Ramasamy, R., Vannucci, S.J., Yan, S.S.D., Herold, K., Yan, S.F. and Schmidt, A.M. (2005) Advanced glycation end products and RAGE: a common thread in aging, diabetes, neurodegeneration, and inflammation. *Glycobiology*, **15**(7), pp. 16R-28R.

Ramasamy, R., Yan, S.F., D'Agati, V. and Schmidt, A.M. (2007) Receptor for Advanced Glycation Endproducts (RAGE): a formidable force in the pathogenesis of the cardiovascular complications of diabetes & aging. *Current Molecular Medicine*, **7**(8), pp. 699-710.

Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., Mason, C.E., Socci, N.D. and Betel, D. (2013) Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome biology*, **14**(9), pp. 3158.

Rhee, S.Y. and Kim, Y.S. (2018). The role of advanced glycation end products in diabetic vascular complications. *Diabetes & metabolism journal*, **42**(3), pp.188-195.

Richter, T. and von Zglinicki, T. (2007) A continuous correlation between oxidative stress and telomere shortening in fibroblasts. *Experimental Gerontology*, **42**(11), pp. 1039-1042.

Risso, D., Ngai, J., Speed, T.P. and Dudoit, S. (2014) Normalization of RNA-seq data using factor analysis of control genes or samples. *Nature Biotechnology*, **32**(9), pp. 896-902.

Ruckriemen, J., Hellwig, A., Schultes, S., Hellwig, M., Hahne, F. and Henle, T. (2018). Studies on the influence of dietary 3-deoxyglucosone on the urinary excretion of 2-keto-3-deoxygluconic acid. *European Food Research and Technology*, **244**(8), pp.1389-1396.

Risso, D., Schwartz, K., Sherlock, G. and Dudoit, S. (2011) GC-content normalization for RNA-Seq data. *BMC Bioinformatics*, **12**(1), pp. 480.

Rivers, C.A., Barton, J.C., Gordeuk, V.R., Acton, R.T., Speechley, M.R., Snively, B.M., Leiendecker-Foster, C., Press, R.D., Adams, P.C. and McLaren, G.D. (2007) Association of ferroportin Q248H polymorphism with elevated levels of serum ferritin

in African Americans in the Hemochromatosis and Iron Overload Screening (HEIRS) Study. *Blood Cells, Molecules, and Diseases*, **38**(3), pp. 247-252.

Roberts, A., Trapnell, C., Donaghey, J., Rinn, J.L. and Pachter, L. (2011) Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome biology*, **12**(3), pp. R22.

Robinson, M.D. and Oshlack, A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome biology*, **11**(3), pp. R25.

Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**(1), pp. 139-140.

Rohlfing, C., Wiedmeyer, H.M., Little, R., Grotz, V.L., Tennill, A., England, J., Madsen, R. and Goldstein, D. (2002) Biological variation of glycoheamoglobin. *Clinical chemistry*, **48**(7), pp. 1116-1118.

Romiguier, J., Ranwez, V., Douzery, E.J. and Galtier, N. (2010) Contrasting GC-content dynamics across 33 mammalian genomes: relationship with life-history traits and chromosome sizes. *Genome research*, **20**(8), pp. 1001-1009.

Roriz-Filho, J.S., Sá-Roriz, T.M., Rosset, I., Camozzato, A.L., Santos, A.C., Chaves, M.L., Moriguti, J.C. and Roriz-Cruz, M. (2009). (Pre) diabetes, brain aging, and cognition. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, **1792**(5), pp.432-443.

Rosenbloom, K.R., Armstrong, J., Barber, G.P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L. and Haeussler, M. (2014) The UCSC genome browser database: 2015 update. *Nucleic acids research*, **43**(D1), pp. D670-D681.

Saaddine, J.B., Fagot-Campagna, A., Rolka, D., Narayan, K.M., Geiss, L., Eberhardt, M. and Flegal, K.M. (2002) Distribution of HbA(1c) levels for children and young adults in the U.S.: Third National Health and Nutrition Examination Survey. *Diabetes Care*, **25**(8), pp. 1326-1330.

Sampson, M. and Hughes, D. (2006) Chromosomal telomere attrition as a mechanism for the increased risk of epithelial cancers and senescent phenotypes in type 2 diabetes. *Diabetologia*, **49**(8), pp. 1726-1731.

Sampson, M.J., Winterbone, M.S., Hughes, J.C., Dozio, N., and Hughes, D.A. (2006) Monocyte telomere shortening and oxidative DNA damage in type 2 diabetes. *Diabetes Care*, **29**(2), pp. 283-289.

