Functional analysis of type 2 diabetes associated transcripts

A thesis submitted to the University of Oxford



in candidature for the degree of Doctor of Philosophy

by

Hannah B. Richards

Green-Templeton College, University of Oxford

Oxford Centre for Diabetes, Endocrinology and Metabolism Radcliffe Department of Medicine, University of Oxford

MRC Mammalian Genetics Unit, Harwell

Michaelmas 2014

Memorandum

The work in this thesis is the original work of the author. Experiments were carried out at the Oxford Centre for Diabetes Endocrinology and Metabolism, University of Oxford, under the supervision of Professor Mark McCarthy, and at the MRC, Harwell, under the supervision of Professor Roger Cox.

Funding was provided by the Clarendon Award and by the Nuffield Prize Studentship.

I hereby state that no part of this thesis has been submitted for any other degree at this or any other university.

This thesis is approximately 50,000 words.

Functional analysis of type 2 diabetes associated transcripts

Hannah Richards, Green-Templeton College

Submitted in candidature for the degree of Doctor of Philosophy, Michaelmas term 2014

Abstract

Genome wide association studies (GWAS) have transformed the study of the heritability of complex diseases such as type 2 diabetes (T2D), with the current tally of established risk loci close to ninety. Each of these loci has the potential to offer novel insights into the biology of this disease, and opportunities for clinical exploitation. However, the complexity of T2D has often frustrated efforts to achieve these functional and translational advances. This thesis aims to delve into the functional characterisation of two known susceptibility loci, *KLF14* and *ADCY5*, and describe findings relevant to disease pathology.

KLF14 and *ADCY5* are two loci associated with T2D predisposition working through disparate mechanisms. Variants at the maternally imprinted *KLF14* locus are associated with measures of insulin resistance and expression data has implicated *KLF14* as a master regulator of genes in adipose tissue. In contrast, variation at the *ADCY5* locus is associated with impaired beta cell function, high fasting glucose, and low birth weight suggesting *ADCY5* is having an effect on insulin secretion. In this thesis, ENU mouse models of these two genes are investigated functionally to elucidate more about the pathology of common human variation at these loci.

A mouse model was derived with an ENU point mutation at *Adcy5* Y1064C. Phenotyping of this model revealed improved oral glucose tolerance, and secretion studies from isolated islet cells demonstrated impaired glucagon secretion from mice homozygous for the Y1064C mutation in the presence of adrenaline. These results suggest that *Adcy5* is involved in glucagon regulation in the alpha cell. The *Adcy5* Y1064C confers a protective effect against hyperglycaemia in mouse indicating that the T2D risk allele at the *ADCY5* locus in humans may have the opposite direction of effect.

A mouse model containing the ENU point mutation *Klf14* R238L predicted to be disruptive to KLF14 protein function showed no significant difference in body weight, measures of insulin resistance, or blood cholesterol. However, expression of several genes associated in *trans* with variation near *KLF14* in humans was changed in adipose tissue and skeletal muscle when the R238L mutation was inherited maternally compared to mice which had inherited the mutation paternally or carried two wild type alleles. This result suggests a mechanism by which *Klf14* is regulating genes across metabolic tissues.

Acknowledgements

Thank you to my supervisors, Prof Mark McCarthy and Prof Roger Cox for all your support and guidance. I am grateful to my funding bodies, the Clarendon Award and the Nuffield Prize Studentship for the opportunity to pursue my doctorate at the University of Oxford.

I am grateful to all those who trained and mentored me during this project: Anne Clark, Reshma Ramracheya, Dave Wiggins, and Liz Bentley.

I miss and will remember fondly the people at OCDEM who helped me in so many ways: Neelam, Mary, Anne, Nicola, Martijn, Ollie, Lily, Laura, Jana, Saima, Matt, Amy, Mandy, Chris, and Peta. Thank you for all the happy memories.

Thank you to Rosemary for the gift of grace.

A very big thanks to my friends back home and around the globe who have encouraged me and who I can always count on: Becky, Hannah, Kathy, Elizabeth, Christina, Josefin, Katelyn, and Elise.

I am so grateful for my parents and the other members of my family for the gift of unconditional love and endless support.

Thank you to my fiancé, Sean, for undying encouragement and enthusiasm even in the face of my most stormy moods during the writing process. We make a great team.

I could not have come this far without all of you.

List of common abbreviations

ATP adenosine triphosphate

BMI body mass index

bp Base pairs

cDNA Complementary deoxyribonucleic acid

Chr Chromosome

ddNTPs Dideoxyribonucleotide

dH_20 Distilled water

DIAGRAM Diabetes genetics replication and meta-analysis consortium

DNA Deoxyribonucleic acid

dNTPs Deoxyribonucleotide

ENU N-ethyl-nitrosaurea

eQTL Expression quantitative trait loci

GWA(S) Genome-wide association (study)

HKG Housekeeping gene

HOMA-B Homeostatic model assessment of beta-cell function

HOMA-IR Homeostatic model assessment of insulin resistance

IPGTT Intraperitoneal glucose tolerance test

Kb Kilobase

LD Linkage Disequilibrium

MAF Minor allele frequency

MAGIC Meta-analysis of glucose and insulin related traits consortium

Mb Megabase

miRNA Micro ribonucleic acid

MODY Maturity onset diabetes of the young

mRNA Messenger ribonucleic acid

NCBI National Centre for Biotechnology Information

OGTT Oral glucose tolerance test

OR Odds ratio

PBS Phosphate buffered saline

PCR Polymerase chain reaction

QC Quality control

QPCR Quantitative polymerase chain reaction

RAF Risk allele frequency

RNA Ribonucleic acid

RT Reverse transcriptase

SD Standard deviation

SEM Standard error of the mean

SNP Single nucleotide polymorphism

T1D Type 1 diabetes

T2D Type 2 diabetes

TBE Tris/Borate/EDTA

UTR Untranslated region

WHO World health organization

WTCCC Wellcome Trust Case Control Consortium

WT Wild type

Contents

Chapter 1: Introduction 1

1.1 Overview of efforts to characterize the genetic component of type 2 diabetes 2

1.2 The pathophysiology of T2D 3

1.3 The evolutions of genetic studies of T2D 4

1.4 Recent advances in identifying the genetic basis of T2D 9

1.5 The next step: elucidating the biological mechanisms driving association 16

1.6 Loci to be examined in this thesis 21

1.7 *KLF14*: a T2D susceptibility locus working through defective insulin action 21

1.8 Variation in *ADCY5* has multiple effects on metabolism and beta cell function 24

1.9 Thesis aims 26

Chapter 2: Materials and Methods 28

2.1 Standard Solutions and General Reagents 29

2.2 DNA Methods 29

2.2.1 DNA Amplification 29

2.2.1 (a) Primer Design 29

2.2.1 (b) PCR reaction components and cycling conditions 29

2.2.1 (c) Gel electrophoresis 32

2.2.2 Isolation and purification of DNA fragments 32

2.2.2 (a) QiaQuick Gel Extraction Kit (Qiagen) 32

2.2.2 (b) QiaQuick PCR Purification Kit (Qiagen) 33

2.2.3 Genomic DNA extraction 33

2.2.3 (a) DNeasy Genomic DNA Extraction (Qiagen) 33

2.2.3 (b) Maxwell 16 System Tissue DNA Extraction 33

2.2.4 DNA quality and concentration determination 34

2.2.5 Sequence analysis 34

2.3 RNA Methods 35

2.3.1 RNA extraction 35

2.3.1 (a) Human and mouse islet processing and storage for RNA extraction 35

2.3.1 (b) RNA extraction from human and mouse islets 35

2.3.1 (c) RNA extraction from immortalized cell lines 36

2.3.1 (d) RNA extraction from mouse tissue 36

2.3.2 cDNA Synthesis 37

2.3.3 mRNA quantification 37

2.3.3 (a) Taqman probe selection 38

2.3.3 (b) Housekeeping gene selection 38

2.3.3 (c) Generating a standard curve for qPCR normalization 39

2.3.3 (d) Quantitative real-time PCR 40

2.3.3 (e) Analysis 40

2.4 Cell Biology 41

2.4.1 Cell culture 41

2.4.2 Immunocytochemistry 42

2.4.2 (a) Processing of pancreas specimens into wax for light microscopy 42

2.4.2 (b) Immunoperoxidase labeling of pancreas sections 42

2.4.2 (c) Fluorescent labeling of pancreas sections 43

2.4.2 (d) Fixing dissociated islet cells on coverslips 43

2.4.2 (e) Fluorescent labeling of islet cells 44

2.4.3 Electron Microscopy 44

2.4.4 Histological Image Analysis 45

2.4.5 Insulin and Glucagon Secretion from Isolated Islets 46

2.4.5 (a) Isolation of mouse islets 46

2.4.5 (b) Secretion stimulation and collection 48

2.4.6 In situ pancreatic perfusion 48

2.4.7 Radio-immuno assays 49

2.4.7 (a) Insulin assay 49

2.4.7 (b) Glucagon assay 50

2.4.7 (c) Analysis 51

2.5 Animal Husbandry and Metabolic Phenotyping 51

2.5.1 Genotyping 52

2.5.1 (a) Assay design 52

2.5.1 (b) TaqMan PCR 52

2.5.1 (c) Mouse genotype determination from PCR reaction 53

2.5.1 (d) Pyrosequencing 53

2.5.2 Metabolic phenotyping of mice 54
2.5.2 (a) Body composition 54
2.5.2 (b) Adult tissue collection 55
2.5.2 (c) Neonate tissue collection for birth weight measurement 55
2.5.2 (d) Determining bone lengths by X-ray analysis 55
2.5.2 (e) Blood collection and intraperitoneal glucose tolerance test (IPGTT) 55
2.5.3 Plasma analysis of mouse blood samples 56
2.6 Statistical analysis 56
2.6.1 Pathway analysis of *KLF14 trans*-associated transcripts 56
Chapter 3: Screening of the ENU archive at the JAZF1, KLF14, and ADCY5 loci, and re-derivation of mouse models 57
3.1 Introduction 58
3.1.1 The need for functional investigation at T2D-associated loci 58
3.1.2 The benefits of modelling T2D genetic predispositions in mice 58
3.1.3 The challenges of modelling T2D genetic predispositions in mice 60
3.1.4 The ENU method of deriving mouse models 61
3.1.5 Benefits of using the ENU method to generate mouse models of T2D-associated genes 63
3.1.6 Selection criteria for genes to scan in the ENU archive 66
3.1.7 Variants at the *JAZF1* locus are associated with T2D and beta cell phenotypes 68
3.1.8 Variants at the *KLF14* locus are associated with T2D and insulin resistance phenotypes 69
3.1.9 Variants at the *ADCY5* locus are associated with T2D and beta cell phenotypes 70
3.1.10 Experimental aims 72
3.2 Samples and Methods 72
3.2.1 Determination of amplicons for ENU archive screen 72
3.2.2 Screening the ENU archive 73
3.2.3 Sequence alignment 74
3.2.4 Analysis of variants 74
3.2.5 Re-derivation of ENU mutant mice 78
3.3 Results 78
3.3.1 *JAZF1* 78
3.3.2 *KLF14* 79
3.3.3 *ADCY5* 84

3.4 Discussion 92

Chapter 4: Characterization *in vivo* and *in vitro* of the *Adcy5*^{Y1064C} mouse model 95

4.1 Introduction 96

4.1.1 GWAS signals near ADCY5 associate with type 2 diabetes 96

4.1.2 Adenylate cyclases in the pancreatic islet cells 96

4.1.3 Previous mouse models of Adcy5 indicate a metabolic phenotype 98

4.1.4 Adcy5Y1064C may disrupt protein function leading to alpha and/or beta cell dysfunction 99

4.1.5 Aims 100

4.2 Methods 100

4.2.1 Samples used 100

4.2.2 mRNA expression quantification 101

4.2.3 Adcy5Y1064C phenotyping pipeline 102

4.2.4 Birth weight measurements 104

4.2.7 Static hormone secretion assays 107

4.3 Results 107

4.3.1 Quantitative real-time PCR analysis of adenylate cyclases in humans and mice 107

4.3.1 (a) Expression of *ADCY5* and *Adcy5* in humans and mice across relevant tissues 107

4.3.1 (b) Expression of all adenylate cyclases in human and mouse islets 108

4.3.2 Major findings of the *in vivo* and islet studies 111

4.3.3 *Adcy5*^{Y1064C} homozygous mice have lower blood glucose than wild type 111

4.3.3 (a) Mice with 2 copies of the *Adcy5*^{Y1064C} mutation have improved glucose tolerance compared to wild type 112

4.3.3 (b) There was no difference between genotypes after an intraperitoneal glucose tolerance test 117

4.3.3 (c) Fasting insulin and glucagon do not differ between genotypes 122

4.3.3 (d) Body composition does not differ between genotypes 124

4.3.4 The *Adcy5*^{Y1064C} variant has no effect on birth weight or size 124

4.3.5 Islets isolated from mice homozygous for the *Adcy5*^{Y1064C} mutation have altered glucagon secretion 127

4.3.5 (a) Islets from mice homozygous for the *Adcy5*^{Y1064C} mutation secrete less glucagon than wild type in the presence of adrenaline 128

4.3.5 (b) Glucagon islet content does not differ between genotypes 129

4.3.5 (c) Glucose-mediated glucagon inhibition is not altered by the *Adcy5*^{Y1064C} mutation 129

4.3.5 (d) Forskolin-stimulated glucagon secretion is impaired in the *Adcy5*^{Y1064C} mutant mice 130

4.3.6 Insulin secretion from isolated islets in response to low glucose is elevated in cohort 1 Y1064C allele carriers 136

4.3.6 (a) Glucose-stimulated insulin secretion is not consistently different between genotypes 137

4.3.6 (b) Insulin content is significantly reduced in *Adcy5*^{Y1064C} allele carriers 138

4.3.6 (c) There were no consistent changes in insulin secretion in response to forskolin or adrenaline between genotypes 138

4.3.7 Insulin secretion in high glucose + GLP-1 is different between Y1064C homozygote mice and wild type 147

4.4 Conclusions 151

4.4.1 *ADCY5* is expressed in pancreatic islets and in adipose tissue 151

4.4.2 *Adcy5*^{Y1064C} homozygote mice have lower blood glucose after an oral glucose tolerance test than wild type mice 151

4.4.3 Insulin secretion from *Adcy5*^{Y1064C} islets is not consistently different from wild type islets153

4.4.4 *Adcy5*^{Y1064C} homozygote mice secrete less glucagon in response to adrenaline 154

4.4.5 Discussion of study limitations 157

Chapter 5: Morphological characterization of the ADCY5 protein in human and mouse islets 159

5.1 Introduction 160

5.1.1 Morphometry of the ADCY5 protein in mice and humans 160

5.1.2 Expression of *ADCY5* in islet cells 160

5.1.3 Beta and alpha cell mass in type 2 diabetes 161

5.1.4 Chapter aims 163

5.2 Methods 163

5.2.1 Immunolabeling of pancreatic sections 163

5.2.2 Immunolabeling of islet cells on coverslips 164

5.2.3 Electron Microscopy 164

5.2.4 Histological Quantitative Analysis 164

5.3 Results 165

5.3.1 Morphological differences in islets between wild type and Y1064C mutants 165

5.3.2 Islet quantification in wild type and Y1064C mutants showed no differences between genotypes 166

5.3.3 Visualizing ADCY5 in mouse isolated islet cells....169
5.3.4 Visualizing ADCY5 in mouse pancreas sections from *Adcy5*^{Y1064C} mice....169
5.3.5 Structure of islet cells in wild type and Y1064C mutants using electron microscopy 173
5.3.6 Visualizing ADCY5 in human pancreas sections....173
5.4 Conclusions....178
5.4.1 Morphometric analysis of islets from *Adcy5*^{Y1064C} homozygote and heterozygote mice did not show quantifiable differences from wild type....178
5.4.2 The Y1064C mutation alters distribution of the ADCY5 protein in mouse islets....179
5.4.3 ADCY5 localizes to the cytoplasm of alpha and beta cells....180
5.4.4 The pattern of ADCY5 distribution in Y1064C homozygous mouse islets matches ADCY5 distribution in diabetic humans....181
Chapter 6: KLF14: a promising locus for insulin resistance and a master regulator of genes in fat tissue 183
6.1 Introduction....184
6.1.1 Variants near *KLF14* are associated with multiple phenotypes in humans....184
6.1.2 Significance of the imprinting pattern of *KLF14* 185
6.1.3 Structure and biology of Krueppel-like transcription factors 186
6.1.4 Krueppel-like transcription factors and adipocyte biology....187
6.1.5 *KLF14* is a master trans regulator of genes in adipose 188
6.2 Methods....190
6.2.1 Samples used for RT-qPCR analysis 190
6.2.2 Analyses of KLF14 trans-associated transcripts using MAGNTA....191
6.2.2 (a) Analyses for enrichment of T2D risk variants in *KLF14 trans*-network gene set ..191
6.2.2 (b) Pathway analysis of *KLF14 trans*-associated transcripts in gene sets enriched for T2D variants....194
6.3 Results....196
6.3.1 *KLF14* expression profiling in humans 196
6.3.2 *Klf14* expression profiling in mouse....197
6.3.3 Analysis of *KLF14 trans*-associated transcripts in gene sets enriched for T2D-associated variants....200
6.4 Conclusions....203
6.4.1 *KLF14* and *trans*-gene expression in human tissue....205
6.4.2 *Klf14*^{R238L} mouse model re-derivation and phenotyping....205

6.4.3 *Klf14* and *trans* gene expression in metabolic tissues from the *Klf14*^{R238L} mouse model..206
6.4.4 *Klf14* knockout mouse 207
6.4.5 The *KLF14 trans*-network genes are involved in pathways enriched for T2D variants.....208
Chapter 7: General Discussion.....210
7.1 Achievement of Aims.....211
7.2 Selection of transcripts and design of ENU mouse models 212
7.3 Morphometrical examination and quantification of islets from *Adcy5*^{Y1064C} mice 213
7.4 Characterization of the *Adcy5*^{Y1064C} mouse model.....214
7.5 Expression profiling of and gene set analysis of *KLF14* and *trans*-associated genes.....218
7.6 Future directions.....220

Chapter 1:

Introduction

1.1 Overview of efforts to characterize the genetic component of type 2 diabetes

The massive and growing impact of type 2 diabetes (T2D) on global public health has made a priority of efforts to define the pathogenesis of this disease, and explore the mechanistic basis of individual predisposition. The current count of ~350 million people with T2D worldwide is expected to rise to over 500 million by 2030 (World Health Organisation, 2011; International Diabetes Federation, 2011). This rapid rise in overall prevalence is presumed to reflect dramatic shifts in lifestyle and diet, but predisposition is also driven by individual genetic risk. Estimates of the heritability of T2D vary between studies, but figures of 25% (Poulsen, et al., 1999) or more (Florez et al., 2003) are typical.

The search for the DNA sequence variants that underlie the risk of non-autoimmune forms of diabetes (of which T2D is by far the most prevalent) started almost two decades ago, However, linkage studies conducted for more common forms of T2D had little impact, and the same was true (with a few notable exceptions (Altshuler et al., 2000; Gloyn et al., 2003)) for the candidate-gene based association studies that flourished in their place. The advent of genome wide association studies (GWAS) in the mid-2000s, enabled by completion of the HapMap project and developments in genotyping technology, has had a profound effect on discovery efforts, and as of the beginning of 2014, over 90 genetic loci have been shown to be associated with T2D (Mahajan et al., 2014).

While the identification of so many loci offers many potential insights into disease biology, collectively these loci explain only a minor part (around 10%) of the variation in individual predisposition to T2D, and therefore only a minority of that part which is heritable (Morris et al., 2012). In parallel with efforts to explore the biology of the loci already discovered (see below), there is continued interest in uncovering new loci, and in identifying the specific variants that are driving the GWAS loci, in the expectation that this will augment the proportion of variance explained.

This introduction catalogues recent developments in these areas and introduces the two GWAS loci which will be the focus of this thesis. First, I summarize the known pathophysiology of T2D. Second, I review the headway made over the past decade by genetic studies of diabetes. Third, I describe

current progress in expanding the number of T2D susceptibility loci and in defining the variants that may be responsible for them. Fourth, I focus on efforts to define the biology behind these association signals.

1.2 The pathophysiology of T2D

T2D is characterized by the loss of insulin sensitivity, meaning peripheral tissues do not respond to insulin effectively and therefore cannot absorb glucose from the blood stream efficiently, coupled with the failure of beta cells to compensate with sufficient insulin to overcome the resistance. Typically, insulin resistance develops as a result of weight gain, although people of normal body mass index (BMI) can develop the disorder as well (Soumaya, K. 2012). Importantly, most people with insulin resistance do not become type 2 diabetic because their beta cells are able to adequately compensate (Weir and Bonner-Weir, 2013). Genetic as well as acquired defects in the beta or alpha cells lead to dysfunction in insulin and glucagon secretion in response to glucose and other secretagogues. Following the typical course of the disease, serum insulin levels fluctuate and then decrease. Proinsulin processing is impaired leading to higher levels of proinsulin but lower levels of insulin (Porksen et al., 2002). Decreased beta cell mass, possibly due to beta cell apoptosis, and increased alpha cell mass are consistent features of T2D (Butler et al., 2003). Some of the proposed causes of eventual beta cell failure are glucose toxicity, inflammation, endoplasmic reticulum (ER) and oxidative stress and islet amyloid formation (Potter et al., 2014). Genetic variation in mechanisms of beta cell function and insulin signalling may trigger or speed up the course of the disease (Bonnefond et al., 2010). Genetic studies of T2D predisposition are reviewed in the next part of this introduction.

1.3 The evolution of genetic studies of T2D

Initial efforts in the search for DNA sequence variants associated with T2D or other non-autoimmune forms were mostly based around family-based linkage analyses. These studies identified several genes (such as those encoding the beta-cell glucose sensor, glucokinase, and key transcription factors such as hepatocyte nuclear factors 1A and 4A) that contained mutations responsible for monogenic forms of early-onset diabetes (Gragnoli et al., 1997; Furuta et al., 1997; Chèvre et al., 1998). Patients with hepatocyte nuclear factor-1 alpha (*HNF1A)* mutations are now treated with low dose sulphonylureas with considerable success (Pearson et al., 2003). Similarly, patients with mutations in glucokinase (*GCK*) are usually taken off treatment entirely because it is now known that their disease will not progress beyond a relatively benign form (Stride et al., 2002). Substantial progress has been made to identify and treat patients with these rare forms of diabetes from the discovery of alleles in *HNF1A* and *GCK* (Thanabalasingham and Owen, 2011).

While extremely useful in identifying monogenic forms of diabetes, linkage studies proved relatively poor at revealing lower penetrance variants in T2D, and those that were identified as possibly linked could not be replicated or targeted by fine-mapping (Prokopenko et al., 2009). The next studies, candidate gene studies, targeted regions encoding or surrounding genes already presumed to be involved in T2D predisposition. Candidate gene association studies in larger sample sets replaced linkage studies and these revealed two true associations at loci, P12A in peroxisome proliferator-activator receptor gamma (*PPARG*) and E23K in the potassium inwardly-rectifying channel *KCNJ11*; two genes already targeted in T2D treatment (thiazolidinediones for *PPARG* [Altshuler et al., 2000] and sulphonylureas for *KCNJ11* [Gloyn et al., 2003]). Transcription factor 7-like 2 (T-cell specific, HMG box) *TCF7L2* was discovered from association analysis of a region of the genome that had been previous implicated for linkage (Grant et al., 2006). To this day, the variants associated most strongly at this locus have the greatest effect on T2D-susceptibility of any common variant identified.

The candidate gene approach is restricted by reliance upon prior knowledge. To search for true associations when disease pathophysiology is not known, a method for identifying loci was needed. Genome-wide association studies (GWAS) were designed to scan common variation in the genome in large cohorts of people without bias or previous knowledge of biology, and their entrance transformed the search for susceptibility loci for many polygenic diseases. With the advent of the HapMap project and new genotyping technologies, GWAS became possible (The International HapMap Consortium, 2003). The HapMap project made use of public databases of millions of SNPs to characterize linkage disequilibrium (LD) patterns and identify combinations of alleles that associate with each other. With knowledge of the patterns identified by the HapMap project, geneticists did not need to genotype every single nucleotide polymorphism (SNP) in the genome to capture the majority of the genetic information (Eberle et al., 2007). To find common variants in the population, genotyping a set of selected SNPs to serve as markers (called tagging SNPs) is nearly as effective as genotyping every SNP with only a slight loss of statistical power (Eberle et al., 2007). With modern genotyping technologies and decreasing costs, it became possible for many research institutes to genotype half a million SNPs on thousands of DNA samples (Matsuzaki et al., 2004). Fixed content as well as custom made genotyping arrays were made widely available for scanning and fine-mapping of identified regions (McVean et al., 2005). GWAS quickly became the most powerful and efficient study tool for complex diseases.

For T2D in 2007, the known loci at *PPARG, KCNJ11*, and *TCF7L2* were confirmed by GWAS and six more added (Saxena et al., 2007; Scott et al., 2007; Sladek et al., 2007; Steinthorsdottir et al., 2007; Zeggini et al., 2007). The sixth was a locus near fat mass and obesity-associated *FTO* (Frayling et al., 2007) that was associated with T2D only in relation to adiposity, making it the first common variant signal for BMI and risk of obesity.

Chromosome	Locus	IndexSNP	RAF	OR	Population	Reference
3	*PPARG*	rs1801282	0.92	1.14	European	Altshuler(2000)
11	*KCNJ11*	rs5219	0.50	1.14	European	Gloyn(2003)
10	*TCF7L2*	rs7903146	0.25	1.37	European	Grant(2006)
10	*HHEX-IDE*	rs1111875	0.56	1.15	European	Sladek(2007)
17	*HNF1B*	rs757210	0.43	1.10	European	Winkler(2007)
16	*FTO*	rs8050136	0.45	1.17	European	Frayling(2007)
3	*IGF2BP2*	rs4402960	0.29	1.14	European	Zeggini(2007)
4	*WFS1*	rs1001013	0.32	1.12	European	Sandhu(2007)
6	*CDKAL1*	rs7754840	0.31	1.14	European	Zeggini(2007)
8	*SLC30A8*	rs13266634	0.75	1.15	European	Zeggini(2007)
9	*CDKN2A/B*	rs10811661	0.79	1.20	European	Sladek(2007)
1	*NOTCH2*	rs10923931	0.11	1.13	European	Zeggini(2007)
2	*THADA*	rs7578597	0.92	1.15	European	Zeggini(2007)
3	*ADAMT59*	rs4607103	0.81	1.09	European	Zeggini(2007)
7	*JAZF1*	rs864745	0.52	1.10	European	Zeggini(2007)
10	*CDC123*	rs12779790	0.23	1.11	European	Zeggini(2007)
12	*TSPAN8*	rs7961581	0.23	1.09	European	Zeggini(2007)
11	*KCNQ1*	rs2237892	0.61	1.40	E. Asian	Yasuda/Unoki(2008)
11	*MTNR1B*	rs10830963	0.30	1.09	European	Prokopenko(2009)
2	*IRS1*	rs2943641	0.61	1.19	European	Rung(2009)
11	*DUSP8/HCCA2*	rs2334499	0.45	1.08	European	Kong(2009)
19	*GIPR*	rs10423928	0.22	1.04	European	Saxena(2010)
2	*GCKR*	rs780094	0.62	1.06	European	Dupuis(2010)
3	*ADCY5*	rs11708067	0.78	1.12	European	Dupuis(2010)
7	*DGKB*	rs2191349	0.47	1.06	European	Dupuis(2010)
7	*GCK*	rs4607517	0.20	1.07	European	Dupuis(2010)
9	*PTPRD*	rs17584499	0.11	1.57	E. Asian	Tsai(2010)
17	*SRR*	rs391300	0.76	1.28	E. Asian	Tsai(2010)
1	*PROX1*	rs340874	0.50	1.07	European	Voight(2010)
2	*BCL11A*	rs243021	0.46	1.08	European	Voight(2010)
5	*ZBED3*	rs4457053	0.26	1.08	European	Voight(2010)
7	*KLF14*	rs972283	0.55	1.07	European	Voight(2010)
8	*TP53INP1*	rs896854	0.48	1.06	European	Voight(2010)
9	*CHCHD9*	rs13292136	0.93	1.11	European	Voight(2010)
11	*CENTD2*	rs1552224	0.88	1.14	European	Voight(2010)
11	*KCNQ1*	rs231362	0.52	1.08	European	Voight(2010)
12	*HMGA2*	rs1531343	0.10	1.10	European	Voight(2010)
12	*HNF1A*	rs7957197	0.85	1.07	European	Voight(2010)
15	*PRC1*	rs8042680	0.22	1.07	European	Voight(2010)
15	*ZFAND6*	rs11634397	0.56	1.06	European	Voight(2010)
X	*DUSP9*	rs5945326	0.12	1.27	European	Voight(2010)
2	*RBMS1*	rs7593730	0.83	1.10	European	Qi(2010)
3	*UBE2E2*	rs6780569	0.85	1.18	E. Asian	Yamauchi(2010)
15	*C2CD4A/B*	rs7172432	0.58	1.13	E. Asian	Yamauchi(2010)

2	*GRB14*	rs3923113	0.74	1.06	S. Asian	Kooner(2011)
3	*ST6GAL1*	rs16861329	0.75	1.06	S. Asian	Kooner(2011)
10	*VPS26A*	rs1802295	0.26	1.05	S. Asian	Kooner(2011)
15	*AP3S2*	rs2028299	0.31	1.07	S. Asian	Kooner(2011)
15	*HMG20A*	rs7178572	0.52	1.06	S. Asian	Kooner(2011)
20	*HNF4A*	rs4812829	0.29	1.06	S. Asian	Kooner(2011)
3	*PSMD6*	rs831571	0.68	1.09	E. Asian	Cho(2011)
4	*MAEA*	rs6815464	0.58	1.13	E. Asian	Cho(2011)
6	*KCNK16*	rs1535500	0.42	1.08	E. Asian	Cho(2011)
6	*ZFAND3*	rs9470794	0.32	1.12	E. Asian	Cho(2011)
9	*GLIS3*	rs7041847	0.51	1.10	E. Asian	Cho(2011)
19	*PEPD*	rs3786897	0.56	1.10	E. Asian	Cho(2011)
20	*R3HDML-HNF4A*	rs6017317	0.47	1.09	E. Asian	Cho(2011)
20	*GCC1-HNF4A*	rs6467136	0.79	1.11	E. Asian	Cho(2011)
13	*SPRY2*	rs1359790	0.74	1.14	Japanese	Imamura(2011)
18	*BCL2*	rs12454712	0.59*	1.09	Multi-ethnic	Saxena(2012)
19	*GATAD2A*	rs19471596	0.08	1.12	European	Saxena(2012)
2	*RBM43-RND3*	rs7560163	0.14	0.75	African American	Palmer(2012)
10	*TCERG1L*	rs7077836	0.12	1.14	African American	Chen(2012)
10	*GRK5*	rs10886471	0.80	1.12	Chinese Hans	Li(2012)
15	*RASGRP1*	rs7403531	0.37	1.10	Chinese Hans	Li(2012)
5	*ANKRD55*	rs459193	0.70	1.08	European	Morris(2012)
8	*ANK1*	rs516946	0.76	1.09	European	Morris(2012)
9	*TLE1*	rs2796441	0.57	1.07	European	Morris(2012)
10	*ZMIZ1*	rs12571751	0.52	1.08	European	Morris(2012)
12	*KLHDC5*	rs10842994	0.80	1.10	European	Morris(2012)
16	*BCAR1*	rs7202877	0.89	1.12	European	Morris(2012)
18	*MC4R*	rs12970134	0.27	1.08	European	Morris(2012)
19	*CILP2*	rs10401969	0.08	1.13	European	Morris(2012)
7	*PAX4*	rs10229583	0.83	1.18	Multi-ethnic	Ma(2013)
9	*GPSM1*	rs11787792	0.87	1.15	E. Asian	Hara(2014)
11	*MIR129-LEP*	rs791595	0.80	1.17	E. Asian	Hara(2014)
17	*SLC16A13*	rs312457	0.08	1.20	E. Asian	Hara(2014)
2	*DNER*	rs1861612	0.64	1.29	American Indians	Hanson(2014)
17	*SLC16A11*	rs13342232	0.16	1.29	Latin Americans	SIGMA(2014)
1	*FAF1*	rs17106184	0.91	1.10	Multi-ethnic	Mahajan(2014)
3	*LPP*	rs6808574	0.60	1.07	Multi-ethnic	Mahajan(2014)
4	*TMEM154*	rs6813195	0.71	1.08	Multi-ethnic	Mahajan(2014)
5	*ARL15*	rs702634	0.67	1.06	Multi-ethnic	Mahajan(2014)
6	*POU5F1-TCF19*	rs3130501	0.73	1.07	Multi-ethnic	Mahajan(2014)
6	*SSR1-RREB1*	rs9505118	0.73	1.06	Multi-ethnic	Mahajan(2014)
12	*MPHOSPH9*	rs4275659	0.78	1.06	Multi-ethnic	Mahajan(2014)
6	*HLA-B*	rs2244020	0.69	1.11	African Americans	Ng(2014)

11	*INS-IGF2*	rs3842770	0.23	1.14	African Americans	Ng(2014)

Table 1.1: Established common variant T2D susceptibility loci arranged by order of discovery (then chromosome). Gene name assigned from closest or best regional candidate gene and labelled SNP represents strongest association; neither are necessarily causal. RAF = risk allele frequency, or the frequency of the allele which increases T2D risk in the population based on HapMap information. OR = odds ratio, or the per-allele increase in T2D risk. Population is the population group in which the allele was discovered, but most signals have been replicated across multiple populations.

As the sample size of GWAS increased by combining data from European populations in consortia such as the Diabetes Genetics Replication and Meta-analysis (DIAGRAM) consortium (Voight et al., 2010; Zeggini et al., 2008; Morris et al., 2012; Mahajan et al., 2014), the count of confirmed-associated loci reaching genome-wide significance expanded to close to ninety (table 1.1) and the improvement in power exposed variants of smaller effect size and more extreme risk allele frequency.

GWAS in South and East Asian, African American, and American Indians populations (Cho et al., 2011; Kooner et al, 2011; Unoki et al., 2008; Yamauchi et al., 2010; Yasuda et al., 2008; Hanson et al., 2014; Ng et al., 2014) revealed more associated loci, most of which have been replicated in European populations. Despite significant progress in identifying susceptibility loci, these 89 confirmed loci only explain a fraction of the estimated heritability of T2D. Geneticists are continuing to probe new levels of meta-analysis, expand sample sizes, and target different populations to uncover more of the missing heritability and discover new mechanisms of the pathophysiology of T2D.

1.4 Recent advances in identifying the genetic basis of T2D

While some question the rationale for ever larger GWAS studies that seem only to collect growing lists of variants of ever smaller effects (McClellan et al., 2011), expansion of the GWAS approach to access greater sample size and a broader ethnic range has a number of motivations. These include: (a) greater understanding of the overall impact of common variant signals to T2D pathogenesis; (b) increased power to detect associations to low frequency variants, some of which may have larger effects than the common variants detected by "traditional" GWAS; (c) recognition that the modest effect sizes of common risk-alleles have little to do with the potential for the products of the implicated genes to be valuable therapeutic targets (as common T2D-risk variants in *KCNJ11* and *PPARG* clearly illustrate); (d) the expectation that an expanded list of associated loci will make it

possible for analyses to be done that look for evidence that these loci are enriched for genes influencing particular networks or pathways; (e) the need to establish the extent to which the variants and loci involved in predisposition are shared between major ancestral and continental populations; and (f) attempts to identify the variants that are causal for genome wide association signals through fine-mapping.

Several of these features emerge from the T2D GWA meta-analysis, meaning analysis combining several independent studies, performed by the DIAGRAM Consortium in 2012 (Morris et al., 2012). The DIAGRAM Consortium is a group of researchers that have combined 11 datasets of T2D cases and controls of European descent to be able to perform large-scale GWA meta-analyses. Included in the dataset are tissue samples, genetic information, and phenotypic information such as BMI, fasting glucose, etc., for thousands of individuals. An illustration that explains DIAGRAM in more detail is presented as Figure 1.1. Using genotype data from almost 150,000 individuals, the consortium was able to identify 10 novel T2D-susceptibility loci in a European population reaching genome-wide significance, and to highlight several hundreds more that, while failing to reach the stringent criteria typically regarded as proof, are nonetheless highly likely to reflect genuine associations. By combining variation at both the “proven” and “probable” loci, the authors were able to explain ~10.7% of overall variance in risk of T2D (and therefore as much as 40% of the inherited component). Using approaches that take advantage of the complete set of GWAS results, these data revealed that a substantial component (as much as half) of the variation in individual risk of diabetes can be defined by common sequence variants. The expanded list of loci also provided a powerful tool for revealing networks and pathways implicated in T2D pathogenesis.

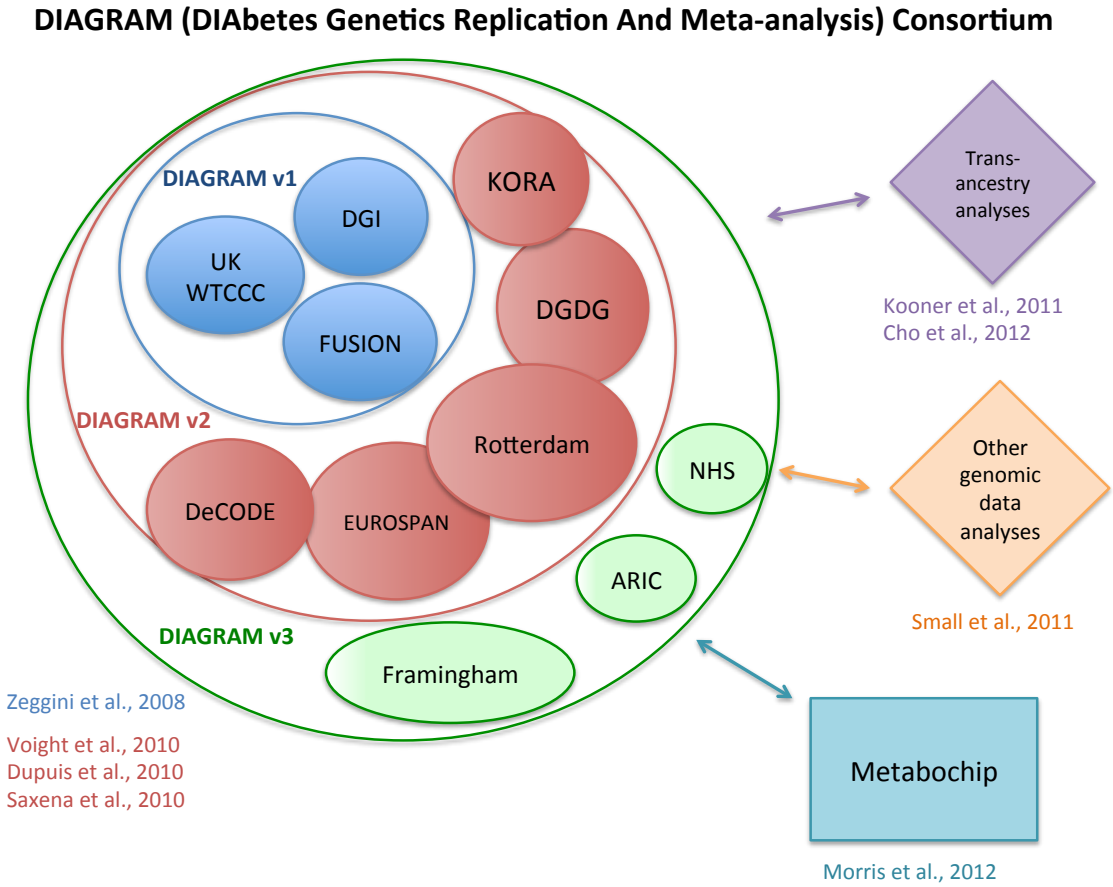


Figure 1.1: The DIAGRAM Consortium started as DIAGRAM v1 in 2008 combining the UK Welcome Trust Case Control Cohort (WTCCC), FUSION dataset, and Diabetes Genetics Initiative (DGI). Subsequently it expanded to include several more studies and become DIAGRAM v2 in 2010, and finally the dataset reached its current size as DIAGRAM v3 in 2012. Consisting of 12,171 cases and 56,862 controls, DIAGRAM v3 is the largest current GWA dataset in samples of European descent. Publications arising from each stage of DIAGRAM are listed, all published in the journal of Nature Genetics. The DIAGRAM v3 dataset was used for selection of SNPs to include in the Metabochip custom array. Likewise, other data analyses have used the DIAGRAM dataset for integration with other genomic data and variant discovery in populations of other ancestries.

The past few years has seen considerable expansion of efforts to identify common variants influencing T2D-risk in other populations. For example, an analysis done by Kooner and colleagues in over 5,500 T2D cases and 14,400 controls from the UK, Singapore and Pakistan, all with origins in the Indian subcontinent, identified six novel association signals including variants near the genes encoding the growth factor receptor-bound protein (*GRB14*) and hepatocyte nuclear factor-4A (*HNF4A*) the latter already implicated in monogenic forms of diabetes (Kooner et al., 2011). Similar efforts in East Asian subjects have added a further sixteen loci to the global tally (Yasuda/Unoki, 2008; Tsai et al., 2010; Yamauchi et al., 2010; Cho et al., 2012; Hara et al., 2014). Studies in individuals of African descent are of particular interest given the population's extensive genetic diversity, and during 2012 the first large-scale association studies from African-American subjects emerged (Palmer et al., 2012; Saxena et al., 2012).

These studies highlighted some of the challenges associated with genetic studies in African-descent populations (such as limited linkage disequilibrium and genetic mixture) but did reveal a number of novel genome-wide significant signals, including those mapping near the rho family GTPase 3 gene (*RND3)* and the B-cell lymphoma 2 gene (*BCL2).* A recent publication from a collaboration of several T2D consortia provided a follow-up meta-analysis in over 100,000 people from four ancestries including East Asian, South Asian, European, Mexican and Mexican American and identified seven new T2D susceptibility loci (Mahajan al., 2014).

The detection of so many novel signals in non-European samples appears to indicate a significant amount of ethnic specificity in the genetic basis of T2D. However, most of the top association signals detected in one ethnic group have been visible (not always at genome-wide significant levels) in others, and the chance that a given signal will be detected for the first time in a particular population is due mainly to random variation in effect size. In an extensive study of trans ethnic associations, Haiman and colleagues (Haiman et al., 2012) tested 20 of the known T2D risk variants across six major racial and ethnic groups from the U.S. (including African Americans, Hispanics, American

Indians, and Native Hawaiians) and found marked overlap in the patterns of association, with almost complete directional consistency of effects, meaning that a variant positively associated with risk of T2D in one population was almost always positively associated with risk of T2D in the others as well. Mahajan et al. (2014) also found very high directional consistency of T2D risk alleles from four ancestry groups. These data provide compelling evidence that common variant associations for T2D are shared between global populations, and that these associations likely reflect the action of causal variants that are common and widely distributed. (The logic runs as follows: if the causal alleles that underlie the common variant signals detected by GWAS were rare, then they would most likely have arisen after the human migration out of Africa, and would not therefore result in broadly consistent patterns of association across diverse ethnic groups).

Efforts to identify the specific causal alleles through fine-mapping approaches have also gained momentum, building on the more complete surveys of regional sequence variation accessible through targeted custom arrays, such as the Metabochip. The Metabochip is an Illumina BeadChip designed to genotype nearly 200,000 SNPs chosen based on GWAS meta-analyses of 23 metabolic traits (Voight et al., 2012). The technology is that thousands of tiny silica beads trapped in holes on the chip are coated with oligonucleotide probes corresponding to the sequences leading up to a SNP of interest. Each bead is coated with thousands of an identical probe. When genomic DNA is added to the chip, it binds the probes, and polymerase adds subsequent bases that have been labelled with a fluorescing tag. The fluorescence emitted corresponds to the allelic ratio of the DNA bound to the bead. The Metabochip was custom designed to genotype replication SNPS, for verification of the top independent association signals, and fine-mapping SNPs to find causal alleles at known loci (Voight et al., 2012). Another method researchers are using to identify causal SNPs is through imputation from sequence-based reference panels, such as those generated by the 1000 Genomes project (a project sequence the genomes of a large cohort of people to provide a resource on human genetic variation [The 1000 Genomes Project Consortium, 2010]). Buyske and colleagues set out to test

whether trans ethnic studies in African Americans would be able to distinguish causal variants shared between ethnicities from their partially-correlated neighbours (Buyske et al., 2012). They were able to show, at the *CELSR2/PSRC1/SORT1* locus known to influence lipid levels, that rs12740374 stood out as the strongest associated SNP for LDL-cholesterol in African-descent subjects, while, in European ancestry samples, this SNP was amongst a group of highly-correlated variants which could not be separated by association analyses. The availability of large multiethnic GWAS is now boosting efforts to fine-map the variants driving the T2D risk loci uncovered by GWAS (Mahajan et al., 2014).

GWAS methodology has, to date, focused mainly on the contribution of common variants to disease susceptibility. However, deep sequencing studies have revealed that the rapid, recent, population growth of mankind has enriched human sequence variation for rare, even private, mutations (Keinan et al., 2012; Nelson et al., 2012). This has generated additional support for the view that low frequency and rare variants are likely to play an important role in defining individual predisposition to common disease (Lupski et al., 2011).

Rapid improvements in the throughput and accuracy of next-generation sequencing now allow for systematic association analysis of variants of all frequencies and all types, and several large-scale efforts are underway to identify and characterise the contribution of low frequency and rare variants in T2D. Steinthorsdottir et al. (2014) identified low frequency and rare variants associated with T2D using whole genome sequencing in a cohort of over 13,000 people. This effort in an Icelandic population yielded four previously unreported low frequency T2D risk variants including one variant, rs76895963[G] in intron 1 of *CCND2*, which reduces risk of T2D by half. Another successful method of achieving high throughput to reveal low frequency variants has been using exome array genotyping. Exome genotyping selectively targets variants within coding regions and is a cheaper alternative to whole genome or whole exome sequencing, though it is only capable of capturing a subset of exomic variation. One study successfully employed this method to identify low frequency

variants associated with fasting proinsulin in known T2D-associated loci as well as several low-frequency variants associated with fasting proinsulin and insulinogenic index in three new genes: *TBC1D30*, *KANK1*, and *PAM* (Huyghe et al., 2013). Before 2013, published research describing low frequency variants was restricted to single gene studies, but these studies served to expose some of the challenges of generating and interpreting rare variant associations.

Bonnefond and colleagues (Bonnefond et al., 2012), for example, focused on resequencing of the *MTNR1B* gene, encoding one of the receptors for the hormone melatonin. Several GWAS have shown that common variants near *MTNR1B* influence fasting glucose levels in non-diabetic individuals, as well as having an impact on T2D risk (Lyssenko et al., 2009; Bouatia-Naji et al., 2009) and there is compelling biological evidence that these are likely to be acting through altered expression of the receptor (Lyssenko et al., 2009). Deep resequencing of *MTNR1B* coding sequence, in more than 7,500 European subjects, revealed a total of 40 nonsynonymous variants, thirty-six of them with minor allele frequency below 0.1%. When considered collectively, these rare variants were seen to be associated with T2D, with estimated mean allelic odds ratios of ~3, far exceeding the equivalent figure for the common variants identified by GWAS (~1.10). To determine the impact of each of these variant on receptor function, the authors functionally characterised each one, distinguishing those with molecular effects from those with none. Reanalysis of the genetic data demonstrated that the T2D association signal was restricted to individuals carrying those functionally defective alleles. These findings confirm *MTNR1B* as the gene mediating the GWAS signal, but also illustrate that prediction of the functional and physiological impact of rare novel alleles – such as those which might be revealed by medical resequencing – is difficult in the absence of direct empirical examination. Rees and colleagues (Rees et al., 2012) reached similar conclusions in their analysis of the effects of rare variants in the *GCKR* gene (encoding the glucokinase regulatory protein) on lipid profiles. Studies such as these have important implications for both the discovery of

rare variant signals, and the clinical use of sequence data for diagnostic and prognostic purposes in common disease.

1.5 The next step: elucidating the biological mechanisms driving association

Once a robust association signal has been identified, the challenge shifts to converting these discoveries into meaningful biological insights. For GWAS signals, this has generally proved a greater challenge than might have been foreseen. There are a number of reasons for this: (a) most of the signals map to non-coding areas of the genome, such that the transcript mediating the signal is typically not obvious; (b) extensive linkage disequilibrium means that the variant(s) responsible for each signal is poorly-specified; and (c) individual effect sizes are modest. Sequence-based studies are expected to start to address these issues by enabling more systematic fine-mapping of the causal alleles and by defining additional risk-alleles at these loci, some of them of lower frequency and larger effect, and some of them mapping within coding sequence (Nejentsev et al., 2009). In addition, the improving annotation of regulatory elements, through the ENCODE project and related efforts in disease relevant tissues (Kellis et al., 2014; Gaulton et al., 2010; Pasquali et al., 2014), will accelerate the interpretation of signals in the non-coding genome.

In the meantime, investigators have been making steady progress using these genetic data to drive biological inference. Broadly speaking, these efforts have followed one of two approaches. In the first, the emphasis has been on exploiting the collective properties of the variants and regions discovered (in the absence of precise information regarding the causal alleles and transcripts) to define the broader phenotypic and genomic attributes of the common variant contribution to T2D. In the second, the focus has been on characterising the mechanisms operating at individual loci.

The studies from the DIAGRAM Consortium provide examples of the former. DIAGRAM investigators (Morris et al., 2012) characterised the common variant architecture of T2D through a case-control

study of ~150,000 European subjects, identifying hundreds of loci that were highly likely (but not absolutely certain) to house further T2D-risk variants. They used this enlarged set of loci to perform a series of "enrichment" studies, asking the following question: is there anything about the transcripts that map near to T2D GWAS signals that is different (from the perspective of functional annotation, protein-protein interactions and the like) from transcripts that do not? These analyses revealed that these T2D loci were collectively enriched for genes involved in: (a) cell cycle regulation (speaking to the importance of beta-cell mass as a determinant of insulin secretion); (b) adipocytokine signalling (reflecting tight links between adipose tissue and glucose metabolism); and (c) CREBBP-interacting transcription factor activity.

Using the same data, the DIAGRAM investigators were also able to extend previous analyses which derive biological insights from the association effects of T2D-risk variants on related traits, such as body mass index, fasting glucose (in non-diabetic individuals) and indices of beta-cell function and insulin action (Morris et al., 2012; Voight et al, 2010; Scott et al., 2012). They were able to confirm that: (a) partial, but not complete, overlap between variants that influence individual risk of T2D, and those that modulate physiological variation in fasting glucose amongst healthy individuals; (b) the only signals which are driven by a primary effect on obesity are those at *FTO* and *MC4R;* and (c) while most risk loci operate via beta-cell dysfunction, a growing number (see table 1.2) exert their T2D-risk effects through an obesity-independent deterioration in insulin sensitivity. This list of "insulin resistance" loci offers interesting insights into key players mediating the actions of insulin in peripheral tissues. In the case of the *GRB14* locus for example, which emerged from GWAS in South Asians as well as Europeans (Kooner et al., 2011), RNA expression data from fat confirms *GRB14* as the strongest transcript candidate: it encodes an adaptor protein that binds to the insulin receptor and facilitates binding of protein complex to inhibit tyrosine kinase signalling (Holt et al., 2009).

The second type of approach to using genetic data to inform biology has been to describe the mechanism operating at individual loci. Investigators have implemented a wide range of approaches,

including association fine-mapping and resequencing in man, and functional examination of variants and candidate transcripts in cellular and animal models, to define the mechanisms whereby the T2D-risk variants at each locus exert their effects on T2D predisposition. These studies aim both to identify specific alleles that drive association (which can deliver insights into the biology upstream of the locus), and the transcripts through which they act (which feeds into downstream mechanisms). This concept is illustrated in Figure 1.2. In this thesis I present data for a T2D locus that illustrates progress in these endeavours.