Saretzki, G. and Von Zglinicki, T. (2002) Replicative aging, telomeres, and oxidative stress. *Annals of the New York Academy of Sciences*, **959**(1), pp. 24-29.

Schmidt, A.M. and Stern, D.M. (2000) RAGE: a new target for the prevention and treatment of the vascular and inflammatory complications of diabetes. *Trends in Endocrinology & Metabolism*, **11**(9), pp. 368-375.

Schmidt, A.M., Du Yan, S., Yan, S.F. and Stern, D.M. (2000) The biology of the receptor for advanced glycation end products and its ligands. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research***, 1498**(2), pp. 99-111.

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M. and Ragg, T. (2006) The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC molecular biology***, 7**(1), pp. 3.

Selvin, E., Halushka, M.K., Rawlings, A.M., Hoogeveen, R.C., Ballantyne, C.M., Coresh, J. and Astor, B.C. (2013) sRAGE and risk of diabetes, cardiovascular disease, and death. *Diabetes***, 62**(6), pp. 2116-2121.

Sakiyama, H., Takahashi, M., Yamamoto, T., Teshima, T., Lee, S.H., Miyamoto, Y., Misonou, Y. and Taniguchi, N. (2006). The internalization and metabolism of 3-deoxyglucosone in human umbilical vein endothelial cells. *Journal of biochemistry*, **139**(2), pp.245-253.

Semba, R.D., Sun, K., Schwartz, A.V., Varadhan, R., Harris, T.B., Satterfield, S., Garcia, M., Ferrucci, L., Newman, A.B. and Health ABC Study (2015) Serum carboxymethyl-lysine, an advanced glycation end product, is associated with arterial stiffness in older adults. *Journal of hypertension***, 33**(4), pp. 797-803;

Serra, V., Grune, T., Sitte, N., Saretzki, G. and Von Zglinicki, T. (2000) Telomere length as a marker of oxidative stress in primary human fibroblast cultures. *Annals of the New York Academy of Sciences***, 908**(1), pp. 327-330.

Sessa, L., Gatti, E., Zeni, F., Antonelli, A., Catucci, A., Koch, M., Pompilio, G., Fritz, G., Raucci, A. and Bianchi, M.E. (2014). The receptor for advanced glycation end-products (RAGE) is only present in mammals and belongs to a family of cell adhesion molecules (CAMs). *PloS one*, **9**(1), p. e86903.

Sexton, A.N., Youmans, D.T. and Collins, K. (2012) Specificity requirements for human telomere protein interaction with telomerase holoenzyme. *The Journal of biological chemistry***, 287**(41), pp. 34455-34464.

Shabalina, S.A. and Spiridonov, N.A. (2004) The mammalian transcriptome and the function of non-coding DNA sequences. *Genome biology***, 5**(4), pp. 105.

Shalev, I. (2012) Early life stress and telomere length: investigating the connection and possible mechanisms. *Bioessays***, 34**(11), pp. 943-952.

Shimizu, I., Yoshida, Y., Katsuno, T., Tateno, K., Okada, S., Moriya, J., Yokoyama, M., Nojima, A., Ito, T. and Zechner, R. (2012) p53-induced adipose tissue inflammation is critically involved in the development of insulin resistance in heart failure. *Cell Metabolism***, 15**(1), pp. 51-64.

Shiroguchi, K., Jia, T.Z., Sims, P.A. and Xie, X.S. (2012) Digital RNA sequencing minimizes sequence-dependent bias and amplification noise with optimized single-molecule barcodes. *Proceedings of the National Academy of Sciences of the United States of America***, 109**(4), pp. 1347-1352.

Singh, R., Barden, A., Mori, T. and Beilin, L. (2001) Advanced glycation end-products: a review. *Diabetologia***, 44**(2), pp. 129-146.

Singh, V.P., Bali, A., Singh, N. and Jaggi, A.S. (2014). Advanced glycation end products and diabetic complications. *The Korean Journal of Physiology & Pharmacology*, **18**(1), pp.1-14.

Škrha Jr, J., Kalousová, M., Švarcová, J., Muravska, A., Kvasnička, J., Landova, L., Zima, T. and Škrha, J. (2012) Relationship of soluble RAGE and RAGE ligands HMGB1 and EN-RAGE to endothelial dysfunction in type 1 and type 2 diabetes mellitus. *Experimental and clinical endocrinology & diabetes***, 120**(05), pp. 277-281.

Skrha Jr, J., Muravska, A., Flekac, M., Horova, E., Novak, J., Novotný, A., Prázný, M., Skrha, J., Kvasnicka, J. and Landová, L. (2014) Fructosamine 3-kinase and glyoxalase I polymorphisms and their association with soluble RAGE and adhesion molecules in diabetes. *Physiological research***, 63**pp. S283-S291.

Skyler, J.S. and Oddo, C. (2002) Diabetes trends in the USA. *Diabetes/metabolism research and reviews***, 18**(3). S21-S26.

Smart, M., Ehnholm, C., Jauhiainen, M., Robciuc, M., Cooper, J., Drenos, F., Humphries, S. and Talmud, P. (2009) Association of telomere length with type 2 diabetes, oxidative stress, and UCP2 gene variation. *Atherosclerosis***, 207**(1), pp. e6.

Smith, Z.D. and Meissner, A. (2013) DNA methylation: roles in mammalian development. *Nature Reviews Genetics*, **14**(3), pp. 204-220.

Smogorzewska, A. and de Lange, T. (2002) Different telomere damage signaling pathways in human and mouse cells. *The EMBO journal*, **21**(16), pp. 4338-4348.

Smogorzewska, A., Karlseder, J., Holtgreve-Grez, H., Jauch, A., and de Lange, T. (2002) DNA ligase IV-dependent NHEJ of deprotected mammalian telomeres in G1 and G2. *Current Biology*, **12**(19), pp. 1635-1644.

Smucker, E.J. and Turchi, J.J. (2001) TRF1 inhibits telomere C-strand DNA synthesis in vitro. *Biochemistry*, **40**(8), pp. 2426-2432.

Sodi, R., McKay, K., Dampetla, S., and Pappachan, J.M. (2018) Monitoring glycaemic control in patients with diabetes mellitus. *BMJ (Clinical research ed.)*, **363**pp. k4723.

Soleimanpour, S.A. and Stoffers, D.A. (2013) The pancreatic β cell and type 1 diabetes: innocent bystander or active participant? *Trends in Endocrinology & Metabolism*, **24**(7), pp. 324-331 .

Sone, H. and Kagawa, Y. (2005) Pancreatic beta cell senescence contributes to the pathogenesis of type 2 diabetes in high-fat-diet-induced diabetic mice. *Diabetologia*, **48**(1), pp. 58-67.

Soranzo N, Sanna S, Wheeler E, Gieger C, Radke D, Dupuis J, Bouatia-Naji N, Langenberg C, Prokopenko I, Stolerman E, Sandhu MS, Heeney MM, Devaney JM, Reilly MP, Ricketts SL, *et al*. (2010) Common variants at 10 genomic loci influence hemoglobin A1c levels via glycemic and nonglycemic pathways. Diabetes **59**(12), pp.3229-3239.

Sozou, P.D. and Kirkwood, T.B. (2001) A stochastic model of cell replicative senescence based on telomere shortening, oxidative stress, and somatic mutations in nuclear and mitochondrial DNA. *Journal of theoretical biology***, 213**(4), pp. 573-586.

Speeckaert, M., Van Biesen, W., Delanghe, J., Slingerland, R., Wiecek, A., Heaf, J., Drechsler, C., Lacatus, R., Vanholder, R., Nistor, I. and Bilo, H. (2014). Are there better alternatives than haemoglobin A1c to estimate glycaemic control in the chronic kidney disease population? *Nephrology Dialysis Transplantation*, **29**(12), pp.2167-2177.

Stern, D., Yan, S. D., Yan, S. F., and Schmidt, A. M. (2002) Receptor for advanced glycation endproducts: a multiligand receptor magnifying cell stress in diverse pathologic settings. Adv Drug Deliv Rev 54, 1615-1625.

Stern, D.L. and Orgogozo, V. (2008) The loci of evolution: how predictable is genetic evolution? *Evolution***, 62**(9), pp. 2155-2177.

Stitt, A.W., Jenkins, A.J. and Cooper, M.E. (2002) Advanced glycation end products and diabetic complications. *Expert opinion on investigational drugs***, 11**(9), pp. 1205-1223.

Stohr, B.A., Xu, L. and Blackburn, E.H. (2010). The terminal telomeric DNA sequence determines the mechanism of dysfunctional telomere fusion. *Molecular cell*, **39**(2), pp.307-314.

Su, Y., Liu, X., Sun, Y., Jin, H., Fu, R., Wang, Y., Wu, Y. and Luan, Y. (2008) The relationship between endothelial dysfunction and oxidative stress in diabetes and prediabetes. *International journal of clinical practice***, 62**(6), pp. 877-882.