One of the first novel T2D-risk signals to emerge from GWAS mapped to the region around *CDKAL1* (Diab et al., 2007; Scott et al., 2007; Steinthorsdottir et al., 2007). It was clear that this effect was mediated via a defect in beta-cell function, and that the risk variants mapped within an intron of *CDKAL1,* a gene of unknown function, its official name (CDK5 regulatory subunit associated protein 1-like 1) having been assigned on the basis of sequence homology alone. However, it was not clear how *CDKAL1* might be mediating T2D risk, nor, indeed whether *CDKAL1* was actually the culprit.

Recent work has established that the beta-cell specific *Cdkal1* knockout mouse displays decreased insulin secretion and impaired glucose tolerance, strongly supporting the candidacy of *CDKAL1* in man (Wei et al., 2011; Wei and Tomizawa, 2011; Ohara-Imaizumi et al., 2010). At the same time, it has emerged that CDKAL1 is not a modulator of cyclin dependent kinase activity, as homology analyses had intimated: instead, it acts as a methylthiotransferase with a crucial role in stabilising the codon-anticodon interaction involving the transfer RNA for lysine in particular, and is consequently involved in the correct incorporation of that amino acid into protein (Wei et al., 2011; Arragain et al., 2010). One possible mechanism for the effect on T2D risk might be that lack of *Cdkal1* in the beta cell causes mistranslation in proinsulin synthesis resulting in defective insulin processing and endoplasmic reticulum stress (Kaufman et al., 2011).

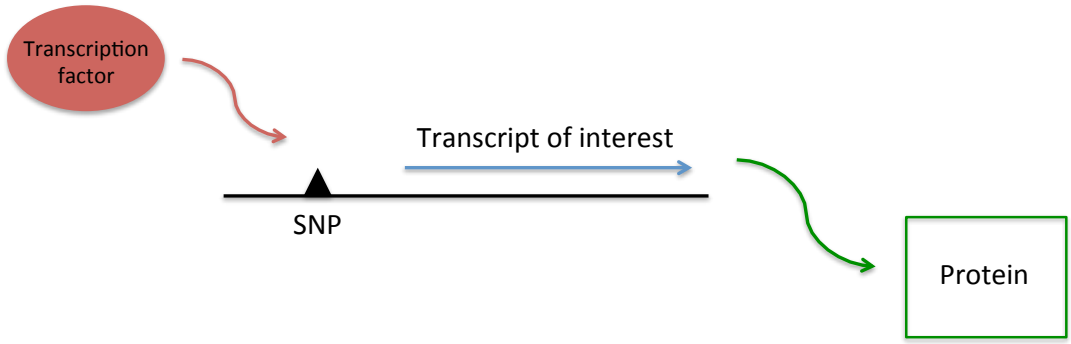


Figure 1.2: An illustration of the aim of using genetic variation data to elucidate biological mechanisms. A SNP implicated by GWAS may reveal a regulatory site, which would allow investigation into the transcriptional regulation of the gene. Or, determining the transcript over which the variant is exerting its effect will allow investigation into the downstream mechanisms such as the protein signaling pathways affected by the transcript.

Gene locus	Associated phenotypes	eQTL data	Phenotypes Influenced
Obesity-independent			
PPARG	T2D[2], BMI[6],CAD[19]	none	adipogenesis, lipid uptake in adipocytes[5]
IRS1	T2D[18], body fat %[11], CAD[47], blood lipids[22], hyperinsulinemia[22]	Cis: adipose[22]	fat distribution[22]; adipocyte biology[22,13]
GCKR	T2D[7], fasting glucose[15], fasting insulin[15], triglycerides[22]	none	glucose sensing, signal transduction[17]
KLF14	T2D[23], HDL[22]	Cis: adipose[23]; Trans: >40 transcripts[21]	adipocyte growth and signaling[21]
ADAMTS9	T2D[10], waist-hip ratio[10]	Cis: omental adipose[24]	unknown
GRB14	T2D[14], waist-hip ratio (in females)[10], HDL[22]	Cis: omental adipose[14]	insulin receptor signaling, regulates insulin action[16]
Primary effect on obesity			
FTO	T2D [20], BMI[8], obesity[8], body fat %[11], waist circum[20], CAD[19]	none	food intake[9]; lypolytic activity in adipose[12]
MC4R	T2D[14], waist circum[4], obesity, weight, BMI, body fat %[4], CAD[19]	none	food intake[1]

Table 1.2 Examples of T2D loci considered to be acting primarily through defects in insulin action. The obesity-independent genes (at the top) are associated with T2D even when obesity is not taken into account. The lower two variants only reach genome-wide significant association with T2D when non-obese individuals are excluded, and therefore these variants exert their primary effect on obesity. References numbered can be found in Appendix after full list of references.

1.6 Loci to be examined in this thesis

In this thesis, I will examine the insulin-resistance associated *KLF14* locus along with the fasting glucose associated locus *ADCY5*. I selected *KLF14* to be a locus of particular interest due to the evidence for its crucial role in adipose regulation and the resulting breakdown in insulin action when *KLF14* is disrupted. Variants near *ADCY5* conversely are associated with fasting glucose and likely to be mediating their effect through islet function. Below I describe the logic behind selection of these loci and more detailed information is given in the introduction sections of chapters 4, 5, and 6.

1.7 *KLF14*: a T2D susceptibility locus working through defective insulin action

Variants near *KLF14* are associated with T2D and a measure of insulin resistance, HOMA-IR (Voight et al., 2008). The T2D-associated alleles are also associated with reduced *KLF14* expression in subcutaneous adipose tissue (Kong et al., 2009; Voight et al., 2010; Small et al., 2011). KLF14 is a maternally-imprinted transcript meaning it is subject to epigenetic effects that limit expression to the maternally-inherited copy of the gene (Parker-Katiree et al., 2007). Furthermore, the T2D-associated signals and expression data demonstrate the same imprinted pattern, with effects restricted to the maternal allele (Kong et al., 2009; Small et al., 2011). These data strongly implicate *KLF14*, which encodes the Krueppel-like factor 14 transcription factor, as the causal transcript at this locus. More recently, Small and colleagues from the MuTHER consortium, again using expression data from subcutaneous fat, were able to show that the same set of SNPs upstream of *KLF14* influencing T2D-risk and modifying the expression of *KLF14* itself, were also driving expression of a suite of several dozen transcripts elsewhere in the genome (that is, in *trans*). The most obvious inference that KLF14 acts as a master regulator of adipocyte expression is supported by the known biology of KLFs, and by the observation that the promoter regions of the *trans* genes are enriched

for KLF binding sites (Small et al., 2011). On the basis of these data, the authors hypothesised that the *KLF14*-associated T2D-risk effect was mediated through the effects of one or more of these *trans*-genes (see Figure 1.3) and found support for this argument in the evidence that several of the *trans*-genes, including *SLC7A10*, themselves harboured variants that were independently associated with a range of insulin-resistant traits (Small et al., 2011). These data suggest that *KLF14* and its set of downstream effectors are intimately involved in the generation of the signals whereby the abnormal function of subcutaneous adipose tissue is communicated to other organs (notably muscle and liver) which are the major sites of insulin action. Further research, involving cellular and animal models, as well as physiological, genomic and biomarker studies in man, is required to establish what these signals might be.

KLF14 is one of 17 Krueppel-like transcription factors. KLFs are widely expressed and involved in multiple processes including haematopoiesis, organ development, and cell growth including adipogenesis (Nuez et al., 1995; Wani et al., 1999; Black et al., 2001; Banerjee et al., 2003). Several of the *KLFs* involved in adipogenesis are known to regulate *PPARγ*, a gene robustly associated with T2D susceptibility and working through insulin action (Altshuler et al., 2000; Banerjee et al., 2003; Wu et al., 2005; Pei et al., 2010). A role for *KLF14* in adipocyte biology has not previously been identified. The only documented target of *KLF14* prior to the study by Small et al. (2011) was the transforming growth factor-β receptor II (Truty et al., 2008). *TGFβ* is a well-characterized growth factor involved in cell proliferation and differentiation suggesting a role for KLF14 in these processes (Forrester and Chytil, 2005). An interesting biological story is taking shape for KLF14 in adipocyte biology and mechanisms of acquiring insulin resistance. Some data collected from further studies of *KLF14* will be presented in this thesis.

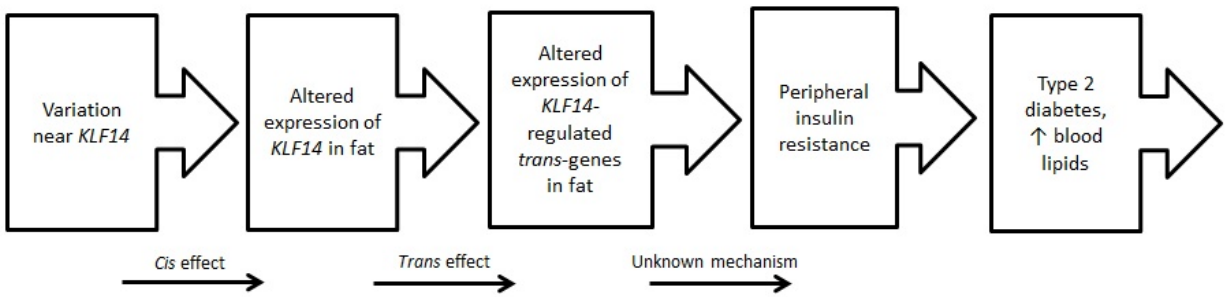


Figure 1.3: Model for KLF14 effects. SNPs upstream of KLF14 influence KLF14 expression in adipose tissue. Since KLF14 is a master regulator of dozens of transcripts in adipose tissue, these same SNPs influence expression of the trans-regulated genes. In turn, altered expression of certain of these transcripts promotes insulin resistance in other key sites of insulin action (through mechanisms yet to be defined) and thereby increases risk of T2D and hyperlipidemia.

1.8 Variation in *ADCY5* has multiple effects on metabolism and beta cell function

Also presented in this thesis are results from research into the metabolic effects of variation at the *ADCY5* locus. Variants in intron 3 within the adenylate cyclase 5 (*ADCY5)* gene have multiple associations with human phenotypes including increased fasting glucose levels, low birth weight, impaired beta cell function, and increased T2D risk (Dupuis et al, 2010; Freathy et al., 2009). These multiple associations suggest an interesting and multi-faceted effect on physiology. Of the genes in the associated region, *ADCY5* is a promising biological candidate the causal transcript because it is implicated in two pathways leading to insulin production and secretion in the beta cell and glucagon production and secretion in the alpha cell: the protein kinase A (PKA) signaling pathway and the exchange protein directly activated by cAMP (Epac) signaling pathway (see figure 1.2). Known islet biology as well as the intronic location of the T2D-associated SNPs implicate *ADCY5* as the likely causal transcript at this locus.

ADCY5 codes for a membrane-bound adenylate cyclase responsible for catalyzing the generation of cAMP. There are 9 types of adenylate cyclases in the body. ADCY5 is calcium-inhibitable and thought to localize to lipid rafts within the plasma membrane (Cooper and Crossthwaite, 2006). Incretins like GLP-1, catecholamines such as adrenaline, and other hormones such as somatostatin and forskolin bind adenylate cyclases and activate the catalyzation of cAMP (Mons et al., 1998). In the cell, cAMP activates PKA which potentiates secretion through several pathways including influx of calcium, release of intracellular calcium from the endoplasmic reticulum (ER), and by acting directly on insulin or glucagon stores (see figure 1.2). There is also evidence that the PKA pathway leads to upregulation of proinsulin formation in the beta cell and proglucagon formation in the alpha cell (Wang et al., 2001; Tengholm, 2012). The Epac pathway is triggered by cAMP generated by adenylate cyclases. Epac acts on the sulfonulyrea receptor (SUR1) on the KATP channel triggering an

influx of calcium leading to depolarization of the cell and exocytosis of insulin or glucagon from granules fusing with the cell membrane (figure 1.4)(Renström et al., 1997; Eliasson et al., 2003).

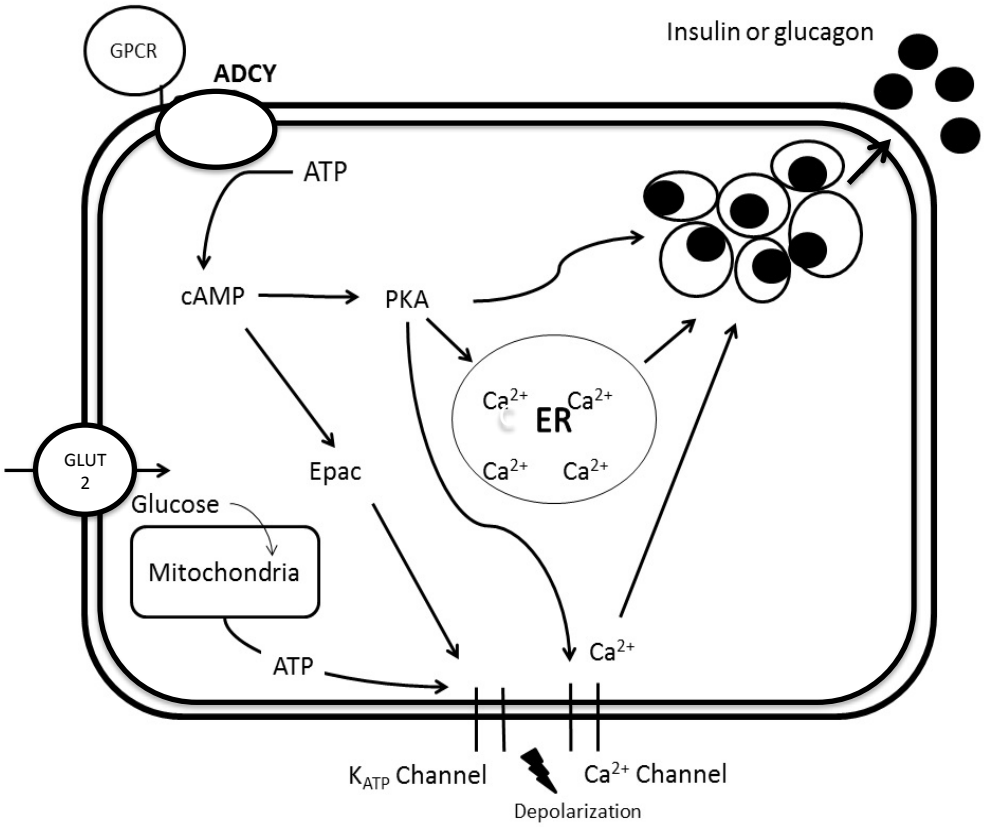


Figure 1.4 Diagram of ADCY5 activation of the PKA and Epac pathways leading to insulin secretion from the beta cell and glucagon secretion from the alpha cell.

Glucagon-like peptide 1 (GLP-1) is an incretin hormone that stimulates insulin secretion after oral glucose ingestion. GLP-1 has been shown to activate both the PKA and Epac pathways through adenylate cyclases in the beta cell to potentiate insulin secretion (Holz, 2004). A closely related adenylate cyclase, *ADCY8*, was shown1` to be down regulated in the beta cell by chronically elevated glucose levels (Roger et al., 2011). The results of down-regulating *ADCY8* were lower levels of cAMP and impaired insulin secretion from the beta cell (Roger et al., 2011).

1.9 Thesis Aims

The studies highlighted in this introduction demonstrate that the application of increasingly powerful genetic and genomic approaches is resulting in the discovery and characterisation of genetic variation that influences individual predisposition to T2D. Furthermore, functional studies arising from implicated variants are, though indisputably challenging, yielding useful insights into disease biology. These studies are motivated by a desire to improve understanding of the pathogenesis of this disease and to exploit the particular value of human genetics to provide causal inference.

Work presented in this thesis uses a variety of approaches in mice and in human cells to elucidate the mechanisms of metabolic dysregulation resulting from variation at two loci: *ADCY5* and *KLF14*. The overall aims are as follows:

- To identify point mutations likely to affect gene function in T2D risk loci *Jazf1*, *Adcy5,* and *Klf14* using the ENU method (Chapter 3).
- To generate a mouse model of *Adcy5* dysfunction and characterise this model for glucose tolerance and birth weight phenotypes *in vivo* as well as measure insulin and glucagon secretion from isolated islets and quantify islet hormone content (Chapter 4).

- To examine and quantify the pancreas morphology of the *Adcy5* Y1064C mouse model and localise the protein in islet cells to better understand an islet phenotype (Chapter 5).
- To profile the expression of *KLF14* in humans and *Klf14* and *trans*-associated genes in mice to inform *Klf14* mouse model characterisation. Further, to complete gene set analyses to look for evidence of enrichment of T2D-associated variants in the *KLF14 trans*-network (Chapter 6).

Chapter 2

Materials and Methods

This chapter describes the general materials and methods common to the thesis. More specific methods, including steps to generate transgenic mice, can be found within the relevant chapters.

2.1 Standard Solutions and General Reagents

Standard solutions were prepared according to Sambrook and Russell (2001). Chemicals were purchased from Sigma Aldrich (Dorset, UK) unless stated otherwise.

2.2 DNA Methods

2.2.1 DNA Amplification

2.2.1 (a) Primer Design

Primers were designed using Primer3 (v0.3.0 and v0.4.0) web-based software (http://frodo.wi.mit.edu/primer3/) and sequences obtained from the University of California Santa Cruz Genome Bioinformatics or National Center for Biotechnology Information websites (http://genome.ucsc.edu/ and http://www.ncbi.nlm.nih.gov/ respectively). Primer pairs were chosen based on length (18-27 bp), melting temperature (57^{o}C - 63^{o}C), guanine and cytosine (GC) content (40-60%), length of amplicon generated (200-600bp), and location free from polymorphic sites. Specific primer sequences are detailed in the appropriate chapters. Primers were synthesized by Eurofins MWG Operon (Ebersberg, Germany) or Eurogentec (Hampshire, UK). Details of all primer pairs can be found in the methods section of the relevant chapters.

2.2.1 (b) PCR reaction components and cycling conditions

Polymerase Chain Reaction (PCR) (*Saiki et al.*, 1985) was used to amplify fragments of DNA sequences of interest. The majority of PCR experiments performed for work described in this thesis were to amplify DNA from the ENU archive of mouse DNA for mutation detection. The protocol is

outlined in section (i) below. The remainder of PCR experiments were performed according to the method outlined in section (ii) below.

(i) PCR to amplify DNA from the ENU archive

PCR was used to amplify selected regions of DNA in order to screen the Harwell ENU archive. A master stock of each primer (100µM) was diluted to a working stock (20µM) with nuclease free water (dH_2O). PCR was performed in the presence of LCGreen (Idaho Technology), a DNA binding dye which saturates the heteroduplexes and allows them to be recognized by high-resolution DNA melting analysis performed on the Idaho Technology Lightscanner by an early decrease in fluorescence (Shimomura et al., Chapter 25). The following PCR conditions were those consistently used:

Reagent	**Concentration in Reaction (mM)**
LCGreen Dye	1
HotShot Mastermix	1
Forward Primer	0.0002
Reverse Primer	0.0002
Distilled Water	n/a

To increase throughput, 2.5ng of archive DNA from four individual mice was pooled together. Master mix of all PCR reagents was made, transferred to 96-well Framestar plates containing 10ng of DNA pools (final PCR volume 10µl), and overlaid with 20µl of mineral oil. PCR was performed using a GRI G-Storm GS4 Thermal Cycler using the following cycling conditions: 95^{o}C for 2 minutes, then 44 cycles of 95^{o}C for 30 sec, primer annealing temperature for 30 sec, 72^{o}C for 30 sec. Following initial

PCR, heteroduplexes were produced by heating samples to 95°C and then rapidly cooling to 15°C. Samples were then centrifuged at 3000rpm for 3 minutes before running on the LightScanner.

(ii) General PCR conditions

For other experiments besides screening the archive, the following PCR protocols were those consistently performed. A master stock of each primer (100µM) was diluted to a working stock (20µM) with nuclease free water (dH_2O).

Reagent	**Volume in 25µl reaction volume (µl)**
AmpliTaq Gold Buffer II (10x)	2.5
2.5mM $MgCl_2$	2.5
1.25 mM dNTPs (Invitrogen)	4
Forward Primer (20µM)	0.25
Reverse Primer (20µM)	0.25
Distilled Water	13.4
1X AmpliTaq Gold Polymerase	0.1
Template DNA (typically 5ng/µl)	2

PCR reaction reagents all sourced from Applied Biosystems unless otherwise stated.

Reaction volumes were prepared in bulk and transferred into individual 96-well microtiter plate wells containing DNA (final volume 25µl). PCR was carried out with a MJ Research PTC 225 thermocycler using the following cycling conditions:

Initial denaturing step 95°C for 10 minutes
13 cycles of:
Denaturing Step 94°C for 20 seconds
Annealing Step *T for 1 minute with a 0.5°C decrease per cycle
Extension Step 72°C for 1 minute
24 cycles of:
Denaturing Step 94°C for 20 seconds
Annealing Step *T for 1 minute
Extension Step 72°C for 1 minute
Final extension step 72°C for 5 minutes

***T = 58°C routinely, but varied depending on specific primer annealing temperature.**

2.2.1 (c) Gel electrophoresis

To visualize the amplified DNA, all PCR products were confirmed using gel electrophoresis with Ethidium Bromide staining. Two percent agarose gels were made using 3 grams agarose (Sigma Aldrich Company Ltd, Gillingham, UK) dissolved in 150ml Tris/Borate/Ethylenediaminetetraacetic acid (TBE) buffer (NBS Biologicals, Huntingdon, UK). The solution was microwaved at 800W for 2 minutes to assist dissolution. To yield a 0.002% concentration of Ethidium Bromide in the agarose solution, 3µl Ethidium Bromide was added after the mixture had been cooled below 50^{o}C. The solution was then poured into a MultiSUB midi gel electrophoresis tank and combs (Appleton Woods Ltd, Birmingham, UK) inserted to create wells at the head of the gel. The gel was allowed to cool for 30 minutes at room temperature until set before filling the tank and covering the gel with 1xTBE buffer. In a separate 96 micro-well plate, 5µl PCR product was vortexed briefly with 5µl purple loading buffer before 10µl of this mixture was loaded into the wells in the gel. In addition, 3.5µl of 100bp or 1kb DNA ladder was loaded into the first well on each row (New England Biolabs UK Ltd, Hitchin, UK and Promega UK Ltd, Southampton, UK respectively), depending upon the expected length of the PCR product. One hundred and twenty volts were passed through the gel for a length of time dependent upon the length of the amplified fragment, and then the elecrophoresed gel was viewed using a Gel Doc 2000 and Quantity One v4.3.1 software (both Bio-Rad Laboratories Ltd, Hemel Hempstead, UK).

2.2.2 Isolation and purification of DNA fragments

DNA fragments generated by PCR were isolated and purified in one of two ways. In both methods, a buffer with a high salt concentration was used to help DNA bind to a silica membrane. Then a series of alcohol washes were done to elute the DNA fragment into 25-40µl dH_2O.

2.2.2 (a) QiaQuick Gel Extraction Kit (Qiagen)

DNA fragments were excised from agarose gel while illuminated with a UV transilluminator. The minimum amount of surrounding agarose was removed and the gel slice was dissolved in a high salt buffer by incubation at 50^{o}C in a water bath. The DNA was isolated using the QiaQuick Gel extraction kit (Qiagen) according to the manufacturer's instructions.

2.2.2 (b) QiaQuick PCR Purification Kit (Qiagen)

PCR product determined by gel electrophoresis to contain only one fragment length was isolated using the QiaQuick PCR purification kit (Qiagen) according to the manufacturer's instructions.

2.2.3 Genomic DNA extraction

DNA was extracted from mouse tissues using one of two methods for genotyping or qPCR analysis: DNeasy Genomic DNA Extraction (Qiagen) or Maxwell 16 system tissue DNA extraction (Promega). DNA was extracted from mouse ear and tail using the DNeasy Qiagen method and from lung, brain, and adipose tissue using the Maxwell 16 system method.

2.2.3 (a) DNeasy Genomic DNA Extraction (Qiagen)

Genomic DNA was extracted from mouse tissues using a DNeasy blood and tissue kit (Qiagen) according to the manufacturer's instructions. The tissue was placed in 180µl of tissue lysis buffer (ATL) with 20µl of proteinase K (PK, 600 mAU/ml solution). This lysate was pipetted onto a silica-based membrane in a spin column. Contaminants and enzyme inhibitors were removed in two wash steps (with AW1 and AW2). The DNA is eluted in H_2O or AE buffer (ear tissue, 100µl; tail tissue, 200µl).

2.2.3 (b) Maxwell 16 system Tissue DNA Extraction

Tissue was placed in the automated Maxwell 16 system with a Tissue DNA Purification Kit (Promega, Madison, USA) according to manufacturer's instructions. Tissue was lysed by shearing and genomic DNA captured on magnetic particles called MagneSil Paramagnetic Particles. The particles and DNA were transferred by magnetism through three wash buffers before being eluted into 300µl elution buffer. Eluted stock DNA was stored at -20^{o}C.

2.2.4 DNA quality and concentration determination

The quality of DNA was assessed and the concentration determined using a NanoDrop 1000 (Thermo-Fischer Scientific, Wilmington, USA) spectrophotometer. The machine was blanked with RNase-free water before 1µl of sample was loaded and assessed for absorbance at wavelengths of 230, 260 and 280nm. The concentration of nucleic acids was determined based on absorbance according to the Beer-Lambert Law (absorbance = wavelength x path length x analyte concentration). The ratios of absorbance at 260:280nm were used to estimate the purity of the samples with an ideal value of 1.8. The concentration of DNA in the sample was determined from absorbance reading at the wavelengths specified in ng/µl.

2.2.5 Sequence analysis

Purified PCR amplification products were sent for Sanger sequencing (Sanger *et al.*, 1977) to either Source BioScience UK Limited (Nottingham, UK), or Geneservice (Cambridge, UK) by courier with 10µM of sequencing primers. Source BioScience UK Limited was used for mutation scanning of DNA pools (Chapter 3), and Geneservice was used for mutation confirmation (Chapter 3). The Sanger method utilizes fluorescently-labelled dideoxynucleotide triphosphates (ddNTPs), which lack a 3'-hydroxyl group and are therefore unable to form a phosphodiester bond with neighbouring nucleotides in an extending DNA sequence. Each ddNTP has a different fluorescence emission spectrum, allowing premature chain termination at each nucleotide to be picked up by capillary

electrophoresis (Rosenblum et al., 1997; Sanger et al., 1977; Smith et al., 1986). Sequencing results were analysed using DNASTAR Seqbuilder and Seqman (Lasergene 7, DNASTAR Inc, Wisconsin, USA), or Mutation Surveyor v3.4 (Soft Genetics, Cambridge, UK).

2.3 RNA Methods

2.3.1 RNA extraction

RNA was extracted from human adult islets to quantify mRNA expression. Details of sample origin, donor characteristics and ethical considerations are provided in the relevant chapters. RNA was also extracted from immortalized cell lines MIN6 and INS1, and from mouse islets, mouse omental, brown and subcutaneous adipose, and mouse fetal tissue.

2.3.1 (a) Human and mouse islet processing and storage for RNA extraction

Islet cells in suspension were transferred to a 15ml centrifuge tube and pelleted by centrifugation at 100G for 5 minutes (4^{o}C). The suspension media was drawn off and 1X phosphate buffered saline (PBS, Sigma Aldrich, Gillingham, UK) added to cover the pellet. Islets were pelleted again and the PBS removed twice more before the islets were re-suspended in 5ml of RNA*later* solution (Ambion, Paisley, UK) to prevent RNA degradation. Samples were stored at 4^{o}C for 24 hours to allow RNA*later* to enter the cells before transferring the samples to -80^{o}C for long term storage.

2.3.1 (b) RNA extraction from human and mouse islets

Extraction was performed using the guanidinium-thiocyanate-phenolchloroform method (Chomczynski and Sacchi, 1987). Islets suspended in RNA*later* were thawed on ice and then centrifuged at 200xg for 5 minutes (4^{o}C) to pellet the cells. As much of the RNA*later* as possible was removed and 1ml Trizol (Ambion, Warrington, UK) was added. The suspension was pushed through a

20 gauge needle (Becton Dickinson, Oxford, UK) 10 times to lyse the cells and then left to incubate at room temperature for 10 minutes to ensure inhibition of RNase activity. The lysates were transferred to 1.5ml RNase-free centrifuge tubes (Applied Biosystems) and 200µl added to the new tubes containing the lysates (FisherScientific, Loughborough, UK). Tubes were shaken vigorously for 30 seconds, left to sit at room temperature for 5 minutes, and then centrifuged at 12,000xg for 15 minutes (4^{o}C) to achieve organic and aqueous phase separation. The clear top layer, or the aqueous phase containing RNA, was removed by pipette to a 1.5ml RNase-free tube along with 500µl isopropanol (FischerScientific, Loughborough, UK) to precipitate RNA. The remaining pink organic and white DNA phases were stored at -20^{o}C for DNA extraction (see section 2.2.3 (b)). The RNA phase and isopropanol were inverted several times, incubated at room temperature for 5 minutes, and then stored overnight at -20^{o}C. Next day, the tubes were centrifuged at 12,000xg for 50 minutes (4^{o}C) and isopropanol removed carefully from the pellet. One milliliter 75% ethanol (Sigma Aldrich, Gillingham, UK) was added to wash, and after centrifugation at 12,000xg for 30 minutes, the wash was repeated with 1ml fresh 75% ethanol. After the last ethanol wash was removed, the RNA pellet was left to dry for 15 minutes before adding 20-100µl RNase-free water for re-suspension (depending upon necessary amount for complete suspension). Sample quality was determined using the NanoDrop (see section 2.2.4) with an ideal value of 2.0 at the ratio of wavelengths 260:280nm before storing at -80^{o}C.

2.3.1 (c) RNA extraction from immortalized cell lines

Cells were pelleted in PBS, supernatant removed, and cells re-suspended in 1ml Trizol. RNA was extracted following the protocol above for human and mouse islets.

2.3.1 (d) RNA extraction from mouse tissue

Samples were thawed on ice and placed in a 2ml tube (Eppendorf, Cambridge, UK) containing a ball bearing (Retsch, Leeds, UK) with 1ml Trizol. The tissue was processed in sections of <100mg. Tissues were homogenized for 5 minutes using a Retsch free-standing homogenizer on the highest frequency. Lysate transferred to a 1.5ml RNase-free tube. The samples were centrifuged at 12,000xg for 20 minutes (4^{o}C) and supernatant was collected through the lipid layer and transferred to a fresh 1.5ml RNase-free tube with 200µl chloroform. Insoluble tissue fragments were discarded. Extraction was carried out according to the protocol detailed above.

2.3.2 cDNA Synthesis

cDNA was generated from RNA to quantify mRNA expression. The QuantiTect reverse transcription kit (Qiagen, Crawley, UK) was used to synthesize single-stranded cDNA from RNA. 1µg of RNA in 12µl ddH_2O was vortexed with 2µl gDNA wipeout buffer. The mixture was incubated at 42^{o}C for 2 minutes to eliminate any remaining gDNA contamination. For each sample, reverse transcription (RT) reactions were carried out in the presence and absence of reverse transcriptase enzyme in order to test for residual gDNA in later expression experiments. Mastermix, 5.8µl, was added to each sample and mixed (components detailed in table 2.1). The samples were briefly centrifuged followed by incubation at 42^{o}C (30 minutes) and then 95^{o}C (3 minutes) on a PTC-225 Peltier Thermal-Cycler (MJ Research, Waltham, Massachusetts, USA) to allow for the reverse transcription reaction and then deactivation of the enzyme. cDNA was stored at 4^{o}C.

2.3.3 mRNA quantification

TaqMan chemistry (Applied Biosystems) was used to carry out quantitative (real-time) PCR (QPCR) to measure mRNA expression. The TaqMan method employs probes with a fluorophore attached to their 5' end and a quencher attached to their 3' end, short enough to keep the fluorophore from fluorescing. Taq polymerase working during each cycle of PCR releases the fluorophore from its

quenched state (Holland et al., 1991). The resulting fluorescent emissions are directly proportional to the quantity of cDNA generated. The cycle number at which fluorescence reaches a critical threshold is proportional to the initial cDNA in the sample and therefore the original mRNA concentration.

2.3.3 (a) Taqman probe selection

Taqman gene expression assays were ordered from Applied Biosystems and selected to amplify as many alternative transcripts of the relevant gene as possible. Wherever possible, probes were selected which crossed an exon-exon boundary (Taqman label: _m1) to prevent gDNA contamination.

Component	RT+ve (µl)	RT--ve
10x RT Buffer	2.0	2.0
25x dNTP Mix	0.8	0.8
10x RT Random Primers	2.0	2.0
Multiscribe Reverse Transcriptase	1.0	0.0
Nuclease free H_2O	4.2	5.2
Total RNA (20-1000 ng)	10.0	10.0
Total	20.0	20.0

Table 2.1: cDNA synthesis reactions from RNA using Quantiftect reverse transcription kit.

For *KLF14* and other single exon genes, _s1 (single exon) probes were required and qPCR experiments included expanded to include RT-ve reactions to verify the absence of gDNA. All assays were ordered fluorescein amitin-labelled (FAM-labelled). Assay inventories are provided in experimental chapters: *ADCY5* qPCR experiments are generated in Chapter 4, and *KLF14* qPCR experiments are detailed in Chapter 6.

2.3.3 (b) Housekeeping gene selection

'Housekeeping genes' are the genes against which an mRNA quantitation is normalized. Housekeeping genes should be uniformly expressed across all samples and therefore make apparent any differences in starting cDNA concentration. Opinion differs on which genes are ideal housekeeping genes for different tissues. Investigation done in our laboratory (data not shown), has previously proposed that in mice, hypoxanthine-guanine phosphoribosylytransferase *(Hprt)* is an effective housekeeping gene to use with islet cDNA as well as across all other tissues. Leukocyte receptor cluster member 8 (*Leng8)* is also particularly appropriate for mouse islet cDNA. Phosphoglycerate kinase 1 (*Pgk1)* was shown to be an effective housekeeping gene when used with rodent adipose tissue. Therefore these genes were chosen as housekeeping genes for expression across the mouse tissue panel and islets.

In humans, qPCR experiments performed in our laboratory have suggested that *HPRT1* is a very successful housekeeping gene for islet cDNA. Beta-2-microglobulin *(B2M)* and beta-actin *(ACTB)* are both established control genes across all tissues. These three genes were used as housekeepers for the human panel.

A mean of the three control genes was used for normalization (Vandesompele et al., 2002), and control genes were VIC dye-labeled and run on the same plate as the FAM-labeled genes of interest.

2.3.3 (c) Generating a standard curve for qPCR normalization

One microliter neat cDNA from each sample was pooled and then diluted serially to generate six concentrations. Concentrations were determined based on the dilution of the samples in the experiment; if the samples were run at a 1:100 dilution, the standards for that experiment were generated at 1:50, 1:100, 1:200, 1:400, 1:800 and 1:1600. Standards were generated for each individual experiment.

2.3.3 (d) Quantitative real-time PCR

Assays were tested before experiments on a 3-5 samples at dilutions (in 0.01M Tris HCl) of 1:5, 1:10, 1:50 and 1:100. The dilution which demonstrated quantifiable cDNA amplification with minimal cDNA use was chosen. Any assay which yielded no detection after 40 amplification cycles at a 1:5 sample dilution was considered nonfunctional and redesigned. The qPCR experiment was designed so that standards for a given assay were always on the same plate as samples for that assay, and all samples and standards were run in triplicate. Quantitative PCR using real-time detection of accumulated fluorescein was performed using TaqMan Gene Expression Assay reagents (Applied Biosystems). Total reaction volume was 10µl: 4µl diluted cDNA, 5µl Taqman gene expression mastermix (Applied Biosystems), 0.5µl assay and 0.5µl dH_2O. Plates were loaded into a 7900 Fast Real-Time PCR System (Applied Biosystems) where the following program was carried out: 50^oC for 2 minutes, 95^oC for 10 minutes followed by 40 cycles of 95^oC for 15 seconds and 60^oC for 1 minute. Fluorescence was read at every cycle and SDSv2.3 software (Applied Biosystems) determined at which cycle number fluorescence level crossed a critical threshold. This value was given as a comparative threshold (C_T) score.

2.3.3 (e) Analysis

The $\Delta\Delta C_T$ method (Pfaffl, 2001) was employed to determine gene expression. In this method, C_T values for genes of interest are compared to a standard at the same dilution (ΔC_T) , and then normalized to the endogenous control C_T value ($\Delta\Delta C_T$). This value was linearized by the following formula:

$$(1 + \text{assay efficiency})^{-\Delta\Delta CT}$$

Assay efficiency was calculated from the standard curve using the following formula:

$$(-1/10^{slope}) - 1$$

Slope was determined by plotting the C_T (x-axis) against the log-dilution (y-axis).

2.4 Cell Biology

2.4.1 Cell culture

Cell lines including MIN6 (mouse insulinoma cell line, Miyazaki et al., 1990, obtained from Dr Nicola Beer, OCDEM, University of Oxford) and INS-1 (rat insulinoma cell line, also obtained from Dr Nicola Beer) were used for this study. Cell culture was performed following aseptic guidelines in a Class II vertical laminar-flow biological cabinet. Cells were cultured in RPMI 1640 11mM glucose plus glutamine (Roswell Park Memorial Institute medium, Invitrogen) supplemented with 10% fetal calf serum and 1% penicillin/streptomycin (Invitrogen) in 75cm^2 flasks with vented caps at 37^{o}C in a 5% CO_2 humidified atmosphere. When cells reached approximately 90% confluence, they were passaged by washing with Dulbecco's phosphate buffered saline (D-PBS) and detached from the flask by swirling with 1ml trypsin (Invitrogen) for 1 minute. Detached cells were then re-suspended in 5ml culture medium to neutralize trypsin before centrifuging at 200xg for 5 minutes to pellet. Pelleted cells were again re-suspended in 10ml medium and 2ml of suspended cells were added to two fresh flasks containing 10ml medium each to split the cells.

Freshly isolated mouse islets were cultured in RPMI 1640 supplemented with 5mM glucose, 10% fetal calf serum and 1% penicillin/streptomycin at 37^{o}C in a humidified 5% CO_2 environment in 5cm^2 cell culture dishes for 5-24 hours before fixation for morphological studies (see section 2.4.2c).

2.4.2 Immunocytochemistry

2.4.2 a) Processing of pancreas specimens into wax for light microscopy

Freshly dissected whole mouse pancreata were fixed in 10% formaldehyde in PBS (phosphate buffered saline, pH 7.4). Specimens were kept in fixative at least overnight at 4^{o}C. When ready for embedding, pancreata were placed in glass heatproof jars. The pancreata were washed in PBS twice for 30 minutes followed by 70% alcohol and then 90% alcohol, each for 45 minutes on the shaker. Subsequently, pancreata were submerged in absolute alcohol 3 times for 45 minutes each, and then 100% Xylene x3 (45 minutes each) and embedded in wax.

2.4.2 b) Immunoperoxidase labeling of pancreas sections

Sections (5 microns thick) were cut of wax embedded mouse pancreata. For immunocytochemistry, sections were de-waxed and rehydrated in alcohol. To retrieve antigens for ADCY5 labeling, slides were microwaved in 0.01M sodium citrate (pH 6.0) for 5 minutes full power and washed in water. For peroxidase substrate labeling, endogenous peroxidase activity was blocked in 0.3% hydrogen peroxide in methanol (30 minutes). Slides were washed 3 times in TRiS buffered saline (TBS, pH 7.4 [5 minutes each]) and put into a wet chamber.

To block non-specific binding of antisera, non-immune animal serum (swine serum) was diluted 1/20 in TRiS buffer and applied to slides for 20 minutes (100µl/slide). Serum was tipped off and the primary antibody applied: anti-insulin (made in a guinea pig) diluted 1/500 in TRIS buffer; mouse anti-glucagon diluted 1/500; anti-ADCY5 (Abcam ab14788 and Abcam 66037, Cambridge, UK) 1/100. Slides were incubated overnight in primary antibody at 4^{o}C. Slides were washed in TBS x3 and the secondary antibodies added diluted 1/50 in Tris buffer. For immunoperoxidase labeling, peroxidase conjugated (HRP) anti-guinea pig (insulin, Dako) or anti-mouse (glucagon, Dako) was used at 1:50 dilutions. For ADCY5, the avidin-biotin system was used to enhance the signal; anti-rabbit biotin

(1:100, Dako) followed by washing and then avidin-HRP (1:50). Antibodies were applied for 30 minutes at room temperature. Slides were washed TBS x3. For visualisation, diamino benzidine hydrochloride (DAB-made up in 10ml dH_2O) was added.

Slides were counterstained with haematoxylin for 30 seconds and then washed in water and de-stained in acid alcohol for 1 minute before a final wash in tap water for 15 minutes. Slides were then dehydrated through graded alcohols and Xylene and mounted with coverslips.

2.4.2 c) Fluorescent labeling of pancreas sections

Double labeling for ADCY5 and insulin or glucagon with fluorescent labels was performed on mouse and human pancreas sections. Sections were de-waxed and rehydrated, antigens were retrieved in sodium citrate and endogenous peroxidase blocked as above. Non-specific binding was blocked with swine serum (1:20) and the primary antibodies applied. Slides were incubated overnight in antibody to ADCY5 (1:100) (Abcam ab14788 or Abcam 66037, Cambridge, UK) washed in TBS x3, followed by anti-rabbit biotin (as above), washing, and then avidin conjugated fluorescein (1:100, Vector) in the dark for 30 minutes incubation. Following washing, hormone labeling for insulin and glucagon (primary antibodies as above) was performed and binding sites identified with anti-guinea pig Texas Red (1:50 Vector) or anti-mouse TRITC (Sigma 1:50). Slides were mounted in Vectashield mounting media and viewed with a confocal microscope (BioRad).

2.4.2 d) Fixing dissociated islet cells on coverslips

For examination of isolated islet cells, isolated islets (50-150 islets) were dispersed by suspension in 1ml dispersion solution (Sigma) and incubated in a water bath at 37^{o}C for 10 minutes. Dispersion was ensured by vigorous mixing (pipetting several times) after incubation, and the reaction was stopped by addition of 9ml cell culture media. The isolated cells were pelleted by centrifugation at

500xg for 5minutes. Media was removed and cells washed with 9ml more media. After second wash, pelleted cells were re-suspended in 900-1600µl PBS (depending upon number of islets used). One hundred and fifty microliters of the cell suspension was pipetted onto 6-12 coverslips coated with poly-l-lysine (Sigma) and allowed to adhere. Cells were fixed for morphological examination by gently adding 2% paraformaldehyde in 0.1M phosphate buffer pH 7.2 (PB) until coverslips were submerged. Fixed cells were stored at 4^{o}C until needed for immunocytochemistry.

2.4.2 e) Fluorescent labeling of islet cells

Dissociated mouse islet cells fixed on coverslips were double immunocytochemically labeled for insulin, or glucagon with ADCY5. A foil covered box was lined with water-soaked paper towels and a layer of parafilm was placed on top to make a wet chamber. The coverslips were placed cell side down on large drops (200µl) of TBS buffer on the parafilm for five minutes and this was repeated in a fresh drop of TBS to wash thoroughly. The cells were permeabilized with TBS + 0.001M Triton X for 5 minutes. Coverslips were washed twice more as described above. Non-specific labeling was blocked with swine serum diluted 1:20 in TBS buffer for 20 minutes. Labeling was performed for insulin+ADCY5 or glucagon+ADCY5 as above (fluorescent labeling of sections). One coverslip was placed in TBS buffer to serve as a 'no primary antibody' control. When labeling was complete, coverslips were mounted on slides with Vectashield mounting media and stored in the dark. Slides were visualized with a confocal microscope (BioRad, Hertfordshire, UK).

2.4.3 Electron Microscopy

Preparation of samples and electron microscopy was done by Dr Anne Clark, University of Oxford. Freshly isolated islets were fixed in 4% paraformaldehyde plus 0.5% glutaraldehyde (electron microscopy grade) in 0.1M phosphate buffer pH 7.2 (PB). Islets were fixed for a minimum of 1 hour, or up to 1 month at 4^{o}C. Pelleted cells were washed in PB 3 times for 10 minutes each, spinning

tubes gently in between to pellet the islets. Cells were then submerged in 1% osmium tetroxide in PB and shaken for 1 hour. Osmium was discarded into a beaker containing 70% ethanol in a fume hood to precipitate it into its less hazardous form, and the cells were washed with distilled water 3 times for 10 minutes each. To block stain the samples, 2% uranyl acetate in acetate buffer was added to cover cells and they were then shaken for 1 hour. Another 3 washes in distilled water for 10 minutes each followed. The islets were submerged in 70% alcohol for 15 minutes, and then 90% alcohol for another 15 minutes, and finally in absolute alcohol 3 times for 10 minutes each. The islets were left overnight in 50% Spurr's resin 50% absolute alcohol. Spurr's resin was made in a fume hood by mixing well 3.5g DER 736 (AGAR Scientific), 13g NSA (AGAR Scientific), 5.0g ERL 4221 (AGAR Scientific), and 0.24ml S1 (AGAR Scientific). The next day, the solution was changed to pure Spurr's resin and left for 6 hours. Fresh pure Spurr was changed again and left overnight on a shaker. Finally, pure Spurr was changed out again and the cells were spun hard and the resin hardened at 50^{o}C overnight. Sections (70nm thick) were cut onto nickel grids, contrast enhanced with uranyl acetate and lead citrate before visualizing on an electron microscope (Jeol 1010, Peabody, Massachusetts, USA).

2.4.4 Histological Image Analysis

Pictures of entire stained pancreas sections were taken on a confocal microscope (BioRad) at two layers for each pancreas. The photos were visualized using ImageJ (Java 1.6.0_20, NIH, Maryland, USA) and pancreas area, islet area, beta cell area, and alpha cell area were counted by tracing the perimeter of each cell, islet, or exocrine area with the picking tool in ImageJ (Java 1.6.0_20) and then calculating the area circled. Areas were used to calculate islet density: (total islet area)/(total pancreas area), % beta cell/islet, %alpha cell/islet, and median islet size. Area of pancreas examined varied between 25 and 50 square millimetres, and between 24-120 islets were measured for each animal.

2.4.5 Insulin and Glucagon Secretion from Isolated Islets

2.4.5 (a) Isolation of Mouse Islets

Pancreas from wild-type and *Adcy5* Y1064C B6/C3H adult mice were inflated in-situ with ca. 3ml of collagenase (Type V, 1.9 U/mg; Sigma) in HANKS Balanced Salt Solution (+$NaHCO_3$ phenol red; Sigma) by back-injecting through the pancreatic duct clamped at the duodenum with a 30-gauge needle (Figure 2.1).

Pancreata were then cut from the abdominal cavity, and each incubated in a 15ml falcon-tube for 13 minutes at 37^{o}C. 12ml of HANKS containing 0.2% bovine spine albumin (BSA, fraction V; Sigma) was added, and the mixture shaken vigorously. The mixture was then left to settle on ice for 5 minutes before the supernatant was removed to 3ml, and 7ml of BSA-containing HANKS added. The tubes were tipped upside down several times, and then left to settle for another 3 minutes on ice. This procedure was repeated another time and then a final suspension made in 10ml HANKS.

The pancreatic islets were isolated from the exocrine tissue by pouring 2ml of the suspension into a cell suspension dish and placing them under a stereo dissecting microscope for picking. Islets were handpicked using a 1ml Gilson pipette coated in RPMI media and transferred to a second cell suspension dish filled with HANKS (BSA free). Another pick was made after this into a 3rd dish filled with RPMI media (-glucose, +glutamine; Gibco) with added Penstrep (10,000 U/ml penicillin/10,000 g/ml streptomycin, Gibco), 10% fetal bovine serum/FBS; Gibco), and containing 10mM glucose.

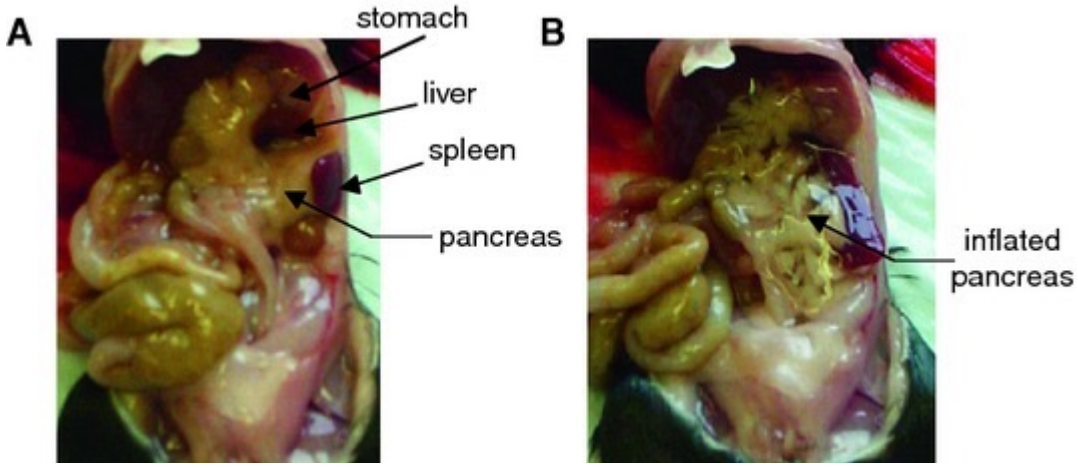


Figure 2.1: Abdominal Cavity before (A), and after (B) inflation of pancreas with collagenase (from Ravier & Rutter in Ward, A. & Tosh, D. (eds.), Mouse Cell Culture, 2010.

2.4.5 (b) Secretion Stimulation and Collection

Secretion studies were performed on isolated mouse islets. Eight to 12 size-matched islets were picked into small tubes containing 300µl RPMI media (-glucose, +glutamine; Gibco) with added penstrep (10,000 U/ml penicillin/10,000 g/ml streptomycin, Gibco), 10% fetal bovine serum/FBS; Gibco). Islets were washed by removing the RPMI media off the pelleted cells and then 300µl was added back. After the pellet re-settled, the media was removed again and 300µl RPMI media containing 1mM glucose solution added. The islets were incubated in the 1mM glucose solution for one hour at 37^{o}C in a humidified chamber, then the solution was removed and 300µl experimental condition solutions were added (more details found in relevant experimental chapters). The tubes were incubated like this for one hour. The supernatant was removed from the pellet (300µl) into a fresh tube, and 100µl of acid ethanol was added to the pellet to lyse the cells. Both were stored at -20^{o}C until concentration could be determined by radio-immuno assay.

2.4.6 In situ pancreatic perfusion

This procedure was done by Nils Rorsman (OCDEM). Mice were culled in the schedule 1 method (cervical dislocation) and secured to a heated operating table (Harvard Apparatus). After opening the abdominal cavity, the renal arteries were sutured followed by the descending aorta above the coeliac branch point. The gastric artery, vein and the esophagus were sutured as well and the posterior end of the abdominal aorta was clamped. A small hole was made in the aorta and a catheter inserted and secured. The portal vein was sutured close to the liver and a small hole cut at the first branch point and a catheter inserted to collect perfusate in a Telecdyne ISCO Foxy R1 fraction collector (Lincoln, Nebraska, USA). Perfusate was kept warm with a Warner Instruments temperature control unit TC-32 4B in conjunction with a tube heater (Warner Instruments P/N 64-

0102, Hamden, Connecticut, USA). The vessels were washed with 6mM glucose for 20 minutes to clear of blood and equilibrate the pancreas. The pancreas was perfused from aorta to portal vein at a rate of ~0.45ml/minutes using an Ismatec Reglo Digital MS2/12 peristaltic pump. Fractions were collected for 1 minute per fraction in a 96 well plate. Samples were assayed by radio-immuno assay (see below).

2.4.7 Radio-immuno assays

2.4.7 (a) Insulin assay

The assay was carried out using the reagents and according to the instructions of the LINCO Rat Insulin Radio-immuno assay kit (LINCO Research, Missouri, USA). All samples (pellet and supernatants) were defrosted, vortexed, and briefly spun down. Pellet samples were diluted 1:1000 and supernatants were diluted 2:3 in assay buffer provided. On day one, 100µl of each sample was pipetted into borosilicate glass tubes (12x75mm). Reference tubes were set up by making serial dilutions from the standard provided of purified rat insulin 10.0ng/ml into concentrations of 10, 5, 2.5, 1.25, 0.625, and 0.312ng/ml. One hundred microliters of each standard was pipetted in duplicate into glass tubes. Two hundred microliters pure assay buffer was added to 2 non-specific binding tubes and the first 2 tubes were left empty as controls according to assay setup instructions. One hundred microliters of hydrated ^{125}I-Insulin was added to all tubes. 100µl of Rat Insulin antibody was added to all tubes except the control tubes (1-2) and non-specific binding tubes (3-4). All tubes were vortexed, covered and incubated overnight (20-24 hours) at 4^{o}C.