Sun, L., Franco, O.H., Hu, F.B., Cai, L., Yu, Z., Li, H., Ye, X., Qi, Q., Wang, J. and Pan, A. (2008) Ferritin concentrations, metabolic syndrome, and type 2 diabetes in middle-aged and elderly Chinese. *The Journal of Clinical Endocrinology & Metabolism***, 93**(12), pp. 4690-4696.

Szwergold, B., Manevich, Y., Payne, L. and Loomes, K. (2007). Fructosamine-3-kinase-related-protein phosphorylates glucitolamines on the C-4 hydroxyl: novel substrate specificity of an enigmatic enzyme. *Biochemical and biophysical research communications*, **361**(4), pp.870-875.

Szwergold, B.S. and Beisswenger, P.J. (2003) Enzymatic deglycation--a new paradigm or an epiphenomenon? *Biochemical Society Transactions***, 31**(Pt 6), pp. 1428-1432.

Szwergold, B.S., Howell, S. and Beisswenger, P.J. (2001) Human fructosamine-3-kinase: purification, sequencing, substrate specificity, and evidence of activity in vivo. *Diabetes*, **50**(9), pp. 2139-2147.

Tamura, Y., Takubo, K., Aida, J., Araki, A. and Ito, H. (2016) Telomere attrition and diabetes mellitus. *Geriatrics & gerontology international*, **16**pp. 66-74.

Tang, F., Barbacioru, C., Nordman, E., Li, B., Xu, N., Bashkirov, V.I., Lao, K. and Surani, M.A. (2010) RNA-Seq analysis to capture the transcriptome landscape of a single cell. *Nature protocols*, **5**(3), pp. 516-535.

Tang, F., K. Lao, and M.A. Surani. (2011). Development and applications of single-cell transcriptome analysis. Nature methods. **8pp.**S6-11.

Tanhäuserová, V., Kuricová, K., Pácal, L., Bartáková, V., Řehořová, J., Svojanovský, J., Olšovský, J., Bělobrádková, J. and Kaňková, K. (2014) Genetic variability in enzymes of metabolic pathways conferring protection against non-enzymatic glycation versus diabetes-related morbidity and mortality. *Clinical chemistry and laboratory medicine*, **52**(1), pp. 77-83.

Teixeira, M.T. (2013). Saccharomyces cerevisiae as a model to study replicative senescence triggered by telomere shortening. *Frontiers in oncology*, **3**pp.101.

Thomas, C.C. and Philipson, L.H. (2015) Update on diabetes classification. *The Medical clinics of North America*, **99**(1), pp. 1-16.

Thomas, M., Söderlund, J., Lehto, M., Mäkinen, V., Moran, J., Cooper, M., Forsblom, C., Groop, P., and FinnDiane Study Group (2011) Soluble receptor for AGE (RAGE) is a novel independent predictor of all-cause and cardiovascular mortality in type 1 diabetes. *Diabetologia*, **54**(10), pp. 2669-2677.

Thomas, M.C., Woodward, M., Neal, B., Li, Q., Pickering, R., Marre, M., Williams, B., Perkovic, V., Cooper, M.E., Zoungas, S., Chalmers, J., Hillis, G.S. and ADVANCE Collaborative Group (2015) Relationship between levels of advanced glycation end products and their soluble receptor and adverse outcomes in adults with type 2 diabetes. *Diabetes Care*, **38**(10), pp. 1891-1897.

Thomas, P., O'Callaghan, N.J. and Fenech, M. (2008) Telomere length in white blood cells, buccal cells and brain tissue and its variation with ageing and Alzheimer's disease. *Mechanisms of ageing and development*, **129**(4), pp. 183-190.

Thornalley, P.J., Langborg, A. and Minhas, H.S. (1999) Formation of glyoxal, methylglyoxal, and 3-deoxyglucosone in the glycation of proteins by glucose. *The Biochemical journal*, **344 Pt 1**pp. 109-116.

Tian, B. and Manley, J.L. (2013) Alternative cleavage and polyadenylation: the long and short of it. *Trends in biochemical sciences*, **38**(6), pp. 312-320.

Tobon-Velasco, J., Cuevas, E. and A Torres-Ramos, M. (2014). Receptor for AGEs (RAGE) as mediator of NF-kB pathway activation in neuroinflammation and oxidative stress. *CNS & Neurological Disorders-Drug Targets (Formerly Current Drug Targets-CNS & Neurological Disorders)*, **13**(9), pp.1615-1626.

Trapnell, C. and Salzberg, S.L. (2009). How to map billions of short reads onto genomes. *Nature biotechnology*, **27**(5), p.455-457.

Trapnell, C., Pachter, L. and Salzberg, S.L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics***, 25**(9), pp. 1105-1111.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M.J., Salzberg, S.L., Wold, B.J. and Pachter, L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology***, 28**(5), pp. 511-515.