On day 2, 1.0ml of cold (4^{o}C) Precipitating Reagent was added to all tubes except control tubes 1-2. All tubes were vortexed and incubated for 20 minutes at 4^{o}C. All tubes except tubes 1-2 were centrifuged for 20 minutes at 2,000xg. Immediately after centrifugation, supernatant was decanted from the tubes by inversion and blotting of excess liquid from lip of tubes. Pellets were allowed to

dry for 5 minutes before loading into a gamma counter (Emphor Fzco, Wizard 2470, Dubai, UAE) for counting of bound insulin.

2.4.7 (b) Glucagon Assay

Samples were vortexed and briefly spun down. The assay was set up according to the procedure manual and using reagents from the EURIA-Glucagon kit (Euro Diagnostica, Malmö, Sweden). Pellet samples were diluted 1/800 in assay buffer. Supernatants were undiluted. Reagents were reconstituted according to kit instructions and standards were diluted from 300 pmol/L standard provided according to the following protocol:

a) 1.00 ml standard 300 pmol/L + 1.00 ml assay diluent = 150 pmol/L

b) 1.00 ml standard 150 pmol/L + 1.00 ml assay diluent = 75 pmol/L

c) 1.00 ml standard 75 pmol/L + 1.00 ml assay diluent = 37.5 pmol/L

d) 1.00 ml standard 37.5 pmol/L + 1.00 ml assay diluent = 18.8 pmol/L

e) 1.00 ml standard 18.8 pmol/L + 1.00 ml assay diluent = 9.4 pmol/L

d) 1.00 ml standard 9.4 pmol/L + 1.00 ml assay diluent = 4.7 pmol/L

g) assay diluent = 0 pmol/L

Two hundred microliters of the standards were pipetted in duplicate in respective tubes as well as 200 µl assay diluent in non-specific binding tubes (NSB). Two hundred microliters of each sample was pipetted into separate tubes. Five hundred microliters anti-glucagon was added to all tubes except the NSB tubes. Five hundred microliters assay diluent was added to the NSB tubes. All tubes were vortexed and incubated for 20-24 hours at 2-8^{o}C. On day two, 500µl ^{125}I-Glucagon was pipetted into all tubes. All tubes were vortexed and incubated for 20-24 hours at 2-8^{o}C. On day 3, 100µl double antibody solid phase was added to all tubes except the TOT tubes. All tubes were vortexed and incubated for 30 minutes at 2-8^{o}C. Tubes were centrifuged for 15 minutes at 4^{o}C at 1700xg.

Supernatant was decanted immediately following centrifugation and the pellets were run through a gamma counter to count the radioactivity (Emphor Fzco, Wizard 2470).

2.4.7 (c) Analysis

The gamma counter generated a standard curve by plotting the bound count per million (CPM) against the concentration of the standards and determined concentrations in the samples and controls from the standard curve. Insulin and glucagon content and secretion were determined according to the following equations.

$$\text{Insulin content } \left(\frac{\text{ng}}{\text{islet}}\right) = \frac{\text{Insulin pellet (ng)} \times \text{volume in assay (0.1mL)}}{\text{\# of islets}}$$

$$\text{Glucagon content } \frac{\text{pg}}{\text{islet}}) = \frac{\text{Glucagon pellet (pg)} \times \text{volume (0.1mL)}}{\text{\# of islets}}$$

$$\text{Insulin secreted } \left(\frac{\text{ng}}{\text{islet}} \div \text{hr}\right) = \frac{\text{Insulin supernatant (ng)} \times \text{volume (0.25mL)}}{\text{\# of islets}} \div 1(\text{hr})$$

$$\text{Glucagon secreted } \left(\frac{\text{pg}}{\text{islet}} \div \text{hr}\right) = \frac{\text{Glucagon supernatant (ng)} \times \text{volume (0.3mL)}}{\text{\# of islets}} \div 1(\text{hr})$$

2.5 Animal Husbandry and Metabolic Phenotyping

Animal studies were carried out under the guidance issued by the Medical Research Council; Responsibility in the Use of Animals for Medical Research (July 1993) and Home Project License Nos. 30/2161 or 30/2642. I was in possession of a personal license from the Home Office to perform mouse studies. Mice were kept in accordance with UK Home Office welfare guidelines and project license restrictions under controlled light (12 hour light and 12 hour dark cycle), temperature (21^{o}C ± 2^{o}C) and humidity (55% ± 10%) conditions. They had free access to water (25 ppm chlorine) and were fed ad libitum on a commercial diet (SDS maintenance chow, RM3, 3.6kcal/g, Essex, UK).

2.5.1 Genotyping

Ear and tail tissue was taken from mice post weaning (around 3-4 weeks of age). DNA extraction (according to section 2.2.3 (a)) and individual genotyping assays performed, the specifics of which are provided in relevant experimental chapters for each mouse line. Genotyping was performed following one of two procedures: using a TaqMan genotyping assay, or using a Pyrosequencing assay.

2.5.1 TaqMan Genotyping

2.5.1 (a) Assay Design

All assays were obtained from Applied Biosystems and the details of the assays can be found in the relevant experimental chapters. If inventoried assays were not available, primers were designed from genomic sequences taken from the National Centre for Biotechnology Information (NCBI) and Primer Express v3.0 (Applied Biosystems). The designed primer/probe mixes were then synthesized by Applied Biosystems.

2.5.1 (b) TaqMan PCR

Genotyping PCR reactions were carried out in a 384-well PCR plate (4titude, Wotten, UK) in 2µl volume. One microliter genotyping mastermix (containing polymerase and dNTPs, KBioscienes, Hoddesdon, UK), 0.98µl distilled water, and 0.02µl assay were added on top of 10ng DNA which had been dried into the plate previously by incubating at 80^{o}C. After brief centrifugation, samples were heated to 95^{o}C for 10 minutes to activate polymerase. Forty cycles of 92^{o}C for 15 seconds (DNA melting) and 60^{o}C for 1 minute (annealing and extension step) followed to amplify the relevant section of the DNA.

2.5.1 (c) Mouse genotype determination from PCR reaction

All mice used in the experiments in Chapter 4 and Chapter 5 were genotyped according to the following method.The plate was placed in an Applied Biosystems 7900HT machine, using SDSv2.3 software and standard analysis settings. Genotypes were assigned using the software algorithms and also checked visually. Some genotype reassignments were made by myself and a postdoctoral colleague (Dr. Amanda Bennett, OCDEM, University of Oxford) where we independently agreed. Samples were re-run if more than 5% of samples appeared to require reassignment. Four or five non-template controls (NTCs) containing dH_2O rather than DNA and four or five sample duplicates were included on each plate. If NTCs displayed DNA amplification or if duplicates did not match, the genotyping was repeated.

2.5.1 (d) Pyrosequencing

The pyrosequencing assay mastermix was made up using the components below except for the template DNA in the volume of the number of samples + 2.8µl of this mastermix was pipetted into each well, and DNA samples were added to each well including a water control. The plate was sealed with a heat-sealed lid and spun briefly before loading onto the thermal cycler and run according to the program laid out below for PCR amplification.

The plate was then run on a PyroMark Pyrosequencer v1.0 (Biotage, Massachusetts, USA). The pyrosequencer adds deoxyribonucleotide triphosphates (dNTPs) one at a time to the PCR template incubated with sequencing primers and enzymes. As the complementary dNTPs are incorporated into the DNA strand by DNA polymerase, pyrophosphate (PPi) is released in equal proportion to the amount of nucleotide incorporated. The PPi is converted to ATP by ATP sulfurylase, and the ATP drives the conversion of luciferin to oxy luciferin which generates visible light. The light is detected by a camera and displayed as a peak which is proportional to the number of nucleotides incorporated. An enzyme degrades ATP and unincorporated dNTPs until the light switches off and

another dNTP is added to repeat the cycle until the complementary DNA strand it built and the light peaks determine the nucleotide sequence.

Component	Volume (µl)
Taq PCR Master Mix (x2)	5
25mM MgCl2	0
Forward Primer (10µM)	0.2
Reverse Primer (10µM)	0.2
Water	2.6
Template DNA (5ng/µl)	2
Reaction Volume	**10**

Initial denaturing step 95^{o}C for 5 minutes	
44 cycles of:	
Denaturing Step	94^{o}C for 15 seconds
Annealing Step	60^{o}C for 30 seconds
Extension Step	72^{o}C for 15 seconds
Final Extension Step	72^{o}C for 5 minutes
Hold 16^{o}C forever	

Table 2.1: PCR Reaction and cycling conditions for pyrosequencing assay.

2.5.2 Metabolic phenotyping of mice

All phenotyping was carried out at the Medical Research Council (Harwell, Oxfordshire) following EMPReSS (European Phenotyping Resource for Standardised Screens from EUMORPHIA; Brown *et al*., 2005; Mandillo *et al*., 2008). Standard operating procedures can be found at http://empress.har.mrc.ac.uk.

2.5.2 (a) Body Composition

Body mass was measured on scales calibrated to 0.01 grams. Dual-energy X-ray Absorptiometry (DEXA, PIXImus, Wisconsin, USA) was used to quantify fat mass, lean mass and bone mineral content and density (Nagy and Clair, 2000). Mice were placed in the DEXA after body weight was collected. The DEXA method employs a high stability X-ray generator to produce photons over a broad

spectrum of energy levels. The software transforms the count data into bone and non-bone components generating bone density as well as soft tissue values.

2.5.2 (b) Adult Tissue Collection

Pancreas and white/brown adipose tissue was collected from adult mice and fixed in neutral buffered formaldehyde for histological analysis (Surgipath Europe Ltd, UK).

2.5.2 (c) Neonate Tissue Collection for birth weight measurement

Pregnant females were watched closely approaching gestation. Litters were culled as close to birth as possible (>2 hours, <24 hours). Neonates were culled by decapitation and the carcasses stored at -20^{o}C for dissection.

2.5.2 (d) Determining Bone Lengths by X-ray Analysis

Neonate carcasses were defrosted and the right hind leg was removed above the hip by severing the joint between the hip and pelvis. Legs were placed on a tray below a strip 1cm in length and X-rayed. X-ray photos were saved and analysed using Adobe Photoshop CS6 software. Lengths of forming tibia and femur were measured using ruler tool and standardizing to length of 1cm strip in photo to account for any differences in zoom.

2.5.2 (e) Blood collection and intraperitoneal glucose tolerance test (IPGTT)

Mice were fasted overnight (12-16 hours) to establish baseline glucose level (time zero). After weighing, a blood sample was collected from the tail vein of each mouse after applying a topical anesthetic (EMLA cream, Eutectic Mixture of Local Anesthetics, Lidocaine/Prilocaine, AstraZeneca, UK). Blood was collected using Lithium-Heparin microvette tubes (CB300, Sarstedt, Nümbrecht, Germany). Each mouse then received an IP injection of 2 grams glucose/kg body weight (20%

glucose in 0.9% NaCl). Blood samples were taken again at 60, 90 and 120 minutes post injection. Fasted blood samples were taken from the tail vein as described for the IPGTT.

2.5.3 Plasma analysis of mouse blood samples

Microvettes containing blood samples were centrifuged at 3000 rpm (800G) for 10 minutes at room temperature. The plasma layer (on top of the red blood cells) was carefully removed and transferred to a 1.5ml tube. Plasma was stored at -20^{o}C. Analysis for plasma glucose content was performed on an Analox Glucose Analyser GM9 (London, UK). Insulin and glucagon levels were measured using a mouse endocrine MILLIPLEX kit (MILLIPLEX MAP, Millipore) and a Bio-Plex 200 system (Bio-Rad). Following the addition of dipeptidyl peptidase IV (DPP-4) inhibitor (Millipore) and serine protease inhibitor mix (Sigma-Aldrich) to plasma to prevent degredation, GLP-1 (active) was measured using a MILLIPLEX MAP mouse gut hormone panel kit (Millipore) with a Bio-Plex 200 system (Bio-Rad).

2.6 Statistical analysis

Linear regression, power calculations, and Student's T-Test analyses were conducted in SPSSv18.0 (PASW statistics 18). Statistical significance was defined as $p<0.05$ for all tests. Results are written as mean ± standard error of the mean (s.e.m.). Genotype groups, usually wild-type, heterozygous, and homozygous were separated by sex.

2.6.1 Pathway Analysis of KLF14 trans-associated transcripts in gene sets enriched for T2D variants

A list of 26,914 human gene transcripts from the UCSC Genome Browser (http://genome.ucsc.edu/) in RefFlat format based on the human March 2006 (hg18) assembly was used in conjunction with the T2D GWA meta-analysis (DIAGRAMv3) (Morris et al., 2012) to test the pathway enrichment of *KLF14 trans*-associated genes with MAGENTA (Segrè et al., 2000).

Specifics of these analyses can be found in the Methods section (section 2) of Chapter 6.

Chapter 3

Screening of the ENU archive at the *JAZF1, KLF14*, and *ADCY5* loci, and re-derivation of mouse models

3.1 Introduction

3.1.1 The need for functional investigation at T2D-associated loci

Genome-wide association studies have established over 90 confirmed loci of type 2 diabetes (T2D) susceptibility, summarized in Chapter 1. These loci provide a foundation from which we can learn more about the pathology of the disease and common predispositions within the population. However, insight into T2D disease mechanisms resulting from the study of these variants has been limited by a variety of factors. One such factor is that many of the variants fall outside of coding regions. There has been limited success explaining the molecular mechanisms by which non-coding variants affect pathology so far. Voight et al. (2010) and others have delved into the molecular mechanisms of T2D associated variants finding an enrichment of variants in cell cycle regulation pathways, metabolic pathways, and other protein-protein interactions (Dimas et al., 2014; DIAGRAM, 2014). However, pinpointing the mechanism connecting the variant to these effects is still a challenge. In order to make use of the vast majority of the data from GWAS, more progress must be made in identifying the physiology by which such non-coding variants impact beta cell function and insulin resistance.

3.1.2 The benefits of modeling T2D genetic predispositions in mice

Mouse models have proven to be extremely successful tools for investigating genetic mechanisms of complex diseases and have remained the primary mammalian model for genetic research over the last century. T2D is a complex disease affecting various pathways and organ systems requiring investigation *in vivo*. The benefits of the mouse as a diabetic model include: availability of tools and resources for manipulation of the mouse genome, comparable physiology to humans, and a short reproduction time for rapid generational studies.

New technology has made it possible to design a mouse line with virtually any genetic manipulation desired. Genes can be 'knocked out' using homologous recombination in ES cells to insert a null allele into the genome, and they can also be knocked out in specific tissues or at specific times using a conditional allele. Genes can be overexpressed by cloning the cDNA into an expression vector with a strong promoter. In recent years, transcription activator like effector nucleases (TALENs) and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) plus a mutated Cas9 gene have made it possible to introduce a desired sequence carrying a tailored mutation into the genome at a desired point (Jiang, et al., 2013; Joung and Sander, 2013). This new technology allowing the introduction of precisely targeted changes in the genome could be useful for introducing human mutations into the mouse genome quickly and easily. Although TALENs and CRISPRs are still developing technologies, another method of inducing point mutations into the genome has been available for decades (Russell and Montgomery, 1982; Russell et al., 1988). This is chemical mutagenesis, or the introduction of mutations into the genome using DNA-damaging chemicals such as *N*-ethyl *N*-nitrosourea (ENU). I will discuss the ENU technique and its implications below.

Although yeast and drosophila are useful for manipulating the cell cycle and studying many developmental processes, they are not good models of immune, endocrine, nervous, cardiovascular, diseases or other systems that mammals share (NIH News Advisory: "The Mouse Genome and the Measure of Man," Dec 2002). Mice are naturally affected by metabolic diseases such as diabetes, and the progression of the disease closely resembles the pathology in humans. There are some key disparities in metabolism such as in cholesterol metabolism, and these limitations must be accounted for when using the mouse for metabolic research (McMurray and Cox, 2011). Diagnostic tests similar to those used in human patients are used in the mouse to achieve comparable results (McMurray and Cox, 2011).

In genetic research, the need to breed multiple cohorts for generational studies requires a model that can reproduce in a short period. The mouse can reproduce as often as every 9 weeks, and

gestation time is 19-21 days. Larger animals with longer gestation periods are less suited be models for genetic studies, as the generational studies would take years rather than months and be significantly more costly.

Most of the T2D loci implicated by GWAS contain genes with no known biology in relation to diabetes or metabolism. The animal model serves as an invaluable whole-body phenotyping tool to explore the effects these genetic modifications may have on various aspects of metabolism.

3.1.3 The challenges of modeling T2D genetic predispositions in mice

Although the mouse provides a valuable model for diabetic phenotyping, it is not without its drawbacks. One challenge is that not all of the GWA T2D genes have orthologues in mice, and furthermore, even a documented orthologue may have a different expression profile or varying functions in the two species. The uncertainty of the function of a human susceptibility gene in mice makes it imperative to replicate results in humans. An islet phenotype found in mice can be replicated in human islets if access to human tissue is available, or in a human beta cell line such as EndoC-βH1 or EndoC-βH2 (Ravassard et al., 2011; Scharfmann et al., 2014). Whole body phenotypes can also be replicated and confirmed in a cohort of patients.

Modification by genetic background and diet may be important in understanding how a gene works. Another challenge researchers have encountered is genetic differences between laboratory mouse strains. For example, when three groups investigated global knockout mouse models of the GWAS-implicated *Slc30a8* gene, a ZnT8 zinc transporter, they found differences in zinc accumulation in β-cells, insulin secretion, and whole body phenotypes including glucose tolerance which were diet, age, and genetic strain dependent (Nicolson et al., 2009; Lemaire et al., 2009; and Pound et al., 2009). Two of these groups detected normal or enhanced insulin secretion in isolated islets

(Nicolson et al., 2009; Lemaire et al., 2009). The third found normal glucose tolerance, but insulin secretion from isolated islets was impaired (Pound et al., 2009). It has been suggested that this issue could be due to variation in the genetic background of the different mouse models as well as gender differences (Pound et al., 2012). A recent study looking for rare variants in *SLC30A8* in a human population of 150,000 identified 12 protein-truncating variants. Those harboring the truncating mutation had collectively 65% lower risk of developing T2D (Flannick et al., 2014). The finding that disabling the SLC30A8 protein in humans protects against diabetes was not predicted by mouse modeling and further suggests that uncontrolled and confounding variables influenced the results. These obstacles in mouse research are being progressively identified and targeted with standardized techniques and more effective reporting by the National Centre for the Replacement, Refinement, and Reduction of Animals in Research (NC3Rs).

3.1.4 The ENU method of deriving mouse models

N-ethyl-N-nitrosourea (ENU) is a long known chemical mutagen that introduces point mutations randomly into the genome at a rate of 1.5-6 mutations per locus per 1000 offspring dependent upon dosage and strain (Hitotsumachi et al., 1985; reviewed in Justice et al., 1999). Treatment with ENU is an efficient way of inducing mutations into the mouse germ-line. ENU alkylates nucleobases in DNA with a preference for A→T base transversions but with the ability to also make G→C transversions (Nolan et al., 2002). ENU reaches all regions of the genome non-preferentially with the mutation rate corresponding to the dose (Coghill et al., 2002). Therefore, ENU mutations can cause a range of mutant effects from complete or partial-loss of function to gain of function. Large-scale mutagenesis efforts in the mouse have generated hundreds of ENU mouse lines (Nolan et al., 2000). These mutagenesis studies can take either a phenotype or a genotype-driven approach. In the former, mutagenized mice can be scanned for phenotypes of interest and when identified, the altered phenotype can be mapped back to a mutation (Thaung et al., 2002). This method has the advantage that the screens are not biased by pre-conceived ideas about gene function. In the latter approach,

genes of interest can be scanned for mutations and a mouse containing the chosen mutation re-derived from frozen sperm and phenotyped. The random nature of the mutations introduced by ENU allows for the discovery of novel genes or gene functions.

The protocol for creating a mouse line with an ENU mutation is simple and relatively fast. Frozen archives of mutagenized mouse sperm facilitate quick selection of a mutation and re-derivation by IVF. MRC Harwell assembled one such archive of frozen mouse sperm from over 10,000 F_1 ENU-mutagenized mice that is available to scan (http://www.har.mrc.ac.uk/services/dna_archive/). To create the archive, ENU was injected intraperitoneally into male BALB/c or C57BL/6 mice. These males were mated to C3H/HeH wild type females and sperm from the F_1 prodigy was frozen and matched with a DNA sample for mutation scanning (Quwailid, 2004; Coghill, 2003). One way to screen the archive in a region of interest is by amplifying that region using PCR and melting the product in a LightScanner which detects mutations based on the shape of the DNA melting curves.

The probability of finding functional mutations in the archive depends upon the number of base pairs screened. A test scan of 27.4 Mbp of DNA over the exons of 15 genes from the archive across 6000 mice showed a mutation rate of 1 in 1.01 Mbp of which 1 in 1.82 Mbp were coding mutations (Quwailid et al., 2004). An alternative way to estimate mutation rates that takes into account any functional mutations (mutations within a locus capable of causing a phenotype) is to use a specific locus test. The benefit of finding the specific locus mutation rate is to know the likelihood of confounding mutations and be able to take steps to reduce this risk. Thaung et al. (2002) were able to very precisely estimate a specific locus mutation rate for the Harwell ENU archive. DNA available in the archive is the product of mutagenized BALB/c or C57BL/6 male mice crossed with C3H females which harbour the recessive retinal degeneration allele (rd1) of the *Pde6b* gene. In order to find the specific locus mutation rate, the authors scanned the archive for any mutation that could give rise to the retinal degeneration phenotype (when the other allele already carries the recessive rd1 mutation). In 6,500 mice screened, seven alleles were identified at the rd1 locus giving a specific

locus mutation rate of 0.00108. This is to say that ~1/1000 mice will have a functional mutation at any one locus. Using this rate, the number of individuals to screen can be estimated in order to find functional alleles at a specific locus (Figure 3.1).

The F_1 animals carry all of the ENU mutations on the paternal chromosome. Using the rate of potentially functional mutations in coding DNA of 1 in 1.82 Mb calculated by Quwailid et al., (2000) and assuming the amount of coding DNA in a set of chromosomes in a male mouse is 54.89 Mb (Ensembl), each F_1 mouse will have 30 functional mutations. After 5 backcrosses the mice will contain 97% of the recipient genome and harbour approximately 1 functional mutation, therefore diminishing the risk of other functional mutations contributing to a phenotype. Linked confounding mutations present a problem, as a recombination event is necessary to remove the confounding mutation. However, the probability of deriving a mouse with a linked confounding mutation is very small; when the candidate region is 5Mb $p<0.002$ (Keays et al., 2005).

The Harwell ENU archive can be screened for mutations in any amplified region of the genome. Once a mutation is located, the fragment of interest is sequenced and the type of the mutation and its predicted effect can be determined using modelling or predictive genetic software such as PYMOL (14.1, Schrodinger) Polyphen (Adzhubei et al., 2010) and SIFT (Kumar et al., 2009). Live animals containing the mutation of interest are recovered by IVF (a core service at MRC Harwell).

3.1.5 Benefits of using the ENU method to generate mouse models of T2D-associated genes

The inherited component of T2D is complex turning up hundreds of genes related to the pathology of the disease. The majority of alleles linked to T2D are in or near genes previously unstudied in relation to T2D or metabolism. Laboratories around the world are now investigating the implicated genes. There are a variety of approaches used by human geneticists to try and locate the underlying genes, as well as cellular and animal modelling alongside and in conjunction with the human research (see Chapter 1 for details).

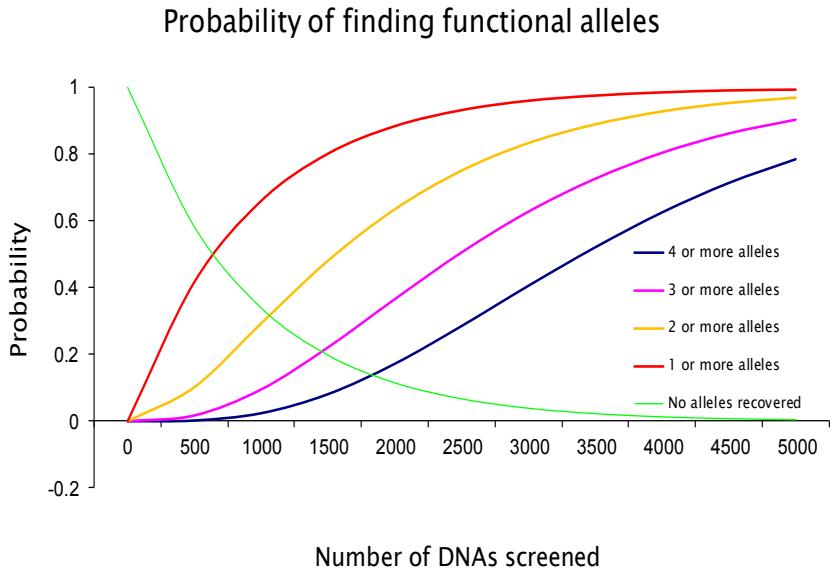


Figure 3.1: The probability of finding functional alleles increases with the number of individuals screened based on an estimated mutation rate of .00108 per locus (adapted from Coghill, E, 2002).

There are several benefits to ENU mutations over knockout models. The first is that knocking out a whole gene can disrupt flanking sequences containing regulatory elements. Introducing a single point mutation poses less risk of unintentionally disrupting a non-coding sequence. Secondly, an ENU allele is sometimes capable of mimicking a natural mutation with a subtler phenotype than a knockout by deletion. Knockouts and over-expression alleles are useful for creating a drastic phenotype, easily detectable and helpful for initial investigation into the function of a gene. However, rarely is a gene silenced by homozygosity of a mutation or by compound heterozygosity in nature as it is in a knockout model. The complete loss of function can be helpful for rapidly determining the function of a gene or testing a predetermined hypothesis, but it is problematic for two reasons: a) when investigating complex genetic diseases such as T2D, a 100% loss of function mutation may be too extreme to reveal the subtle phenotype which contributes to disease pathology, b) a knockout could be lethal.

An example of an ENU mutation in comparison with a knockout mutation is in the fat mass and obesity- associated gene *FTO* gene. Church et al. (2009) found a mutation at *Fto*^{I367F} which is in a highly conserved region at the C terminus of the protein. The mice with this mutation exhibited decreased adiposity and body weight and increased metabolism similar to the knockout mice (Church et al., 2009). However, the weight of *Fto*^{I367F} mice started to diverge after weaning unlike the knockout mice that weighed less on day 2 after birth. The *Fto*^{I367F} mice only weighed 10% less than wild type fully grown, while the knockout mice weighed 30-40% less. In humans, common alleles which map near *FTO* confer up to a 3kg increase in body weight, so the more moderate weight difference in the I367F mice more closely resembles observed human variations (Church et al., 2009; Frayling etl a., 2007). Furthermore, *Fto* knockout mice are shorter than wild type in addition to weighing less whereas the I367F mice showed no difference in body length. Common alleles that map near *FTO* in humans have not been linked to any differences in height, so the I367F mutation is

again more consistent with traits in humans (Church et al., 2009). Analysis of the catalytic activity of the enzyme with a mutation at the C terminus showed activity to be 40% of wild type, and the mutated protein could not dimerise accounting for the moderate in vivo phenotype (Church). The ENU mutant showed an *Fto* phenotype more similar to that seen in humans than the knockout model.

3.1.6 Selection criteria for genes to scan in the ENU archive

I selected these three genes implicated in T2D GWA loci: *Jazf1, Klf14* and *Adcy5* to screen in the Harwell ENU archive based on the following criteria:

1. *Cis*-eQTL data

   A variant associated with expression of a proximal gene in a particular tissue is called a *cis*-eQTL (expression quantitative trait locus). A transcript implicated by *cis*-eQTL data is very likely to be the causal transcript. Most of the T2D GWA genes did not have *cis*-eQTL data when I began this project. The SNPs near *KLF14* were associated with expression of *KLF14* in adipose tissue, which was an important factor in my selection. *ADCY5* and *JAZF1* did not have relevant *cis*-eQTL data when I began my DPhil.

2. Location of associated SNPs in relation to gene

   One of the biggest challenges to attempting to study the SNPs implicated by GWAS studies is determining which gene the SNP is influencing. Often implicated variants are not located within the coding region of a gene. When *cis*-eQTL data are not available, SNP location is often an important factor in predicting the gene of interest. Although not a firm predictor, often SNPs intronic within or nearby the upstream or downstream boundaries of a gene are expected to be within regulatory elements of the proximal gene (Shin et al., 2014). I selected genes where the associated SNPs were either within the gene or within 110 kb upstream of the gene's most extreme transcript start site or 40 kb downstream of the most extreme transcript end site. These boundaries were chosen

because they encompass the 99th percentile of the distances of cis-eQTLs from their associated gene's transcript site (Veyrieras et al., 2008). Reasonable confidence of the correct gene was crucial when selecting genes to investigate.

3. Genome-wide significant association with T2D or fasting glucose

   I selected genes near common variants with p values of association with T2D or FG exceeding the genome-wide threshold value of p value < 5×10^{-8}.

4. Novelty of gene

   Several of the T2D GWAS candidate genes had already been researched in depth and modeled for metabolic traits such as *TCF7L2* and *IRS1* (Villareal et al., 2009; Hart et al., 2010; Bruning et al., 1997; Kido et al., 2000). I selected from the list of T2D GWA loci that had not been studied in the context of metabolism before implication as a T2D-associated locus with the aim of discovering new targets and novel biological mechanisms of the pathophysiology of T2D.

5. Previous knowledge of the gene or family

   Although novelty was an important factor, some previous knowledge of the genes or family of genes was preferred in order to begin with a basic understanding of the biology of that gene. Understanding of the upstream and downstream effectors of the protein or protein family and a logical link to how the gene could be involved in metabolism were factors I considered when selecting transcripts. Many of the SNPs associated with traits such as fasting glucose are near genes not previously implicated in diabetes but with complex roles in various aspects of human biology.

6. Previous knowledge of animal models

   Since my approach involved modeling these genes in mice, I made sure that the genes I chose to study had not already been modeled and fully investigated in vivo for metabolic phenotypes. I did not exclude those genes that had been previously studied in mice but investigated for other traits such as neurological or behavioural traits.

7. Length of gene, number of exons

   The length of the gene and number of exons also had an impact on my selection. Only about 500bp can be amplified and scanned at a time in the ENU archive due to PCR amplification limitations. While I chose only to scan coding regions, a huge gene with 30 or so exons may be prohibitively time consuming to scan. However, if only a few of the exons were in an important catalytic domain or binding site, it would be possible to only scan those exons.

8. Multiple phenotypes in humans

   As mentioned in #5, I favored genes near variants that were associated with multiple traits above those only associated with T2D. Multiple associations could indicate an interesting and multifaceted effect on biology.

9. Orthologue gene in mouse

   A requirement of the animal model approach is that the genes selected have a known orthologue in mice with similar structure and protein function. Differences in expression may still exist and these were accounted for in my results.

Below I describe how I applied these criteria to select the genes studied in this thesis.

3.1.7 Variants at the *JAZF1* locus are associated with T2D and beta cell phenotypes

Researchers from the DIAGRAM consortium discovered SNPs intronic within *JAZF1* to be strongly associated with T2D and impaired beta cell function. SNP rs864745 is associated with T2D with a P value of 5.0×10^{-14} (Zeggini et al, 2008) (Table 3.1). The intronic location of these SNPs within the gene could mean that *JAZF1* is mediating the effect of these variants though this is not at all certain. The *JAZF1* locus is also associated with risk of prostate cancer, stature, and skeletal frame size in humans, but *JAZF1* had never been studied functionally in relation to T2D (Thomas et al., 2008; Frayling et al., 2008; Johansson et al., 2009; Soranzo et al., 2009). Juxtaposed with another zinc finger 1, *JAZF1*, encodes a transcriptional repressor of nuclear receptor subfamily 2, group c,

member 2, *NR2C2,* and a knockout mouse model of *NR2C2* showed low *IGF1* serum levels, hypoglycaemia and growth retardation (Collins et al., 2004; Nakajima et al., 2004). The mechanism by which this transcription factor regulating *NR2C2* could be affecting all of these phenotypes would be an intriguing story to study.

JAZF1 has never been studied in relation to T2D; however, *JAZF1* has been linked with prostate cancer and height/skeletal frame size in humans (Johansson et al., 2009; Thomas et al., 2008). These multiple associations suggest an interesting and multi-faceted biology involving *JAZF1*.

3.1.8 Variants at the *KLF14* locus are associated with T2D and insulin resistance phenotypes

The rs972283 SNP 50kb upstream of *KLF14* was found to be associated with T2D and insulin sensitivity (Voight et al., 2010) (Table 3.1). Although insulin resistance is one of the hallmarks of T2D, only 9 out of about 90 GWAS loci have yet associated with an index of insulin sensitivity (Richards and McCarthy, 2012). Kong et al. (2009) showed that the SNPs at the *KLF14* locus are parent-of-origin specific for association with T2D and insulin sensitivity, meaning their effect is dependent upon whether the SNP was inherited from the mother or father. Their data are consistent with the fact that *KLF14* is an imprinted gene only expressed from the maternal allele (Parker-Katiraee et al., 2007). *Cis*-eQTL data showed the same SNPs implicated in T2D and insulin resistance to be also associated with expression levels of *KLF14* in adipose tissue confirming *KLF14* as the transcript of interest (Voight et al., 2010; Small et al., 2011). This effect is maternal specific for all of the signals which is consistent with the imprinting pattern of *KLF14* (Kong et al., 2009).

Teslovich et al. (2010) published an association between the same variants at the *KLF14* locus and blood cholesterol levels, further connecting this gene with lipid biology. Researchers using data from the MUTHER consortium found at least 40 genes associated in trans with *KLF14* in adipose (Small et al., 2011) which has been increased to 140 by new RNA-Seq data (unpublished, 2014). The relationship between *KLF14* and these genes remains to be demonstrated functionally, but a likely

explanation is that *KLF14* is regulating these genes in adipose. The *KLF14* region also harbours signals for basal cell carcinoma implicating it as an important transcript with many effects on human biology (Stacey et al., 2009). This progression of evidence establishes *KLF14* as a key player in the regulation of adipose gene expression and in part responsible for downstream mechanisms that connect adipocyte dysfunction to peripheral insulin resistance and T2D.

3.1.9 Variants at the *ADCY5* locus are associated with T2D and beta cell phenotypes

Variants at the *ADCY5* locus have multiple effects on human biology including impacting fasting glucose levels, birth weight, and T2D risk (Dupuis et al., 2009; Freathy et al., 2010). The leading T2D SNP rs11708067 is within intron 3 of the *ADCY5* gene. The A allele is associated with a higher risk of T2D (OR: 1.12[95%CI: 1.04-1.15]; P=9.9 x 10^{-21}) and higher fasting glucose (0.027mmol/l [95%CI:0.021-0.033]; P= 7.1 x 10^{-22})(Dupuis et al., 2009). Another allele from the same signal is associated with lower birth weight, and this allele is in LD with the A-allele associated with T2D and fasting glucose (Freathy et al., 2010). These multiple associations suggest a complex effect on physiology (Table 3.1).

Of the genes in the associated region, *ADCY5* is the most promising biological candidate for a T2D gene because it is implicated in two pathways leading to insulin production in the beta cell. *ADCY5* codes for an adenylate cyclase responsible for catalyzing the generation of cAMP. In the beta cell, cAMP activates protein kinase A (PKA) which potentiates proinsulin formation and eventually insulin production. The *Epac* signaling pathway in beta/alpha cells could also be impacted by *ADCY5* catalysis of cAMP. Epac proteins are cAMP-regulated guanine nucleotide exchange factors. They are activated by cAMP and assist Ca^{2+} signaling and insulin secretion in the beta cell. Impaired adenylate cyclase function in the beta cell could affect either or both the PKA and *Epac* signaling pathways.

	JAZF1	***KLF14***	***ADCY5***
Location of T2D-associated SNPs	Intronic in *JAZF1*	50 kb upstream	Intronic in *ADCY5*
Lead SNP	rs849134	rs972283	rs11708067
Association with T2D (P value)	5.0×10^{-14}	2.2×10^{-10}	8.4×10^{-22}
Association with FG (P value)	0.08	0.003	7.1×10^{-22}
Effect Size for T2D	1.10	1.07	1.12
Cis-eQTL associations	In adipose	In adipose	None
Novelty	yes	yes	yes
Previous knowledge of gene	Transcription factor	Transcription factor regulates TGF-β, imprinted gene	Adenylate cyclase activates the PKA and Epac pathways
Previous modelling in mouse	None	None	Coordination and balance
Length of gene	5 exons	1 exon	21 exons
Phenotypes in humans	T2D, beta cell function, prostate cancer, height	T2D, insulin sensitivity, blood cholesterol, basal cell carcinoma	T2D, fasting glucose, birth weight, beta cell function
Orthologue in mouse	yes	yes	yes

Table 3.1: Summary of selection criteria for genes examined in this thesis. *JAZF1, KLF14*, and *ADCY5* met the conditions for strong association with disease and interesting biology to qualify for investigation in mice. Information from the table is from Dupuis et al., 2010; Voight et al., 2010; Zeggini et al., 2008; Johansson et al., 2009.

At the time that I began my thesis*, ADCY5* was a novel gene to study in relation to diabetes. Although researchers had studied *ADCY5* in relation to heart disease, it had not been functionally characterized previously in relation to islet function. (Hu et al., 2009; Okumura et al., 2008).

The basic biology and structure of the ADCY5 protein is quite well understood facilitating interpretation of functional results.

3.1.10 Experimental Aims

The objective of this chapter is to design and derive ENU mouse models of variation in the three selected genes *Jazf1, Klf14* and *Adcy5* in order to investigate the roles of these genes in the pathophysiology of T2D. Three specific aims were necessary to achieve this: a) select amplicons to scan for mutations in the ENU archive for *Jazf1, Klf14* and *Adcy5,* b) scan the archive for mutations in the selected amplicons, and c) evaluate the mutations and select promising candidates to re-derive as live mice and phenotype.

3.2 Samples and Methods

3.2.1 Determination of amplicons for ENU archive screen

To optimize my chances of finding functional alleles in my three genes of interest, I chose to scan about 3,300 bp over a total of 3 genes in 9,696 individual DNAs. A scan of that size would be expected to yield between 28-30 alleles based on the ENU locus mutation rate of 0.00108, 15-18 of which would be functional, according to the predicted rate of functional mutations of 1 in 1.82Mbp. Therefore I would have a selection of several functional mutations in each gene. Amplicons were selected for the genes of interest within and flanking coding regions. Regions of catalytic activity, DNA binding, and other areas crucial for protein function were targeted. PCR was performed to

amplify the regions selected regions in 9,696 mice in the Harwell ENU-DNA archive (http://www.har.mrc.ac.uk/services/dna_archive/).

Jazf1 is a five-exon transcription factor containing a zinc finger. The binding site of the zinc finger is contained within exon 5. I chose to screen only exon 5 because this was the most conserved part of the gene and the most crucial for protein function. I was able to scan exon 5 in one amplicon of 276 base pairs.

Adcy5 is a 21 exon adenylate cyclase containing two catalytic domains. I chose to scan all of the exons that make up the catalytic domains. The remaining parts of the protein are low complexity domains or short transmembrane domains responsible for anchoring the protein to the membrane; a mutation in one of these domains may be less likely to affect protein function because these domains are less conserved across species (Section 3.3.3 and Figure 3.9 demonstrate species alignment). I scanned a total of seven amplicons in seven exons (exons 2, 3, 4, 5, 14, 15 and 18) of *Adcy5* totaling 2,195 base pairs.

Klf14 is a one-exon gene encoding a transcription factor containing three zinc finger domains. I chose to scan the whole exon in two amplicons of 402 and 417 base pairs.

3.2.2 Screening the ENU archive

The process of screening the Harwell ENU archive went as follows: a) primers were designed to amplify the selected sequences, b) PCR was performed to amplify the fragments and integrate LCGreen Dye into the DNA, c) PCR products were loaded onto a LightScanner and heated from 60°C - 98°C to dissociate the DNA strands and record fluorescence. Mutations were detected by shifts in the melting curve to a lower fluorescence. A mismatch in a heteroduplex causes the DNA to denature at a lower temperature. Denaturation releases the dye and lowers fluorescence (Figures 3.2 and 3.3). Mouse DNAs were scanned in pools of 4 mice. Following detection of a mutation in a

DNA pool, the individual DNAs were then tested alongside at least 12 wild type DNA controls. Final PCR products from individual DNAs containing a mutation were purified and sequenced.

Primers for all amplicons were designed using Primer Select. Primer pairs were selected with a size between 20-23 bp, an annealing temperature of 58-60^{0}C, a guanine and cytosine (GC) content of 40-60%, and location outside polymorphic sites. The primers used to screen the amplicons of *Jazf1, Klf14* and *Adcy5* can be found in Table 3.2.

The PCR protocol can be found in Chapter 2 section 2.2.1(b[i]) and 2.2.1(c). DNA purification and sequencing protocols can be found in Chapter 2 sections 2.2.2 and 2.2.5 respectively.

3.2.3 Sequence Alignment

All genomic transcripts and peptide sequences were exported from Ensembl (http://www.ensembl.org/index.html). The sequence alignments of *ADCY5* and *KLF14* peptide sequences were performed using Vector NTI advance 10-AlignX software (Invitrogen) and by BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&BLAST_SPEC=blast2seq&LINK_LOC=blasttab).

3.2.4 Analysis of Variants

Variants found were modeled in PYMOL (14.1, Schrodinger) Polyphen (Adzhubei et al., 2010) and SIFT (Kumar et al., 2009) to predict the effect of the mutation on protein function.

Gene	Primer Name	Sequence 5'-3'
Jazf1		
	Jazf1_exon5_F	GACAGGACATTTGGTGACCC
	Jazf1_exon5_R	TCAGCAACTGCTGGTGAGG
Adcy5		
	Adcy5_exon2_F	GGGTGTGAACTTGCTAAGATGC
	Adcy5_exon2_R	CCGTTTGCTTGTGTGTCTG
	Adcy5_exon3_F	ATATACAGGTTCCGTGGGTCTAAG
	Adcy5_ exon3_R	CATTAGGAGGAAGAAGAGAGGCT
	Adcy5_exon4_F	CCTTACAACAGACATTAGGATTT
	Adcy5_exon4_R	CTAGATTACCGTCCTCAACC
	Adcy5_exon5_F	AGTGACTCTCTGTTTCTCATTT
	Adcy5_exon5_R	GAAGACAGTGGCTTGGTTC
	Adcy5_exon14_F	GCTTGGCTTTTAGGGGCTG
	Adcy5_exon14_R	AATGTCTGCCTGCCTCTGCTG
	Adcy5_exon15_F	GGTGAGGAGGGGCCAGTAA
	Adcy5_exon15_R	CCTAGGGAGACGGAAGACATTT
	Adcy5_exon18_F	TTGGCACCCTGAGTGTTTC
	Adcy5_exon18_R	ATCAGAGCCTCTGGCAAAT
Klf14		
	Klf14_amp1_F	CGAGCGCTTTGGCTGATTTA
	Klf14_amp1_R	TGTGGGTACGCTGGTGTGA
	Klf14_amp2_F	ATGGCTGCAACAAGGCCTA
	Klf14_amp2_R	GGTTGTGCTTGAGGGAAAGAT

Table 3.2: Primer pairs for amplicons screened in Harwell ENU-DNA archive.

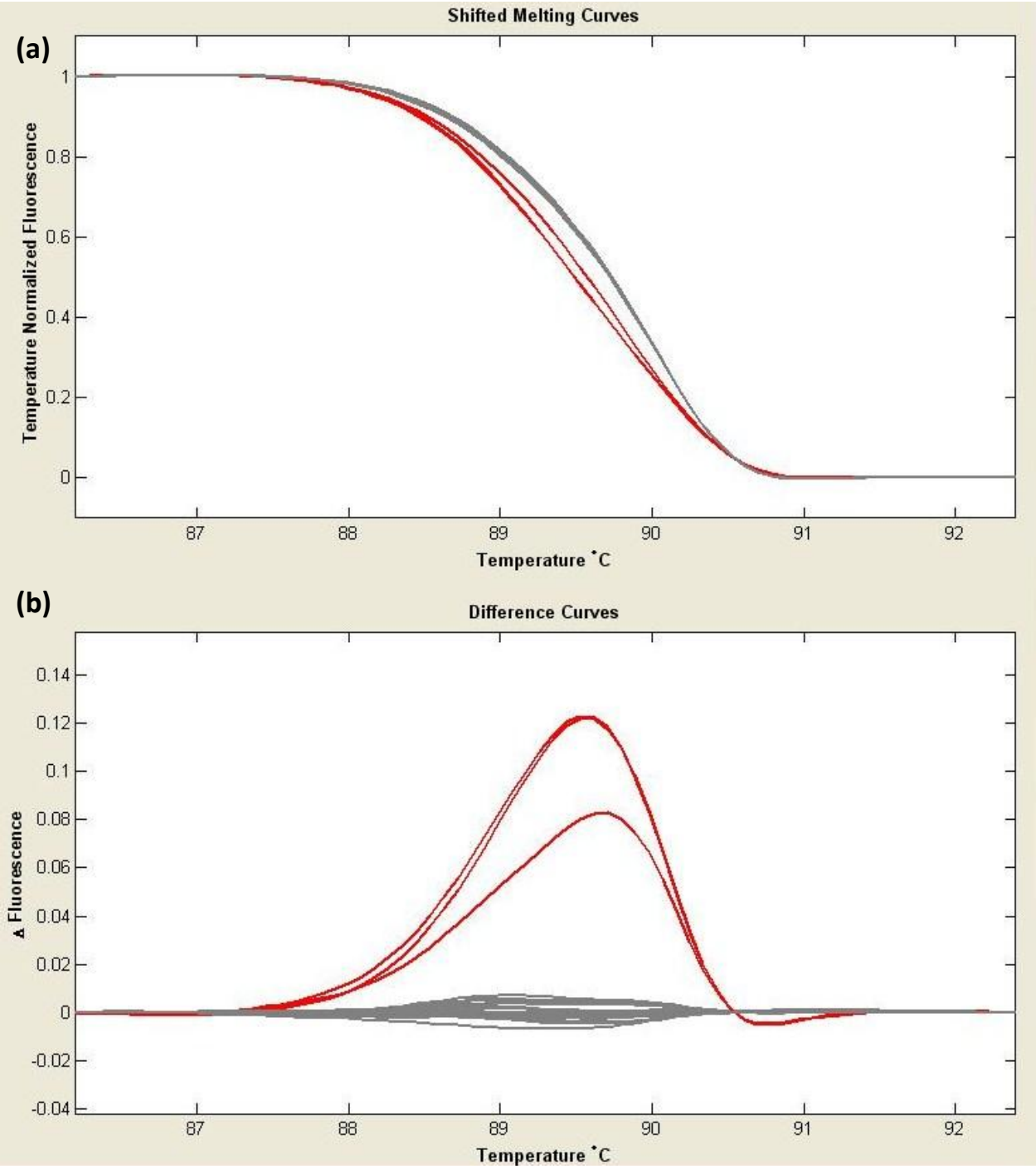


Figure 3.2: Both panels (a) and (b) show a fragment from *Adcy5* exon 3 from 32 mice melted on the LightScanner. Three mutations are detected (in red). The top panel (a) shows the mutations detected as a shift in fluorescence from the other DNA fragments (grey). The bottom panel (b) shows the same data expressed as delta (Δ) fluorescence.

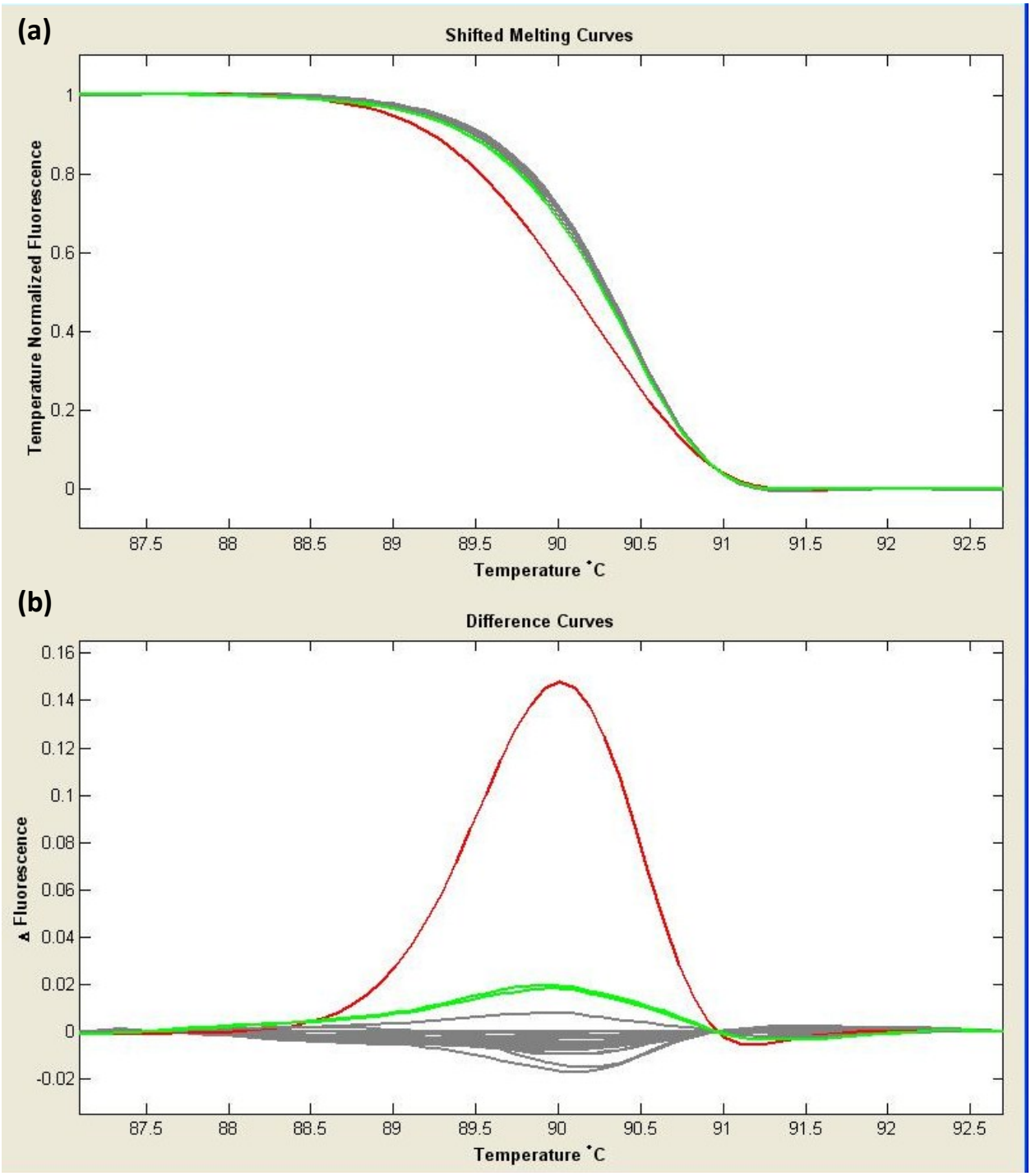


Figure 3.3: Both panels (a) and (b) shows a fragment from *Adcy5* exon 4 in 96 pools of 4 mice melted on the LightScanner. The top panel (a) shows the distinct shift in fluorescence of one strand containing a mutation (in red), and several strands with moderately lower fluorescence (in green). The bottom panel (b) shows the same data as delta (Δ) fluorescence. The ambiguous fragments were re-PCRed with individual DNA and re-melted for verification.