Turro, E., Su, S., Gonçalves, Â., Coin, L.J., Richardson, S. and Lewin, A. (2011) Haplotype and isoform specific expression estimation using multi-mapping RNA-seq reads. *Genome biology***, 12**(2), pp. R13.

Ulrich, P. and Cerami, A. (2001) Protein glycation, diabetes, and aging. *Recent progress in hormone research***, 56**pp. 1-21.

Uppu, S., Krishna, A. and Gopalan, R.P. (2016). A review on methods for detecting SNP interactions in high-dimensional genomic data. *IEEE/ACM transactions on computational biology and bioinformatics*, **15**(2), pp.599-612.

Uziel, O., Singer, J.A., Danicek, V., Sahar, G., Berkov, E., Luchansky, M., Fraser, A., Ram, R. and Lahav, M. (2007) Telomere dynamics in arteries and mononuclear cells of diabetic patients: effect of diabetes and of glycemic control. *Experimental Gerontology***, 42**(10), pp. 971-978.

Valente, L. and Nishikura, K. (2005) ADAR gene family and A-to-I RNA editing: diverse roles in posttranscriptional gene regulation. *Progress in nucleic acid research and molecular biology*, **79**pp. 299-338.

Van Schaftingen, E., Collard, F., Wiame, E. and Veiga-da-Cunha, M. (2012) Enzymatic repair of Amadori products. *Amino acids*, **42**(4), pp. 1143-1150.

Veiga da-Cunha, M., Jacquemin, P., Delpierre, G., Godfraind, C., Theate, I., Vertommen, D., Clotman, F., Lemaigre, F., Devuyst, O. and Van Schaftingen, E. (2006) Increased protein glycation in fructosamine 3-kinase-deficient mice. *The Biochemical journal*, **399**(2), pp. 257-264.

Vlassara, H. and Striker, G.E. (2013) Advanced glycation endproducts in diabetes and diabetic complications. *Endocrinology and metabolism clinics of North America*, **42**(4), pp. 697-719.

Vlassara, H., Cai, W., Crandall, J., Goldberg, T., Oberstein, R., Dardaine, V., Peppa, M. and Rayfield, E.J. (2002) Inflammatory mediators are induced by dietary glycotoxins, a major risk factor for diabetic angiopathy. *Proceedings of the National Academy of Sciences of the United States of America*, **99**(24), pp. 15596-15601.

Vogel, C. and Marcotte, E.M. (2012) Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics*, **13**(4), pp. 227-232.

Von Zglinicki, T. (2000) Role of oxidative stress in telomere length regulation and replicative senescence. *Annals of the New York Academy of Sciences*, **908**(1), pp. 99-110.

Von Zglinicki, T. (2002) Oxidative stress shortens telomeres. *Trends in biochemical sciences*, **27**(7), pp. 339-344.

von Zglinicki, T., Martin-Ruiz, C.M. and Saretzki, G. (2005) Telomeres, cell senescence, and human ageing. *Signal Transduction*, **5**(3), pp. 103-114.

Vos, F.E., Schollum, J.B. and Walker, R.J. (2011) Glycated albumin is the preferred marker for assessing glycaemic control in advanced chronic kidney disease. *Nephrology Dialysis Transplantation Plus*, **4**(6), pp. 368-375.

Wagner, G.P., Kin, K. and Lynch, V.J. (2012) Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory in biosciences*, **131**(4), pp. 281-285.

Wai, L.K. (2004). Telomeres, telomerase, and tumorigenesis--a review. *Medscape General Medicine*, **6**(3), pp. 19.

Wajid, B. and Serpedin, E. (2012) Review of general algorithmic features for genome assemblers for next-generation sequencers. *Genomics, proteomics & bioinformatics*, **10**(2), pp. 58-73.

Wang, C.Y. (2018). Asynchronous Shortening of Telomere Length and Cardiovascular Outcomes. *JACC: Basic to Translational Science,* **3**(5), pp. 601-603.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P. and Burge, C.B. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature*, 456(7221), pp. 470-476.

Wang, F., Podell, E.R., Zaug, A.J., Yang, Y., Baciu, P., Cech, T.R. and Lei, M. (2007) The POT1-TPP1 telomere complex is a telomerase processivity factor. *Nature***, 445**(7127), pp. 506-510.

Wang, J., Dong, X., Cao, L., Sun, Y., Qiu, Y., Zhang, Y., Cao, R., Covasa, M. and Zhong, L. (2016). Association between telomere length and diabetes mellitus: A meta-analysis. *Journal of International Medical Research*, **44**(6), pp.1156-1173.