3.2.5 Re-derivation of ENU mutant mice

After selection, C57BL/6J x C3H/HeH F1 (in the case of *Adcy5* Y1064C) and BALB/C x C3H/HeH F1 (in the case of *Klf14* R238L) were re-derived by *in vitro* fertilization (IVF) from frozen sperm using C3H/HeH eggs. The strain difference arose from the fact that earlier ENU-treated mice had been of the BALB/c strain and the later additions to the archive were C57BL/6J mutagenized males.

The IVF was performed by Frozen Embryo and Sperm Archive department (FESA; http:/www.har.mrc.ac.uk/services/fesa/, at the Mary Lyon Centre MRC Harwell; Fray, 2009a; Fray, 2009b).

The first generation IVF mice contain about one mutation per two million base pairs in addition to the targeted mutation. To remove confounding mutations, the IVF progeny were backcrossed 3 times to wild type C3H/HeH mice resulting in a BC3 generation which is only 6.25% of the original ENU treated mouse and unlikely to harbour a confounding ENU functional mutation based on the specific mutation rate (see section 3.1.5). Further backcrosses were completed as time would allow, but BC3 was determined to be the first generation suitable for inter-crossing based on a balance of risk vs. resources. Inter-crosses were then bred from the BC3 generation and cohorts derived of male and female mice carrying zero, one or two copies of the mutant allele. All breeding and colony maintenance was done by the MRC staff at the Mary Lyons Centre, Harwell.

3.3 Results

3.3.1 *JAZF1*

I found just one SNP, S151G, in my screen of exon five of *Jazf1*. According to the mutation rate of 1 functional mutation in 1.82Mb coding DNA scanned across 9696 mice, scanning 276 base pairs

would yield 2.7[95%CI: 2.5-2.8], so my discovery of one mutation is below expected yield. Modeling in PYMOL predicted this mutation to be benign, as did Polyphen. I did not re-derive any *Jazf1* mutants and do not plan to do any further work on *JAZF1*.

3.3.2 *KLF14*

Amplicon 1 yielded zero mutations. I found four mutations in Amplicon 2 of the *Klf14* exon: D231A, R238L, M284K, and R287R which is a silent mutation (Table 3.3). I would expect to find 5.29[95%CI: 5.03-5.54] mutations in the Amplicon 1 of 555 bp, and 4.2[95%CI: 4.0-4.5] mutations in the second amplicon of 493 bp based on the estimated mutation rate of 1 potentially functional mutation in 1.82Mb coding DNA across 9696 mice. The 4 mutations found in Amplicon 2 were within the expected outcome, but Amplicon 1 yielded 5 fewer mutations than expected. Although I scanned mainly coding DNA, the amplicons did extend into the flanking regions for optimal primer binding purposes. Therefore, I can expect my yield of functional mutations to be slightly lower than that predicted by the mutation rate in coding DNA.

The region containing the four mutations in Amplicon 2 is highly conserved, and all three missense mutations change amino acids that are conserved across multiple species (Figure 3.4). The biology of the amino-acid chains was also taken into account for mutation predictions. For instance, a small amino acid may be substituted for another small or nucleophilic amino acid often without much disruption to the protein. However, if an amino acid with a hydrophobic side chain is inserted, the folding of the protein may be altered substantially (Charton and Charton, 1982). For the missense mutations in KLF14, all of the amino acid substitutions could alter protein folding as the biology of the side chains change for each mutation. R238L is especially likely to cause disruption as the hydrophobic amino acid leucine is replaced with a large basic amino acid, arginine that is hydrophilic.

Gene	Exon	Coding Region	Animals Screened	# Mutations Found	Mutation Protein Changes
Klf14	amp1	555	9696	0	0
	amp2	493	9696	4	D231A, R238L, M284K, R287R (silent)

Table 3.3: Mutations found from screening the single exon gene, *Klf14.*



Figure 3.4: All of the *Klf14* missense mutations detected in the ENU archive fall in a highly conserved part of the protein. The three amino acids are conserved in human, macaque monkey, mouse, and rat. R238 is not conserved in dog where it is P238. D231 and R238 fall within a zinc finger domain. The M284 amino acid is not within an identified functional domain.

The PAM (Point Accepted Mutation) matrix is a bioinformatics tool which predicts how well one amino acid can perform the functions of another based on how often one amino acid replaces another in nature (Dayhoff et al., 1978). According to the PAM matrix, alanine (A) replaces aspartic acid (D) and lysine (K) replaces methionine (M) about as often as chance. However, leucine (L) replaces arginine (R) 3 times less often than chance. The lower frequency of this replacement suggests that this mutation may be deleterious.

I used the program PyMOL (14.1, Schrodinger) to model the effect each mutation might have on protein structure. R238L stood out as the most disruptive mutation. PyMOL modeling did not show disruption of the protein from the M284K mutation (Figure 3.5).

Modeling of the D231A mutation showed a possible disruption of an alpha helix downstream of the mutation (Figure 3.5). However, D231 has no polar bonds and change to an alanine does not form any bonds (Figure 3.6).

Conversely, modeling the R238L mutation showed a loss of a polar bond involved in DNA binding and suggested that changing this arginine to a leucine would inhibit the third zinc finger of *Klf14* from binding DNA (Figure 3.7).

Mutation prediction software Polyphen (Adzhubei et al., 2010) and SIFT (Kumar et al., 2009) confirmed the probability of damage that this mutation would cause, while Polyphen and SIFT predicted that the D231A mutation would be benign and the M284K mutation would be possibly damaging. Taking into account the evidence from PyMOL, Polyphen, SIFT, species alignment, and biological principles, I concluded that the R238L mutation would most likely disrupt DNA binding and therefore be a likely model of *Klf14* dysfunction. I decided to re-derive a mouse line carrying the *Klf14* R238L mutation.

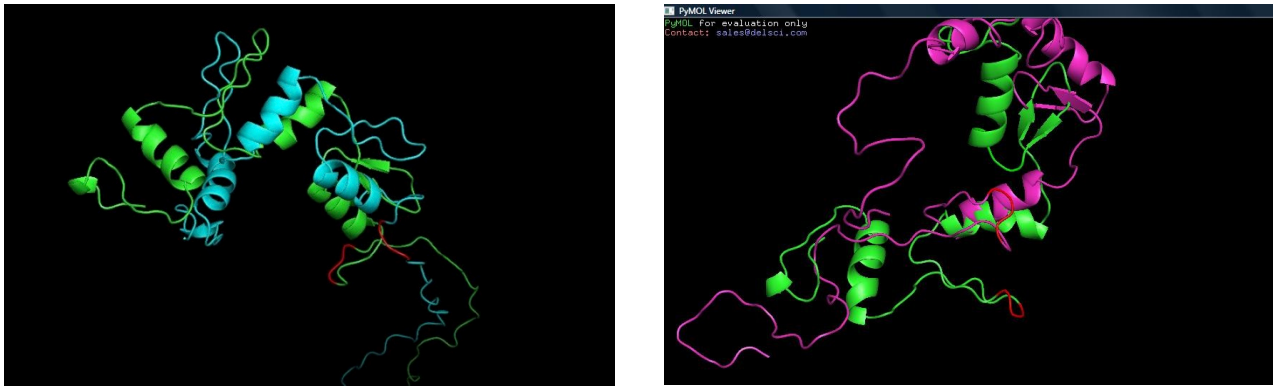


Figure 3.5: Modelling in PyMOL shows the M284K mutation (left, wild type in green, mutant in blue) does not appear to disrupt protein folding. The D231A mutation (right, wild type green, mutant purple) may disrupt an alpha helix present in wild type but not in the mutant.

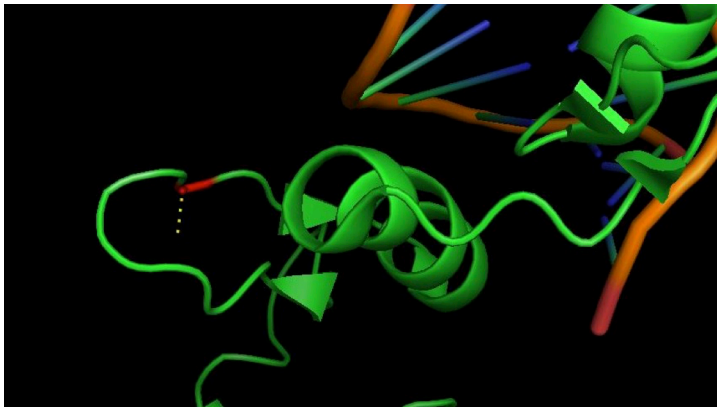


Figure 3.6: Further modelling of the D231 residue demonstrates that this amino acid (highlighted in red) makes no polar bonds and therefore the D231A mutation may not affect protein folding in a dramatic way.

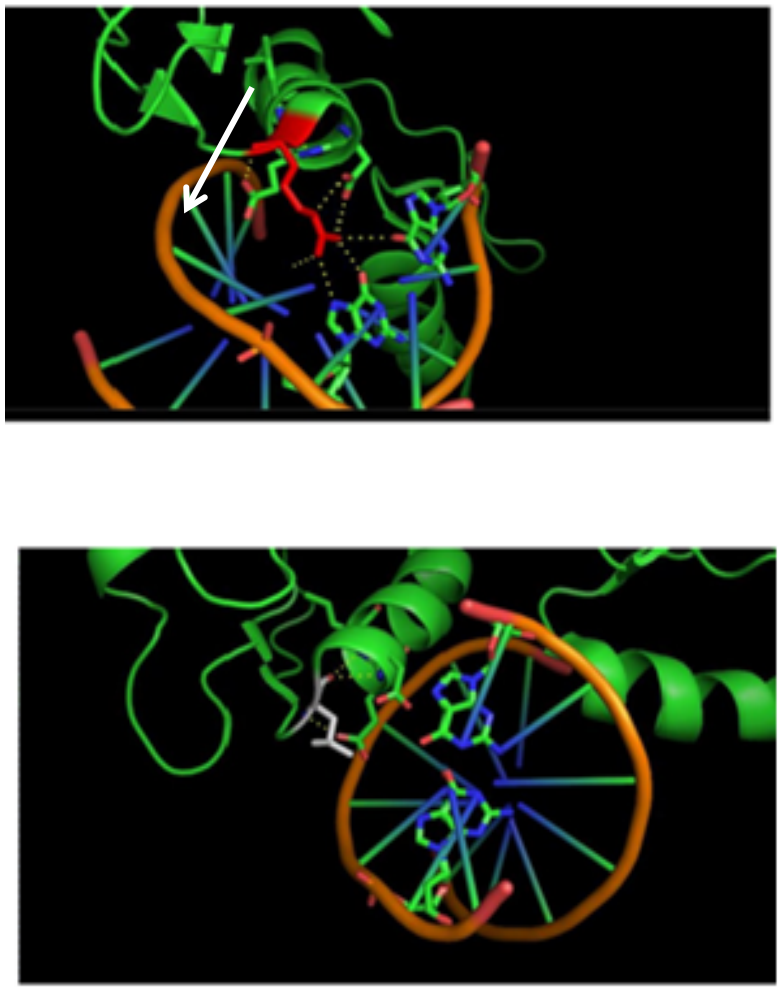


Figure 3.7: *Klf14* R238 is located in the binding region of one of the zinc fingers of this transcription factor. The arginine (in red) would normally form several polar bonds to bind DNA (top). However, when the arginine 238 residue is changed to a leucine (shown in white), the bonds are broken and DNA binding is disrupted (bottom).

3.3.3 *ADCY5*

The exons I screened in *Adcy5* fell within the cytoplasmic domains of the protein which are catalytically active (Figure 3.8). I scanned the majority of exons 2, 3, 4, 5, 14, 15, and 18. Overall, I found 15 mutations in 2,195 base pairs. According to the specific mutation rate of 1 potentially functional mutation in 1.82Mb of DNA across 9696 mice, the expected yield was 21.3[95%CI: 20.1-22.3] mutations, around 6 mutations above actual yield. While I attempted to scan as much coding DNA as possible, 8 of the mutations I found were in the intronic regions bordering the exons I chose. For these mutations, I used online software such as Ensembl genome browser to find the coordinates of promoters, splice sites, and splice regulatory sites. None of the intronic mutations I found fell within a splice site or known promoter region. I found a total of 7 coding mutations in seven separate mice of which 1 was silent (see Table 3.4). To better understand how these mutations might disrupt protein function, I (a) studied the biology of the amino acid substitutions, (b) used PyMol to model the wild type ADCY5 mouse protein and then any changes that could occur when the mutation was present, (c) also consulted Polyphen and SIFT for further analysis of the effect of the mutations, and finally (d) made alignments of the *Adcy5* gene in several species and compared how conserved the mutated residue was between species.

The biology of the amino acid changes in all of the missense mutations found except for L1054M in which a hydrophobic amino acid is exchanged for another hydrophobic amino acid. The most likely substitution to affect biology is the Y1064C mutation as a cysteine is capable of forming very strong di-sulphide bonds that a tyrosine cannot form. The PAM matrix predicted M543K to be the most disruptive mutation with a score of -4. The D533Y and Y1064C mutations were each scored as 0 by the PAM matrix (Table 3.5). Both the *Adcy5* exon 5 region and the exon 18 region are highly conserved across species (Figures 3.9 and 3.10). All mutations are conserved across mammalian species and chicken, and the 3 mutations in exon 18 were conserved in Xenopus as well.

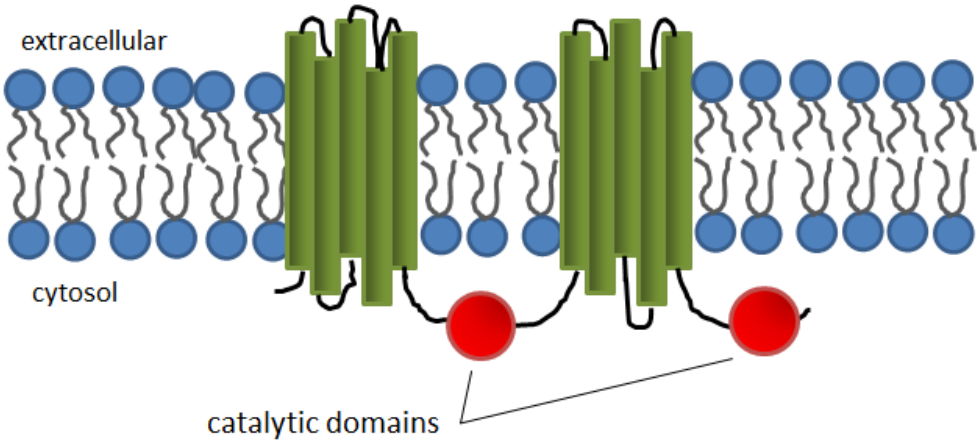


Figure 3.8: The structure of ADCY5. The cytoplasmic domains are catalytically active. All of the coding mutations found fell within the two catalytic domains. (Image adapted from http://www.vivo.colostate.edu/hbooks/molecules/cyclase.html)

Modeling in PyMOL showed few protein structure changes in the D533Y, M543K, G1094S, and L1054M mutations (Figures 3.11 and 3.12). However, G1094S modifies polar contacts in the region at the start of an alpha helix. Disruption to the alpha helix could affect protein function (Figure 3.13). Y1064C also disrupts polar contacts and appears to be damaging according to PyMOL modeling (Figure 3.14). Polyphen and SIFT mutation prediction programs were inconclusive for D533Y and M543K (Table 3.6). Both Polyphen and SIFT predicted mutations G1094S and Y1064C to be probably damaging, but the Y1064C mutation had the highest PSIC (Position-Specific Independent Counts) score difference (difference of scores calculated for each amino acid based on how conserved it is and structural predictions). When the scores are the most different for each amino acid, then the mutation is more likely to be damaging. Polyphen predicted the Y1064C mutation to be the most likely damaging of all the mutations.

When all of the data were consolidated, the most interesting mutation was Y1064C in exon 18 of *Adcy5*. Y1064C is part of a highly conserved block of amino acids near the end of the protein. The tyrosine residue at position 1064 forms two polar bonds in the wild type protein. When replaced by a cysteine, at least one of these bonds is broken, which appears to disrupt the formation of a beta sheet (see Figure 14). Both Polyphen and SIFT place this mutation in their “most damaging” category. I chose to re-derive the mouse carrying the *Adcy5* Y1064C mutation from the archive.

Gene	exon	Coding region	Animals screened	# mutation found	Mutations
Adcy5	2	233	9696	0	0
	3	152	9696	3	All intronic (not splice sites)
	4	183	9696	3	2 intronic (not splice sites), T421T(silent)
	5	121	9696	2	D533Y, M543K
	14	152	9696	3	2 intronic, S885P
	15	166	9696	1	intronic
	18	254	9696	3	G1094S, Y1064C, L1054M

Table 3.3: Mutations in *Adcy5* from the Harwell ENU mouse DNA archive.

Mutation	PAM score
D533Y	0
M543K	-4
S885P	1
G1094S	1
Y1064C	0
L1054M	4

Table 3.4: *Adcy5* mutation PAM matrix scores. Scores from PAM table (Appendix 1) showing how often the second amino acid replaces the first in nature.



Figure 3.9: Species alignments for *Adcy5* exon 5 mutations. D533 and M543 are conserved in all mammals shown as well as zebra fish. Both residues fall within a catalytic domain of the protein.



Figure 3.10: *Adcy5* exon 18 mutation alignments. L1054, Y1064, and G1094 are conserved in all species shown and fall within the second catalytic domain of the protein.

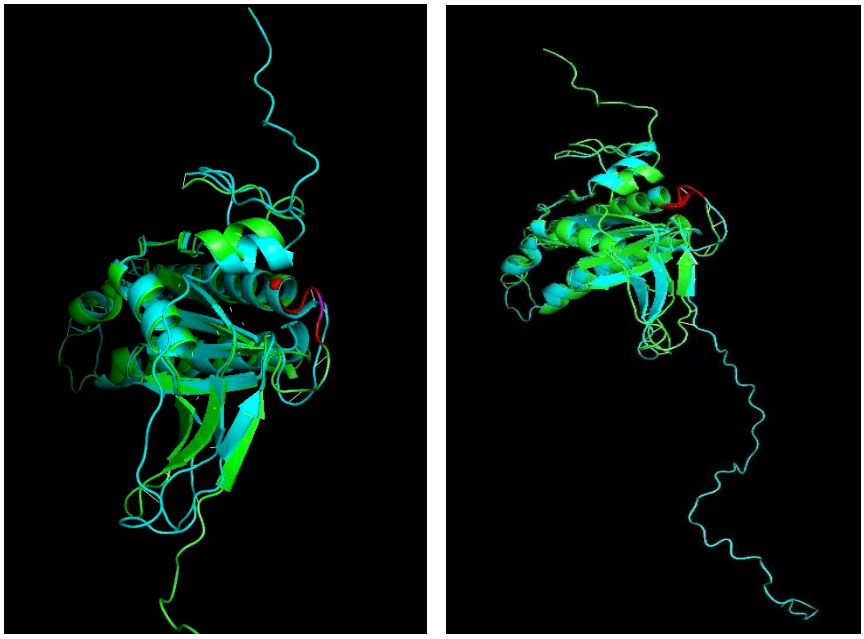


Figure 3.11: PyMOL modeling showed no changes in protein structure for *Adcy5* D533Y (left, wild type shown in green, mutant shown in blue) or for M543K (right, wild type green, mutant blue).

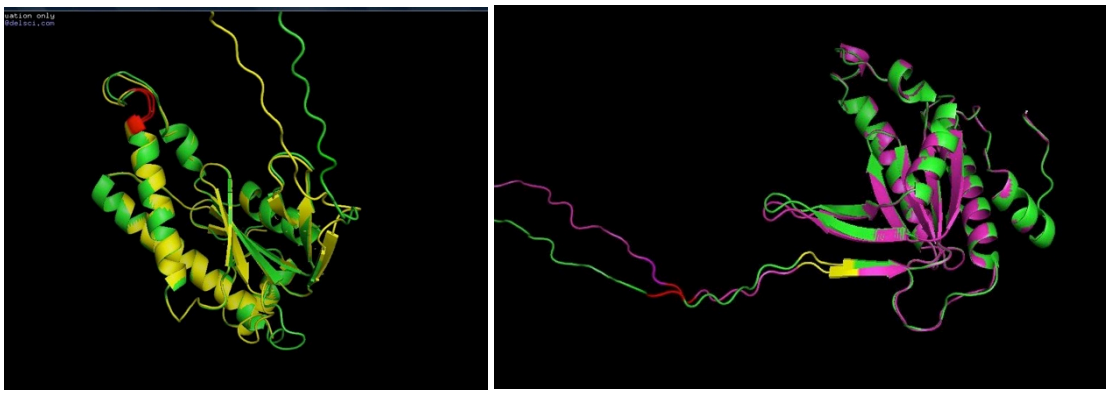


Figure 3.12: PyMOL modelling showed no change in protein structure for the G1094S mutation (left, wild type shown in green, mutant in yellow). The L1054M mutation appeared to possibly lengthen a beta sheet (right, wild type green, mutant purple, yellow highlighted region possibly changed).

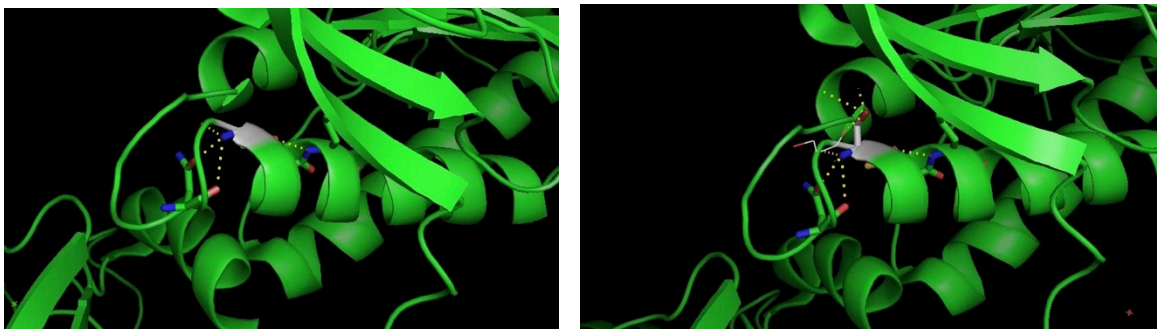


Figure 3.13: *Adcy5* G1094 polar contacts in wild type (above) and then in G1094S mutant (below). The G1094S mutation adds a polar contact which is not present in wild type.

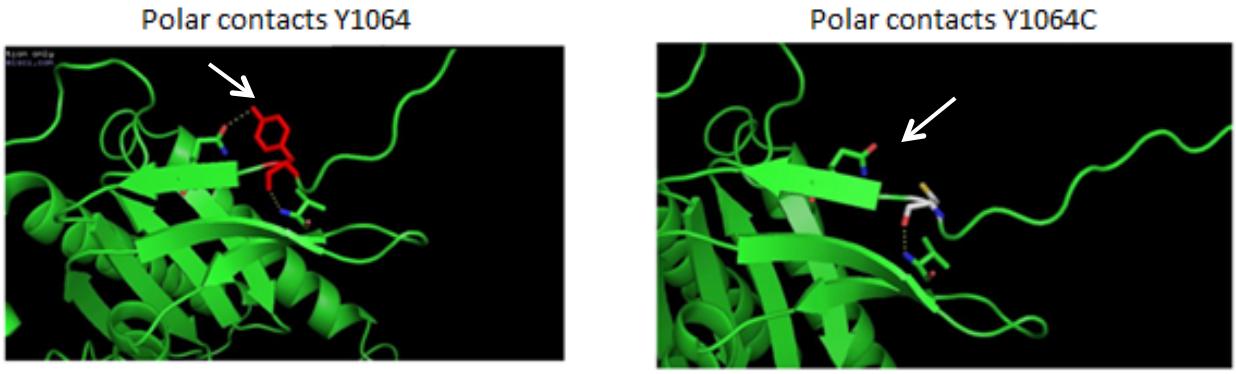


Figure 3.14: PyMOL modelling of Y1064 and Y1064C. One polar bond is broken when Y is changed to C (signalled by white arrows), and this appears to de-stabilize an alpha helix.

Mutation	Polyphen Prediction	SIFT prediction
D533Y	Probably damaging	No effect on protein function
M543K	Possibly damaging	No effect on protein function
S885P	Benign	No effect on protein function
L1054M	Benign	No effect on protein function
Y1064C	Probably damaging	Affect protein function
G1094S	Probably damaging	Affect protein function

Table 3.6: Amino acid substitution analysis results from Polyphen and SIFT. The programs agreed that the Y1064C and G1094S mutations would be damaging to protein function.

3.4 Discussion

The ENU archive at the MRC Harwell provided me with the opportunity to identify functional mutations in my genes of interest. Rather than making a knockout model, I chose to create mouse models with ENU point mutations that more closely resemble naturally occurring mutations found in humans. The major risk of the ENU approach was that the mutations recovered are not guaranteed to be functional or cause a measurable phenotype in the animal.

I selected *JAZF1, ADCY5*, and *KLF14* from the list of genome-wide associated T2D genes based on an extensive list of selection criteria described in section 3.1.7 of the introduction to this chapter. I chose amplicons coding for critical catalytic or binding domains in each of these genes and scanned for mutations in the MRC Harwell ENU archive. The estimated mutation rate of 1 functional mutation per 1.82Mb coding DNA scanned predicted I would find 32.0[95%CI: 30.4-33.6] in the 3,300 bp across 9696 mice that I scanned. I found a total of 20 mutations, out of which 11 were functional. This falls short of my predicted mutation rate and predicted functional mutation rate by about 65%. However, although I attempted to scan mainly coding regions, I scanned ~500bp of intronic regions due to the location of optimal primer annealing sites and to detect any splice site variants. Therefore, the 1 in 1.82 functional mutation rate should only be applied to around 2,800bp in which I would expect to find ~27 functional mutations. The unexpectedly low number of mutations I found may be due to chance, an anomaly in the genes I scanned, or the hit rate may be inaccurate. It could also be that mutations were missed when scanning. Mutations could have been overlooked if the PCR did not amplify the DNA consistently and effectively, or if the DNA was melted at too high or low temperature on the LightScanner. I found several disruptive mutations in *Klf14* and *Adcy5*, but no promising mutations in *Jazf1*, so I chose to discontinue work on this gene.

I found a mutation at *Adcy5* Y1064C in the ENU archive, which I predicted would change the structure of the ADCY5 protein and affect its function. The Y1064C mutation falls in the active catalytic domain of *Adcy5*. The tyrosine 1064 residue is highly conserved across a range of species, and modeling in PyMOL predicted possible de-stabilization of the protein when this residue is changed. This mutation was predicted probably damaging by Polyphen and SIFT software. I determined that the Y1064C mutation was most likely of the mutations located in *Adcy5* to affect function of the protein.

The IVF team at the MRC Harwell completed the re-derivation. The mutagenized sperm from the mouse carrying the Y1064C mutation was inseminated into a wild type C3H/HeH female and 7 pups resulted, 3 carrying the mutation. These F1 heterozygotes were crossed with wild type C3H/HeH mice and 3 more backcrosses completed before intercrossing to produce the full range of genotypes. The mutation did not result in an embryonic lethal phenotype, and the pups appeared normal. I expect to see altered function of ADCY5 in these mice likely leading to a metabolic phenotype. Whether this mutation is a gain or loss of function cannot yet be predicted.

Variants near *KLF14* associate with T2D and insulin resistance. *Cis*-eQTL data as well as cholesterol data, parent of origin data, and *trans*-associations data has confirmed *KLF14* as the gene of interest. *KLF14* appears to be a master-regulator of genes in adipose and a link from adipose to insulin resistance. I found a mutation in the ENU archive which appears to disrupt DNA binding in one of the zinc fingers of *Klf14.* The one exon gene *Klf14* is highly conserved throughout, so all of the 3 functional mutations I found in the latter part of the exon were conserved across multiple species. Two of the mutations modelled in PyMOL showed no change in protein structure or polar bonds. The R238L mutation appeared to disrupt a bond between the zinc finger domain of the protein and a DNA molecule, possibly damaging the ability of the protein to bind DNA and inhibit or enhance transcription. PolyPhen and SIFT also predicted this mutation to be the most damaging of the 3.

The re-derivation of R238L was problematic. There were no pups from the R238L IVF after two attempts. To bypass the problem of defective sperm, the IVF team at MRC Harwell attempted laser drilling into the zona pellucida and injected the ENU sperm directly, but this yielded only five pups, none of which harbored the mutation. Subsequently, a sample of sperm was sent to collaborators in Japan, Dr Naomi Nakagata and team at Kumamoto University, who performed intracytoplasmic sperm injection (ICSI) and attempted to recover live animals. They were successful and sent 16 live pups from Japan. Genotyping confirmed the presence of 10 heterozygotes for the R238L mutation in these litters. The mice arrived at the MRC Harwell where they spent almost a year in quarantine and backcrossing to remove confounding mutations. Unfortunately, this wait time ruled out the *Klf14* R238L mice from being a part of my thesis and I prioritized other experiments.

During the difficulty with the re-derivation, the International Knockout Mouse Consortium (KOMP) created a knockout model of *Klf14* (Project ID VG11504, *Klf14*$^{tm1(KOMP)Vlcg}$). The knockout is a simple deletion with *lacZ* replacement allele. The knockout allele was integrated into embryonic stem cells by homologous recombination deleting both copies of the single exon *Klf14* gene. The knockout model has potential drawbacks as removing an entire gene could disrupt flanking regions and regulatory sequences. However, *KLF14* has repeatedly turned up in genetic association studies recently peaking interest in this gene – a knockout model may be able to help us shed some light on the *KLF14* adipose phenotype alongside the ENU model. The knockout model was ordered in the form of cryogenically frozen embryos from the KOMP repository, Mouse Biology Program at University of California, Davis.

Chapter 4

Characterization *in vivo* and *in vitro* of the *Adcy5*^{Y1064C} mouse model

4.1 Introduction

4.1.1 GWAS signals near ADCY5 associate with type 2 diabetes

Variants intronic within *ADCY5* are associated with T2D, beta cell function, birth weight, and fasting glucose indicating an interesting and complex mechanism of action affecting multiple aspects of human biology (Dupuis et al., 2009; Freathy et al., 2009). *ADCY5* is a promising biological candidate for involvement in T2D pathology because it is involved in two pathways leading to insulin secretion in the beta cell and glucagon secretion in the alpha cell (see figure 4.1). Adenylate cyclases catalyze the generation of cAMP which activates the protein-kinase A (PKA) pathway and the exchange proteins activated by cAMP (Epac) pathway. Both pathways lead to insulin secretion in the beta cell or glucagon secretion in the alpha cell: the PKA pathway by acting directly on proinsulin/proglucagon transcription, and Epac through the influx of calcium leading to upregulation of hormone production. In 2011, Wagner et al. assessed insulin secretion, proinsulin to insulin conversion, and insulin resistance in 1,782 nondiabetic individuals at increased risk of developing T2D. They found that the fasting glucose associated allele at SNP rs11708067 in *ADCY5* also associated with impaired conversion of proinsulin to insulin (Wagner et al., 2011). The other genes in the region around the implicated variants are *SEC22A*, a vesicle trafficking protein, *PTPLB*, involved in fatty acid synthesis, *MYLK*, which facilitates contraction in smooth muscle, and *PDIA5*, which catalyzes disulphide bonds. The location of the markers within an intron of *ADCY5* makes *ADCY5* the most promising causal transcript in the region.

4.1.2 Adenylate cyclases in the pancreatic islet cells

The *ADCY5* gene encodes a membrane-bound adenylate cyclase responsible for catalyzing the generation of cAMP. There are 9 types of adenylate cyclases in mammals. ADCY5 is calcium-inhibitable and, like all membrane-bound adenylate cyclases, thought to localize to lipid rafts within the plasma membrane or other membranes in the cell (Cooper and Crossthwaite, 2006). Hormones

such as incretins like GLP-1, catecholamines such as adrenaline, and others such as somatostatin and forskolin bind receptors activating coupled adenylate cyclases and catalyzing the generation of cAMP (Pierre et al., 2009). In the beta and alpha cells, cAMP triggers PKA- independent and dependent signaling pathways (Renström et al., 1997). Both pathways lead to exocytosis (Renström et al., 1997). When cAMP activates PKA, PKA triggers a series of events leading to calcium influx and release of insulin or glucagon granules (Holz and Habener, 1992). PKA also activates PDX-1 which translocates to the nucleus and upregulates transcription of the insulin gene (Wang et al., 2001). Alternatively, cAMP can activate the Epac pathway. Epac2 interacts with the insulin granule-associated sulfonylurea receptors (gSUR and SUR1), which regulate chloride channels important in granule priming (Eliasson et al., 2003). GLP-1 has been shown to activate both the PKA and Epac pathways through adenylate cyclases in the beta cell to potentiate insulin secretion (Holz, G., 2004).

4.1.3 Previous mouse models of *Adcy5* indicate a metabolic phenotype

ADCY5-dependent cAMP signaling is critical in the heart and striatum of the brain and may be involved in metabolic processes. Male mice homozygous for an *Adcy5* knockout allele exhibit decreased body mass, increased food intake, less adiposity, lower fasting blood glucose, and increased longevity suggesting a metabolic phenotype (Yan et al., 2007; Ho DZX, et al., 2010; Li et al., 2006). Li et al. (2006) propose that *Adcy5* disruption leads to protection against oxidative stress resulting in increased longevity. *ADCY5* has a known role in cardiac function and in behavioral traits. In the heart, *ADCY5* disruption results in beta-adrenergic signaling dysfunction (Tepe et al., 1999). Cardiac-specific *Adcy5* knockout mice develop resistance to chronic pressure overload caused by adrenaline stress (Ho et al., 2010).

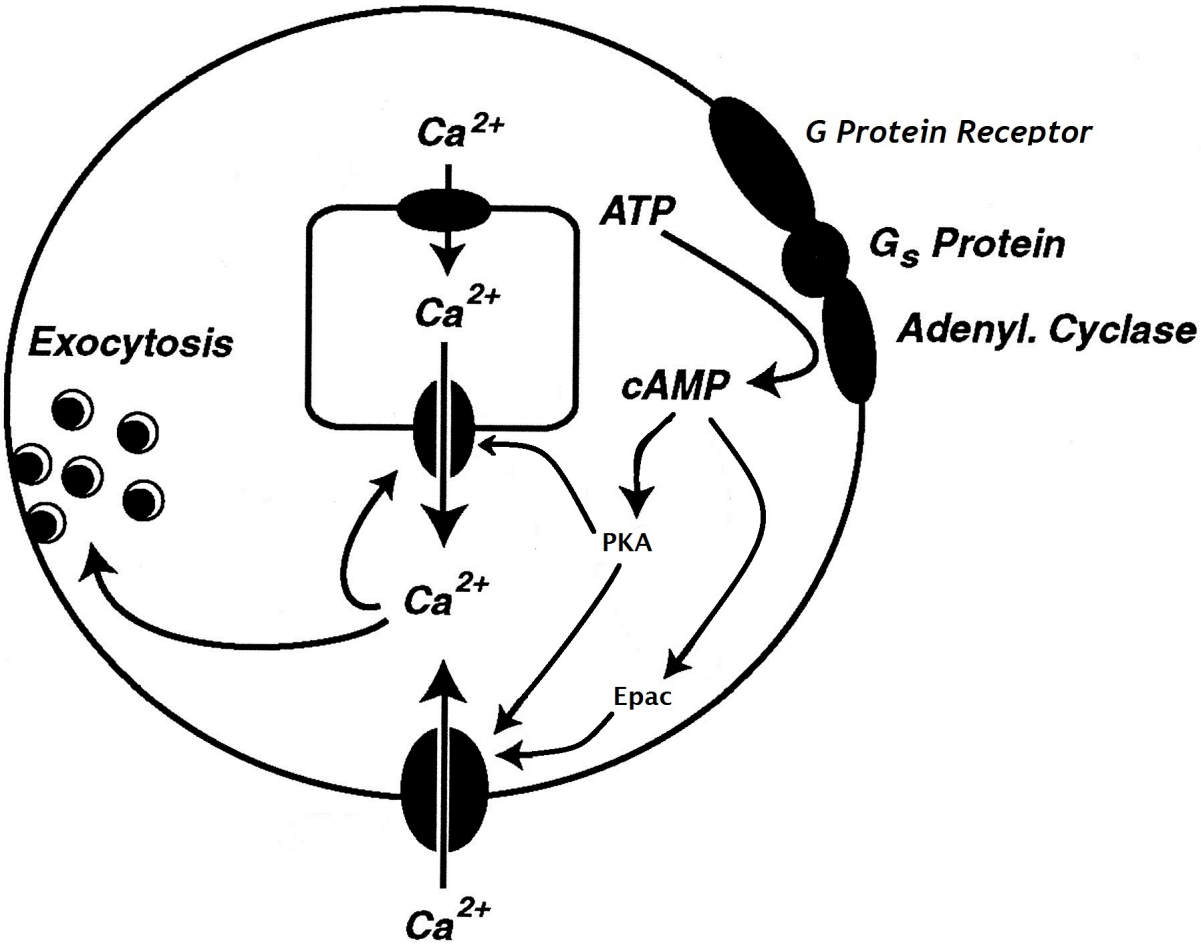


Figure 4.1: Adenylate cyclases in the beta cell catalyze the generation of cAMP which activates both the PKA and Epac pathways. Epac2 potentiates calcium release from intracellular stores and PKA triggers extracellular calcium influx. Increased calcium leads to insulin exocytosis.

ADCY5 is also highly expressed in the brain, particularly the dorsal striatum and nucleus accumbens (Levey et al. 1993), areas which play a key role in addiction, anxiety and stress. ADCY5 activity in the brain is enhanced by ethanol (Yoshimura and Tabakoff, 1995), and cAMP levels are increased by administration of ethanol (Galassetti et al., 2000). *Adcy5* knockout mice exhibited increased ethanol consumption and sensitivity to ethanol is decreased (Kim et al., 2010).

The dorsal striatum is also critical in procedural learning, such as for motor skills and stimulus-response habits (Packard et al., 1989; Packard and White, 1991). Mice lacking *Adcy5* had impaired dorsal striatum learning (Iwamoto et al., 2003; Kheirbek et al., 2008; Kheirbek et al., 2009). *Adcy5* knockout mice could not learn procedural tasks such as navigating mazes and rota-rod skills as easily as wild type mice due to impaired cAMP signaling in the striatum (Kheirbek et al., 2009). In summary, *ADCY5* mouse models have shown a role for the gene in cardiac and brain function. When this thesis commenced, it was the first functional investigation into *ADCY5* function in pancreatic islets.

4.1.4 *Adcy5*^{Y1064C} may disrupt protein function leading to alpha and/or beta cell dysfunction

I found 15 mutations in scanning 7 exons of *Adcy5* in the ENU archive at the MRC Harwell. These results are described in Chapter 3, and I will review briefly here. Out of the 15 mutations found, I predicted the *Adcy5* ^{Y1064C} mutation in exon 18 to be the most likely to be damaging to protein function. The Y1064C falls in a catalytic domain of *Adcy5* and appears to be important for correct protein folding from modelling in PyMOL (14.1, Schrodinger) and predictions from Polyphen and Sift mutation prediction software. The Y1064C mutation may lead to loss of function but could also be hypomorphic or hypermorphic and present a subtler phenotype.

The genome-wide associated signals from within *ADCY5* in humans suggest that, if *ADCY5* is the causal gene, mutations in its coding region could affect one or all of the following traits: fasting glucose, glucose tolerance, beta cell function and birth weight. The directional effect of the T2D-risk allele at rs11708067 in humans cannot be predicted – it may be a gain of function mutation or a loss

of function mutation. However, my modelling of ADCY5 protein structure suggests that the Y1064C allele is loss of function mutation in this case. Confirmation of the directional effect of the Y1064C allele and phenotypic data from the mice may help determine the direction of the effect of the human variant.

4.1.5 Aims

Based upon the reported associations of variants near *ADCY5* with T2D, impaired beta cell function, high fasting glucose, and birth weight, I chose to characterize the *Adcy5* ^{Y1064C} model for similar phenotypes to test the hypothesis that *ADCY5* is the causal gene associated with these traits. Furthermore, my experiments were designed to test whether *Adcy5* is involved in important signaling cascades in islet cells.

The research described in this chapter was done to address the following aims:

(i) To evaluate the expression profile of *ADCY5* and other adenylate cyclases in mice and humans with the objective of determining the location and expression level of the genes. This information can be used to inform the most targeted phenotyping pipeline and *in vitro* experiments.

(ii) To characterize metabolic phenotypes in the *Adcy5*^{Y1064C} homozygote, heterozygote and wild type mice *in vivo* to understand the role, if any, of *Adcy5* in metabolism.

(iii) To characterize insulin and glucagon secretion phenotypes in the *Adcy5*^{Y1064C} homozygote, heterozygote and wild type islets *in vitro*.

4.2 Methods

4.2.1 Samples Used

Quantitative real-time PCR was carried out in human and in mouse tissues to investigate expression levels of *ADCY5* and the other 8 adenylate cyclase isoforms across a range of tissues including pancreatic islets to inform further characterization *in vivo* and *in vitro*. Human total RNA samples (skeletal muscle, brain, liver and heart) were purchased as part of a commercially available tissue panel (Clontech, Panel #1409300, Oxford, UK), with each sample comprising a pool of either 2 or 3 donors (Table 4.1). Omental adipose tissue (2 donors) and subcutaneous adipose tissue (1 donor) were obtained through Oxford University's MolSURG project, collected as part of the MolPAGE (Molecular, Phenotyping to Accelerate Genomic Epidemiology) consortium (MolPAGE, 2012). Human islet RNA was isolated from islets collected from 3 non-diabetic organ donors post-mortem. Tissue samples were collected with full research consent (Oxfordshire Regional Ethics Committee B) from participants undergoing elective surgery, snap frozen and stored as whole tissue in liquid nitrogen prior to extraction. Pancreases were digested in collagenase and islets picked, lysed, and stored in TRIzol (Life Technologies) prior to RNA isolation.

Mouse total RNA samples (skeletal muscle, whole brain, liver and heart) were purchased as part of a commercially available tissue panel (Clontech, Panel #1504211, Oxford, UK), with each sample comprising a pool of between 12-600 mice (Table 4.2). Omental adipose tissue, brown fat and skeletal muscle were collected at the MRC Harwell from C3H/HeH mice at 12 weeks of age (n=3), snap frozen and stored as whole tissue in liquid nitrogen prior to extraction. Ins-1 cells (rat insulinoma line, obtained from Dr. Nicola Beer, OCDEM, University of Oxford) and Min-6 cells (mouse insulinoma cell line, Miyazaki et al., 1990) were cultured at OCDEM following procedures described in Chapter 2, section 2.4.1. Islet RNA was collected from pooled islets from 6 mice. Cells were pelleted and lysed and RNA extracted following the procedure in section 2.3.1.

4.2.2 mRNA expression quantification

RNA was extracted from tissues not purchased as part of an RNA panel as described in Chapter 2 section 2.3.1, and cDNA was synthesized from RNA as described in Chapter 2 section 2.3.2. Total gene expression was quantified using TaqMan gene expression assays (Applied Biosystems) as described in section 2.3.3. Inventoried assays were selected to cover the maximum number of transcripts and to span an exon junction (_m1 suffix). Assay details provided in Table 4.3. All cDNA samples were diluted to one in 50 parts (1:50) in 0.01M Tris and standard curves were generated by serially diluting a pool of all cDNA samples as described in section 2.3.

Real-time PCR was carried out on an Applied Biosystems 7900HT machine as described in section 2.3.3. All samples were run in triplicate in the presence of a reverse-transcriptase negative control. No samples showed amplification in negative control wells.

For expression quantification within one gene across many tissues (section 2.3.3 (e)) analysis was performed using the ΔΔCt method. Expression was determined with respect to the 1:50 pooled sample standard and normalized to the geometric mean of the 3 endogenous controls (B2M, HPRT and ACTB).

4.2.5 *Adcy5*^{Y1064C} Phenotyping Pipeline

Animal studies were carried out under the guidance issued by the MRC: Responsibility in the Use of Animals for Medical Research (July 1993) and Home Office Project Licence No. 30/2642. I obtained a Home Office Personal License for mouse studies. Mice were kept in accordance with U.K. Home Office welfare guidelines and project license restrictions under controlled light (12 hr light and 12 hr dark cycle), temperature (21 °C ± 2 °C) and humidity (55 % ± 10 %) conditions. They had free access to water (25 ppm chlorine) and were fed *ad libitum* on a commercial diet (SDS maintenance chow, RM3, 3.6 kcal/g, Essex, U.K.).

Tissue	N donors	Donor age range (years)
skeletal muscle	3	43-46
heart	3	30-39
liver	3	29-51
whole brain	2	47-55

Table 4.1: Human total RNA panel donor details

Tissue	Source	Age (weeks)
whole brain	pooled from 600 BALB/c mice	9-12
heart	pooled from 105 male BALB/c mice	10-12
liver	pooled from 12 BALB/c mice	8-9

Table 4.2: Mouse total RNA panel details

Human Assays		
Gene ID	Full gene name	Assay ID
ADCY1	adenylate cyclase 1	Hs00299832_m1
ADCY2	adenylate cyclase 2	Hs01058848_m1
ADCY3	adenylate cyclase 3	Hs01086502_m1
ADCY4	adenylate cyclase 4	Hs00934099_m1
ADCY5	adenylate cyclase 5	Hs00766287_m1
ADCY6	adenylate cyclase 6	Hs00209600_m1
ADCY7	adenylate cyclase 7	Hs00936808_m1
ADCY8	adenylate cyclase 8	Hs00905034_m1
ADCY9	adenylate cyclase 9	Hs00181599_m1
Mouse Assays		
Adcy1	adenylate cyclase 1	Mm01187829_m1
Adcy2	adenylate cyclase 2	Mm00467874_m1
Adcy3	adenylate cyclase 3	Mm00460371_m1
Adcy4	adenylate cyclase 4	Mm00475491_m1
Adcy5	adenylate cyclase 5	Mm01343348_m1
Adcy6	adenylate cyclase 6	Mm00475772_m1
Adcy7	adenylate cyclase 7	Mm00545780_m1
Adcy8	adenylate cyclase 8	Mm00507722_m1
Adcy9	adenylate cyclase 9	Mm00507743_m1

Table 4.3: Applied Biosystems inventoried assays used for mRNA quantification.

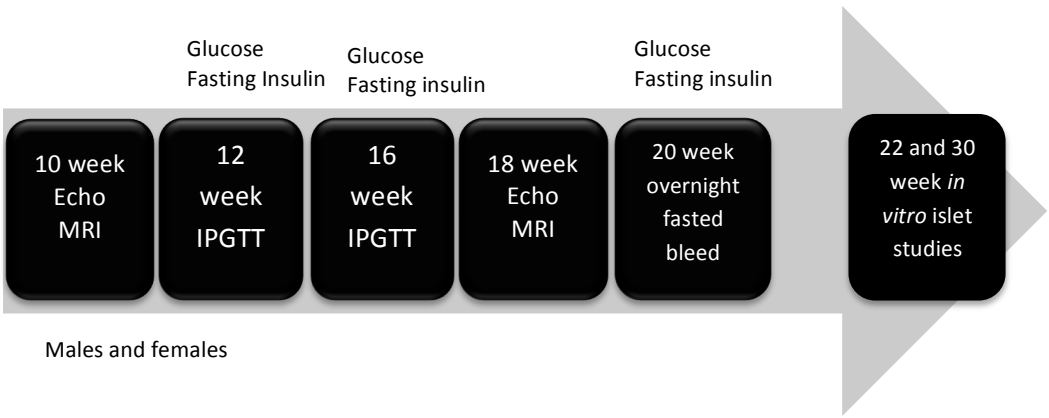
Mice were characterized in one of three metabolic phenotyping pipelines (Figure 4.2). Details of phenotyping procedures can be found in Chapter 2 section 2.5. Cohort 1 mice of both genders underwent the first pipeline that focused on body composition, intraperitoneal glucose tolerance, and fasting glucose; islets were used for *in vitro* characterization (sample sizes of cohorts 1, 2, and 3 are given in Table 4.4).

A second pipeline was devised to test oral glucose tolerance, repeat the IPGTT in more mice, and to generate more islets for *in vitro* studies (Figure 4.2, Table 4.4). Due to problems with the glucose analyzer, nearly half of the mice had to be excluded from the 20-week oral glucose tolerance test. For further investigation into the oral glucose tolerance phenotype, a third cohort was bred and underwent a shorter pipeline including an OGTT and a fasting bleed (Figure 4.2). Females were not included in cohort 3 because the OGTT phenotype was only apparent in the males. Both genders were used for *in vitro* islet studies. Throughout this chapter I will refer to the first cohort and therefore the mice that underwent the first pipeline as cohort 1. I will refer to the second cohort as cohort 2 and the third cohort as cohort 3 (Figure 4.2). Oral glucose tolerance tests (OGTTs) and fasted bleeds were performed by Dr. Liz Bentley (MRC Harwell). I performed EchoMRI body composition tests and IPGTTs.

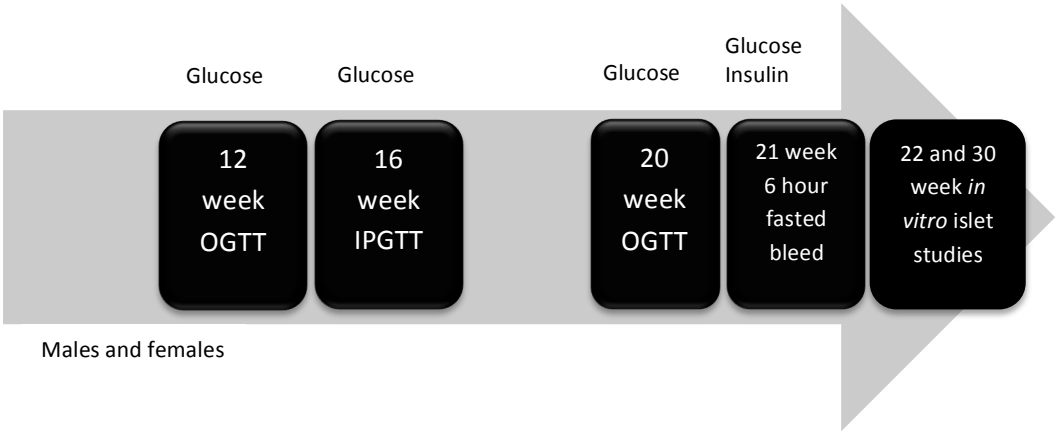
4.2.6 Birth weight Measurements

A pilot cohort was bred of 8 *Adcy5*^{Y1064C} homozygote mice, 16 heterozygote mice, and 9 wild type mice. Pups were weighed and measured as described in Chapter 2 section 2.5.2(c&d). Sample size was calculated from pilot results to achieve 80% power using University of British Columbia Power Sample Size Calculation Software (http://www.stat.ubc.ca/~rollin/stats/ssize/n2.html). Power calculations indicated ≥11 mice were needed for each genotype for a significance cut-off of 0.05. A separate intercross was bred from backcross 7 mice to create a cohort of 13 wild type, 37 *Adcy5*^{Y1064C} heterozygote, and 12 *Adcy5*^{Y1064C} homozygote mice comprising both males and females. The mice were culled as neonates and weight and bone lengths were measured as in the pilot cohort.

Cohort 1



Cohort 2



Cohort 3

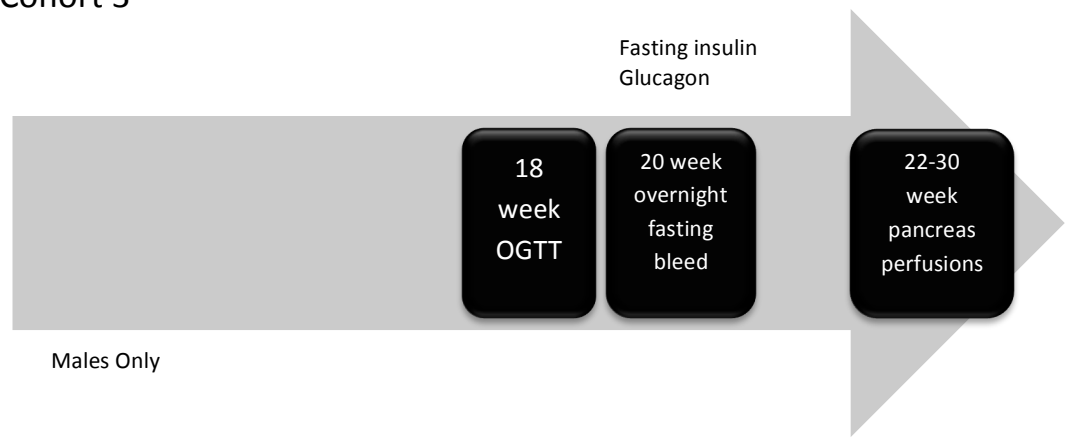


Figure 4.2: Metabolic phenotyping pipelines for cohorts 1, 2 and 3. See table 4.4 for Cohort n numbers. Phentoyping of the females was terminated after Cohort 2 because the phenotype was clearer in males.