Wang, T., Liang, Z.A., Sandford, A.J., Xiong, X.Y., Yang, Y.Y., Ji, Y.L. and He, J.Q. (2012). Selection of suitable housekeeping genes for real-time quantitative PCR in CD4+ lymphocytes from asthmatics with or without depression. *PloS one*, **7**(10), p.e48367.

Wang, Z., Gerstein, M., and Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews genetics***, 10**(1), pp. 57-63.

Wautier, M.P., Chappey, O., Corda, S., Stern, D.M., Schmidt, A.M. and Wautier, J.L. (2001) Activation of NADPH oxidase by AGE links oxidant stress to altered gene expression via RAGE. American journal of physiology.Endocrinology and metabolism, 280(5), pp. E685-694.

Wei, N., Zheng, H. and Nathan, D.M. (2014). Empirically establishing blood glucose targets to achieve HbA1c goals. *Diabetes Care*, **37**(4), pp.1048-1051.

Wesche, B., Jaeckel, E., Trautwein, C., Wedemeyer, H., Falorni, A., Frank, H., von zur Muhlen, A., Manns, M.P. and Brabant, G. (2001) Induction of autoantibodies to the adrenal cortex and pancreatic islet cells by interferon alpha therapy for chronic hepatitis C. *Gut***, 48**(3), pp. 378-383.

Weykamp, C., John, W.G. and Mosca, A. (2009) A review of the challenge in measuring hemoglobin A1c. *Journal of diabetes science and technology,* **3**(3), pp.439-445.

Wheeler, E., Leong, A., Liu, C., Hivert, M., Strawbridge, R.J., Podmore, C., Li, M., Yao, J., Sim, X. and Hong, J. (2017) Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. *PLoS medicine***, 14**(9), pp. e1002383.

Wiame, E., Delpierre, G., Collard, F., and Van Schaftingen, E. (2002) Identification of a pathway for the utilization of the Amadori product fructoselysine in Escherichia coli. *The Journal of biological chemistry***, 277**(45), pp. 42523-42529.

Wilcox, G. (2005) Insulin and insulin resistance. *The Clinical Biochemist.Reviews*, **26**(2), pp. 19-39.

Wild, S.H., Roglic, G., Green, A., and Sicree, R. (2004) Global prevalence of diabetes: estimates for the year 2030 and projections for 2030: response to Rathman and Giani. *Diabetes Care*, ***27***(10), pp.1047-1053.

Wilhelm, B.T. and Landry, J. (2009) RNA-Seq—quantitative measurement of expression through massively parallel RNA-sequencing. *Methods*, **48**(3), pp. 249-257.

Wittkopp, P. (2005) Genomic sources of regulatory variation in cis and in trans. *Cellular and Molecular Life Sciences CMLS*, **62**(16), pp. 1779-1783.

World Health Organization (2006) Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia: report of a WHO/IDF consultation.

Wu, L., Multani, A.S., He, H., Cosme-Blanco, W., Deng, Y., Deng, J.M., Bachilo, O., Pathak, S., Tahara, H. and Bailey, S.M. (2006) Pot1 deficiency initiates DNA damage checkpoint activation and aberrant homologous recombination at telomeres. *Cell*, **126**(1), pp. 49-62.

Wu, Y.Y., Xiao, E. and Graves, D.T. (2015). Diabetes mellitus related bone metabolism and periodontal disease. *International journal of oral science*, **7**(2), p.63-72.

Xin, H., Liu, D., Ma, W., Safari, A., Kim, H., Sun, W., O'connor, M.S. and Songyang, Z. (2007) TPP1 is a homologue of ciliate TEBP-[beta] and interacts with POT1 to recruit telomerase. *Nature*, **445**(7127), pp. 559-562.

Xu, H., Barnes, G.T., Yang, Q., Tan, G., Yang, D., Chou, C.J., Sole, J., Nichols, A., Ross, J.S., Tartaglia, L.A. and Chen, H. (2003) Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance. *The Journal of clinical investigation*, **112**(12), pp. 1821-1830.

Yamagishi, S. and Matsui, T. (2010) Soluble form of a receptor for advanced glycation end products (sRAGE) as a biomarker. *Frontiers in bioscience (Elite edition)*, **2**pp. 1184-1195.

Yan, S.F., Barile, G.R., D'Agati, V., Du Yan, S., Ramasamy, R. and Schmidt, A.M. (2007) The biology of RAGE and its ligands: uncovering mechanisms at the heart of diabetes and its complications. *Current diabetes reports*, **7**(2), pp. 146-153.