Sample size: biological replicates

	Cohort 1		Cohort 2		Cohort 3	
	Males	Females	Males	Females	Males	Females
Adcy5$^{Y1064C+/+}$	11	6	24	19	15	0
Adcy5$^{Y1064C+/-}$	26	21	21	24	22	0
Adcy5$^{Y1064C-/-}$	11	12	21	14	15	0

Table 4.4: Summary of sample sizes for cohorts 1, 2 and 3. Zero female mice were included in the characterization of cohort 3 because the phenotype was more pronounced in the males.

4.2.7 Static hormone secretion assays

Islets from cohort 1 and cohort 2 mice were isolated by collagenase digestion as described in Chapter 2 section 2.4.5(a). Secretion studies were performed using the protocol in Chapter 2 section 2.4.5(b). Samples were assayed for insulin and glucagon by radioimmunoassay as explained in Chapter 2 section 2.4.7.

4.3 Results

4.3.1 Quantitative real-time PCR analyses of adenylate cyclases in humans and mice

4.3.1 a) Expression of *ADCY5* and *Adcy5* in humans and mice across relevant tissues

I performed qRT-PCR for *ADCY5* expression over a range of rodent and human tissues to pinpoint where it is most highly expressed in the body, particularly focusing on metabolic tissues and islets. Information about tissues and number of donors can be found in methods section 4.2.1. House-keeping genes (HKGs) hypoxanthine phosphoribosyltransferase (*Hprt/HPRT*), beta-2-microglobulin (*B2m, B2M*), and beta actin (*Actb/ACTB*) were chosen because of their known expression in islet cells (described in Methods section 2.3.3).

In the human tissues analyzed, *ADCY5* is expressed most highly in brain, omental fat, and heart (>40% of HKG expression). *ADCY5* is expressed in islets at low levels (around 3% of averaged HKG expression). The results are shown in Figure 4.3. This data confirms the presence of *ADCY5* in islets. ADCY5 expression levels are comparable to those published on the GTEx Portal (Broad Institute). The tissue with highest expression of *ADCY5* on the GTEx Portal is brain followed by heart. Brain and heart are the tissues I found to most highly express *ADCY5* as well. GTEx reports expression of *ADCY5* in adipose tissue at about 30% of the levels of expression in heart which is also comparable to my data. GTEx shows low expression of *ADCY5* in pancreas; islet expression data is not available.

Adcy5 is expressed highly in mouse brain, mouse omental fat, mouse heart, and Ins1 cells (Figure 4.3). MIN6 cells (mouse-derived beta cell line) showed only about 1.5% of the levels of *Adcy5* expression as the rat-derived beta cell line INS-1. Isolated mouse islets also displayed low expression, about 10% of the expression level of *Adcy5* in INS-1 cells and 5% of the levels of expression relative to HKGs (Figure 4.3). The probes used will equally amplify both mouse and rat sequences, so the distinct difference in *Adcy5* expression between the mouse islets, MIN6 cells and INS-1 cells may point to functional differences in mouse vs. rat beta cells. Another explanation for the divergence in *Adcy5* expression between MIN6 and INS-1 insulinoma cell lines is that they may have accumulated genetic and epigenetic changes over generations of cell culture that has altered their function and patterns of gene expression.

Although expressed at low levels in human and mouse islets *ADCY5* may still be functionally important. To further understand the role of *ADCY5* in islets I measured the expression of other adenylate cyclase genes in islets.

4.3.1 b) Expression of all adenylate cyclases in human and mouse islets

There are 9 adenylate cyclases in both human and mouse. In order to measure their relative expression in islets I performed qRT-PCR for all 9 isoforms in human and mouse islets to identify which isoforms are present in addition to *ADCY5* and *Adcy5.* A pool of islets from 3 nondiabetic donors and 6 mice were used. In human islets *ADCY3* and *ADCY9* are expressed most highly. All other adenylate cyclases are expressed at lower levels (Figure 4.4). My data confirm that *ADCY5* is expressed in islets at about 15% of *ADCY9* mRNA expression. *ADCY8,* expressed at similar levels to *ADCY5* in islets, has been shown to be integral to GLP-1 signaling in human beta cells (Willoughby et al., 2010; Roger et al., 2011).

In mouse islets, *Adcy4* and *Adcy6* are the most highly expressed. *Adcy5* is expressed at low levels, about 5% of *Adcy4* mRNA expression (Figure 4.4). *Adcy4* and *Adcy6* have no established role in the islets. *Adcy6* is highly expressed in pancreatic exocrine cells (Sabbatini et al., 2013).

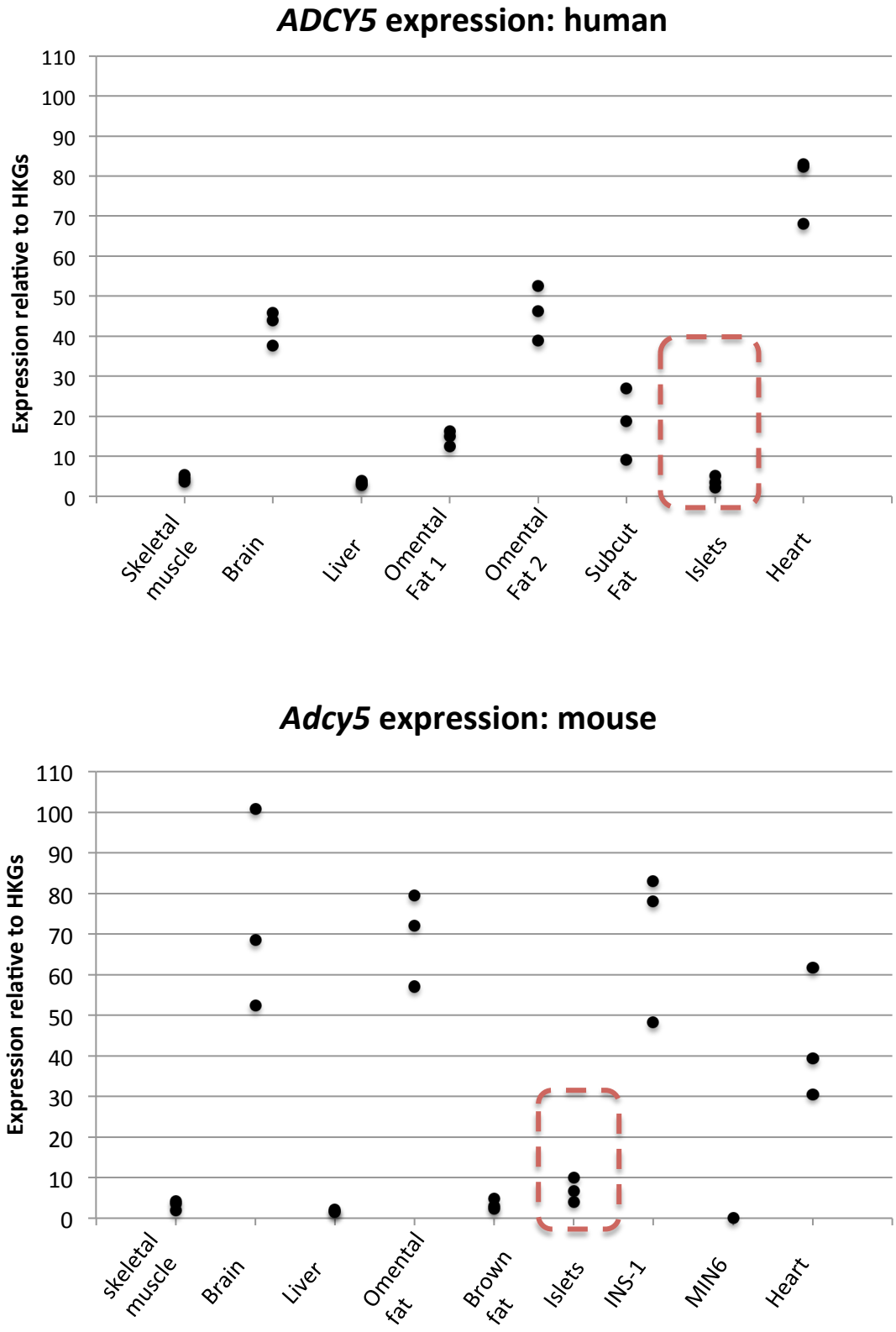


Figure 4.3 *ADCY5* expression across relevant tissues in human (top) and *Adcy5* expression in mouse tissues (bottom) relative to the mean expression of 3 housekeeping genes (*HPRT/Hprt, B2M/B2m, ACTB/Actb*). Three replicates for each tissue were tested. Omental fat 1 (human) belongs to female donor with BMI 19.1kg/m^2, and omental fat 2 belongs to a male donor of BMI 31.1kg/m^2.

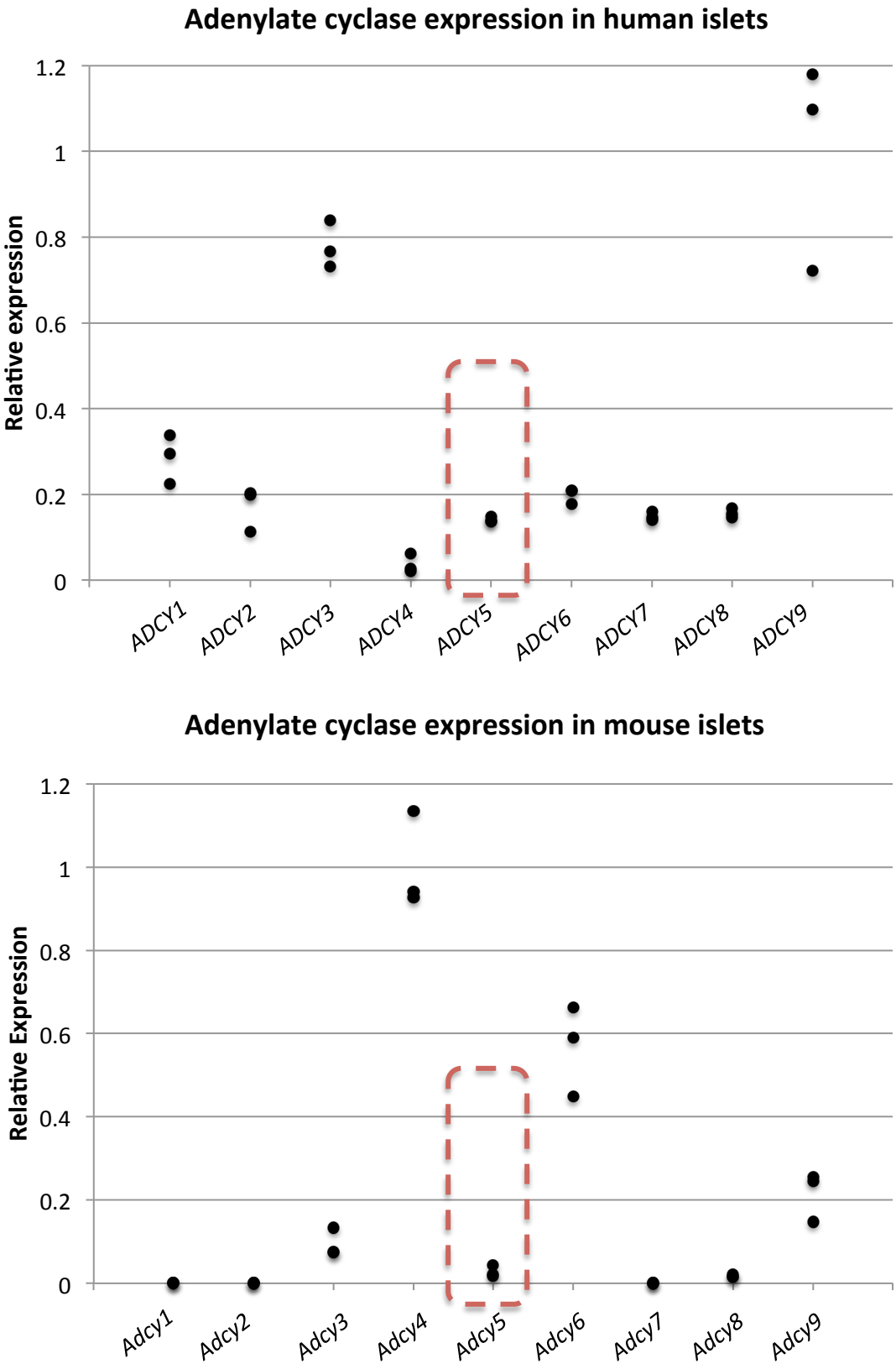


Figure 4.4 Expression of all adenylate cyclases in human (top) and mouse (bottom) islets relative to the most highly expressed gene and normalized to house-keeping genes. A pool of islets from three mice and three humans was tested in 3 technical replicates.

This analysis demonstrates that while *Adcy5* is present in mouse islets, it is not the most highly expressed adenylate cyclase gene. In the mouse, *Adcy4* and *6,* and the other expressed adenylate cyclases *Adcy9* and *3*, may be performing roles in cAMP signaling in addition to those of *Adcy5*. Therefore it is important to identify the specific role of *Adcy5* through functional investigation.

4.3.2 Major findings of the *in vivo* and islet studies

In the next section I present data from *in vivo* mouse studies and from isolated islets. Although there were some conflicts and inconsistencies in some of the data, particularly from isolated islets, there were three main findings from this work:

a) whole body characterization showed mice with two copies of the *Adcy5* ^{Y1064C} allele have lower blood glucose than wild type;
b) birth weight is unaffected by the *Adcy5* ^{Y1064C} mutation;
c) glucagon secretion is altered in isolated islets from mice homozygous for the Y1064C mutation.

These results are presented in the following sections.

4.3.3 *Adcy5*^{Y1064C} homozygous mice have lower blood glucose than wild type

In vivo metabolic characterisation of the *Adcy5*^{Y1064C} mutant mice revealed there is a metabolic phenotype in the homozygote mice. To summarize the full results presented below: the mice with two copies of the Y1064C mutation have improved oral glucose tolerance compared to wild type mice. Further phenotyping showed that the genotypes responded the same to an intraperitoneal glucose tolerance test, and there was also no difference in fasting blood insulin or glucagon to suggest metabolic hormone secretion at basal levels is altered. Body composition did not differ between genotypes confining the metabolic phenotype detected solely to an improvement in oral glucose tolerance.

4.3.3 a) Mice with 2 copies of the *Adcy5* ^{Y1064C} mutation have improved oral glucose tolerance test compared to wild type

Liz Bentley (MRC Harwell) performed oral glucose tolerance tests in cohort 2 *Adcy5* ^{Y1064C} male and female wild type, heterozygote and homozygote mice at 12 and 20 weeks and in cohort 3 male mice of all genotypes at 18 weeks (see section 4.2.5 for cohort information).

Administering glucose orally stimulates gut-mediated hormones such as GLP-1 that bind a G-protein receptor on the beta cell membrane and potentiate insulin secretion. If *Adcy5* is coupled to one of the gut mediated hormones, *Adcy5*^{Y1064C} allele carriers and wild type may have a different insulin or glucagon secretion response to orally-administered glucose.

There was a difference in oral glucose tolerance between genotypes in the male mice. At 12 weeks, the cohort 2 homozygote and heterozygote males had borderline-significantly lower plasma glucose 60 minutes after oral dosage than wild type males (males: wild type mean(standard deviation) =12.1(1.5)mmol/L; heterozygote 10.9 (1.8)mmol/L; homozygote mean = 10.8(1.2)mmol/L (p=0.05), Figure 4.5[a]). At 20 weeks the homozygote male mice had lower blood glucose at all time points following glucose administration (T_0 (before glucose dosage) p=0.013; T_{60} p=0.0004; T_{120} p=0.006) (Figure 4.5[b]). There was a problem with the glucose analyzer during the 20 week OGTT, so roughly one third of the samples could not be analyzed. To confirm that the same mice exhibiting a phenotype at 20 weeks were not different from wild type at 12 weeks, I re-analyzed the OGTT at 12 weeks including only the mice that were measured in the 20 week OGTT. There was no difference between genotypes in these mice at 12 weeks, but by 20 weeks, the genotypes had diverged.

Cohort 3 male mice also showed improved glucose tolerance when tested at 18 weeks (Figure 4.6). For the OGTT at 18 weeks, 4 time points were included (0, 30, 60, and 120 minutes after dosage) in order to more precisely identify when the genotypes diverged. We found that plasma glucose was lower in the male homozygote mice at 30, 60 and 120 minutes after dosage ($_{T30}$ p=0.002, T_{60}

p=0.0001, T_{120} p=0.001) with the most significant difference occurring at 60 minutes after glucose dosage.

The female heterozygote and homozygote mice did not differ from wild type at any time point during the OGTT at 12 weeks of age (Figure 4.7[a]). Plasma glucose in the female mice did not differ between genotypes during an OGTT at 20 weeks of age, although there was a trend toward the homozygote mice exhibiting lower plasma glucose (homozygote mice vs. wild type at 60 minutes after glucose dosage: p = 0.06, Figure 4.7[b]). We eliminated the females from phenotyping in the 3rd cohort because the oral glucose tolerance phenotype was more evident in the males in cohort 2.

To summarize, at 12 and 18 weeks all genotypes started at T0 with the same overnight fasting glucose levels, but the cohort 2 mice at 20 weeks had lower fasting glucose before dosage at T0. The oral glucose tolerance data from male mice at 18 and 20 weeks showed a robust difference in plasma glucose levels between Y1064C homozygous and wild type mice supporting the inference that Adcy5$^{Y1064C-/-}$ mice are protected against hyperglycemia.

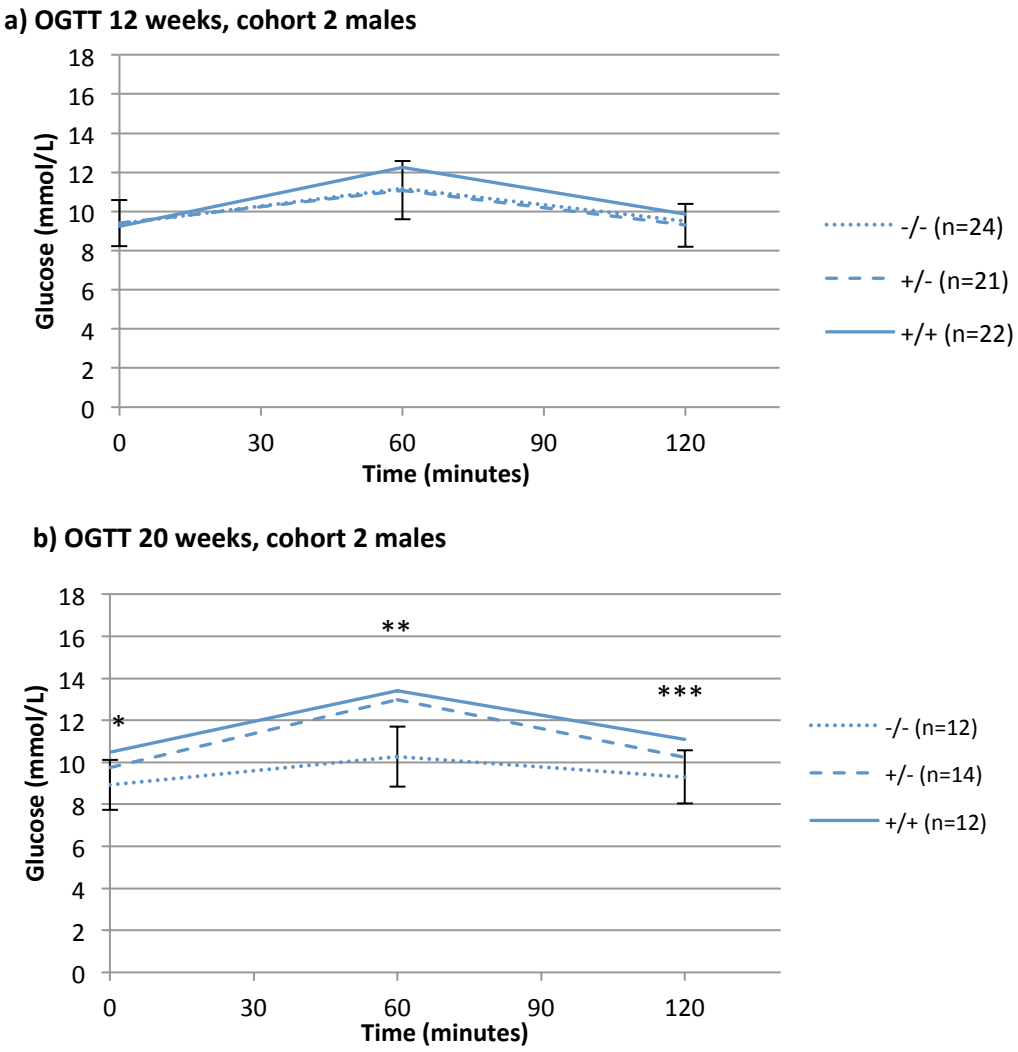


Figure 4.5: OGTT in male mice at 12 weeks (a) and 20 weeks (b). Error bars (standard deviation) are only shown for homozygote (-/-) mice for clearer visualization. Standard deviation was similar for all genotypes. Significance is indicated by *p<0.05, **p≤0.01, ***p≤0.001. At 12 weeks, male mice homozygous for the Y1064C mutation had borderline-significant lower plasma glucose 60 minutes after oral dosage (p=0.05). At 20 weeks male homozygote mice had lower fasting plasma glucose than wild type as well as lower plasma glucose 60 and 120 minutes after glucose dosage.

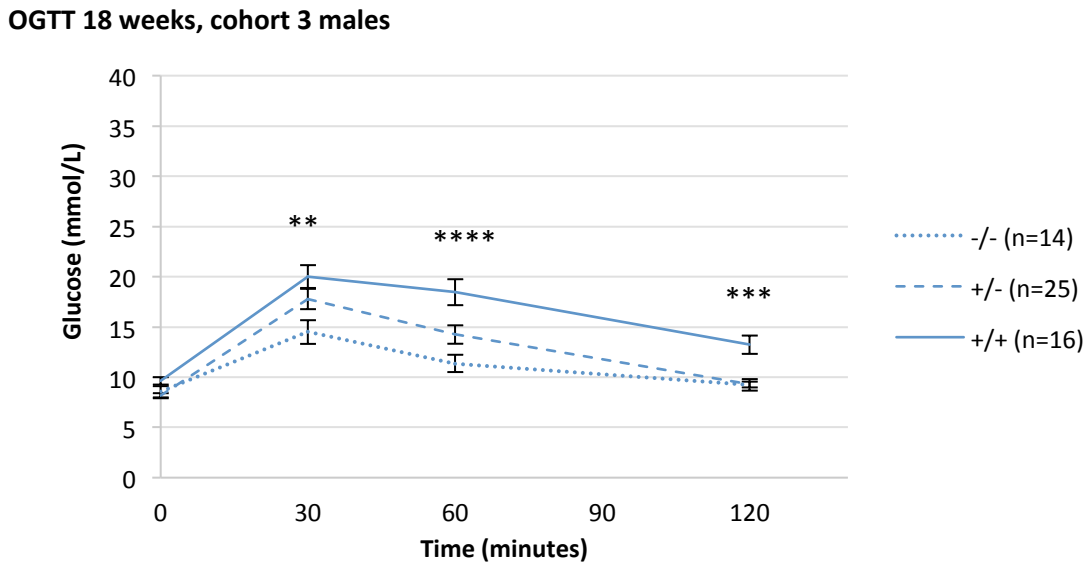


Figure 4.6: At 18 weeks, an extra collection point at 30 minutes was added to determine plasma glucose closer to dosage. Male homozygote animals had lower plasma glucose than wild type at every point after dosage.

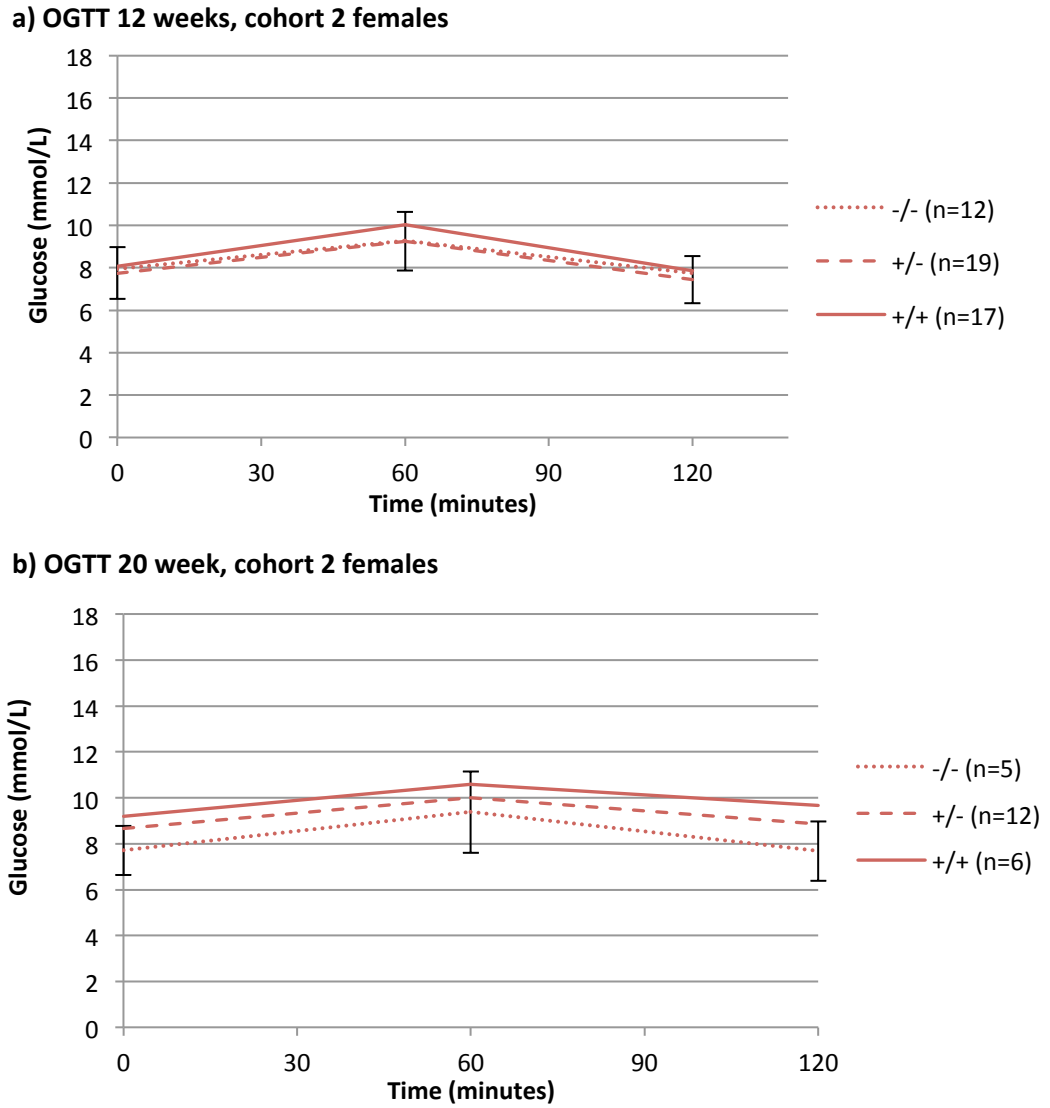


Figure 4.7: OGTT in female mice at 12 (a) and 20 weeks (b) showed no difference between genotypes. Females were excluded from cohort 3 because the phenotype was stronger in the males. Error bars (standard deviation) are only shown for the homozygote (-/-) animals for clearer visualization.

4.3.3 b) There was no difference between genotypes after an intraperitoneal glucose tolerance test

Intraperitoneal glucose tolerance was measured in cohort 1 mice at 12 weeks 0, 60 and 120 minutes after injection and at 16 weeks 0, 10, 20 and 30 minutes after injection. The shorter time frame was meant to capture peak glucose levels after injection. At 12 weeks cohort 1 wild type, heterozygote, and homozygote male mice did not differ at any time point (Figure 4.8). At 16 weeks there was also no difference in plasma glucose between genotypes and peak glucose may not have been reached by 30 minutes post injection when plasma insulin averaged ~30mmol/L (Figure 4.9). It would have been interesting to measure peak insulin to determine if the genotypes were secreting the same amount of insulin, but this data was not collected at the time. Variation in sample size (n) is due to random error such as failure to collect adequate blood samples at all time-points.

Another IPGTT was performed in cohort 2 at 16 weeks; blood was collected up to 120 minutes post injection and a collection time point was added at 30 minutes to capture a full range of blood glucose levels from 30 minutes to 2 hours post injection (Figure 4.10[a]). There was no significant difference between genotypes at any time points. The cohort 2 male mice did show a difference in blood glucose at 20 weeks during an OGTT as described in the previous section. To verify that the specific mice included in the OGTT experiment were not exhibiting an IPGTT phenotype when separated from the larger cohort, I analyzed the mice that were included in both experiments (Figure 4.10[b]). Again there were no differences between genotypes confirming that a phenotype was only observed after an oral glucose challenge.

No phenotype was observed in the female mice after an IPGTT at 12 or 16 weeks (Figure 4.11, 4.12, and 4.13). At 16 weeks, an IPGTT capturing the glucose profile in the first 30 minutes after injection showed peak glucose was reached for all genotypes at 10 minutes post injection and averaged ~22 mmol/L (Figure 4.12). This demonstrates a clear sexual dimorphism in glucose tolerance as the males had not yet reached peak glucose at ~30mmol/L (see above and Figure 4.9). Sex differences in glucose tolerance and diabetes susceptibility in mice are well documented (Vital et al., 2006).

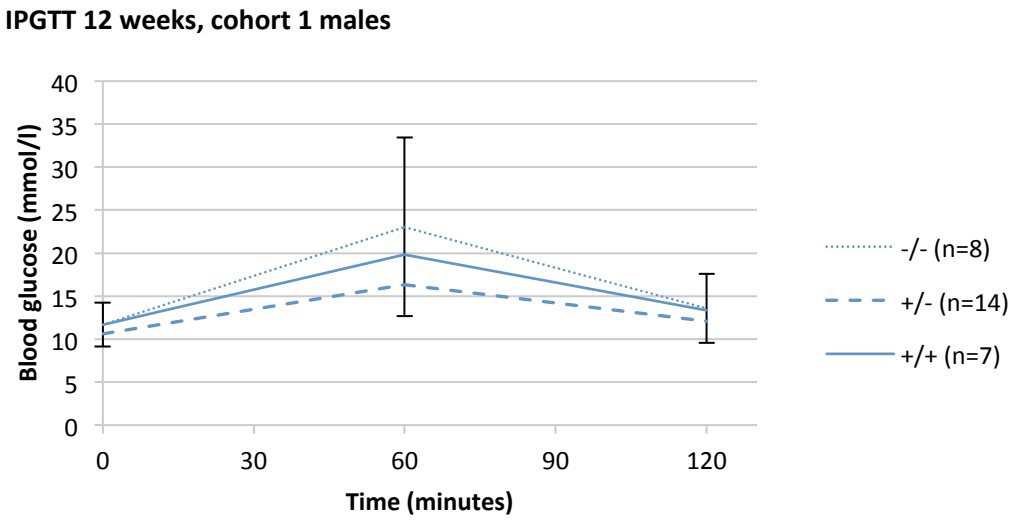


Figure 4.8: IPGTT in male mice at 12 weeks of age showed no difference between genotypes. Error bars (standard deviation) are only shown for the homozygote (-/-) mice for clearer visualization.

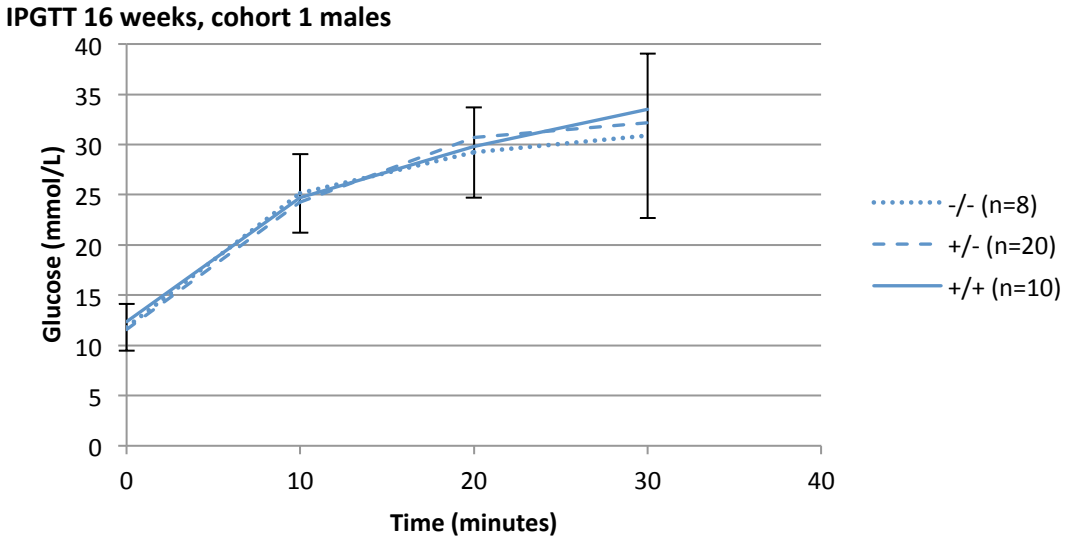


Figure 4.9: IPGTT for the first 30 minutes after injection in male mice at 16 weeks. This test was devised to determine if the genotypes processed glucose differently during the first 30 minutes after an intraperitoneal glucose bolus. The genotypes did not differ at any point. Glucose had not yet peaked by 30 minutes after injection. Error bars (standard deviation) are only shown for homozygote mice for visualization purposes.

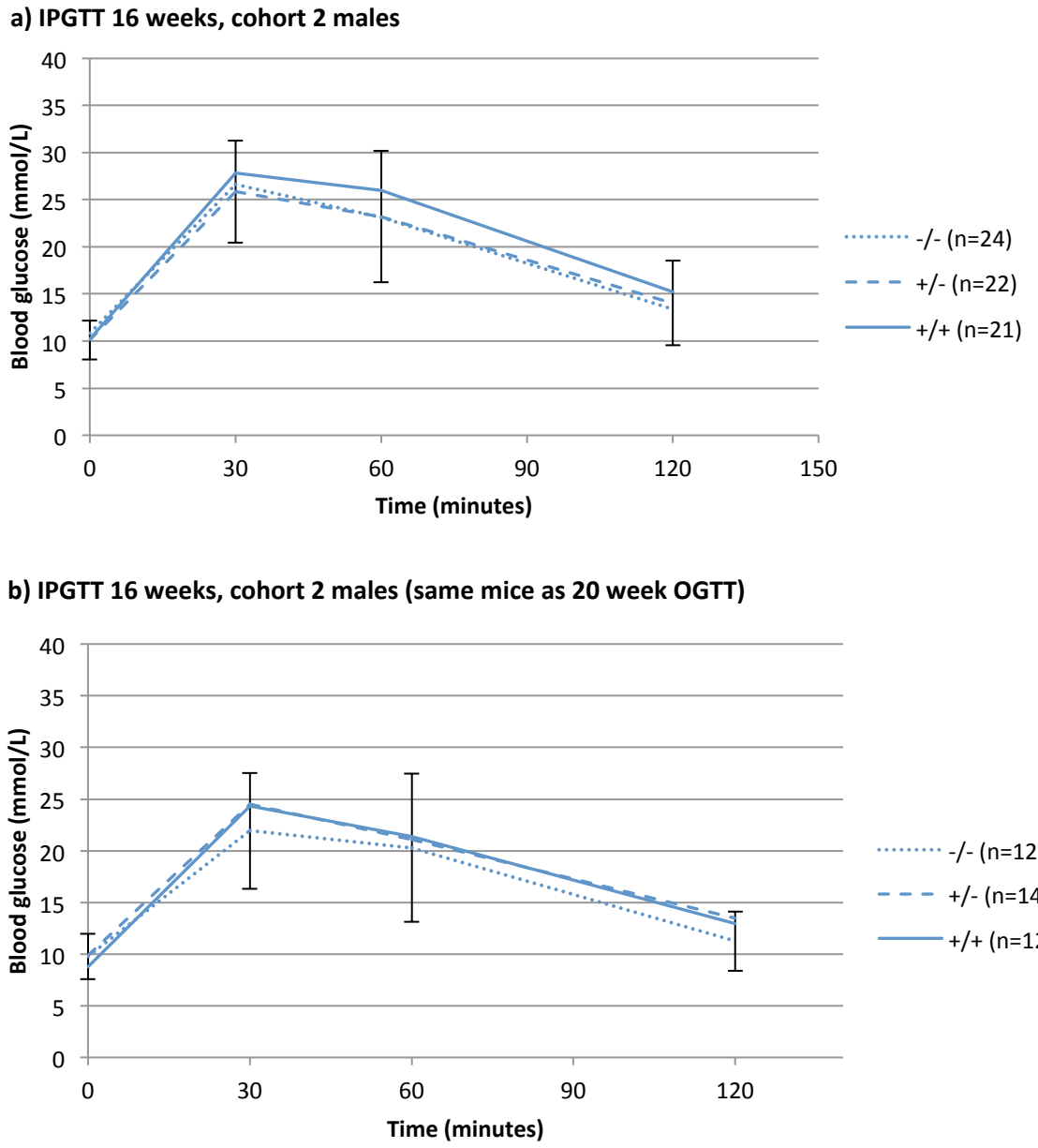


Figure 4.10: IPGTT results for male mice at 16 weeks show no difference between genotypes (a). A time point at 30 minutes was added to measure plasma glucose closer to injection. Panel (b) includes only the mice that were also included in the 20 week OGTT (see figure 4.5 above) to demonstrate that the mice were not exhibiting an IPG phenotype at 16 weeks. Error bars showing standard deviation are only displayed for homozygote mice for clearer visualization.

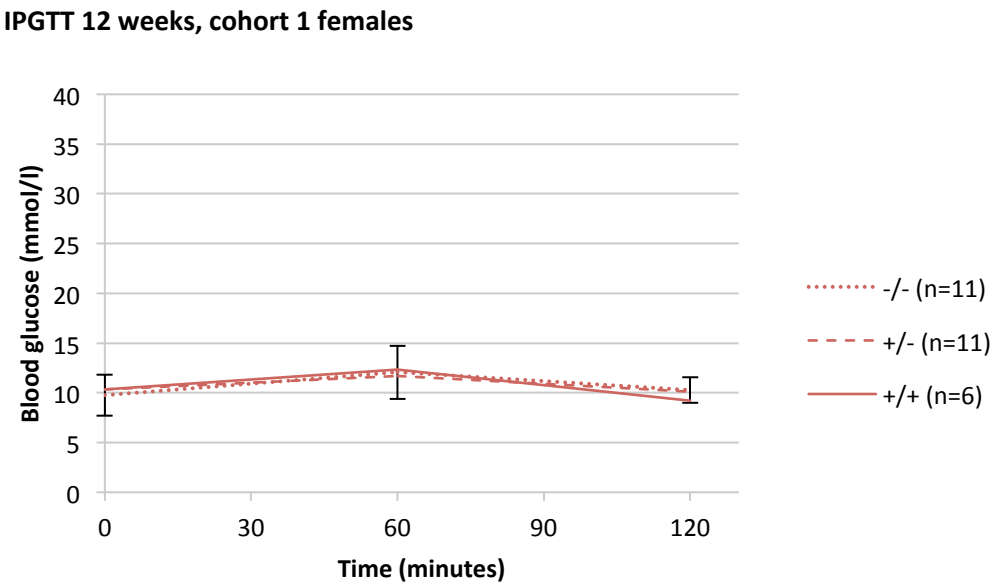


Figure 4.11: IPGTT in female mice at 12 weeks of age shows no difference between genotypes. Error bars (standard deviation) are only shown for homozygote (-/-) animals for clearer visualization.

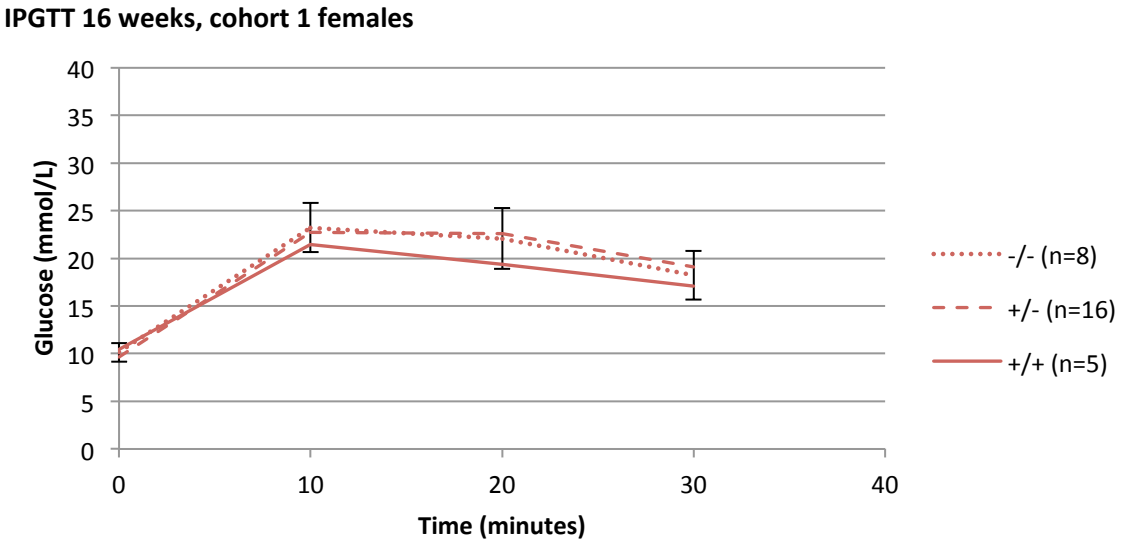


Figure 4.12: IPGTT for first 30 minutes after injection in female mice. As for the males, the genotypes did not differ. The female mice cleared glucose from the bloodstream quicker than the male mice with a peak glucose concentration at 10 minutes (see figure 4.12). Error bars showing standard deviation are only shown for homozygous mice for clearer visualization.

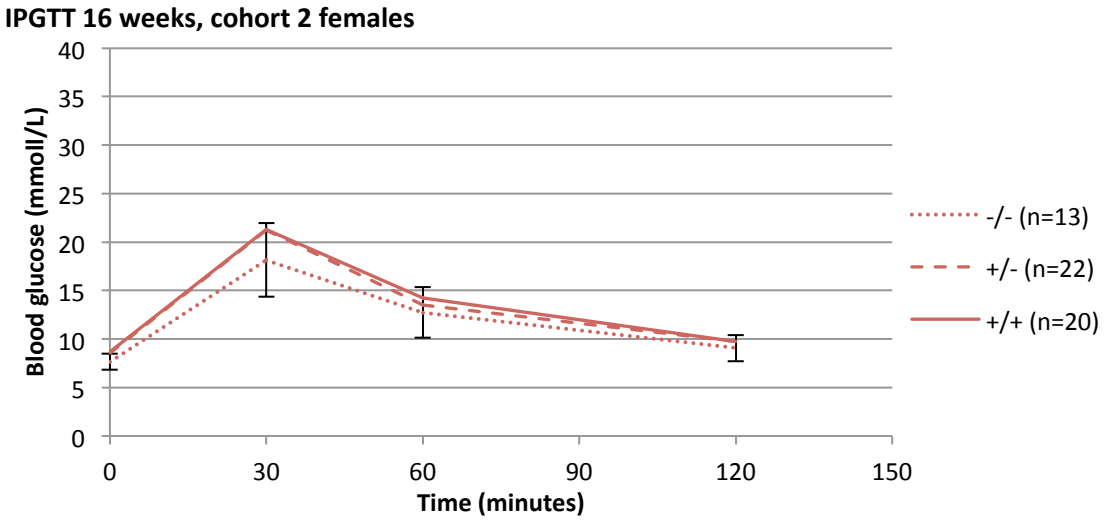


Figure 4.13: IPGTT results for females at 16 weeks. There were no differences between genotypes. Error bars (standard deviation) are only shown for homozygote mice (-/-) for clearer visualization.

4.3.3 c) Fasting insulin and glucagon do not differ between genotypes

Fasting insulin was measured at 16 weeks in both genders and all three genotypes in cohort 1 mice and then at 18 and 20 weeks in males in cohort 3 (13 wild type, 21 heterozygote, 14 homozygote mice [Figure 4.14]). All measurements were done by Daniel Andrews (MRC Harwell). The ages in cohort 3 were selected to correspond with the oral glucose tolerance tests where there were differences in glucose tolerance between genotypes. There was no difference in fasting insulin between the wild type, heterozygous and homozygous cohort 1 mice at 16 weeks. Fasting insulin also did not differ between genotypes in the cohort 3 mice at 18 or 20 weeks (Figure 4.14[c&d]).

Fasting plasma glucagon levels from cohort 3 male mice were measured at 20 weeks on a BioPlex 200 (Bio-Rad) by Daniel Andrew (MRC Harwell), but most of the values were undetectably low (below 100pg/mL) in all genotypes. The samples were re-analyzed by radioimmunoassay (RIA) by David Wiggins (OCDEM, University of Oxford). When the samples were re-analyzed about 50% of the samples were detectable, while 50% were too low to be measured on the gamma counter (Figure 4.15). Circulating blood glucagon levels are typically very low (<200pg/mL) and difficult to measure (Yang et al., 2012). No differences in glucagon were seen between genotypes. Overall there were no detected differences in fasting insulin or glucagon.

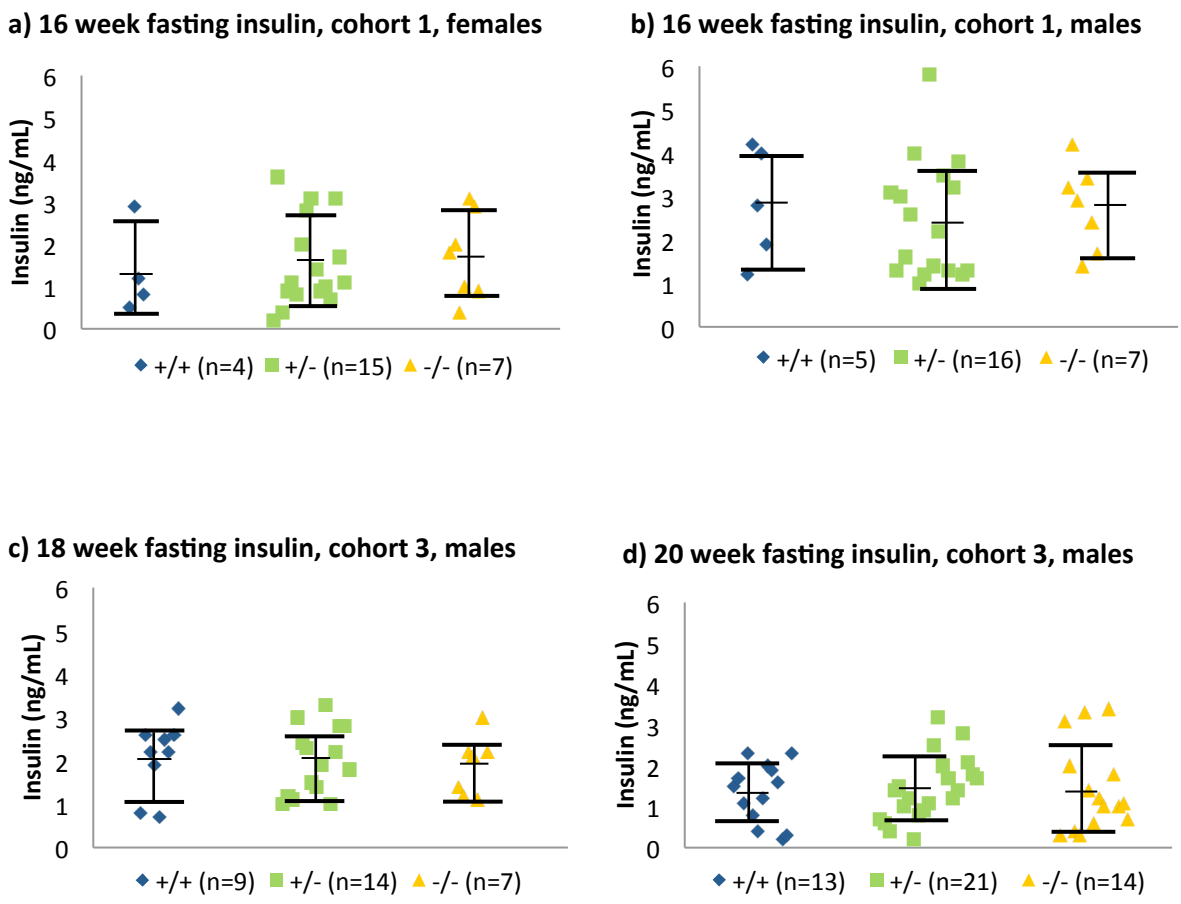


Figure 4.14: Fasting insulin for a) females at 16 weeks, and males at b) 16 weeks, c) 18 weeks, and d) 20 weeks. Mean is shown, and error bars are standard deviation. There are no differences between genotypes at any age.

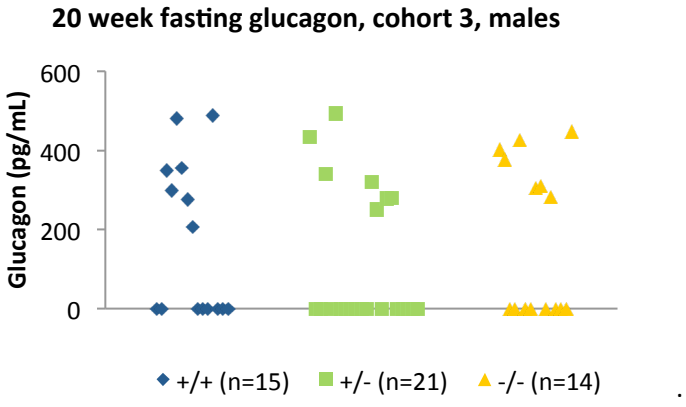


Figure 4.15: Fasting glucagon in males at 20 weeks of age. Radioimmunoassay detected glucagon levels in around 50% of samples. No mean or error bars are shown due to skew from unreadable low values. No difference was seen between genotypes.

4.3.3 d) Body composition does not differ between genotypes

I measured body composition (% fat, lean and water) of cohort 1 male and female mice of all three genotypes by EchoMRI at 10 and 18 weeks. There were no differences between genotypes in male mouse body composition at either age (Figure 4.16). At 10 weeks, the wild type females had a lower percentage lean mass of total body mass (mean = 11.5[2.8] grams) than female heterozygote mice (mean = 15.8[3.4] grams; p=0.04) and female homozygote mice (mean = 16.8[2.5] grams; p=0.009 [Figure 4.17a]). Female homozygote mice also had a higher percentage of fat mass compared to female wild type (mean = p=0.006 [Figure 4.17b]). At 18 weeks these difference had become non-significant (p>0.05 [Figure 4.17c&d]). Overall there were no robust changes in body composition in Y1064C allele carriers.