Yan, S.F., Ramasamy, R. and Schmidt, A.M. (2008) Mechanisms of disease: advanced glycation end-products and their receptor in inflammation and diabetes complications. *Nature Reviews. Endocrinology*, **4**(5), pp. 285-293.

Yan, S.F., Ramasamy, R. and Schmidt, A.M. (2009) The receptor for advanced glycation endproducts (RAGE) and cardiovascular disease. *Expert reviews in molecular medicine*, **11**.

Yan, S.F., Ramasamy, R. and Schmidt, A.M. (2010) The RAGE axis: a fundamental mechanism signaling danger to the vulnerable vasculature. *Circulation research***,** **106**(5), pp. 842-853.

Yáñez-Mó, M., Siljander, P.R.M., Andreu, Z., Bedina Zavec, A., Borràs, F.E., Buzas, E.I., Buzas, K., Casal, E., Cappello, F., Carvalho, J. and Colás, E. (2015). Biological properties of extracellular vesicles and their physiological functions. *Journal of extracellular vesicles*, **4**(1), p.27066.

Yonekura, H., Yamamoto, Y., Sakurai, S., Petrova, R.G., Abedin, M.J., Li, H., Yasui, K., Takeuchi, M., Makita, Z., Takasawa, S., Okamoto, H., Watanabe, T. and Yamamoto, H. (2003) Novel splice variants of the receptor for advanced glycation end-products expressed in human vascular endothelial cells and pericytes, and their putative roles in diabetes-induced vascular injury. *The Biochemical journal*, **370**(Pt 3), pp. 1097-1109.

Yu, E.Y., Wang, F., Lei, M. and Lue, N.F. (2008) A proposed OB-fold with a protein-interaction surface in Candida albicans telomerase protein Est3. *Nature structural & molecular biology***,** **15**(9), pp. 985-989.

Zee, R.Y., Castonguay, A.J., Barton, N.S., Germer, S. and Martin, M. (2010) Mean leukocyte telomere length shortening and type 2 diabetes mellitus: a case-control study. *Translational research***,** **155**(4), pp. 166-169.

Zemlin, A.E., Matsha, T.E., Hassan, M.S. and Erasmus, R.T. (2011) HbA1c of 6.5% to diagnose diabetes mellitus—does it work for us? —the Bellville South Africa study. *PloS one*, **6**(8), pp. e22558.

Zeng, W. and Mortazavi, A. (2012) Technical considerations for functional sequencing assays. *Nature Immunology*, **13**(9), pp. 802-807.

Zglinicki, T.v. and Martin-Ruiz, C. (2005) Telomeres as biomarkers for aging and age-related diseases. *Current Molecular Medicine*, **5**(2), pp. 197-203.

Zhang, L., Postina, R. and Wang, Y. (2009) Ectodomain shedding of the receptor for advanced glycation end products: a novel therapeutic target for Alzheimer's disease. *Cellular and molecular life sciences*, **66**(24), pp. 3923-3935.

Zhang, M., Kho, A.L., Anilkumar, N., Chibber, R., Pagano, P.J., Shah, A.M. and Cave, A.C. (2006) Glycated proteins stimulate reactive oxygen species production in cardiac myocytes: involvement of Nox2 (gp91phox)-containing NADPH oxidase. *Circulation*, **113**(9), pp. 1235-1243.

Zhang, Y., Sturgill, D., Parisi, M., Kumar, S. and Oliver, B. (2007) Constraint and turnover in sex-biased gene expression in the genus Drosophila. *Nature*, **450**(7167), pp. 233-237.

Zhao, S. (2014). Assessment of the impact of using a reference transcriptome in mapping short RNA-Seq reads. *PLoS One*, 9(7), p.e101374.

Zheng, F., He, C., Cai, W., Hattori, M., Steffes, M. and Vlassara, H. (2002) Prevention of diabetic nephropathy in mice by a diet low in glycoxidation products. *Diabetes/metabolism research and reviews*, **18**(3), pp. 224-237.

Zheng, W., Chung, L.M. and Zhao, H. (2011) Bias detection and correction in RNA-Sequencing data. *BMC Bioinformatics*, **12**(1), pp. 290.

Zhu, H., Belcher, M. and van der Harst, P. (2011) Healthy aging and disease: a role for telomere biology? *Clinical science (London, England: 1979)*, **120**(10), pp. 427-440.

Zieman, S.J., Melenovsky, V., Clattenburg, L., Corretti, M.C., Capriotti, A., Gerstenblith, G., and Kass, D.A. (2007) Advanced glycation end-product cross-link breaker (alagebrium) improves endothelial function in patients with isolated systolic hypertension. *Journal of hypertension*, **25**(3), pp. 577-583.

Appendices

Appendix 1

1.1. Specialised kit, chemical, and reagents for clinical research

1.1.1. The specialised kits for clinical research

Duoset Elisa kit for Human sRAGE	R&D systems Europe Ltd Abingdon Science Park, UK
QIAamp DNA Blood Mini Kit	QIAGEN Lloyd Street North Manchester, UK
TaqMan Gene Expression Assay	ThermoFisher Scientific
Norgen Bioteck RNA purification kit	Norgen, Thorold, On, Canada
RNase-Free DNase 1 kit	Norgen, Thorold, On, Canada

RiboGreen RNA assay kit (Invitrogen)	ThermoFisher Scientific
Ribo-Zero rRNA removal kit	Epicentre an Illumina Biotechnology Company, Research Park, Madison, USA
TruSeq standard total RNA kit	Illumina Madison Epicentre, USA

1.1.2. Reagents and Chemicals for clinical research

PBS: 137 mM NaCl, 2.7 mM KCl, 8.1 mM Na_2HPO_4, 1.5 mM KH_2PO_4, pH 7.2 - 7.4, 0.2 µm filtered

Wash Buffer: 0.05% Tween 20 in PBS, pH 7.2-7.4

Reagent Diluent: 1% BSA in PBS, pH 7.2-7.4, 0.2 µm filtered. Quality of BSA is critical.

Substrate Solution: 1:1 mixture of Color Reagent A (H_2O_2) and Color Reagent B (Tetramethylbenzidine).

Stop Solution - 2 N H_2SO_4

Appendix 2

Table: Serial dilution of the standard stock solution

Standard No.	Standard Volume	2 X Assay Buffer	Concentrations
Stock	500 µl	-------	4000 pg/ml
# 1	500 µl of stock	500 µl	2000 pg/ml
# 2	500 µl of # 1	500 µl	1000 pg/ml
# 3	500 µl of # 2	500 µl	500 pg/ml
# 4	500 µl of # 3	500 µl	250 pg/ml
# 5	500 µl of # 4	500 µl	125 pg/ml
# 6	500 µl of # 5	500 µl	62.5 pg/ml
# 7	-------	500 µl	0 pg/ml

Appendix 3

Consent form

Study Title: The enzymatic deglycation pathway may be an explanation for the variation in glycated HbA_{1c} - Exploring the glycation gap

Name of Researcher: [redacted]
Telephone: [redacted] (Office), [redacted] (mobile)
E-mail: [redacted]

Please initial box

1. I confirm that I have read and understand the information sheet dated……(version 1.2 dated5/5/2011……) for the above study and have had the opportunity to ask questions and have had these answered satisfactorily. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time; without giving a reason, without effecting my treatment or my legal rights. ☐

3. I understand that relevant section of my medical notes and data collected during the study, may be looked at by the researcher and supervisor(s) at New cross hospital, from regulatory authorities or from the NHS trust, where it is relevant to my taking in his research. I give permission for these individuals to have access to my records. ☐

4. I understand that the project using the sample I give will include DNA extraction and genetic research aimed at understanding the genetic influences on factors that may influence variation in HbA_{1c} levels but that the results of these investigations may be unlikely to have any implications for me personally. ☐

5. I agree that samples (genetic or otherwise) I have given, and the information gathered about me can be stored and possibly used for future projects in diabetes, if these projects are approved by local research ethics committee. ☐

6. I understand that I will not benefit financially if this research leads to the development of a new treatment or medical test ☐

7. I agree to take part in the above study. ☐

Thank you for agreeing to participate in this research

Name of patient	Date	Signature
Name of person taking consent (if different from researcher)	Date	Signature

Each individual who signs this document must PERSONALLY date his or her signature.

When completed, 1 for patient; 1 for researcher site file; 1 (original) to be kept in medical notes

Appendix 4

Conference presentations

1. Abstract: 468-P 08 June 2017; Simon J. Dunmore, Amr S. Alderawi, Aruna Narshi, Fakhra Naseem, James E.P Brown, Alan Nevill, Ananth Nayak, Baldev M Singh: Evidence of the Role of Fructosamine-3-Kinase in the Glycation Gap and in Protection against Complications in Patients with Diabetes.

2. Abstract: 0494-P 06 December 2017; Simon J. Dunmore, Fakhra Naseem, James E.P Brown, Adam Watkins, Paul Kirkham, Ananth Nayak, Baldev M Singh: Single Nucleotide Polymorphisms (SNPs) in the Fructosamine-3-Kinase gene as a possible explanation of variation in the glycation gap between individuals.