4.3.4 The *Adcy5* Y1064C variant has no effect on birth weight or size

Variants within *ADCY5* are associated with low birth weight in humans. In a separate cohort from the 3 phenotyping cohorts, I weighed 13 wild type, 37 heterozygote, and 12 homozygote mouse neonates of both genders as close to birth as possible. For explanations of power calculations to detect a difference see section 4.2.6 in Methods above. All neonates were weighed between 6 and 24 hours after birth. I found no difference between genotypes in the birth weight of the *Adcy5* ^{Y1064C} mice (Figure 4.18[a]). To rule extraneous factors, such as feeding close to birth, that would affect weight, I also dissected and measured the length of the leg bones (femur and tibia) of the neonates. I X-rayed the right hind leg of each mouse and measured bone lengths using Photoshop (Adobe). I found no difference in either femur or tibia lengths in these neonates between genotypes (Figure 4.18[b&c]).

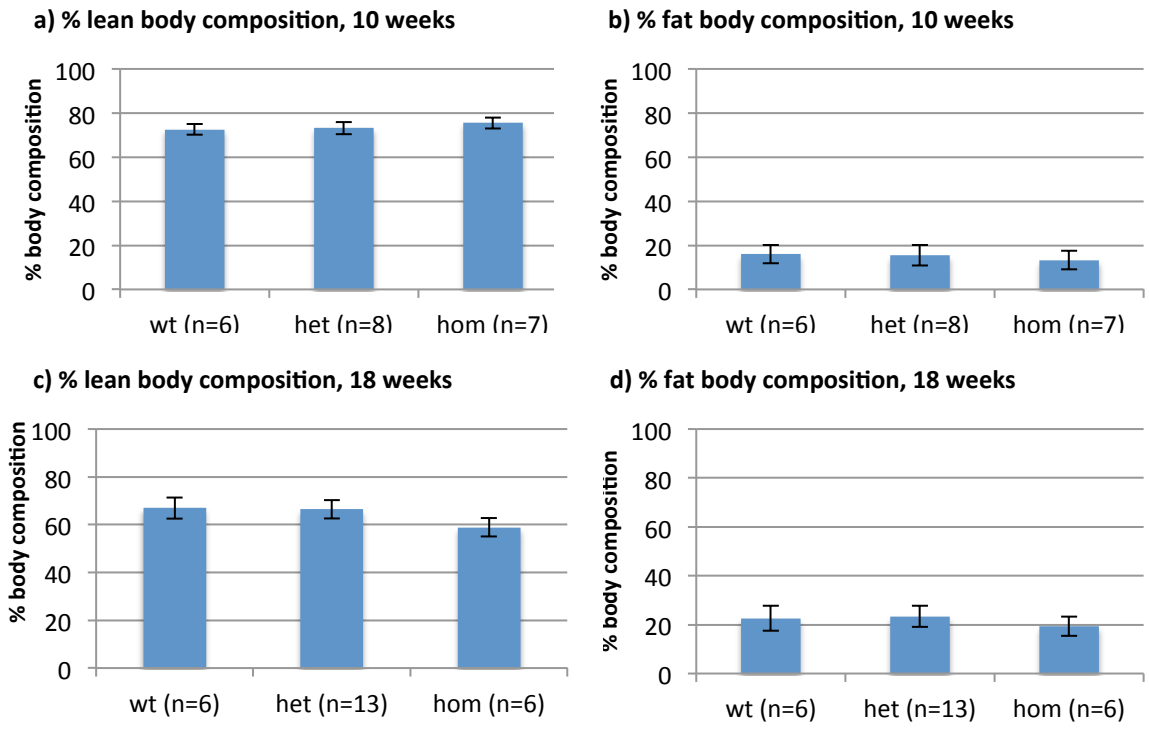


Figure 4.16: Body composition determined by EchoMRI at 10 weeks (a&b) and 18 weeks (c&d) of age in male mice. There were no differences between genotypes in fat or lean mass.

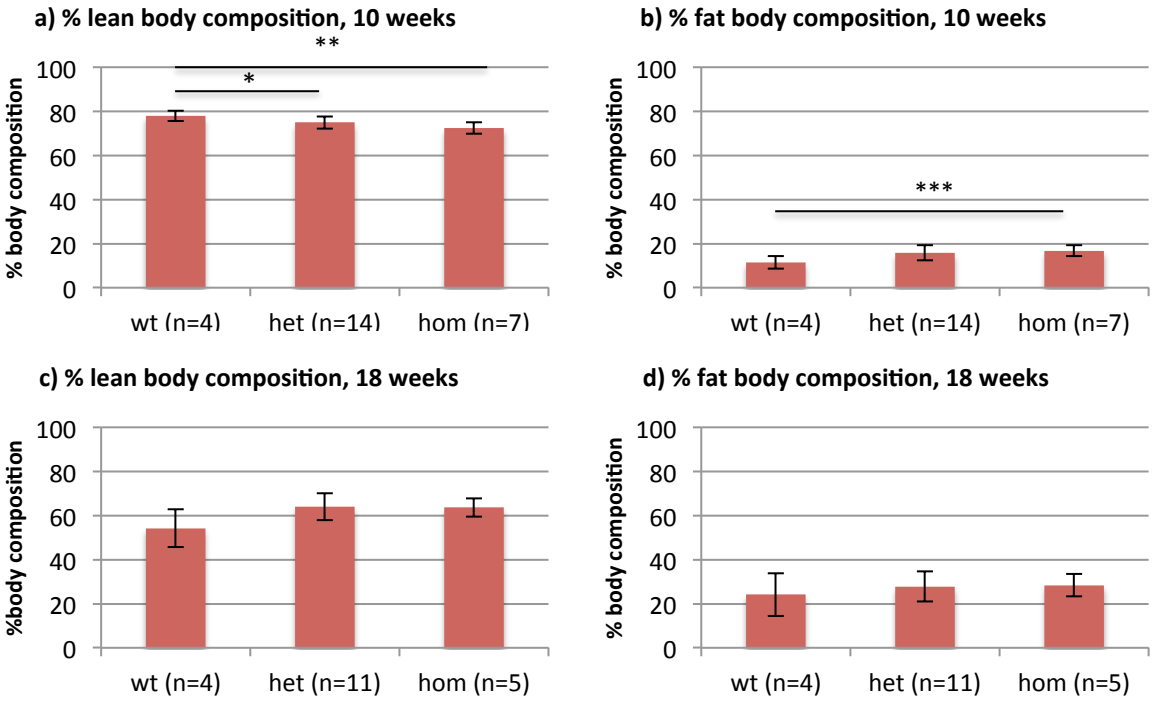


Figure 4.17: Body composition determined by EchoMRI at 10 weeks (a&b) and 18 weeks (c&d) of age in female mice. Significance is indicated by *p<0.05, **p≤0.01, ***p≤0.001. Female heterozygote and homozygote mice had lower % lean mass than female wild type mice at 10 weeks, and female homozygote mice had higher % fat mass than wild type mice at 10 weeks. All differences were gone by 18 weeks.

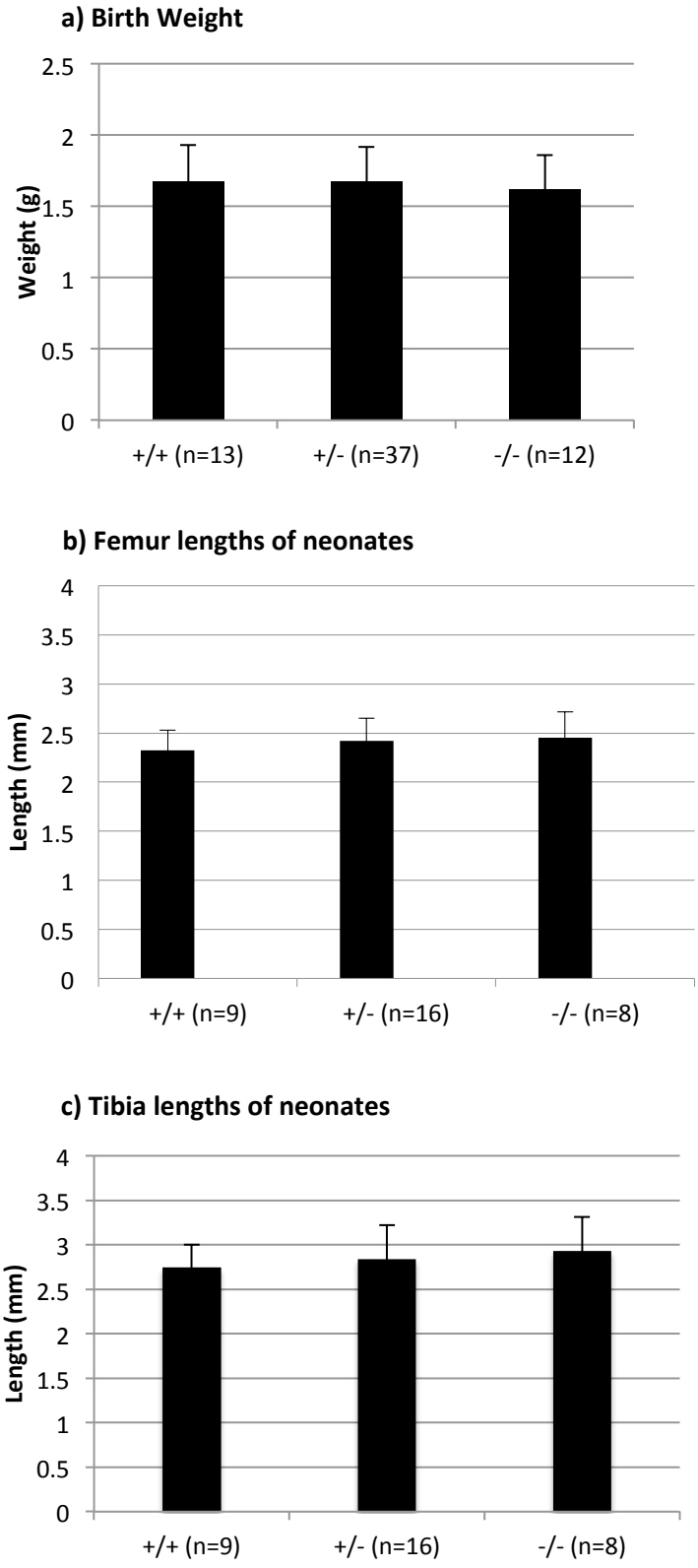


Figure 4.18: Birth weight was not different between wild type, heterozygote and homozygote *Adcy5*^{Y1064C} mice (a). Panel (b) and (c) show leg bone lengths, femur and tibia, of neonates. Leg bone length, a measurement of birth size, did not differ between genotypes.

4.3.5 Islets isolated from mice homozygous for the *Adcy5*^{Y1064C} mutation have altered glucagon secretion

To characterize islet function in the *Adcy5*^{Y1064C} mice, I measured insulin and glucagon secretion from islets isolated from Cohort 1 mice at 22 weeks of age (wild type n= 7; heterozygote n =4; homozygote n=4) and Cohort 2 mice at 22 (wild type n= 4; heterozygote n =3; homozygote n=3) and 30 weeks of age (wild type n=8; heterozygote n=7; homozygote n=9) under a variety of conditions. The incubation conditions were as follows:

a) basal glucose (1mM glucose) where one would expect to see low insulin secretion and high glucagon secretion in wild type mice;

b) high glucose (20mM glucose) where one would expect high insulin and low to moderate glucagon secretion in wild type mice;

c) 20mM glucose + forskolin where one would expect to see very high insulin and glucagon secretion in wild type mice because forskolin is a stimulator of insulin and glucagon secretion;

d) 1mM glucose + adrenaline where one would expect to see high glucagon stimulation and only minimal insulin stimulation in wild type islets because adrenaline is an inhibitor of insulin secretion but a glucagon agonist.

There was no difference in secretion between genders, so all male and female data is combined below. Islets from wild type mice secreted within the normal range for both insulin and glucagon. The first 3 mice of each genotype were completed with the help of Dave Wiggins (OCDEM, University of Oxford) and Reshma Ramracheya (OCDEM, University of Oxford). Reshma picked size-matched islets for the experiment as part of my training, and Dave performed islet isolations while training me. The remainder of the experiments was completed entirely by me.

4.3.5 a) Islets from mice homozygous for the *Adcy5*^{Y1064C} mutation secrete less glucagon than wild type in the presence of adrenaline

The most significant finding from the secretion experiments was that islets from mice homozygous for the *Adcy5* ^{Y1064C} mutation did not secrete glucagon in response to adrenaline. Islets from male and female 22 week old (cohort 1: wild type n= 7; heterozygote n =4; homozygote n=4) wild type and heterozygous mice secreted more glucagon in the presence of 5μM adrenaline in basal glucose than in basal glucose alone (wild type $p=6x10^{-5}$, heterozygotes p=0.0008), but islets homozygous for the Y1064C mutation were not stimulated by adrenaline (Figure 4.19, summarized results in Table 4.5). The islets from homozygote mice secreted only one third of the glucagon that islets from wild type and heterozygote mice did at 1mM glucose + 5μM adrenaline (wild type vs. homozygote mice p = 0.0007, heterozygote vs. homozygote mice p = 0.0009 [Figure 4.19]).

In vivo plasma glucose was lower in homozygote mice compared to wild type after an oral glucose tolerance test, and the difference was starker in older mice compared to younger ones (Section 4.3.2). Therefore, I chose to carry out a second set of experiments in cohort 2 mice at two ages: 22 and 30 weeks. Four wild type, 3 heterozygote and 3 homozygote mice were tested at 22 weeks of age to expand the sample size of 22-week-old mice. Eight wild type, 7 heterozygote and 9 homozygote mice were tested at the older time point of 30 weeks of age. These sample sizes were chosen to yield a power of 80% to detect a small effect (one fifth of the standard deviation) at a significance cut-off of 0.05 for each age group.

When analyzing the results, the overall levels of hormone secretion in experiments were higher from the cohort 1 mice possibly due to selection of larger islets or precision with the radioimmunoassay (RIA), so I did not combine the 22-week data from cohort 2 with the 22 week data from cohort 1 for any of the experiments presented below. The same trend was seen in 22-week-old male and female Cohort 2 mice, but the islets from homozygote mice did not secrete significantly less glucagon than those from wild type and heterozygote mice (Figure 4.20, Table 4.5). This may because of the

smaller sample size from cohort 2 at this age, or it may be that adrenaline failed to stimulate glucagon due to the poor response to glucagon in any of the genotypes. Islets from wild type cohort 2 mice at 22 weeks secreted only about 30% of the glucagon secreted by the islets from wild type cohort 1 mice. This is unlikely to be a problem with the islet isolation or experimental conditions as islets from several experiments performed similarly. It could be a problem with the RIA as all the samples were analyzed as part of the same RIA. Glucagon degrades quickly when left at room temperature. The samples may have been left out of the freezer longer than the cohort 1 samples. When I repeated these experiments in 30 week old mice from Cohort 2, I again found that islets from Y1064C homozygous mice secreted significantly less (about 60%) glucagon than islets from heterozygote mice and wild type mice in the presence of low glucose and adrenaline ($p=0.008$; $p=0.02$ [Figure 4.21, Table 4.5]). Again the glucagon levels secreted are lower than from the cohort 1 mice possibly due to glucagon degradation before assaying, but the difference between genotypes is still clear. The finding that glucagon secretion is impaired in homozygote mice in response to adrenaline is consistent in cohort 1 twenty-two week old mice and cohort 2 thirty week old mice.

4.3.5 b) Glucagon islet content does not differ between genotypes

Differences in glucagon islet content between genotypes may impact secretion and point to a mechanism of impaired glucagon production. Secretion can be interpreted as a percentage of content if islet contents differ dramatically. To measure glucagon content, I lysed the islet cells and measured glucagon released by RIA using the EURIA-Glucagon kit (Euro Diagnostica, Malmö, Sweden). Glucagon content did not differ between genotypes in islets isolated from cohort 1 22 week old mice (wild type [n=7]: mean[SEM] glucagon content 1321.8[158.91]pg/islet; heterozygote [n=4]: 1116.21[237.30]pg/islet; homozygote [n=4]: 1045.52[60.77]pg/islet [Figure 4.22]). Since islet content is not different, the results from the glucagon secretion experiments in this chapter are presented as picogram glucagon per islet per hour (pg/islet/hr) and are not normalized to content.

4.3.5 c) Glucose-mediated glucagon inhibition is not altered by the *Adcy5*^{Y1064C} mutation

Genotype did not influence glucose mediated glucagon inhibition. Islets from cohort 1 mice of all genotypes had similar glucagon secretion in both basal (1mM) glucose and 20mM glucose (Figure 4.23). Islets from cohort 2 mice at 22 weeks had similar glucagon secretion in basal, 8mM and 20mM glucose (Figure 4.24). Islets from cohort 2 mice at 30 weeks had similar glucagon stimulation in basal glucose and high glucose (15mM) (Figure 4.25). In the last experiment the concentration of high glucose had been changed from 20mM glucose to 15mM glucose to identify whether glucagon stimulation is restored by 15mM glucose as it is at 20mM glucose. Normal glucagon inhibition by glucose occurs most significantly between 5mM and 10mM glucose (Rorsman et al., 2008). Glucagon stimulation in 8mM glucose is not apparent in cohort 2 mice at 30 weeks. My results may not capture the effect fully due to small sample size and imprecision of islet size matching. It is also possible that 8mM glucose was not low enough to capture the inhibitory effect, and I may have seen an effect at 5mM glucose. However, the experiments were powered to detect any significant change between genotypes, and there was no difference. Results are summarized in Table 4.6.

4.3.5 d) Forskolin-stimulated glucagon secretion is impaired in the *Adcy5*^{Y1064C} mutant mice

Forskolin works directly to stimulate adenylate cyclases in the islet cells and is a stimulator of glucagon secretion (Hermansen, 1985). I tested glucagon secretion from isolated *Adcy5*^{Y1064C} islets from cohort 1 mice in the presence of 10µM forskolin. In basal glucose conditions in the presence of 10µM forskolin, islets from wild type and heterozygote mice were stimulated to secrete more glucagon than in basal glucose ($p=0.031$), but islets from homozygote mice did not respond to 10µM forskolin (Figure 4.26).

		Sample size	Cohort 1 Glucagon (pg/islet/hr)	Cohort 2 Glucagon (pg/islet/hr)	
		(biological reps x technical reps)	22 weeks mean(SEM)	22 weeks mean(SEM)	30 weeks mean(SEM)
1mM glucose	+/+	n=7x3, 4x3, 8x3	2.80(0.34)	1.62(0.13)	1.18(0.03)
	+/-	n=4x3, 3x3, 7x3	3.79(0.44)	1.27(0.18)	2.53(0.10)
	-/-	n=4x3, 3x3, 9x3	3.64(0.39)	1.83(0.14)	1.03(0.02)
1mM glucose + 5µM adrenaline	+/+	n=7x3, 4x3, 8x3	15.85(1.9)	7.11(0.65)	10.94(1.04)
	+/-	n=4x3, 3x3, 7x3	23.23(3.19)	6.52(0.51)	6.83(0.18)
	-/-	n=4x3, 3x3, 9x3	6.18(1.16)	5.55(0.27)	4.44(0.81)

Table 4.5: Summary of glucagon secretion in basal glucose and basal glucose + adrenaline.

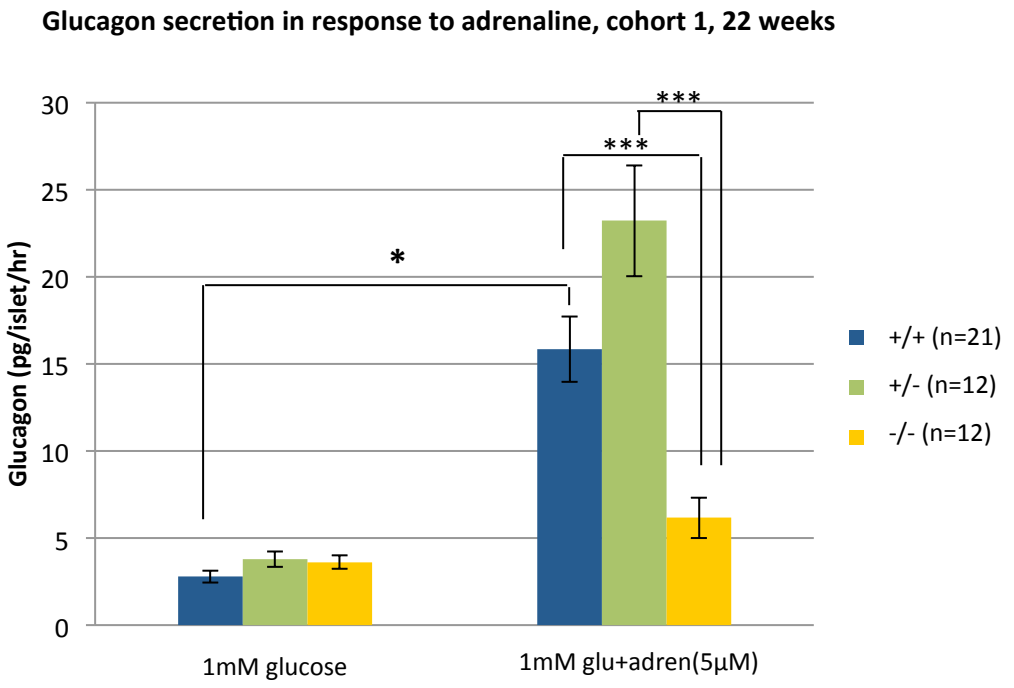


Figure 4.19: Glucagon secretion from isolated islets from cohort 1 mice at 22 weeks of age in basal glucose and basal glucose + adrenaline. Sample sizes are 21 wild type comprising 7 biological replicates and 3 technical replicates, 12 heterozygote comprising 4 biological replicates and 3 technical replicates, and 12 homozygote comprising 4 biological replicates and 3 technical replicates. Significance is indicated by *p<0.05, **p≤0.01, ***p≤0.001. Adrenaline is a significant stimulator of glucagon secretion in wild type and heterozygotes. Glucagon secretion from islets homozygous for the Y1064C mutation in the presence of adrenaline was significantly impaired compared to wild type and heterozygote mice.

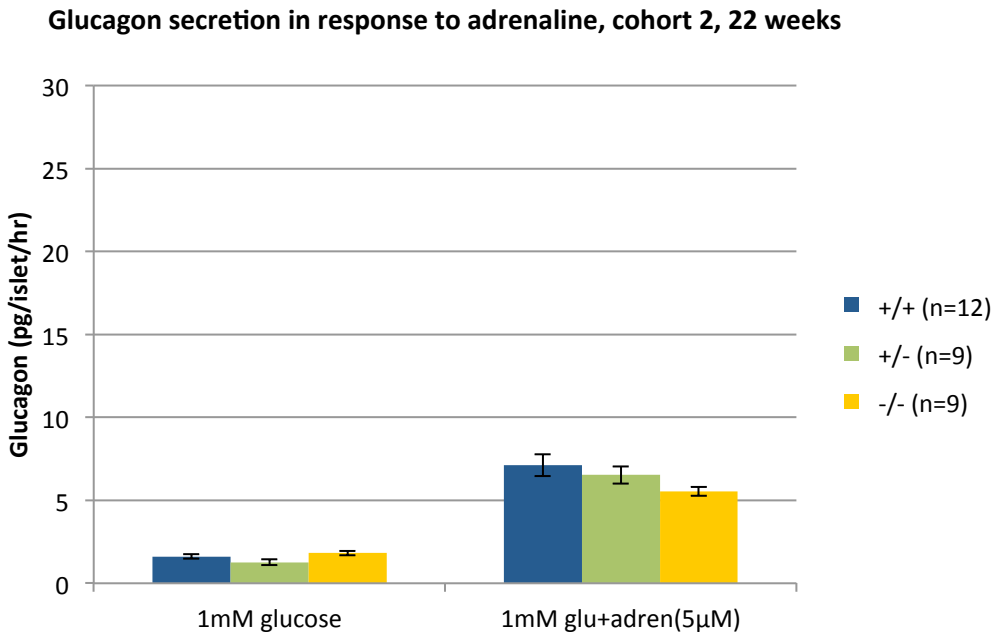


Figure 4.20: Glucagon secretion in cohort 2 mice 22 weeks of age in response to 1mM glucose + 5µM adrenaline is not different between genotypes. Sample size consists of biological replicates x technical replicates. Technical replicates = 3 for all experiments.

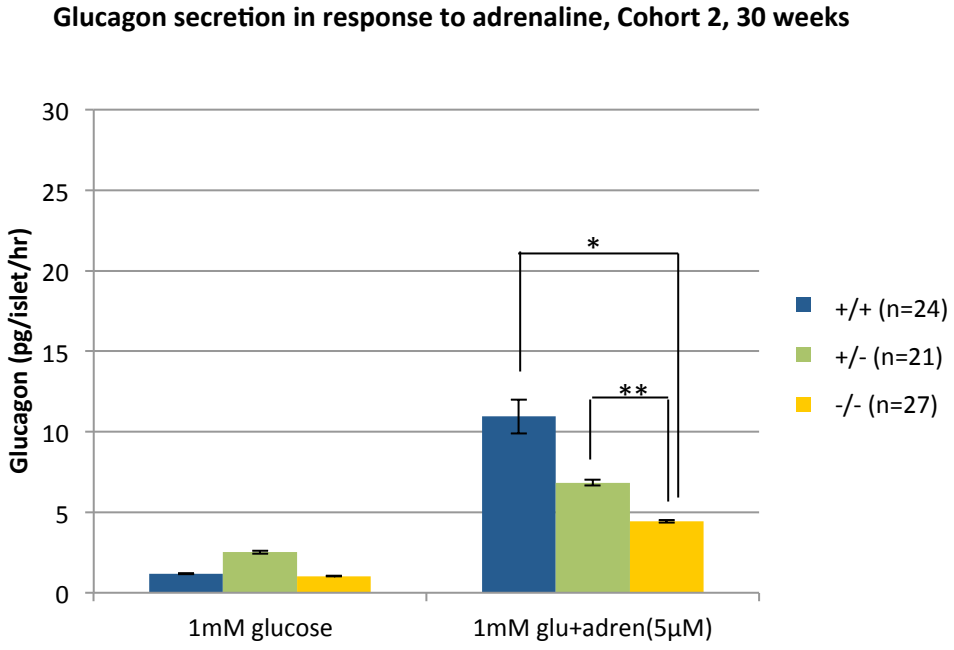


Figure 4.21: Glucagon secretion in cohort 2 mice 30 weeks of age in response to 1mM glucose + 5µM adrenaline is impaired in homozygote mice compared to heterozygote, and in homozygote mice compared to wild type. Significance is indicated by *p<0.05, **p≤0.01. Sample size consists of biological replicates x technical replicates with technical replicates equal to 3 in all experiments.

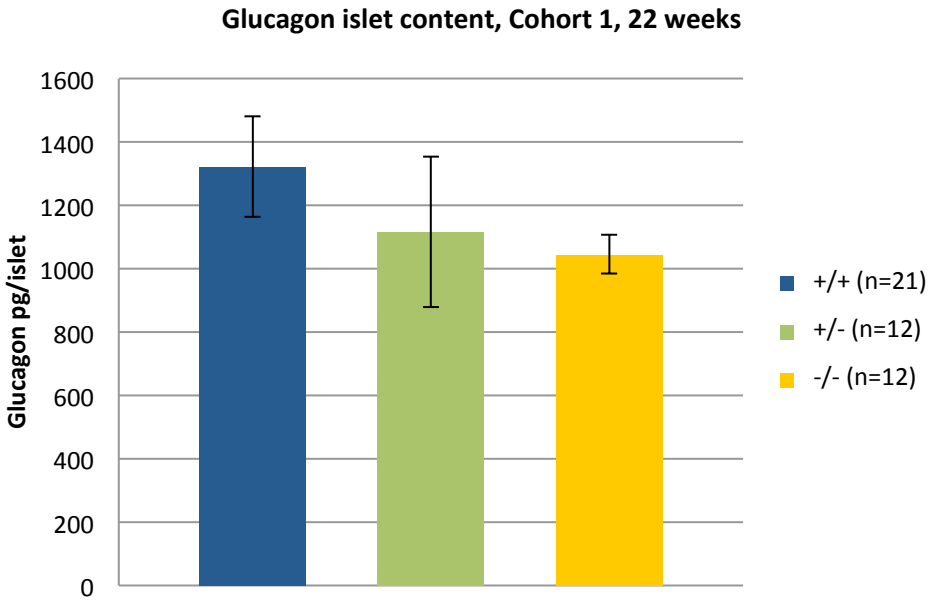


Figure 4.22: Glucagon content in isolated islets does not differ between genotypes. Sample size consists of biological replicates x technical replicates with technical replicates equal to 3 in all experiments.

		Sample size	Cohort 1 Glucagon (pg/islet/hr)	Cohort 2 Glucagon (pg/islet/hr)	
		Biological reps x technical reps	22 weeks mean(SEM)	22 weeks mean(SEM)	30 weeks mean(SEM)
1mM glucose	+/+	n=21,12,24	2.80(0.34)	1.62(0.13)	1.18(0.03)
	+/-	n=12,9,21	3.79(0.44)	1.27(0.18)	2.53(0.10)
	-/-	n=12,9,27	3.64(0.39)	1.83(0.14)	1.03(0.02)
8mM glucose	+/+	n=-,12,-		1.58(0.09)	
	+/-	n=-,9,-		1.94(0.20)	
	-/-	n=-,9,-		0.96(0.12)	
15mM glucose	+/+	n=-,-,24			2.04(0.07)
	+/-	n=-,-,21			3.84(0.15)
	-/-	n=-,-,27			2.61(0.07)
20mM glucose	+/+	n=21,12,-	1.63(0.20)	2.47(0.13)	
	+/-	n=12,9,-	2.52(0.34)	1.55(0.09)	
	-/-	n=12,9,-	2.03(0.22)	2.40(0.39)	

Table 4.6: Summary of glucagon secretion data from wild type, heterozygote and homozygote *Adcy5*^{Y1064C} mice in response to 1mM, 8mM, 15mM and 20mM glucose. Sample sizes correspond to each cohort and age group of mice tested in that condition and consist of biological replicates x technical replicates with technical replicates equal to 3 for all experiments.

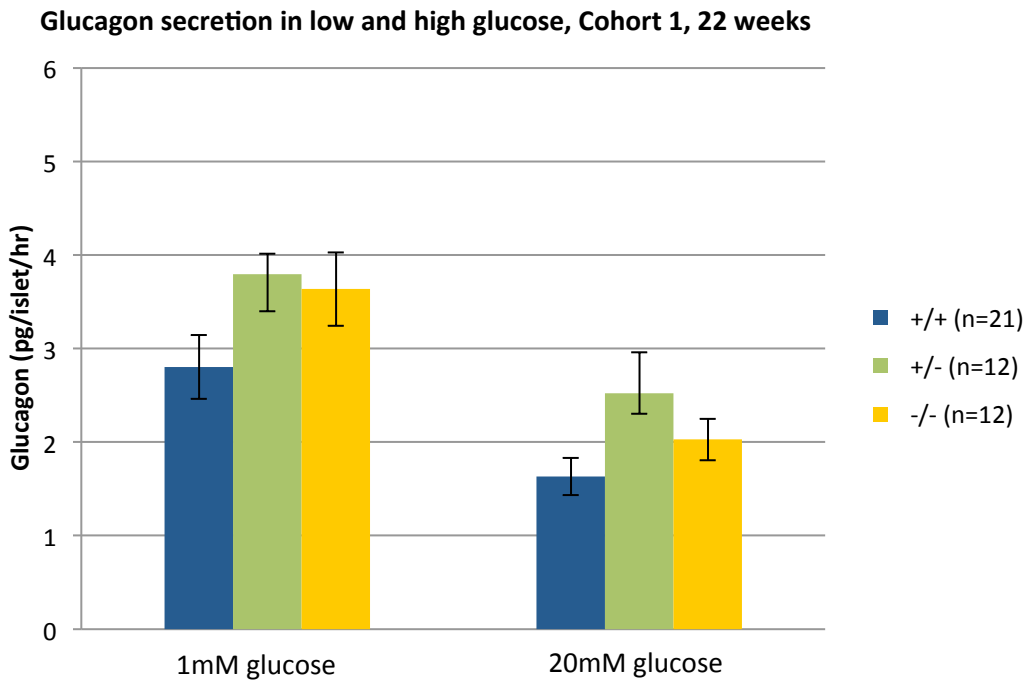


Figure 4.23: Glucagon secretion from isolated islets from cohort 1 mice in low and high glucose. Glucagon inhibition by glucose is not different between genotypes. Sample size consists of biological replicates x technical replicates with technical replicates equal to 3 for all experiments.

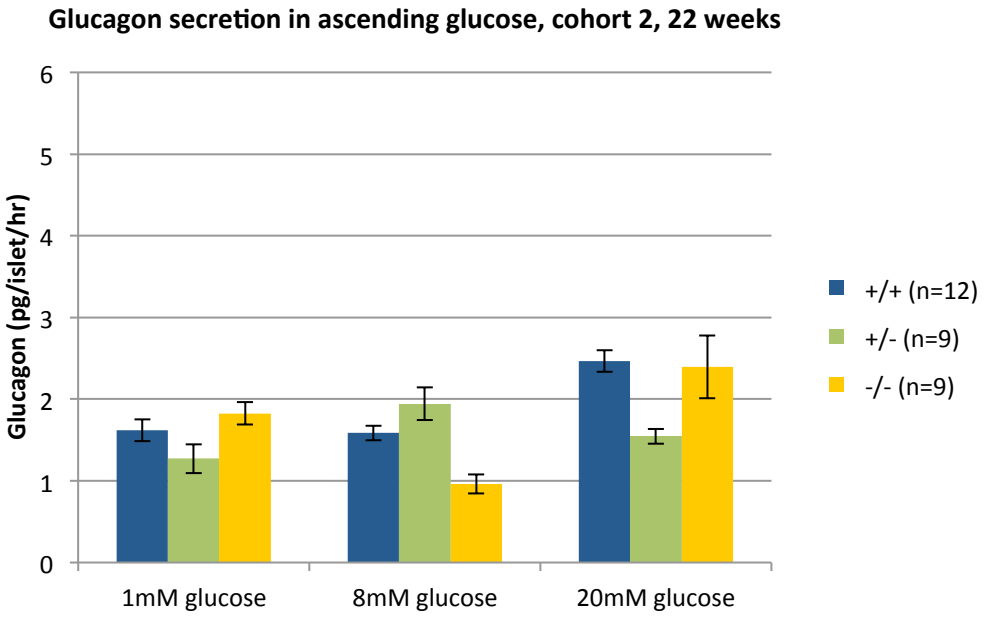


Figure 4.24: Glucagon secretion in islets from 22 week old mice in response to 3 concentrations of glucose is not difference between genotypes. Stimulation of glucagon secretion was not achieved at basal glucose in these experiments. Sample size consists of biological replicates x technical replicates(3).

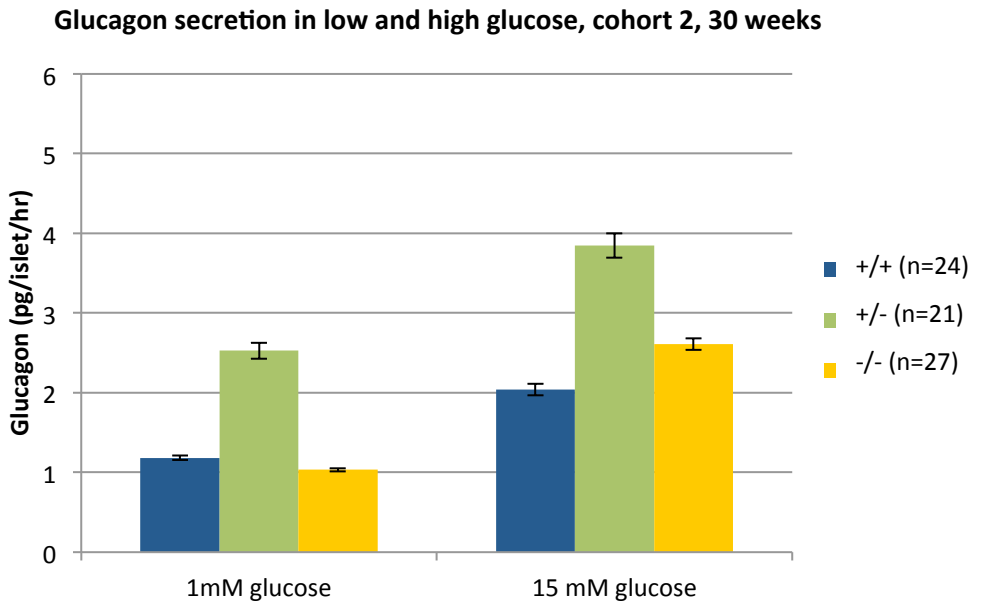


Figure 4.25: Glucagon inhibition by glucose in 30 week old mice did not differ between genotypes in response to 1mM glucose or 15mM glucose. Stimulation of glucagon secretion was not achieved at low glucose. Sample size consists of biological replicates x technical replicates (3).

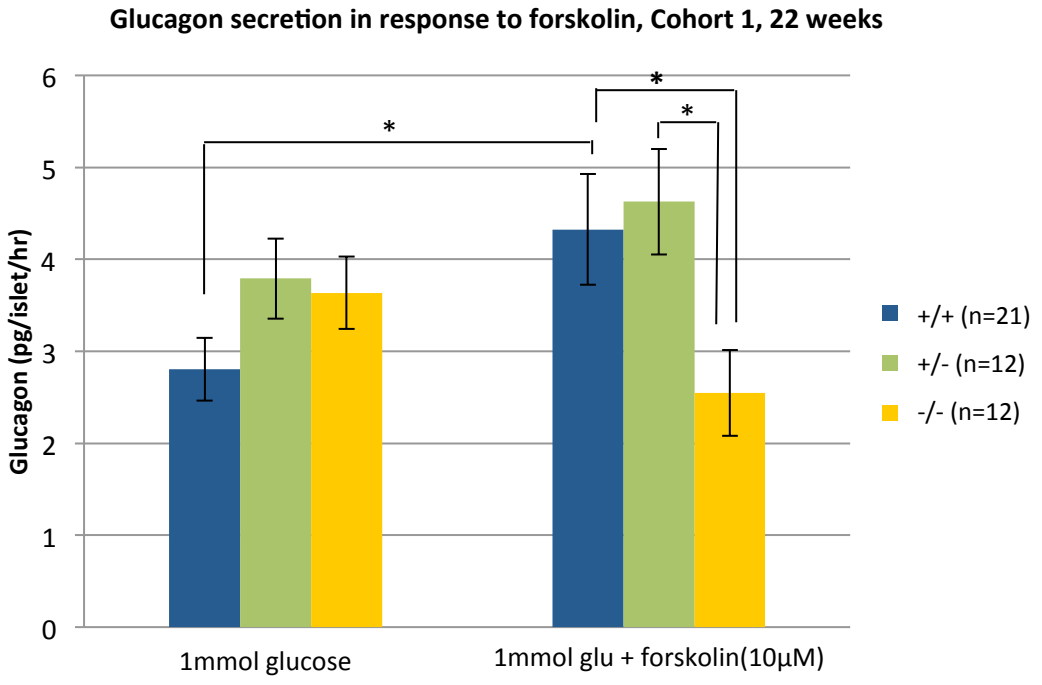


Figure 4.26: Glucagon secretion in the presence of basal glucose and basal glucose + forskolin. Significance is indicated by *p<0.05, **p≤0.01. Forskolin did not stimulate glucagon secretion in islets from Y1064C homozygous mice. Islets from homozygote mice secreted significantly less glucagon than islets from wild type and heterozygote mice. Sample size consists of biological replicates x technical replicates (3).

Islets from cohort 1 *Adcy5* ^{Y1064C} homozygote mice secreted nearly 50% less glucagon than wild type and heterozygote mice in basal glucose + 10μM forskolin (wild type [n=7 biological replicates, 3 technical replicates]: mean [SEM] glucagon 4.32 [0.60] pg/islet/hr; heterozygote [n=4, 3]: 4.63 [0.57] pg/islet/hr; homozygote [n=4, 3]: 2.55 [0.47] pg/islet/hr; wildtype vs homozygote, t-test p=0.034; heterozygote vs homozygote, p=0.036).

In summary, there is evidence of impairment in glucagon secretion in response to adrenaline in the *Adcy5*^{Y1064C} homozygous mice at 22 weeks in cohort 1 and at 30 weeks in cohort 2. Response to forskolin in homozygous mice is also impaired. I found no evidence of a difference in glucose-mediated glucagon inhibition between genotypes. Glucagon content did not differ between genotypes.

4.3.6 Insulin secretion from isolated islets in response to low glucose is elevated in cohort 1 Y1064C allele carriers

In addition to glucagon secretion, I measured insulin secretion in response to basal (1mM) glucose, middle and high glucose, forskolin, and adrenaline in cohort 1 mice at 22 weeks and at 22 and 30 weeks in cohort 2 mice in response to glucose and adrenaline. Most of the mice are the same as those included in the glucagon experiments with the exception of a) there are more mice included in the insulin secretion experiments from cohort 1 because several experiments were carried out before the decision was made to measure glucagon in addition to insulin; and b) one experiment for which glucagon was successfully measured in 22 week old mice from cohort 2 was unable to be measured for insulin because of an error with the gamma reader during RIA. Insulin was measured by RIA in ng/islet/hr. Some of the results of the insulin secretion experiments were conflicting between mice of different ages.

4.3.6 a) Glucose-stimulated insulin secretion is not consistently different between genotypes

At basal (1mM) glucose in 22 week old mice, insulin secretion was significantly elevated in 10 heterozygote mice compared to 8 wild type mice (p=0.009) and in 8 homozygote compared to 8 wild type mice (p=0.007), Table 4.7 and Figure 4.27. Elevated basal secretion may be the result of overstimulated islets, insulin leakage caused by damage to the cell membrane, an alteration in beta cell function, or it may be the result of a difference in islet insulin content. Overstimulation is the result of prolonged exposure to very high levels of glucose, but these mice are not hyperglycemic, so it is unlikely the islets would be overstimulated (Bjorklund et al., 2000). Leakage would release large amounts of insulin and values would not be consistent between replicates. The elevation in basal insulin secretion in the heterozygote and homozygote mice was subtle and consistent between replicates, so it is more likely to be the result of a discrepancy in islet insulin content or a change in beta cell function. It is possible that the difference seen is an experimental error because the levels of insulin secreted at basal glucose are extremely low and difficult to measure reliably. There was no difference in insulin secretion between genotypes at 20mM glucose (Table 4.7, Figure 4.27).

The data from cohort 2 mice in response to glucose did not confirm the results from cohort 1 mice. Nearly all insulin secretion values collected from cohort 2 mice were nearly 10x the values collected from cohort 1. This could be due to selection of larger islets than were selected from cohort 1 mice. The data were analyzed for comparison between genotypes with the caveat of unexplained highly discrepant raw values. Twenty-two week old cohort 2 mice (3 wild type, 2 heterozygote, 2 homozygote) showed no differences between genotypes in response to glucose at basal (1mM), middle (8mM) or high (20mM) glucose concentrations (Table 4.7, Figure 4.28). The sample size for this experiment was lower than the experiment in cohort 1 mice, so the phenotype could have gone undetected. In cohort 2 mice at 30 weeks, the islets from 5 wild type, 6 heterozygote, and 7 homozygote mice were exposed to 1mM glucose and 15mM glucose. The islets from heterozygote mice released significantly more insulin than the wild type in 15mM glucose (p=0.011, Table 4.7,

Figure 4.29). This may be the result of a leaking islet or poorly size-matched islets since the result was not seen in the other experiments. The Y1064C homozygote mice did not differ from wild type. In summary, basal insulin secretion from heterozygote and homozygote mice was elevated in cohort 1 mice at 22 weeks, but I did not see this result again in cohort 2 mice. It is possible that the result in the first cohort was an error or that it was related to differences in islet insulin content. I tested the latter by measuring insulin content in these islets.

4.3.6 b) Insulin content is significantly reduced in *Adcy5* ^{Y1064C} allele carriers

Islets from 8 cohort one 22 week old *Adcy5* Y1064C homozygous mice contained ~20% less insulin than islets from 8 cohort one 22 week old mice wild type mice (p=0.041,) and heterozygous islets from 10 cohort one 22 week old mice contained ~30% less insulin than wild type islets from 8 mice (p=0.0008) (Figure 4.30). The lower insulin content in the mutant islets should be taken into account when interpreting secretion results as these islets have 20-30% less insulin stores to secrete. When content was taken into account for cohort 1 mice (8 wild type, 10 heterozygote, 8 homozygote), basal insulin secretion is still significantly elevated in islets from Y1064C heterozygote and homozygote mice compared to wild type, and p values are more significant than when secretion is not normalized to content (wild type vs. heterozygote: p=0.006, wild type vs. homozygote: p=0.004 [Figure 4.31). The islets with lower insulin content may be working to secrete a greater percentage of their content than wild type. I was unable to collect insulin content data from cohort 2 mice because islet pellets were left at room temperature for 48 hours and the insulin allowed to degrade.

4.3.6 c) There were no consistent changes in insulin secretion in response to forskolin or adrenaline between genotypes

When stimulating with forskolin, I found no repeatable difference in insulin secretion between genotypes in the cohort 1 mice tested (Figure 4.32). Cohort 2 mice were not tested in the presence of forskolin because no interesting differences were seen in cohort 1. Adrenaline stimulated wild type islets to secrete insulin under low glucose conditions ($p=9.2 \times 10^{-5}$) (Figure 4.33). This was an

unexpected result because adrenaline is generally understood to suppress insulin secretion. Adrenaline promotes the fight-or-flight response and prepares the body for physical exertion. Conversely, insulin is secreted as part of the parasympathetic nervous system when the body is at rest. In all secretion experiments from isolated islets, insulin was consistently stimulated by adrenaline, not inhibited. Several subtypes exist of the alpha and beta-adrenergic receptors, all triggering disparate pathways and changes to metabolism (Arnall et al., 1986). It is possible that in these islets, the receptors inhibiting insulin release (namely the alpha-adrenergic receptors) are not functioning, or adrenaline is stimulating other receptors more effectively. Nevertheless, the result is consistent in wild type mice in all experiments, so the physiology is likely a strain-specific trait rather than a result of the *Adcy5* Y1064C mutation.

While islets from wild type mice secreted more insulin in the presence of adrenaline, islets from Y1064C allele carriers had no response to adrenaline. The islets from heterozygote and homozygote mice were already hyper secreting at basal glucose, and no further stimulation from adrenaline was seen. Due to the discrepancy in islet content discussed above, I normalized the cohort 1 data to insulin content and found the same result; islets from Y1064C allele carriers did not respond to adrenaline while wild type islets responded by secreting more insulin (p=0.0006, Figure 4.34).

When repeated with islets from Cohort 2 mice (22 weeks old), all genotypes responded to adrenaline by secreting more insulin in the presence of 1mM glucose + adrenaline than in 1mM glucose (wild type p=0.008, heterozygote mice p=0.04, homozygote mice p=0.0005 [Figure 4.27]).

In 30-week islets, only wild type and islets from Y1064C homozygote mice responded to adrenaline (wild type p=0.0008, *Adcy5*$^{Y1064C-/-}$ p=0.001 [Figure 4.28]). The results are inconsistent between the experiments for insulin secretion in response to adrenaline, but I found no repeatable evidence of a difference in insulin secretion between the *Adcy5* ^{Y1064C} mutation carriers and wild type. Islets from all genotypes may be stimulated to secrete more insulin in the presence of adrenaline, but allele

carriers with already elevated basal secretion did not respond in a meaningful way to the subtle stimulation of adrenaline.

In summary, islets from cohort 1 *Adcy5* Y1064C allele carriers had elevated insulin secretion in basal glucose compared to wild type, but this result was not confirmed in cohort 2 mice at 22 or 30 weeks. There was a difference in insulin content between the genotypes. Islets from heterozygote mice contain ~30% less insulin than wild type islets, and islets from homozygote mice contain ~20% less than wild type. When normalized to insulin content, the difference in basal insulin secretion between genotypes is even starker. Unfortunately, insulin content could not be collected from cohort 2 islets due to sample degradation. Forskolin had the same effect on wild type, heterozygote and homozygote islets. Stimulation in response to adrenaline was varied, but I found no repeatable differences between genotypes.

		Sample size	Cohort 1 Insulin (ng/islet/hr)	Cohort 2 Insulin (ng/islet/hr)	
		Biological reps x technical reps	**22 weeks** mean(SEM)	**22 weeks** mean(SEM)	**30 weeks** mean(SEM)
1mM glucose	+/+	n=24,12,15	0.009(0.001)	0.039(0.004)	0.071(0.003)
	+/-	n=30,6,18	0.028(0.005)	0.064(0.008)	0.160(0.011)
	-/-	n=24,6,21	0.033(0.007)	0.018(0.002)	0.068(0.003)
8mM glucose	+/+	n=-,9,-		0.057(0.003)	
	+/-	n=-,6,-		0.128(0.016)	
	-/-	n=-,6,-		0.028(0.001)	
15mM glucose	+/+	n=-,-,15			0.226(0.007)
	+/-	n=-,-,18			0.545(0.042)
	-/-	n=-,-,21			0.339(0.003)
20mM glucose	+/+	n=24,9,-	0.063(0.014)	0.595(0.045)	
	+/-	n=30,6,-	0.047(0.007)	0.577(0.084)	
	-/-	n=24,6,-	0.069(0.008)	0.342(0.031)	

Table 4.7: Insulin secretion in response to varying concentrations of glucose in cohort 1 and 2 mice. Sample sizes correspond to the cohorts and age groups shown in the final three columns and consists of biological replicates x technical replicates when technical replicates is equal to 3 in all experiments.

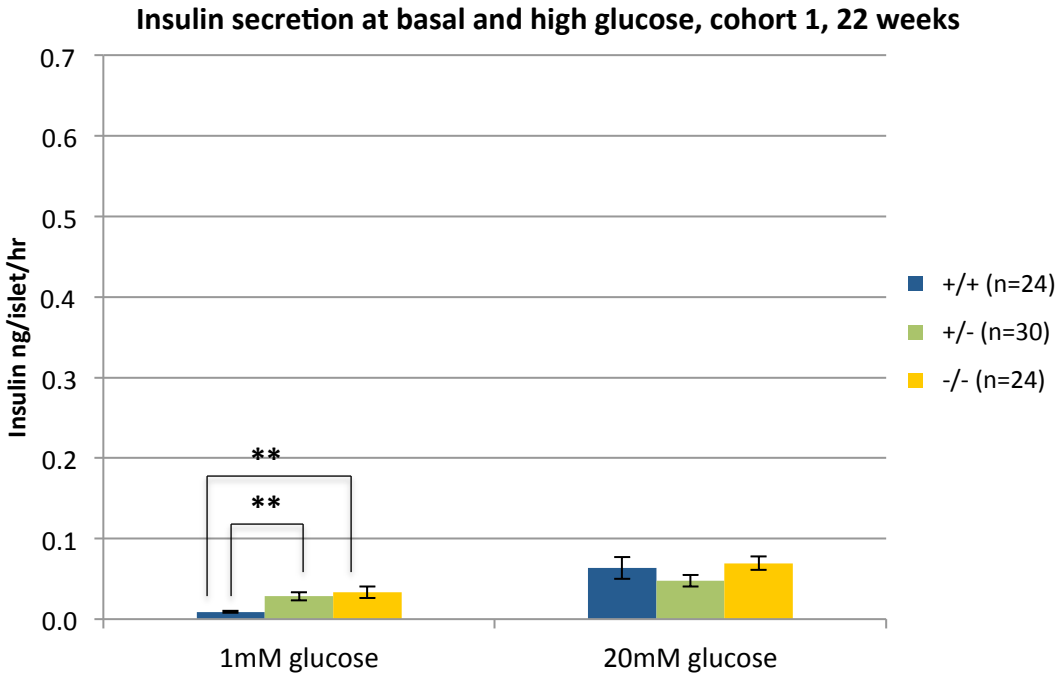


Figure 4.27: Basal insulin secretion is significantly elevated in heterozygous and homozygous *Adcy5*^{Y1064C} mice. Significance is indicated by *p<0.05, **p≤0.01). Secretion at high glucose is the same in all genotypes. Sample size consists of biological replicates x technical replicates (3).

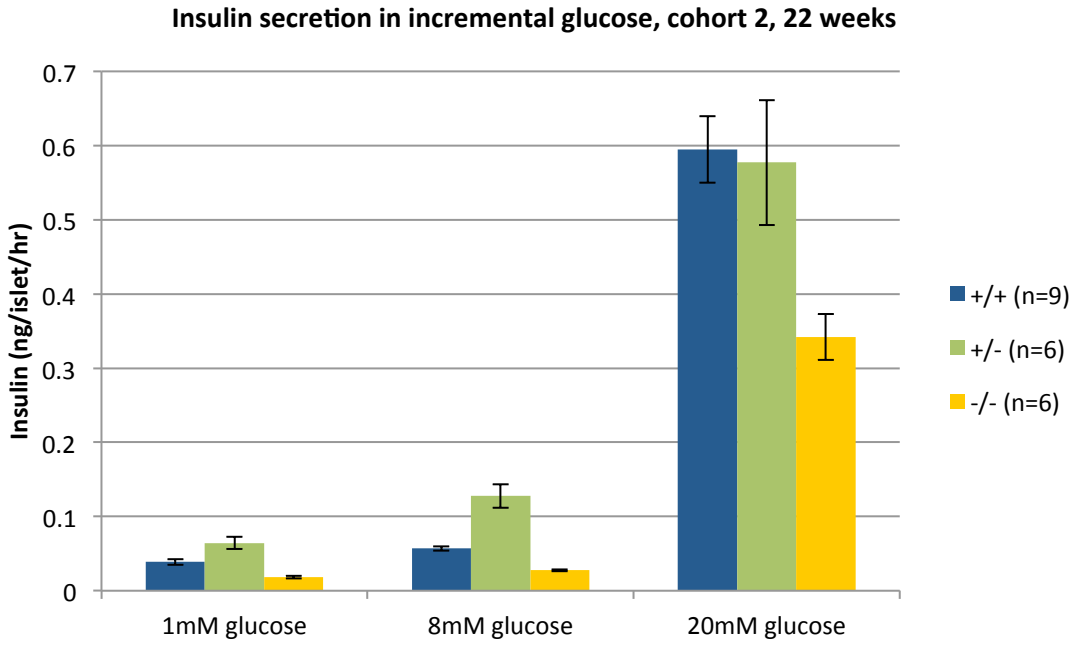


Figure 4.28: Insulin secretion in 22 week old mice is not different between genotypes at low, middle, or high glucose. Insulin secretion values measured were ~10x greater than were measured in the same conditions from cohort 1 mice.

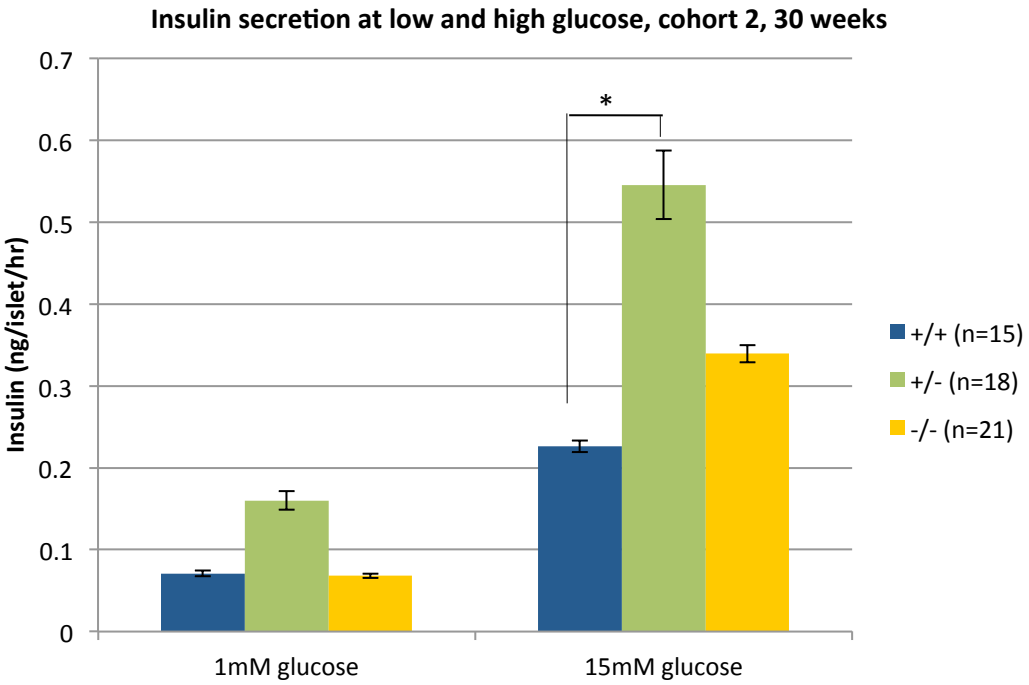


Figure 4.29: Insulin secretion in 30-week-old cohort 2 mice is elevated in heterozygotes in response to 15mM glucose. Significance is indicated by *p<0.05, **p≤0.01. Sample sizes are comprised of biological replicates x technical replicates where 3 technical replicates were used for all experiments.

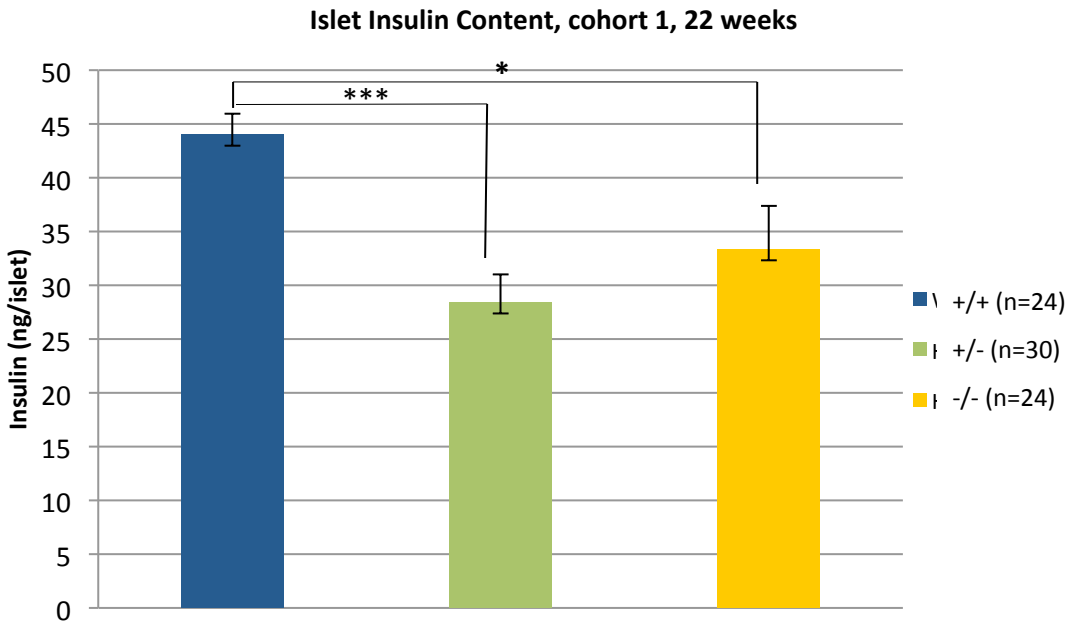


Figure 4.30: Insulin content from isolated islets of 22 week old mice is lower in *Adcy5* Y1064C heterozygote and homozygote mice. Significance is indicated by *p<0.05, *p≤0.01, ***p≤0.001. Islets from *Adcy5* Y1064C heterozygote and homozygote mice contain almost 30% less insulin than wild type islets. Sample sizes are comprised of biological replicates x technical replicates (3).

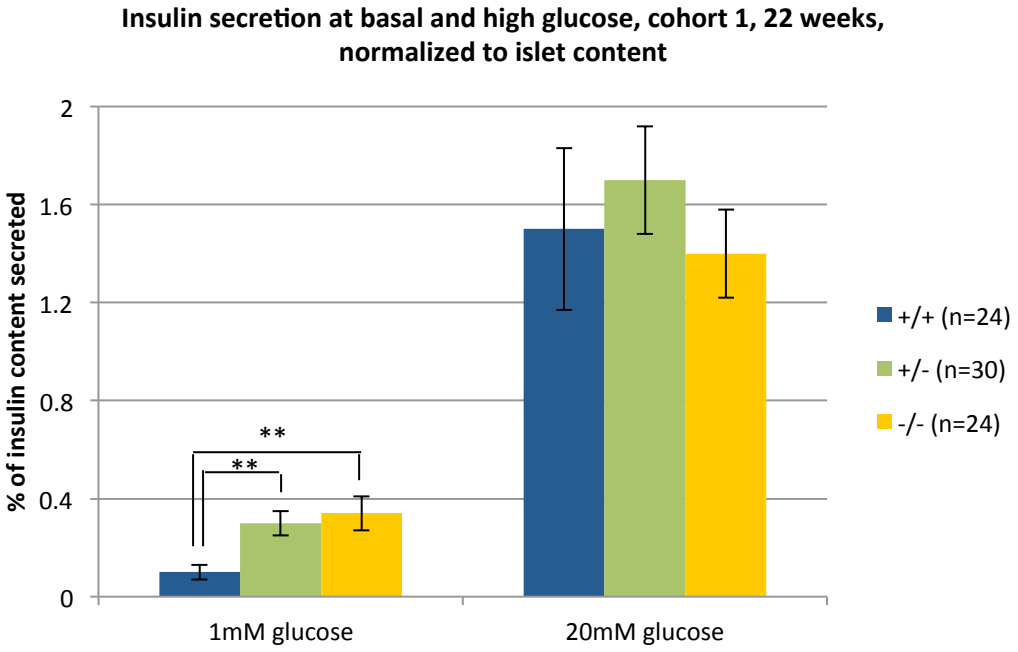


Figure 4.31: When insulin secretion is normalized to islet content, basal insulin secretion is elevated in islets from Y1064C allele carriers. Significance is indicated by *p<0.05, **p≤0.01. Sample sizes are comprised of biological replicates x technical replicates (3).

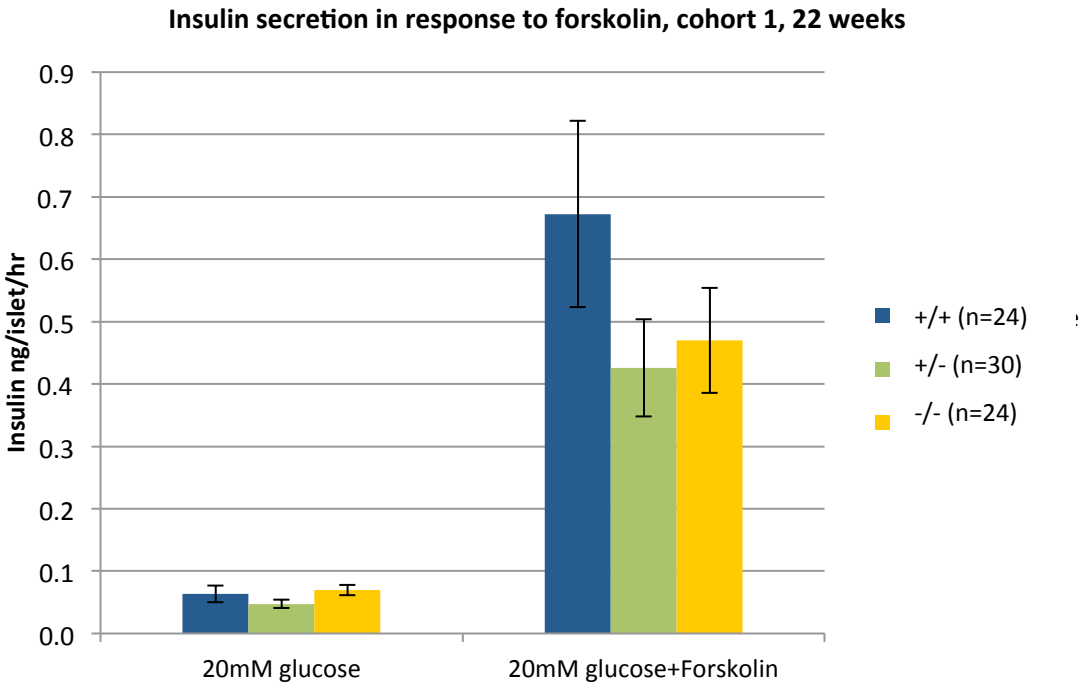


Figure 4.32: Insulin secretion from isolated islets stimulated with forskolin. There was no measurable difference between genotypes. Sample sizes are comprised of biological replicates x technical replicates (3).

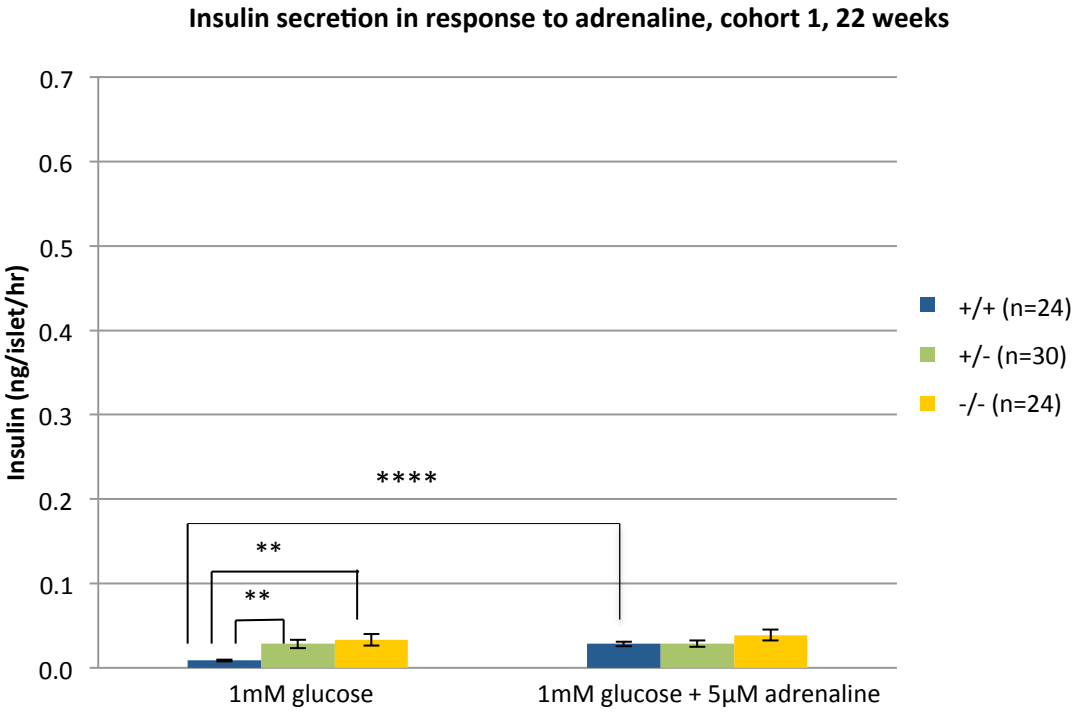


Figure 4.33: Insulin secretion in response to adrenaline from islets from 22 week old cohort 1 mice. Significance is indicated by *p<0.05, **p≤0.01, ***p≤0.001, ****p≤1x10^{-4}. Sample sizes consist of biological replicates x technical replicates (3). Wild type isolated islets were significantly stimulated by adrenaline in low glucose, but islets from Y1064C allele carriers did not respond to adrenaline by secreting more insulin (1mM glucose is same as in Figure 4.15).

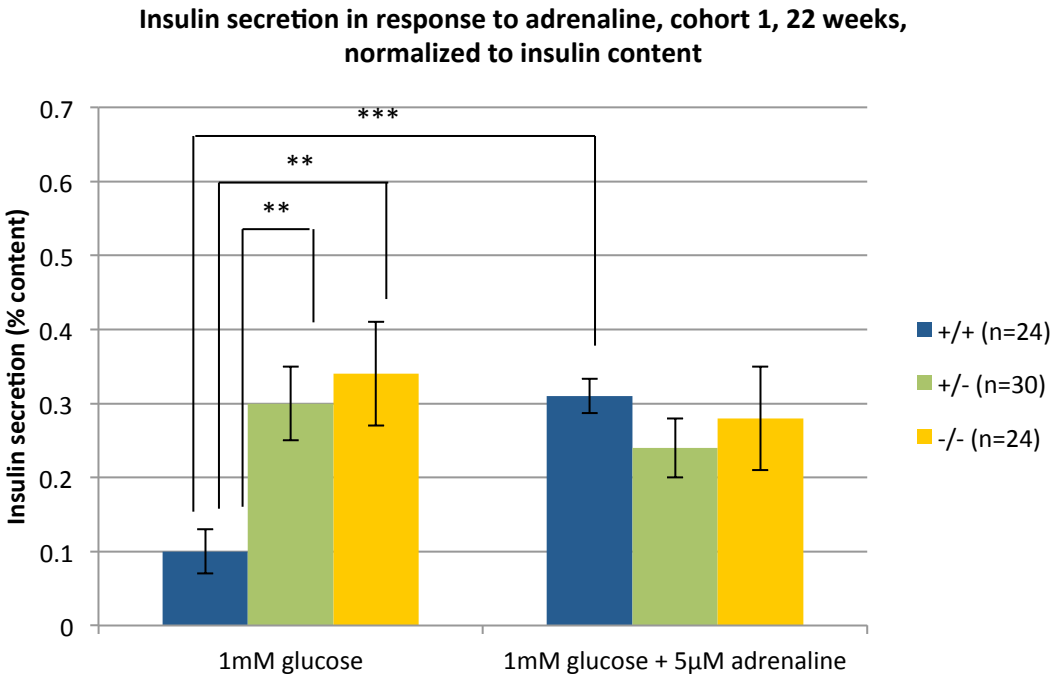


Figure 4.34: Insulin secretion from cohort 1 mice, expressed as percent insulin islet content. Significance is indicated by *p<0.05, **p≤0.01, ***p≤0.001. Sample sizes consist of biological replicates x technical replicates (3). Insulin secretion is stimulated by 5µM adrenaline in basal glucose from wild type islets, but islets from Y1064C allele carriers do not respond to adrenaline. Insulin secretion in basal glucose is already elevated from homozygote and heterozygote islets.

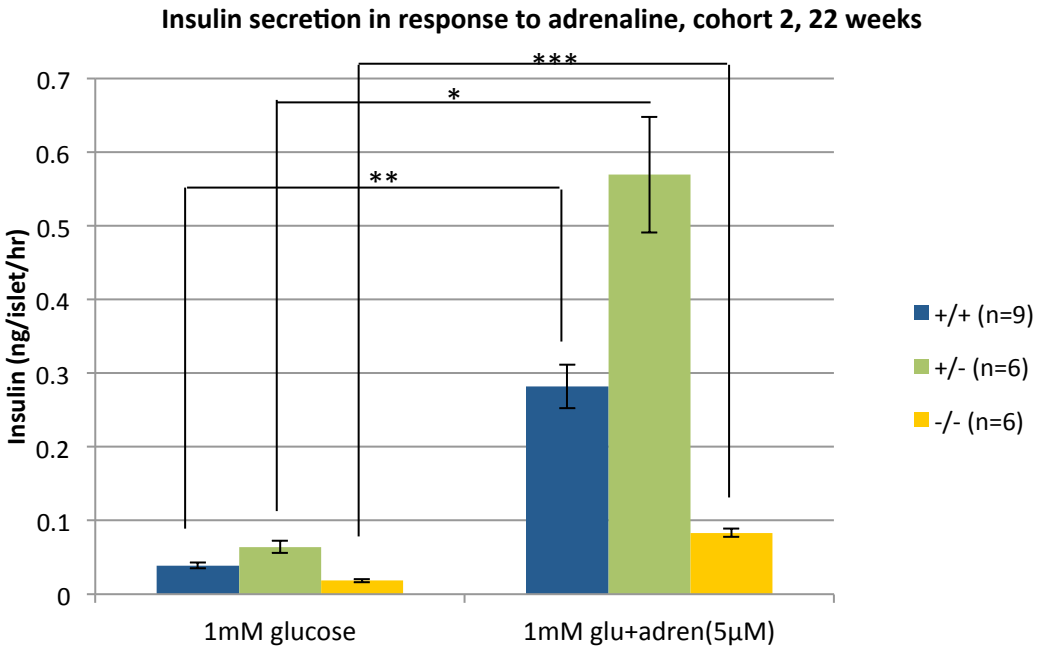


Figure 4.35: Insulin secretion in islets from 22-week-old mice in response to adrenaline. Significance is indicated by *p<0.05, **p≤0.01, ***p≤0.001. All genotypes secreted more insulin in response to adrenaline. Insulin secretion levels were 7-10x higher from cohort 2 mice than from cohort 1 mice (Figure 4.33).

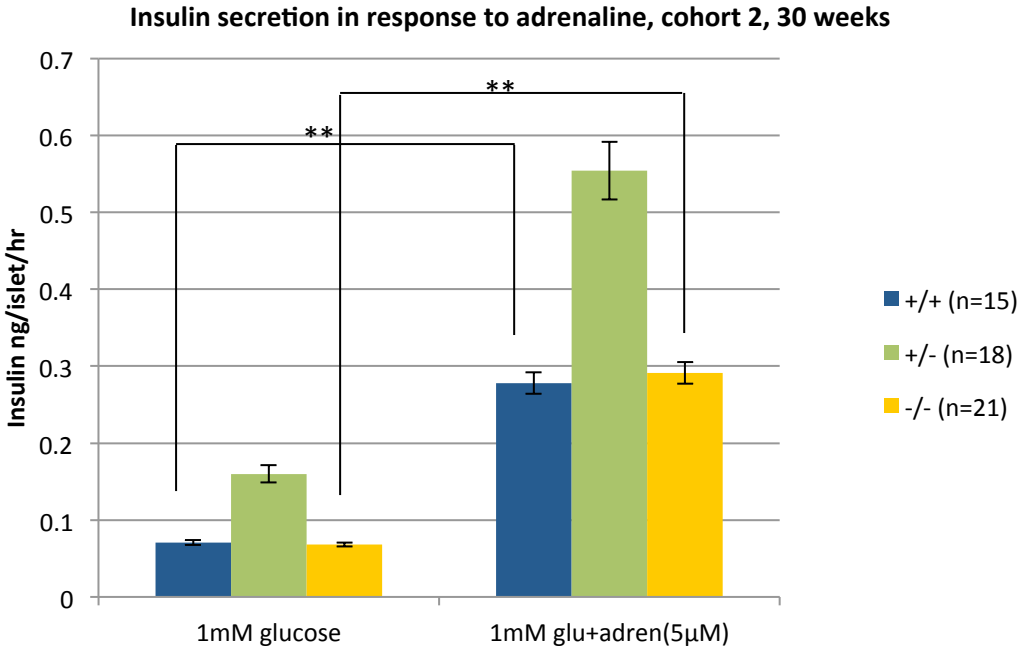


Figure 4.36: Insulin secretion in 30-week-old mice in response to adrenaline. Significance is indicated by *p<0.05, **p≤0.01. Wild type and Y1064C homozygote mice 30 weeks of age secrete more insulin in response to adrenaline. Heterozygotes did not secrete more insulin in response to adrenaline but standard error was high. Insulin secretion levels from cohort 2 mice were 7-10x higher than from cohort 1 mice (Figure 4.33). This difference was taken into account during analysis.

4.3.7 Insulin secretion in high glucose + GLP-1 is different between Y1064C homozygote mice and wild type

A logical hypothesis for the role of *Adcy5* would follow that *Adcy5* is involved in incretin hormone signaling in islet cells because the *in vivo* OGTT data (section 4.3.3) shows that *Adcy5* ^{Y1064C} homozygous mice are more tolerant of glucose than wild type after oral glucose dosing but not after glucose injection. Only when the incretin pathway was activated were the allele carriers different from wild type. Therefore, I chose to test the effect of the most abundant incretin hormone glucagon-like peptide 1 (GLP-1) on islets from Y1064C allele carriers and wild type. GLP-1 is a potent stimulator of insulin secretion and inhibitor of glucagon secretion.

Islets from cohort 2 mice at 22 weeks were tested in basal glucose (1mM) + 10µM GLP-1 and in 8mM glucose + 10µM GLP-1 (sample size: n=3 wild type, 2 heterozygote, 2 homozygote for insulin experiment; n=4 wild type, 3 heterozygote, 3 homozygote for glucagon experiment). Islets from cohort 2 mice at 30 weeks were tested in 1mM glucose + 10µM GLP-1 and in 15mM glucose + 10µM GLP-1 (sample size: n=5 wild type, 6 heterozygote, 7 homozygote for insulin experiment; n=8 wild type, 6 heterozygote, 9 homozygote for glucagon experiment). I chose these conditions to observe the effect of the incretin hormone at basal glucose as well as the expected pattern of amplification for insulin secretion or inhibition for glucagon secretion when glucose is present.

GLP-1 + 1mM glucose potentiated insulin secretion in wild type and Y1064C homozygous mice cohort 2 mice at 22 weeks old compared to secretion in plain 1mM glucose (wild type p=0.03, *Adcy5*$^{Y1064C-/-}$ p=0.04) and in all genotypes at 30 weeks of age (wild type p=0.015, heterozygote mice p=0.014, homozygote mice p=0.0009)(Figures 4.37 and 4.38). In the 8mM glucose + GLP-1 condition, 22 week old cohort 2 mice of all genotypes were not stimulated to secrete more insulin than in 1mM glucose + GLP-1, however there was a trend of higher insulin secretion in this condition (Figure 4.37). An 8mM glucose concentration may not have enough of a stimulatory effect to make a significant difference to secretion in the presence of the strongly stimulating GLP-1. However, islets from wild

type mice at 30 weeks did respond to 15mM glucose + GLP-1 by secreting more insulin than in basal glucose + GLP-1 (p=0.036) while islets from heterozygote and homozygote mice did not respond by secreting more insulin than in the basal glucose + GLP-1 condition (Figure 4.38). In 15mM glucose + GLP-1, islets from Y1064C homozygote mice secreted less insulin than wild type (p=0.03) suggesting that insulin secretion in the presence of high glucose + GLP-1 may be impaired in mice with two copies of the *Adcy5* Y1064C mutation (Figure 4.38).

GLP-1 is an inhibitor of glucagon secretion (de Marinis et al., 2010). Glucagon secretion was inconsistent in the GLP-1 conditions. Islets from cohort 2 twenty-two week old heterozygote and homozygote mice secreted less glucagon compared to wild type mice in 1mM glucose + GLP-1 ($Adcy5^{Y1064C+/-}$ p=0.046, $Adcy5^{Y1064C-/-}$ p=0.020 [Figure 4.39]). However, in the 30-week-old mice all genotypes secreted the same amount of glucagon in the same condition (Figure 4.40). Islets from mice 22 weeks of age did not show inhibition of glucagon secretion in 8mM glucose + GLP-1, although the wild type and heterozygote mice trended towards lower secretion. The sample size may not have been sufficient to capture this result, or the GLP-1 may have been degraded in this experiment. In 30 week old mice, islets from wild type and heterozygote mice secreted more glucagon in high glucose(15mM) + GLP-1 than in low glucose(1mM) + GLP-1 (wild type p=0.038, heterozygotes p=0.043 [Figure 4.40]). This result also suggests there was a problem with the integrity of the GLP-1 stock because GLP-1 is a potent inhibitor of glucagon secretion independent of glucose levels (Hare, 2010). It was not possible to repeat this experiment as no mice from this litter remained.

To summarize the GLP-1 experiments, there is some evidence in the 30-week-old mice that GLP-1 stimulation of insulin in high glucose is impaired in $Adcy5^{Y1064C-/-}$ mice. However, since insulin content data could not be measured for this cohort, the result cannot be confirmed by normalizing to insulin content.

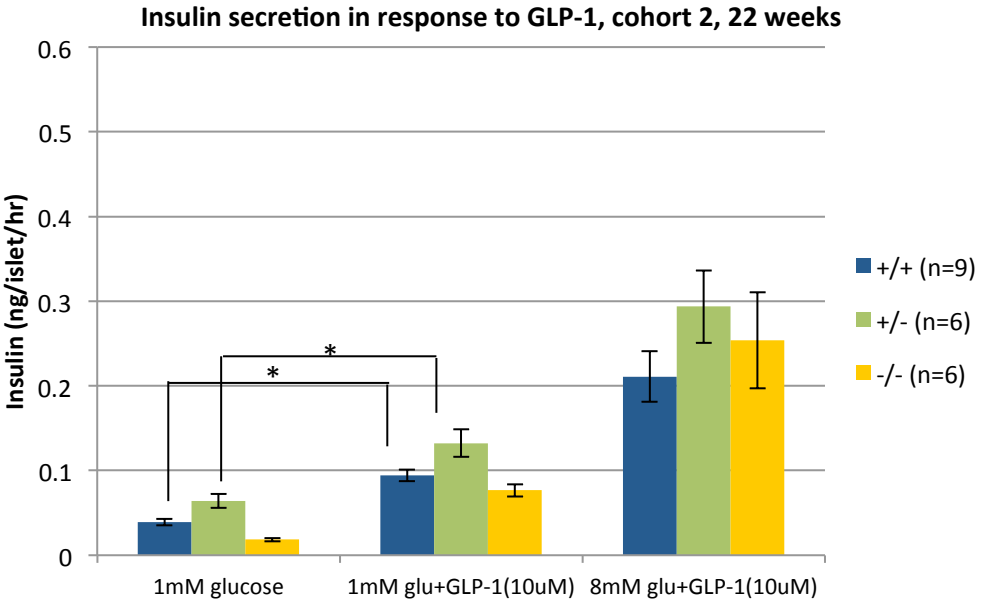


Figure 4.37: Insulin secretion in response to GLP-1 in 22-week-old mice. Significance is indicated by *p<0.05, **p≤0.01. Sample size (n) = biological replicates x technical replicates (3). Wild type and *Adcy5* Y1064C homozygote mouse islets secrete more insulin in response to GLP-1 in basal glucose, while heterozygotes are not stimulated by GLP-1. None of the genotypes are significantly stimulated to secrete more insulin from basal glucose + GLP-1 to 8mM glucose + GLP-1.

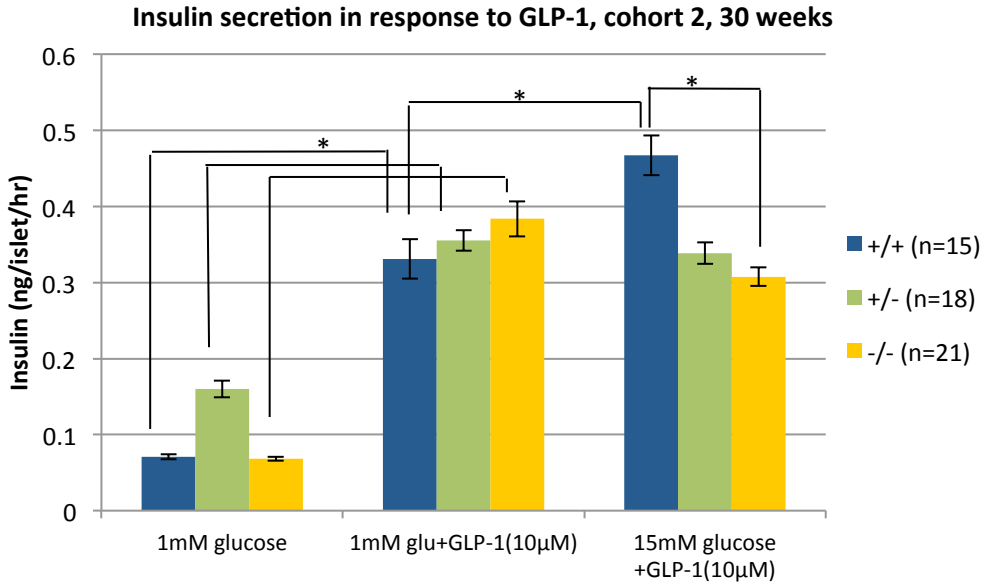


Figure 4.38: Insulin secretion from 30-week-old mice in response to GLP-1. Significance is indicated by *p<0.05, **p≤0.01. Sample size (n) = biological replicates x technical replicates (3). All genotypes were stimulated by 1mM glucose + GLP- from 1mM glucose. Islets from wild type mice secreted more insulin in the 15mM glucose + GLP-1 condition compared to in the 1mM glucose + GLP-1 condition, but islets from Y1064C heterozygote and homozygote mice were not stimulated to secrete more insulin in 15mM glucose + GLP-1. Furthermore, in the 15mM glucose + GLP-1 condition, islets from homozygote mice did not secrete as much insulin as wild type.

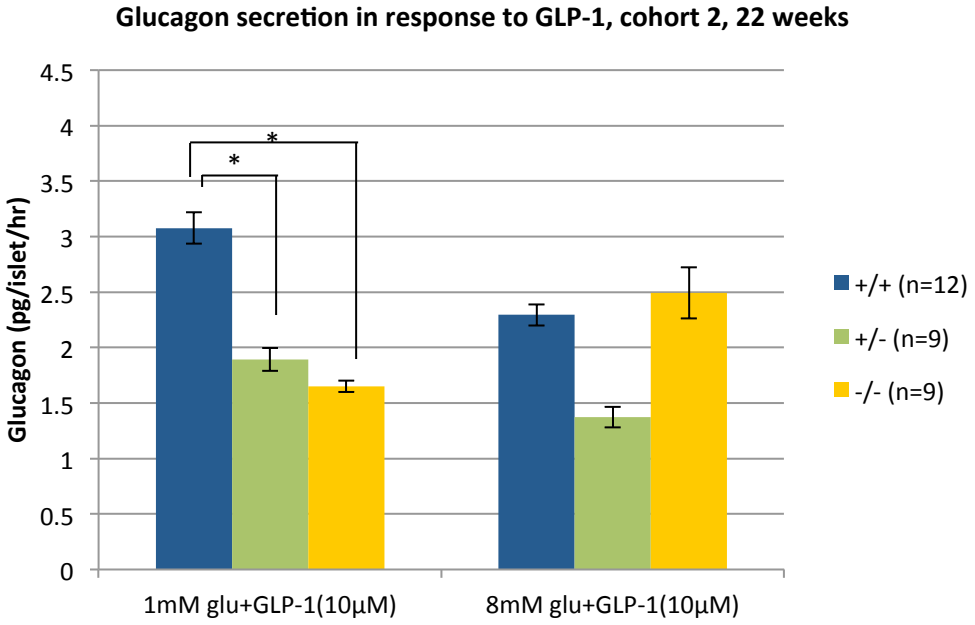


Figure 4.39: Glucagon secretion in response to GLP-1 in mice 22 weeks of age. Significance is indicated by *p<0.05, **p≤0.01. Sample size (n) = biological replicates x technical replicates (3). The *Adcy5* Y1064C heterozygote and homozygote mice secrete less glucagon in response to 1mM glucose + GLP-1 than wild type. Secretion did not differ between genotypes in response to 8mM glucose + GLP-1.

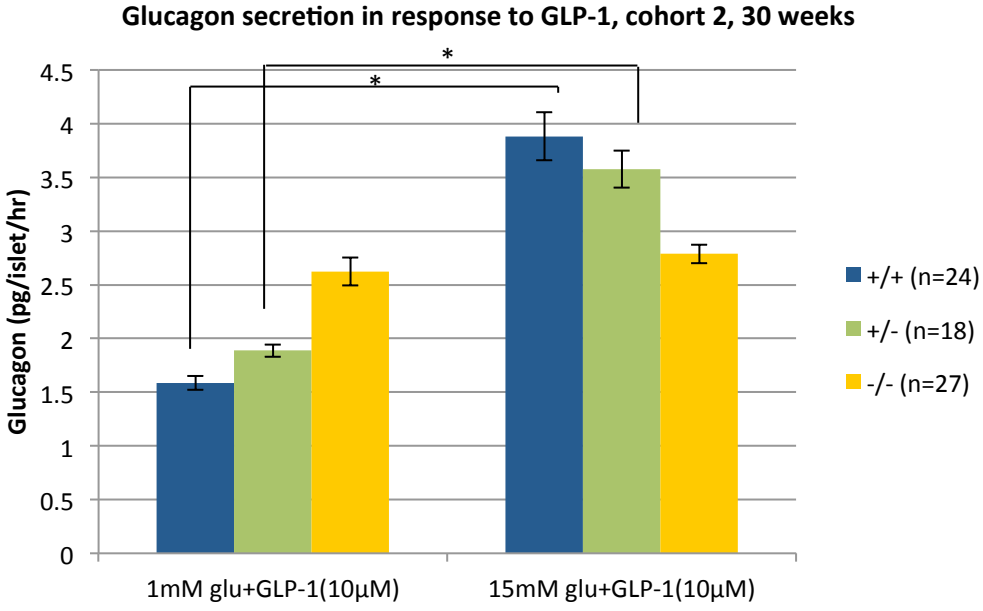


Figure 4.40: Glucagon secretion in response to GLP-1 in 30-week-old mice. Significance is indicated by *p<0.05, **p≤0.01. Sample size (n) = biological replicates x technical replicates (3). Wild type and *Adcy5*Y1064C heterozygote mice secrete more glucagon in response to 15mM glucose + GLP-1, but Y1064C homozygote mice do not respond to 15mM glucose + GLP-1. There was no difference between genotypes within the conditions.

4.4 Conclusions

4.4.1 *ADCY5* is expressed in pancreatic islets and in adipose tissue

This is the first investigation to report the presence of *ADCY5* in adipose tissue. I confirmed its presence in both subcutaneous and omental adipose tissue. I showed that *ADCY5* is expressed in human islets at low levels along with all of the other 9 adenylate cyclase genes. The presence of multiple adenylate cyclases in human islet cells suggests a complex cAMP-signaling network in involving *ADCY5* and other adenylate cyclases.

Similarly, *Adcy5* is expressed in mouse islets at low levels along with 5 other adenylate cyclases. Another very lowly expressed adenylate cyclase, *Adcy8,* has been shown to have an important role in mouse islets mediating GLP-1 signaling (Roger, et al., 2011). Confirmation of expression in both mouse and human islets led us to mouse modeling of *Adcy5* dysfunction.

4.4.2 *Adcy5*^{Y1064C} homozygote mice have lower blood glucose after an oral glucose tolerance test than wild type mice

Male mice expressing two copies of the *Adcy5*^{Y1064C} mutation had lower blood glucose levels than wild type mice 30, 60, and 120 minutes after an oral glucose tolerance test at every age tested: 12, 18 and 21 weeks. The mice with one copy of the mutation generally exhibited an intermediate phenotype. At 21 weeks, fasting glucose (before oral dosage) was greatly decreased in *Adcy5* Y1064C homozygote mice compared to wild type. Results from phenotyping of an *Adcy5* knockout mouse also showed lower fasting blood glucose in homozygote mice (Ho et al., 2010; Okumura et al., 2003; Hu et al., 2009) supporting my prediction that Y1064C is a loss of function mutation.

Glucose tolerance when tested by intraperitoneal injection was not different between the *Adcy5*^{Y1064C} mice and wild type. The difference in male mice was robust and reproducible when glucose was administered orally which may implicate *Adcy5* in a pathway activated by ingestion.

Various gastrointestinal hormones are activated when glucose is administered orally including gastrin, secretin, cholecystokinin (CCK), GIP (glucose-dependent insulinotropic peptide) and GLP-1 (glucagon-like peptide-1) (Unger and Eisentraut, 1969). All of these hormones are known to activate feed-forward mechanisms that enhance insulin secretion anticipating the rise in blood glucose following ingestion of nutrients. This mechanism is called the incretin effect (Unger and Eisentraut, 1969). While gastrin, secretin and CCK have relatively modest effects on insulin secretion in response to glucose loading, GLP-1 and GIP elicit substantial insulin release and glucagon inhibition (Rehfeld, 2011). Agonists of incretin hormones GLP-1 and GIP have been of particular interest in the treatment of T2D as enhancers of insulin secretion when normal beta cell function is impaired.

4.4.3 Insulin secretion from *Adcy5*^{Y1064C} islets is not consistently different than from wild type islets

Isolated islets from Cohort 1 *Adcy5*^{Y1064C} heterozygote and homozygote mice secrete more insulin in basal glucose (1mM) than wild type mice, but this difference was not seen in Cohort 2 mice. The hyper-secretion in Cohort 1 mice is still significant when islet insulin content is taken into account. Islets from *Adcy5*^{Y1064C} homozygote and heterozygote mice have lower insulin content than wild type; hence the percent content secreted from the Y1064C allele carriers is decidedly more than wild type to secrete a greater insulin output. The amounts of insulin secreted in low glucose are extremely small and difficult to measure, so the result in Cohort 1 could be due to imprecision of the assay. Likewise, the result in Cohort 2 could be a lack of detection of the phenotype due to imprecise measurements. More experiments would be necessary to verify that the islets of *Adcy5* Y1064C allele carriers hyper-secrete in basal glucose.

If the Cohort 1 result is true, hyper secretion in low glucose could be a result of increased sensitivity to glucose or sensitivity to another stimulatory hormone such as glucagon, adrenaline, or GLP-1. Hypersensitivity to glucose or another mechanism causing increased insulin secretion at basal conditions may help explain the in vivo phenotype. In vivo, the mice homozygous for the Y1064C mutation are protected against hyperglycemia after an oral glucose tolerance test. Higher circulating

levels of blood insulin from hyper secreting islets would keep blood glucose levels lower in these mice. However, I found no evidence of increased blood insulin levels after an overnight fast in any genotype.

ADCY5 is bound to a G-protein receptor and could be activated by a number of G-proteins including incretin hormones GLP-1 and GIP. My data show that genotype does not impact insulin secretion in response to basal glucose + 10μM GLP-1; however, in a high glucose condition + 10μM GLP-1, isolated islets from 30 week old Y1064C homozygote mice do not secrete as much insulin as islets from wild type and heterozygote mice. *Adcy5* may be involved in GLP-1 stimulation of insulin secretion in the presence of high glucose. Roger et al. (2010), found that by knocking down *ADCY8*, human and rat islets exposed to >5.5mM glucose for 3 days were not stimulated by GLP-1 to secrete insulin. Re-expression of *ADCY8* rescued GLP-1 signaling in glucotoxicity. My results suggest *Adcy5* could be performing an analogous function to *Adcy8* in these mice, but the experiment would need to be repeated and insulin secretion values normalized to islet insulin content to verify the phenotype. Insulin content data was not available for these mice, so the data is not adjusted for the islet hormone content present. In cohort 1 mice carrying two copies of the Y1064C allele, insulin content was lower than in wild type mice, and if this is true in cohort 2 mice, the Y1064C homozygote mice may be secreting the same percentage insulin of islet content as the wild type and heterozygote mice. A possible future experiment would be to incubate isolated mouse islets in varying glucose concentrations as Roger et al. (2010) have done before performing secretion experiments.

I found no effect on forskolin-stimulated insulin secretion from the Y1064C mutation, although there was a trend toward the heterozygote and homozygote mice secreting less insulin than wild type. Forskolin directly stimulates adenylate cyclases, so the lack of a phenotype would suggest that either the Y1064C mutation is not changing protein function or that *Adcy5* does not have a substantial role in insulin secretion.

Adrenaline acted as a stimulant to insulin secretion in the islets from all mouse cohorts tested. Typically, adrenaline inhibits insulin secretion by binding the alpha-adrenergic receptors and inactivating adenylate cyclases (Debuyser et al., 1991). This is an important physiological function as it frees up energy for action rather than diverting energy to digestion. The insulin secretion phenotype of these islets is an anomaly and suggests that the alpha-adrenergic receptors may not be functioning normally.

4.4.4 *Adcy5*^{Y1064C} homozygote mice secrete less glucagon in response to adrenaline

Glucagon secretion is impaired in the Y1064C homozygous mice in response to adrenaline. Islets isolated from mice homozygous for the Y1064C mutation secreted 40-60% less glucagon than islets from wild type or heterozygote mice in response to basal glucose plus adrenaline. I reproduced these results in 30-week-old mice from cohort 2. The younger mice from cohort 2 showed the same trend in a small sample size. This suggests that *Adcy5* function could be necessary for transmitting the signal from beta-adrenergic receptors on the alpha cell membrane to the glucagon production and secretion pathways.

Nils Rorsman (Oxford Centre for Diabetes, Endocrinology and Metabolism) performed *in situ* pancreas perfusion in order to replicate these results using a more sensitive method that more closely resembles the *in vivo* environment than static islet secretion measurements. The pancreata of 5 freshly culled wild type and 7 homozygous mice were perfused with 1 mM glucose, 14 mM glucose and 18 mM glucose, and then 18 mM glucose + isoprenaline which is a beta-adrenergic receptor agonist. Glucagon was measured at one-minute time points throughout the experiment. Islets from homozygous mice performed the same as those from wild type mice when challenged with these conditions. The islets from homozygote mice secreted slightly less glucagon than wild type throughout the experiment, but there were no significant differences seen at any condition (Figure 4.41). These data are challenging to interpret alongside the static incubations, but isoprenaline was only tested at high glucose during the perfusion experiments, while my secretion

studies tested adrenaline stimulation in 1mM glucose. There may be a phenomenon whereby catecholamine stimulation in low glucose, when the alpha cells are already secreting maximum glucagon, is impaired by the Y1064C mutation, while stimulation at high glucose, when the alpha cells are inhibited by glucose, is not impacted.

Nils Rorsman also analyzed the samples from the *in situ* perfusion experiments for an insulin phenotype in the *Adcy5*^{Y1064C} mice (Figure 4.34). The wild type mice and Y1064C allele carriers secreted the same amount of insulin in basal glucose (1mM), but wild type secreted more insulin in response to high glucose (18mM) than the Y1064C homozygous mice. I found no difference in insulin secretion at high glucose, but Hodson and colleagues (Hodson et al., 2014) collected calcium-imaging and insulin secretion data from human islets after *ADCY5* was knocked down and found that *ADCY5* is necessary in coupling glucose to insulin secretion. The data from the perfusion experiment replicates the finding from Hodson and colleagues in mouse islets and suggests that the *Adcy5*^{Y1064C} mutation is a loss of function allele.

The glucagon phenotype I observed in the static secretion experiments may help explain the protective effect this mutation has against hyperglycemia *in vivo*. If glucagon secretion is impaired, the mice may be clearing glucose from their blood without replacing it from glycogen stores in the liver. The results of the OGTT showing that Y1064C homozygote mice have lower blood glucose at every time point after an oral glucose tolerance test are explained by this hypothesis. My assays for glucagon in the bloodstream did not pick up lower blood glucagon in Y1064C homozygote mice. These assays may not be sensitive enough, or the mice may not have been in a state (hungry, stressed, etc.) where the mutation was having an effect on glucagon. These questions require further testing of blood glucagon levels under various conditions and with larger cohorts.

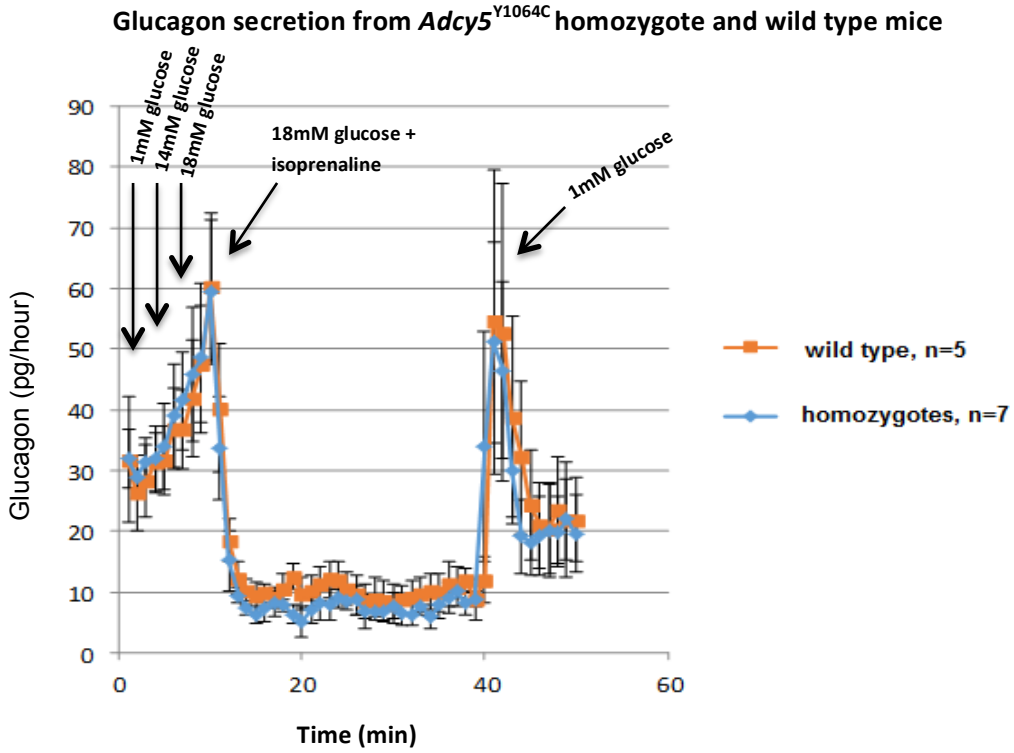


Figure 4.41: Work performed by Nils Rorsman (University of Oxford). Glucagon secretion from in situ pancreas perfusion of *Adcy5*^{Y1064C} mice. Both genotypes reacted to low, mid and high glucose and isoprelanine conditions similarly.

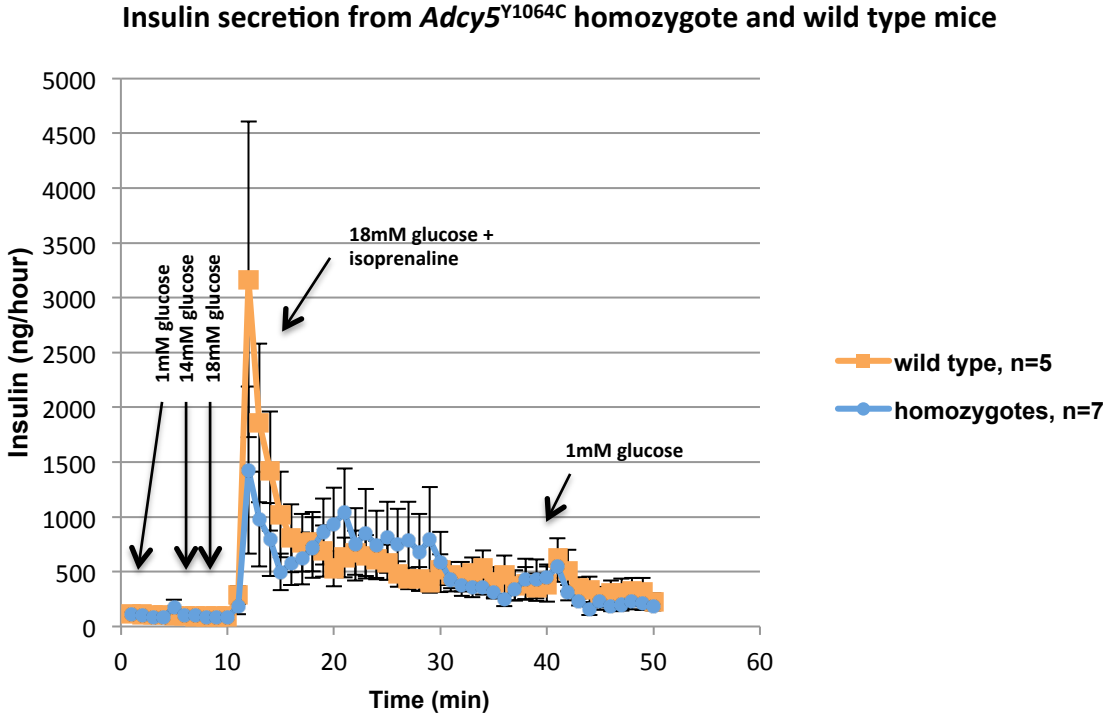


Figure 4.42: Work performed by Nils Rorsman (University of Oxford). Insulin secretion from *in situ* pancreas perfusion of *Adcy5*^{Y1064C} mice. Mice homozygous for the Y1064C allele have impaired insulin secretion at 18mM glucose compared to wild type.

4.4.5 Discussion of study limitations

There were several problems and limitations with the *in vivo* and *in vitro* characterization of the *Adcy5* Y1064C mouse model. In retrospect, the in vivo experiments were not designed as efficiently as possible, and there were issues with degraded samples and lack of positive controls in the secretion experiments. If I were able to re-do the in vivo experiments, I would devise phenotyping pipelines that facilitate direct comparison of the IPGTTs and OGTTs at each age and with the same collection time points. IPGTTs were collected at 12 and 16 weeks, while OGTTs were collected at 12, 18 and 20 weeks. The most pronounced phenotype in the Y1064C allele carriers was seen during the 20 week OGTT, but the most comparable IPGTT is at 16 weeks. The OGTT phenotype is very clear by 18 weeks, while the IPGTT is unaffected at 16 weeks suggesting oral glucose tolerance is altered while intraperitoneal glucose tolerance is not. However, if the IP phenotype is subtle, I may have missed it by not testing the older mice.

Furthermore, while in vitro insulin secretion is studied extensively, in vivo insulin phenotypes could be probed further. It would have been interesting to collect plasma insulin concentrations at all time points during an OGTT and an IPGTT to determine whether peak insulin is different between wild type and mutant mice. If insulin secretion is truly impaired in high glucose + GLP-1, the peak insulin concentration may be lower in the mutant mice. Another mechanism could be involved such as insulin sensitivity in peripheral tissues. I would have, with more time and resources, carried out an insulin tolerance test to determine if the mice carrying the mutation are more or less insulin sensitive. Insulin tolerance can be extremely correlated with body weight. In retrospect, it would have made more sense to collect body weights at every time point than to collect body composition at 10 and 18 weeks. A body weight phenotype might be subtle and only become apparent at an older age. The *Adcy5* null mice published by Yan et al. (2007) exhibit lower fasting glucose as well as lower body weight and lower fat mass around 12-16 weeks of age. The *Adcy5* Y1064C homozygote mice also have lower fasting glucose at 20 weeks, and they may present a similar but subtler

phenotype to the *Adcy5* null. Body weight and body composition measurements at an older age would help to verify this hypothesis. Challenging the mice with a high fat diet would also be an interesting future direction. The high fat diet would put pressure on the alpha and beta cells and the peripheral tissues to secrete and take up insulin and glucagon effectively, and any glucose processing impairments may show up more starkly in a body weight, insulin tolerance, or glucose tolerance phenotype.

The *in vitro* experiments point out a deficiency in glucagon secretion in response to adrenaline. Less glucagon circulating in the bloodstream would prevent the liver from converting glycogen stores to glucose therefore leading to lower circulating blood glucose. This mechanism may help to explain the diabetes-protective effect *in vivo*. The experiments would have benefitted from the addition of a positive control of glucagon secretion. Low glucose is a stimulator of glucagon secretion, but basal glucose on its own was not sufficient to stimulate glucagon to the level that adrenaline did. Inclusion of alanine or arginine, two amino acids known to stimulate glucagon secretion, may have been a more effective positive control. The addition of one of these compounds could have provided a comparable stimulator of glucagon for validation that the islets were behaving as expected and responding to normal stimuli.

Possible beta cell mechanisms impacted by the Y1064C mutation include hyper secretion of insulin in low glucose, and impairment of insulin secretion in high glucose + GLP-1. The results of these experiments varied between the mouse cohorts and between the ages of the mice. Repeating these experiments with more mice and collecting insulin islet content values would be helpful to decipher the secretion phenotype. Moreover, adding another control for the inhibition of insulin secretion besides low glucose may have been helpful to determine if the islets are hyper secreting in low glucose. It is possible that *Adcy5* has a role in all of these pathways depending upon where it is located in the islets. In the following chapter I will examine localization of the ADCY5 protein in the alpha and beta cells as well as overall homology of the *Adcy5*^{Y1064C} mouse pancreas.

Chapter 5

Morphological characterization of the ADCY5 protein in human and mouse islets

5.1 Introduction

5.1.1 Morphometry of the ADCY5 protein in mice and humans

To better understand the *in vivo* and *in vitro* phenotyping data generated in Chapter 4, it is necessary to locate ADCY5 in the islet and pinpoint its location intracellularly. Furthermore, changes in the distribution of ADCY5 in the islet cells between the wild type mice and *Adcy5*^{Y1064C} allele carriers may help to clarify its downstream and upstream effectors causing lower blood glucose in allele-carriers (section 4.3.2) and altered hormone secretion (sections 4.3.3 – 4.3.7). In this chapter I will give background on what was known about ADCY5 localization prior to this thesis, and then present my work to characterize the morphology of ADCY5 in mouse and human islets.

5.1.2 Expression of *ADCY5* in islet cells

In the previous chapter (section 4.3.1), quantitative real-time PCR analysis revealed that *ADCY5* is expressed in human and *Adcy5* is expressed in mouse islets (see Chapter 4 section 4.3.1 and 4.3.2). Both are expressed in various other tissues as well: most highly in heart, brain and adipose and at lower levels in islets (section 4.3.1). Many adenylate cyclases are present in islets in addition to *ADCY5* (section 4.3.2). Therefore we can expect to locate the protein in islet cells.

Hodson et al. (2014) confirmed that alleles associated with T2D at the *ADCY5* locus also associate with expression of *ADCY5* in islets. In islets isolated from human donors the T2D risk allele at rs11708067 (intronic within *ADCY5*) lowers *ADCY5* mRNA expression levels by 2 fold in islets (Hodson et al., 2014). The *Adcy5*^{Y1064C} allele is a coding variant, so it is unlikely to have an effect on expression of the protein.

ADCY5 likely exerts its effect on insulin and/or glucagon secretion through the amplifying pathways of insulin secretion in beta cells and glucagon secretion in the alpha cells (Henquin et al., 2006; Delmeire, et al., 2003). As explained previously in section 4.1.2, adenylate cyclase 5 is membrane-bound and couples a G-protein receptor. When the receptor is activated, ADCY5 catalyzes the

generation of cAMP leading to activation of the PKA and/or Epac pathways and potentiating secretion of insulin in beta cells or glucagon in alpha cells (see Figure 4.1). To validate and expand upon this hypothesis, it will be helpful to know if ADCY5 anchors to the outer cell membrane or an inner membrane such as the nuclear and mitochondrial membranes. If ADCY5 anchors to the outer cell membrane, the list of activators will be different than if it binds to the mitochondrial or nuclear membrane and it may regulate distinct signalling pathways. Beta-adrenergic receptors, opioid receptors, and angiotensin receptors are located both on the outer cell membrane as well as on the nuclear membrane in some cell types and trigger different signalling cascades dependent upon the location of the receptor (Boivin et al., 2006; Tadevosyan et al., 2011). Little is known about the function of adenylate cyclases which locate to the endoplasmic reticulum and mitochondrial membranes, but there is evidence that ADCYs are present in both membranes (Cheng and Farguhar, 1976; Fine et al., 1982). Both of the immediate downstream effectors of ADCY5, PKA and Epac, localize to the cytoplasm and may cluster in discrete pockets where the cAMP signal is carried directly to them without diffusing through the greater cytoplasm (Willoughby et al., 2010).

5.1.3 Beta and alpha cell mass in type 2 diabetes

Abnormal insulin and glucagon secretion are features of T2D. Other consistent features of the disease are reduced beta cell mass and increased alpha cell mass in man (Butler et al., 2003) as well as in animal models of diabetes (Goto et al., 1988; Ostensen et al., 1996). In mice lacking insulin-receptor substrate 1 (IRS1), growth slows, but diabetes does not develop because the mice are able to increase insulin secretion to overcome the mild insulin resistance (Tamemoto et al., 1994). Conversely, in mice lacking the insulin-receptor substrate 2 (IRS2), the liver becomes highly resistant to insulin and beta cells are not able to compensate resulting in a severe diabetic phenotype (Withers et al., 1988). In these IRS2-/- mice the beta cells do not grow or multiply, but instead beta cell mass decreases up to 25% (Kubota et al., 2000). Meanwhile in IRS1-/- mice, beta cell mass increases up to 85% resulting in very mild to no diabetic symptoms (Hennige et al., 2005; Kulkarni et

al., 1999). The Goto-Kakizaki (GK) rat model, a non-obese model of diabetes which develops hyperglycaemia and impaired insulin secretion early in life, has lower beta cell mass than wild type from birth. The islets of these mice also have an irregular starfish shape due to disrupted structure (Ostensen et al., 1996). Without genetic alteration, prolonged exposure to hyperglycaemia will also result in decreased beta cell mass and increased alpha cell mass (Schuit et al., 2002), and beta cells may convert to alpha cells and begin to produce and secrete glucagon rather than insulin (Brereton et al. 2014). The foremost hypothesis for why beta cell mass decreases is through disrupted balance of cellular turnover which is thought to be balanced in rodents by a continuous process of beta cell apoptosis and renewal by neogenesis or cell division (Weir and Bonner-Weir, 2013). When insulin signaling in the islet cells is disrupted, beta cell proliferation slows down significantly and the cells are arrested in G1 phase (Folli et al., 2011). However, the relative contribution of decreased beta cell function vs. decreased beta cell mass is still highly debated (Clark et al., 2001; Costes et al., 2013).

While beta cell mass decreases in response to insulin resistance, the opposite result happens in response to obesity. In the ob/ob mouse, a model of obesity, overall islet size increases due to expanding beta cell mass, and resulting in hyperinsulinemia (Tomita et al., 1992). The number of islets may increase in addition to their size in obesity. The BSB mouse, a model of multigenic obesity, exhibits increased size and number of islets compared to controls (Slavin et al., 2010). Quantifying islet size and number as well as beta and alpha cell proportion in the *Adcy5*^{Y1064C} mice is important to relate the *in vivo* (glucose tolerance data) and ex vivo phenotyping (insulin and glucagon secretion) to what is happening in the pancreas in terms of changes in *Adcy5* localisation and beta and alpha cell mass. The protective effect of the mutation in the mice against hyperglycaemia (4.3.2) suggests either altered function of islet cells or a potential increase in islet cell size or number, but the impaired glucagon secretion from isolated islets (Chapter 4.3.3-4.3.7) could be the result of decreased alpha cell size or number. In this chapter I will investigate the possibility of altered islet size, number, and composition in *Adcy5*^{Y1064C} allele carriers by histologically examining mouse pancreata from *Adcy5*^{Y1064C} mutants and controls and from human diabetic patients and controls.

5.1.4 Chapter Aims

The aims of this section were:

1. To determine morphological alterations in islets and cell distribution in mutant mice of different ages compared with wild type, and to relate these to changes in glucose tolerance.
2. To determine any subcellular and morphological changes in islet cells arising as a result of the *Adcy5* Y1064C mutation in mice.
3. To localize ADCY5 to mouse islets by immunohistochemistry in order to understand where ADCY5 may be acting in the cell.
4. To visualize intra-cellular localization of ADCY5 in mouse and human islets using confocal microscopy.

5.2 Methods

Methods for histological analyses are summarized here. More detailed methods can be found in section 2.4.2.

5.2.1 Immunolabeling of pancreatic sections

Freshly dissected whole mouse pancreata from WT, heterozygous or homozygous mutant mice were fixed in 10% formaldehyde in PBS (phosphate buffered saline, pH 7.4). Human pancreas specimens from adult non-diabetic (n=3) and T2DM patients (n=2) were obtained at OCDEM (Oxfordshire, UK) post-mortem following appropriate ethical consent and consent from the relatives. Specimens were kept in fixative at least overnight at 4^{o}C and then embedded in paraffin wax following the procedure detailed in Chapter 2 section 2.4.2(a).

To identify islet structures and specific hormone- containing cells in islets, pancreata were immunolabelled using immunoperoxidase staining and also fluorescent labeling. The protocol for both procedures can be found in Chapter 2 section 2.4.2(b&c).

5.2.2 Immunolabeling of islet cells on coverslips

To identify *Adcy5* in cells which were subjected to minimal disruption (mild fixation and minimal loss of protein structure), isolated islet cells were dispersed and attached to coverslips for immunolabeling according to the protocol in Chapter 2 section 2.4.2(d). Dissociated mouse islet cells fixed on coverslips were double immunocytochemically labelled for insulin or glucagon with ADCY5. The protocol can be found in Chapter 2 section 2.4.2(e).

5.2.3 Electron Microscopy

To determine if cellular ultrastructure was altered by the mutation, islets were prepared for electron microscopy. This work was done by Dr. Anne Clark, University of Oxford. Freshly isolated islets were fixed in 4% paraformaldehyde plus 0.5% glutaraldehyde (EM grade) then block stained in 2% uranyl acetate in acetate buffer and hardened in Spurr's resin before visualization on an electron microscope (Jeol 1010, Massachusetts, USA). Full methods can be found in Chapter 2 section 2.4.3. Images were viewed and analyzed by me.

5.2.4 Histological Quantitative Analysis

To determine if there were quantitative differences in islet and islet cell distribution as a result of the mutation, quantitative morphometric analysis was undertaken. Pancreas sections were cut at two different depths of the block 150µm apart to obtain enough islet sections (minimum 25) for quantification. Sections were labelled as in section 5.2.2 for insulin or glucagon. Sequential photomicrograph images were taken of the entire sections (Olympus) at x10 magnification and visualized using ImageJ (Java 1.6.0_20, National Institutes of Health, Bethesda, Maryland). Area of

the pancreas was measured as the total area of exocrine pancreas, using only the lobes visualized in the photographs. Total islet area per pancreas was measured from insulin labelled sections. Exocrine pancreas area, islet area, beta cell area, and alpha cell area were counted by tracing and then calculating area traced (all using ImageJ). Islet density was calculated from 'total islet area' divided by 'total pancreas area.' Percent beta cell per islet and percent alpha cell per islet were also calculated by dividing beta cell area or alpha cell area by total islet arear. Median islet size was determined from all islets visible in the pancreas sections. Area of pancreas examined varied between 25 and 50 square millimeters, and between 24-120 islets were measured for each animal.

5.3 Results

5.3.1 Morphological differences in islets between wild type and Y1064C mutants

Initial observations from collagenase-digested hand-picked islets led to quantification of islet size and density. Islets from mice carrying two copies of the Y1064C allele appeared to be larger on average than islets from wild type mice and yielded more extra-large (>2mm diameter) islets suggesting an increased islet size in the mice homozygous for the allele. However, examination of the islet in tissue sections which were labelled for insulin with immuno-peroxidase suggested that islets from 5 homozygous mice were of similar size but had less densely labelled insulin compared to islets from 5 wild type mice; a representative example is shown in Figure 5.1 (a) and (b). To quantify the islet size, pancreatic islet density, beta cell and alpha cell proportion in the mice, I designed quantifiable experiments, the results of which are shown below in section 5.3.2.

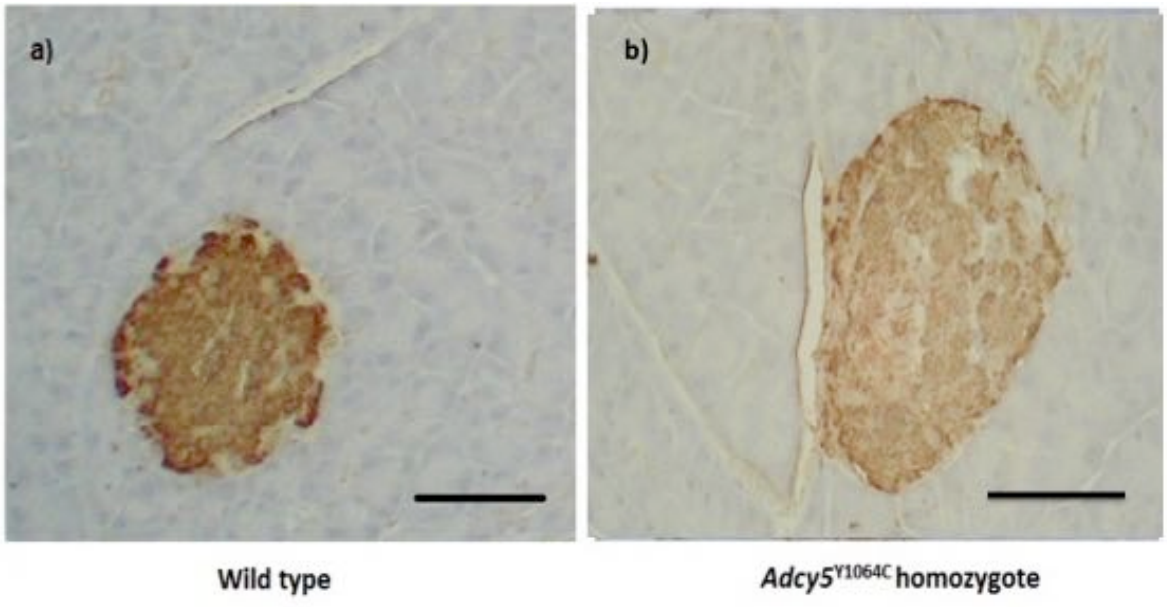
Typically, alpha cells in normal mouse islets localize to the periphery of mouse islets. Alpha cells in the islets of *Adcy5*^{Y1064C} homozygous mice were notably spread throughout the islet very different from the wild type islets in >100 islets viewed for each genotype (see Figure 5.1[c and d]). The proportion of alpha cells per islet is quantified in section 5.3.2.

5.3.2 Islet quantification in wild type and Y1064C mutants showed no differences between genotypes

Sections were analyzed from wild type (n=6) *Adcy5*^{Y1064C} heterozygotes (n=4) and homozygotes (n=5) of both genders aged 22 and 30 weeks (Table 5.1). Analysis was done blind to the genotype of the mice. Since only a small section of the islet was transected by the histological section, it was important to use the median for estimation of islet sizes. It was important to sample enough islets to obtain a median representative of the sample and >25 islet sections was calculated to be adequate with a power of 80% at significance cut-off 0.05. Median islet size was highly variable within genotypes, but there was a trend towards smaller median size in the heterozygote mice of both ages (Figure 5.2). This may be an artifact of the small sample size; especially since the homozygote mice do not appear to share this feature. Sample size would need to be scaled up to determine if there is an islet size difference.

Alpha and beta cell mass were estimated by measuring the area of alpha cells in sections through the pancreas labelled for glucagon and the area of beta cells in sections labelled for insulin. The alpha or beta cell population in the islets was not different between Y1064C allele carriers and wild type mice at any age (Table 5.1). Despite the disruption in the islet architecture with alpha cells located more towards the center of the islet in mice carrying two copies of the Y1064C allele (described in section 5.3.1), the numbers of alpha cells were not different between genotypes.

High fasting glucose did not correspond with larger islet size or greater islet density (Table 5.1). However, fasting glucose data was only available for the 22 week old mice. Body weight was not available for these mice, but an explanation for the mice with particularly large islets or high islet densities may be that they had higher body weights. High body mass corresponds with increased beta cell mass in mice and humans as explained in section 5.1.3 of the introduction to this chapter.



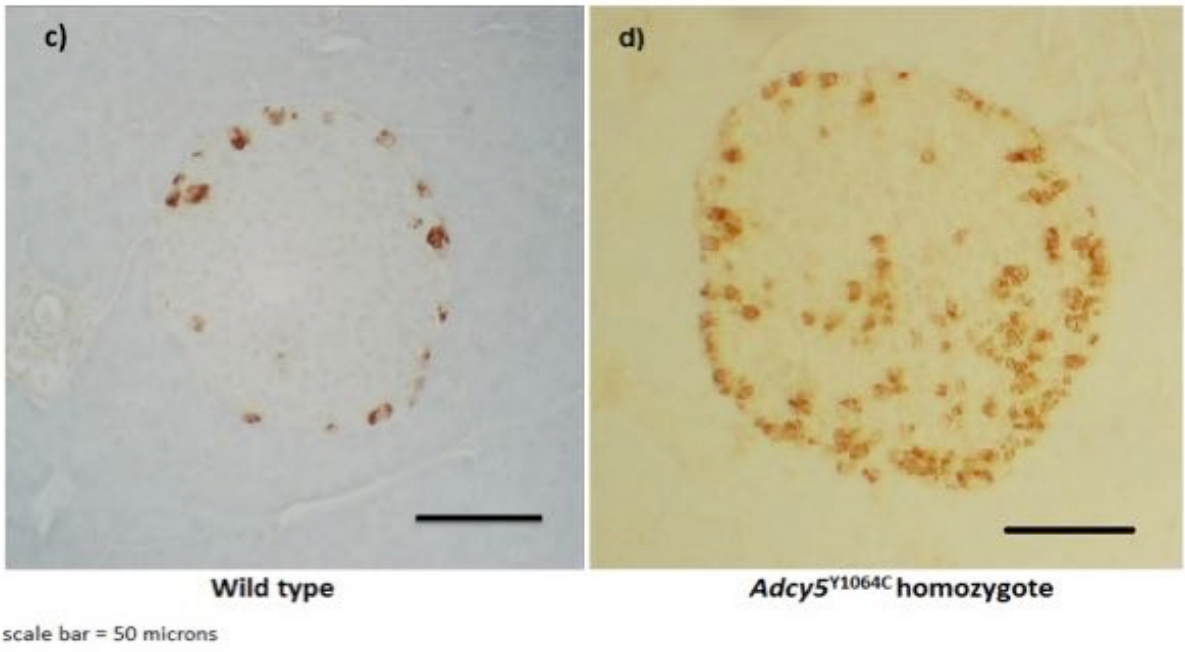


Figure 5.1: A representative image of an islet from wild type mouse ID #8 (a) and homozygote mouse ID #13 (b) labelled for insulin, and wild type #9 (c) and homozygote mouse #13 islet (d) labelled for glucagon. The insulin in the homozygous islet is less densely labelled and appears granulated (b). The alpha cells are localized to the periphery of the islet in the wild type mouse but many are spread throughout the islet in the homozygote (d).

Mouse ID	Age	Mouse genotype	Mouse sex	# Islets	Median size (µm2)	Islet density (islet area/panc area)	β–cells/ islet area %	α-cells/islet area %	Fasting Glucose 20 weeks (mmol/L)
Male mice									
1	22	+/+	male	26	11,500	0.01	84.50%	15.73%	10.73
2	22	+/+	male	46	18,809	0.037	84.50%	17.93%	10.77
3	22	+/-	male	71	4,268	0.06	75.90%	20.12%	12.99
5	22	-/-	male	44	10,744	0.021	88.00%	19.81%	11.99
7	30	+/+	male	105	22,933	0.028	85.87%	9.98%	
8	30	+/+	male	96	17,218	0.043	85.19%	10.30%	
9	30	+/+	male	130	17,361	0.059	83.81%	6.47%	
10	30	+/+	male	59	9,337	0.017	85.97%	12.88%	
11	30	+/-	male	40	5,567	0.01	76.50%	22.63%	
12	30	+/-	male	57	4,091	0.038	86.50%	12.74%	
13	30	-/-	male	31	5,074	0.016	80.20%	12.83%	
14	30	-/-	male	28	16,845	0.03	83.60%	13.72%	
Female mice									
4	22	+/-	female	36	3,947	0.013	75.50%	18.76%	8.91
6	22	-/-	female	28	12,363	0.015	83.00%	26.83%	
15	30	-/-	female	47	20,841	0.038	82.30%	11.34%	

Table 5.1: Quantification of stained pancreas slides. Mice sacrificed at 22 weeks of age are highlighted in grey while mice sacrificed at 30 weeks of age are not highlighted. There are no significant differences between genotypes for median islet size, density, beta cell density, alpha cell density, or fasting glucose.

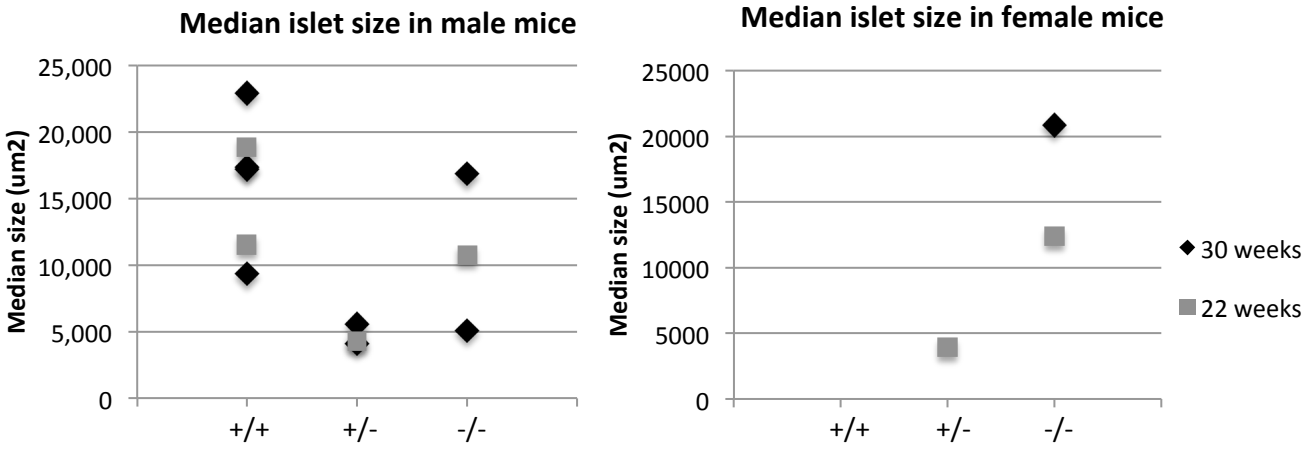


Figure 5.2: Median islet size in male mice (left) and female mice (right) from Table 5.1 plotted by genotype. The black markers represent mice sacrificed at 30 weeks, and the grey markers represent mice sacrificed at 22 weeks. There are no significant differences in median islet size, but the heterozygote mice show a trend of smaller islets.

5.3.3 Visualizing ADCY5 in mouse isolated islet cells

Dispersed islet cells from 6 wild type and 6 *Adcy5*^{Y1064C} homozygous mice were mounted on coverslips and labeled for ADCY5, insulin and glucagon. The islets had been cultured in RPMI 1640 + 5mM glucose for 5-7 hours before dispersion and fixation. ADCY5 was present in 46 out of 50 beta cells in wild type mice and 48 out of 50 beta cells in *Adcy5*^{Y1064C} homozygous mice (Figure 5.3). ADCY5 was present with glucagon in 30 out of 35 alpha cells viewed in wild type mice and all 35 alpha cells viewed from homozygous mice. A representative image from each genotype is shown in Figure 5.3. Labeling in the dispersed cells localized ADCY5 throughout the cytoplasm in beta and alpha cells rather than on the outer cell membrane. The protein may be clustering in distinct locations in the cytoplasm as is evident in the punctuate pattern of staining rather than filling the cytoplasm in a smooth and ubiquitous distribution.

5.3.4 Visualizing ADCY5 in mouse pancreas sections from *Adcy5*^{Y1064C} mice

Sections of whole pancreas from 2 wild type, 2 heterozygous and 2 homozygous *Adcy5*^{Y1064C} mice were immunohistochemically labelled for ADCY5, insulin and glucagon (Figure 5.4). The pattern of labeling in the whole pancreas sections differed from the nearly consistent labeling of ADCY5 in dispersed islet cells. In the pancreata of the wild type mice, ADCY5 was present in beta cells (labelling was consistently high), but it was only present at low levels in alpha cells (labelling was present in some alpha cells but fluorescence was not bright). Islets from heterozygous mice also exhibited very little ADCY5 in alpha cells. Conversely, ADCY5 was present at very low levels in beta cells in mice homozygous for the Y1064C allele and showed stronger immunofluorescence in alpha cells (Table 5.2, Figure 5.4). Figure 5.3 also demonstrates little immunolabeling of ADCY5 with glucagon in an islet from a wild type and a heterozygous mouse but higher fluorescence in the alpha cells of the islet from the Y1064C homozygous mouse. The labelling in mouse pancreas sections differs from labelling in isolated islet cells discussed in section 5.3.3; dispersed alpha and beta cells both

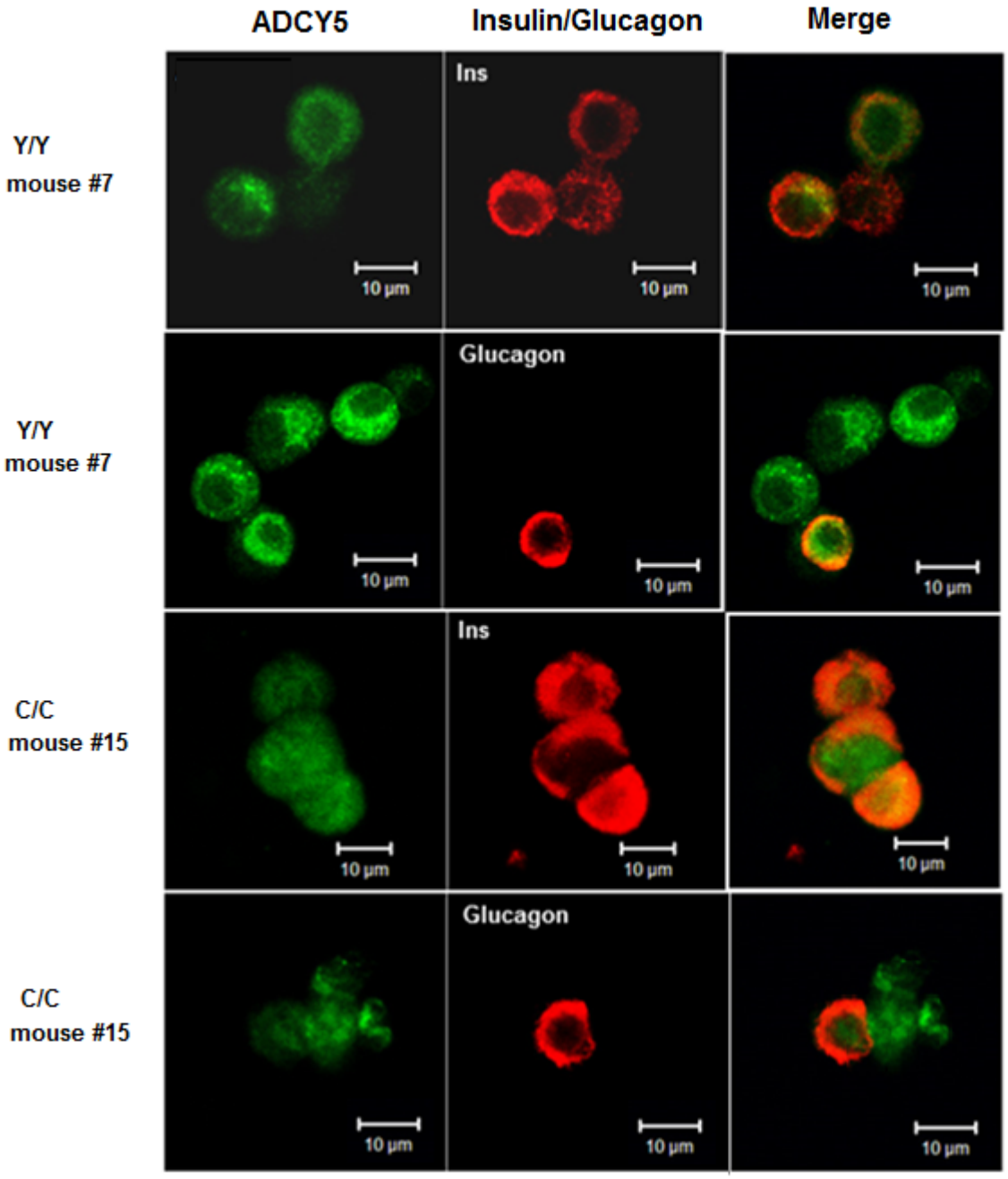


Figure 5.3: Dispersed islet cells from 2 wild type (Y/Y) and 2 homozygous (C/C) *Adcy5*^{Y1064C} mice. These images are representative of all islets observed in both genotypes. ADCY5 is present in most alpha and beta cells in both wild type and homozygous mice. The ADCY5 protein localises to the cytoplasm of the cells.

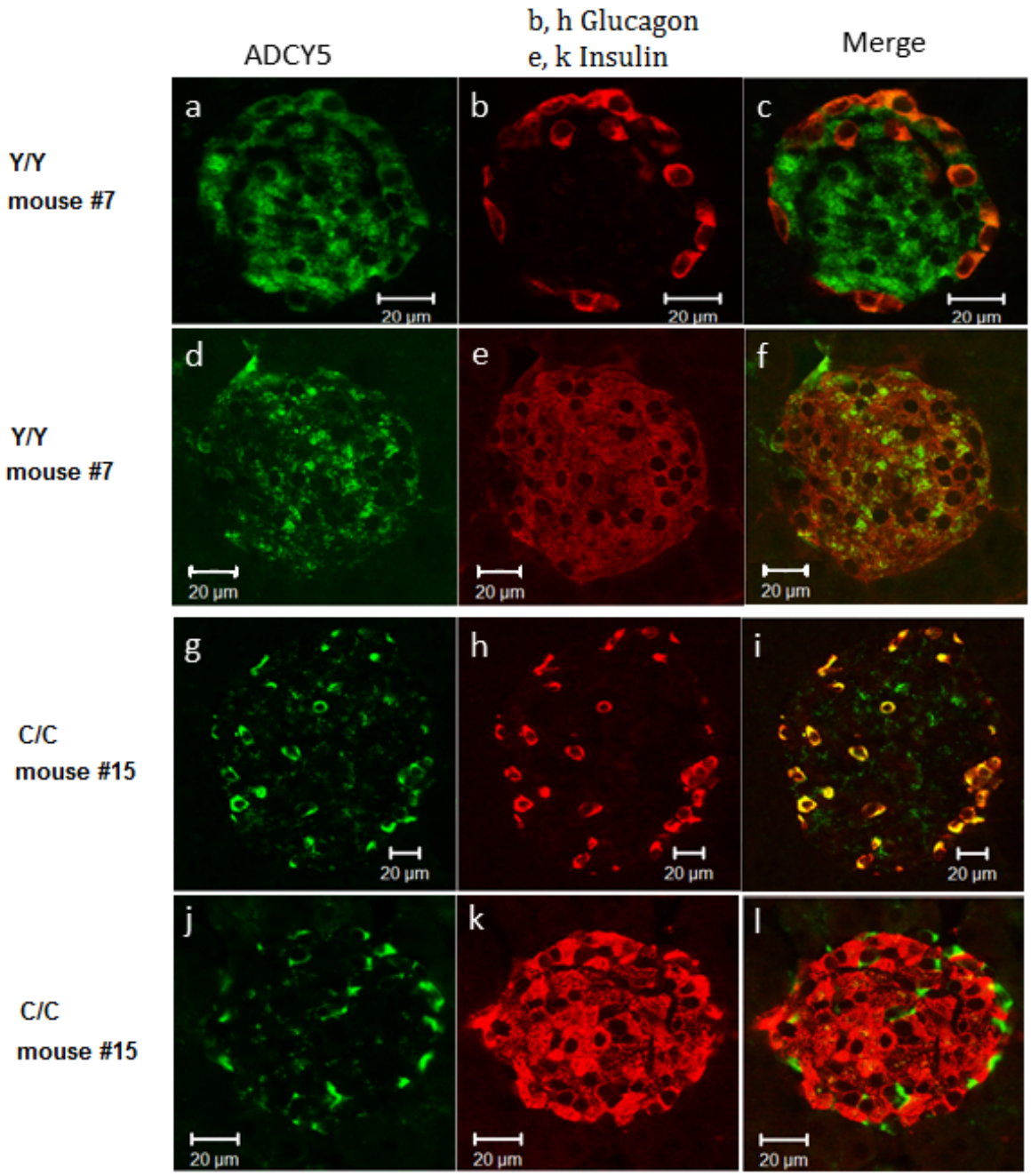


Figure 5.4: Representative islets from 1 wild type (mouse ID #7) and 1 *Adcy5*^{Y1064C} homozygous (mouse ID #15) mouse immunolabeled for ADCY5, insulin and glucagon. In a wild type mouse, ADCY5 is present with glucagon only in some alpha cells (a,b,c), but is clearly present in most beta cells (d,e,f). Alpha cells where glucagon and ADCY5 do not co-localize are marked by arrows. In the islet from a homozygous mouse (#15), ADCY5 is present all alpha cells (g,h,i) but very sparse in beta cells (j,k,l). ADCY5 is distributed throughout the cytoplasm of the cells in all photos.

ID	Genotype	ADCY5 + Insulin (beta cells)	ADCY5 + Glucagon (alpha cells)
1	Y/Y	High immunolabeling	Very low immunolabeling
7	Y/Y	High immunolabeling	Low immunolabeling
4	Y/C	Speckled immunolabeling	Low immunolabeling
11	Y/C	Moderate immunolabeling	Low immunolabeling
14	C/C	Low immunolabeling	Moderate immunolabeling
15	C/C	Low immunolabeling	High immunolabeling

Table 5.2: Summary of immunoreactivity in mouse pancreas sections for ADCY5, insulin and glucagon. ADCY5 was present with insulin in beta cells of wild type mice (Y/Y) and mice heterozygous (Y/C) for the Y1064C mutation, but it was very sparse in alpha cells. In homozygous mice (C/C), the reverse was true: ADCY5 was abundantly present in alpha cells, but only present in a few beta cells. The heterozygote mouse islets were similar to wild type.

consistently label for ADCY5, but in pancreas sections, labelling of ADCY5 in alpha cells was limited. It is likely that this is due to different fixation and preparation protocols. Single cells on coverslips are not subjected to stringent fixation and heat which can affect protein structure and therefore antigen availability. Alternatively, the culture conditions of the islets prior to fixation (RPMI + 5mM glucose) may have altered gene expression.

5.3.5 Structure of islet cells in wild type and Y1064C mutants using electron microscopy

Isolated islets from 3 wild type and 3 homozygous *Adcy5*^{Y1064C} mice were fixed as described in section 5.2.5 of this chapter, and I took them for analysis. There were no differences in organelle structure between wild type and homozygous mice (Figure 5.5). Particular attention was paid to mitochondrial structure, nuclear evidence for apoptosis (nuclear condensation), endoplasmic reticulum dilation (signs of ER stress), and abnormalities in granule structure. Observations on at least 5 islets per genotype suggested that the beta and alpha cells were of normal appearance with no evidence of severe degranulation of insulin and glucagon secretory granules in both genotypes. There is no evidence of the reduced insulin content phenotype discussed in Chapter 4 from visualization by electron microscopy. However, insulin content was only ~30% lower than wild type which may not be evident at this level of magnification.

5.3.6 Visualizing ADCY5 in human pancreas sections

Human pancreas sections from 2 type 2 diabetic patients and 3 controls (see Methods section 5.2.1 for specimen details) were stained for ADCY5, insulin and glucagon and visualized under a confocal microscope. ADCY5 consistently co-localized with insulin in the beta cells of non-diabetic patients. In patients with type 2 diabetes, the presence of ADCY5 in the beta cells was inconsistent and was only detected in about 30% of beta cells (see Figure 5.6 and Table 5.3). ADCY5 was minimally present in

alpha cells in two out of three control subjects and absent in the third. However, in the islets from type 2 diabetic patients, there was significant labelling of ADCY5 with glucagon in the alpha cells (see Figure 5.6).

Under high magnification in Figure 5.7, ADCY5 is visibly dispersed through the cytoplasm of the beta cells of an islet from a control subject. This is a representative image ofADCY5 immunolabeling in all patients examined. As in the fixed islet cells in the previous sections, ADCY5 is not locating to the outer cell membrane but is distributed throughout the cytoplasm.

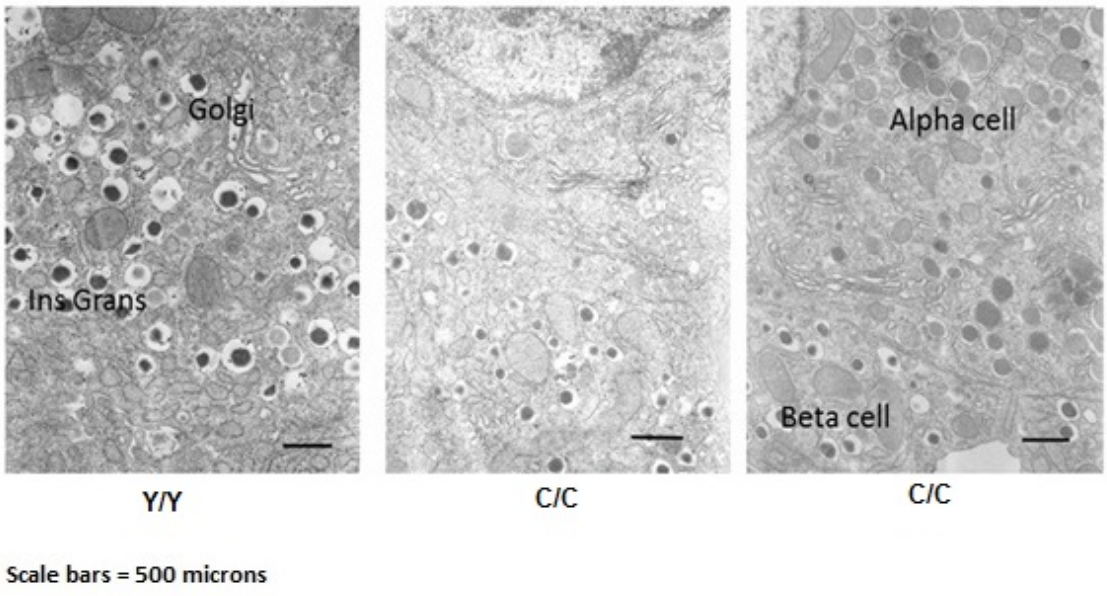


Figure 5.5: Wild type (Y/Y) and homozygous (C/C) *Adcy5*^{Y1064C} mouse pancreas sections viewed under an electron microscope. There are no visible differences between wild type and homozygous mouse pancreata. Beta and alpha cells appear normal with normal insulin and glucagon granules. All organelles are intact.

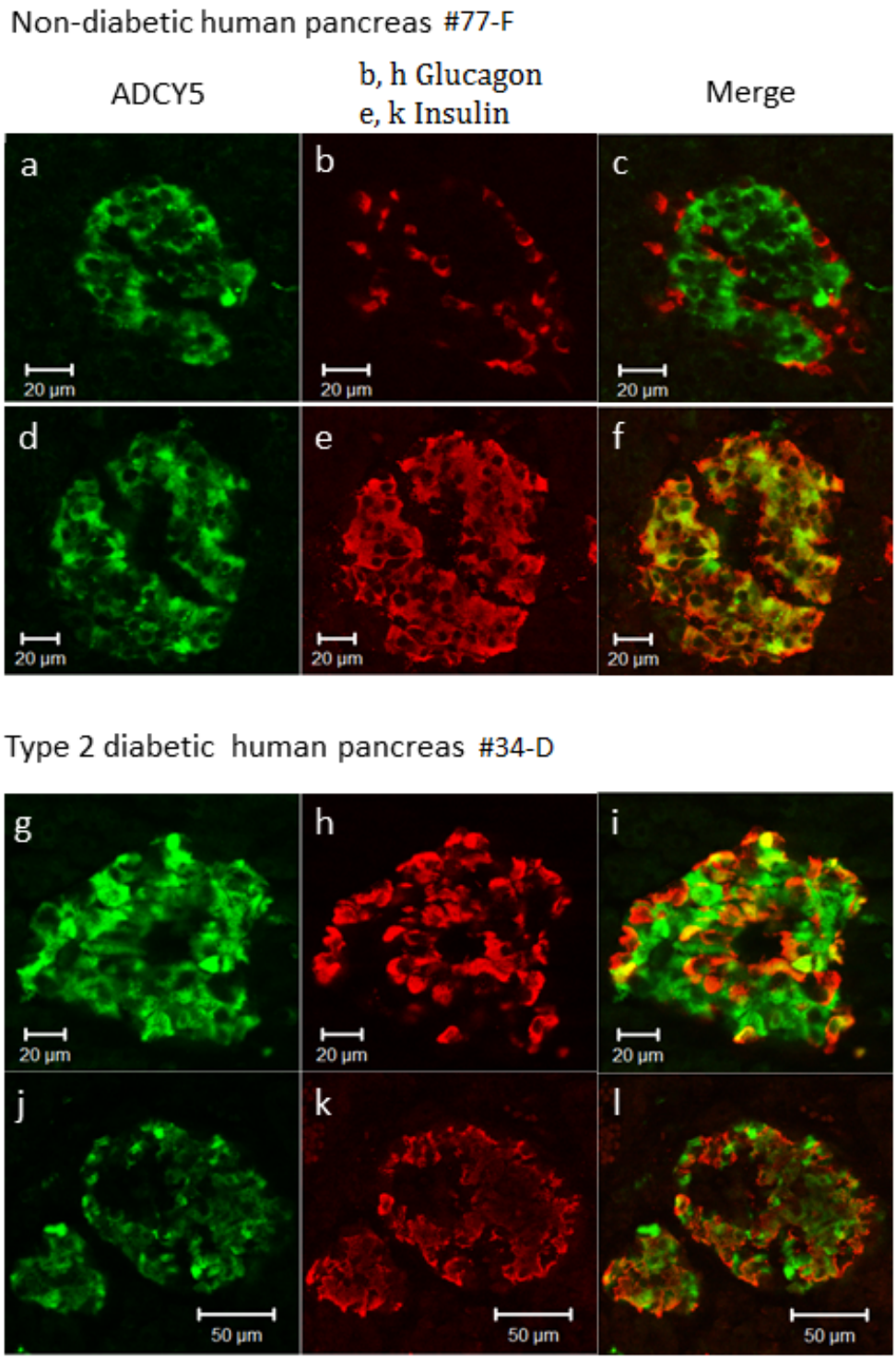


Figure 5.6: Staining for ADCY5 in human islets from pancreas sections of non-diabetic patients and patients with type 2 diabetes. In control patients, ADCY5 is not present in alpha cells (a,b,c) but is present in beta cells (d,e,f). Type 2 diabetic patients show more co-localization with glucagon in the alpha cells than controls (g,h,i), but there is less ADCY5 in the beta cells of T2D patients than in controls (j,k,l).

ID	Type	ADCY5 + Insulin	ADCY5 + Glucagon
77-F	Control	Co-localization	Very low immunolabeling
43-F	Control	Co-localization	Low immunolabeling
17-D	Control	Co-localization	Low immunolabeling
34-D	T2DM	Sparse immunolabeling	Co-localization
120-D	T2DM	Very low immunolabeling	Co-localization

Table 5.3: Summary of ADCY5 immunoreactivity in human pancreas sections from control and type 2 diabetic patients. ADCY5 was present with insulin in the beta cells of non-diabetic patients. Co-localization signifies high immunolabeling of endocrine hormone with ADCY5. Only some beta cells were labelled with ADCY5 in diabetic patients. ADCY5 is hardly present in alpha cells only in non-diabetic patients, but in type 2 diabetes patients, there is significant presence of ADCY5 in alpha cells.

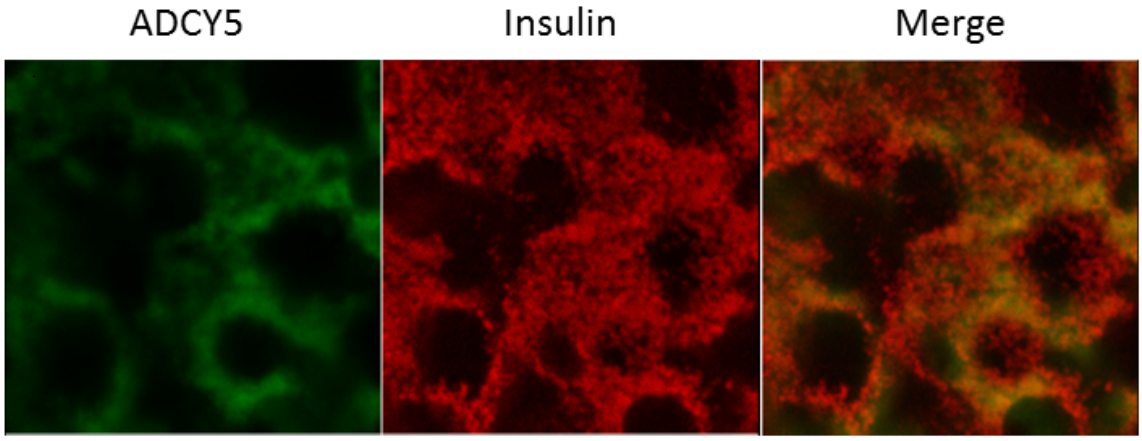


Figure 5.7: A representative islet from a human non-diabetic pancreas section (patient ID 43-F) immunolabeled for ADCY5 and insulin at high magnification to show cytoplasmic location of ADCY5.

5.4 Conclusions

5.4.1 Morphometric analysis of islets from *Adcy5*^{Y1064C} homozygote and heterozygote mice did not show quantifiable differences from wild type

Islet size or density was not changed in mice carrying one or two *Adcy5*^{Y1064C} alleles compared to wild type. The lack of significant change in islet size with age suggests that by 22 weeks, islet size is largely established and no further expansion would occur with age. The islet quantification did not find statistically different alpha and beta cell proportion (alpha cell mass/islet area; beta cell mass/islet area) or islet density. Presented in the *in vivo* phenotyping data (Chapter 4) is data suggesting that the Y1064C homozygote mice may be protected against hyperglycaemia. The *Adcy5*^{Y1064C} mutation may cause a functional change in hormone secretion rather than creating a requirement for compensatory changes in islet beta cell or alpha cell mass.

When comparing this morphometric analysis with the literature, I found that average islet density and median islet size were higher in this analyses for wild type as well as the *Adcy5*^{Y1064C} allele carriers than other authors have reported for wild type and varying transgenic models (Shimomura et al., 2013; Brereton et al., 2014). Shimomura and colleagues (2013) found an average islet size of 4,000 ± 2,100µm^2 in wild type mice at 6 months (~28 weeks). The average islet size for wild type mice at 30 weeks of age in my data was 16,700 ± 5,600µm^2. This fourfold difference is likely a strain difference. Shimomura and colleagues performed their analyses on C57BL/6 mice while my analyses were performed on a C3H/HeH genetic background. I also measured higher islet densities (islet area/ pancreas area) in my mice than Shimomura and colleagues did by about 3-8 fold. However, my analyses were closely in line with measurements of human islet density where islet mass generally makes up about two percent of pancreas mass (Weir and Bonner-Weir, 2013). Islet density in the *Adcy5*^{Y1064C} mice and wild type ranged between 1-6% of pancreas area. In a separate analysis of islet density in C3H/HeH mice the authors found islet densities in six 17 week old male mice ranging

between 1-2% of pancreas area (Goldsworthy et al., 2008). My analysis yielded slightly higher values on average, but the difference is likely explained by age. The mice I counted were 22 or 30 weeks old and may have developed some normal islet hyperplasia as a result of age and weight gain.

The altered distribution of alpha cells in homozygous mouse islets (Section 5.3.1) is typical of other mouse models of diabetes with alpha cells becoming located within the islet rather at the periphery (Gannon et al., 2000; Harndahl et al., 2004). This altered architecture is likely to cause disruption of inter-cellular signaling (paracrine signaling) which could affect regulation of both insulin and glucagon secretion (Bosco et al., 1989; Wendt et al., 2004). Glucose-mediated glucagon inhibition is mediated through GABA signaling from neighboring beta cells (Wendt et al., 2004). The islet architecture is crucial for inter-cell communication and disordered alpha cell distribution in Y1064C allele carriers may disrupt hormone secretion. A further experiment to test this hypothesis would be to measure capacitance of single alpha cells in intact islets from wild type and from Y1064C allele carriers. Membrane capacitance, a measurement of exocytosis, would indicate whether alpha cells interior to the islet were responding to stimuli differently from alpha cells on the periphery of the islet.

5.4.2 The Y1064C mutation alters distribution of the ADCY5 protein in mouse islets

ADCY5 was present in alpha and beta cells of the mice in all genotypes as demonstrated in the immunolabeling of dispersed islet cells in Figure 5.4. However, immunolabeling of pancreas sections revealed that there are some differences in ADCY5 distribution between the genotypes. In wild type islets, ADCY5 is present in all beta cells but is only present in some alpha cells. The reverse is true in the islets of homozygous mice: ADCY5 immunolabeling is strong in alpha cells but lower and sparser in beta cells. The glucagon secretion data (Chapter 4) suggest glucagon secretion is impaired in homozygotes in response to adrenaline. The increased ADCY5 signal in alpha cells may be a compensatory mechanism in attempt to repair glucagon response.

5.4.3 ADCY5 localizes to the cytoplasm of alpha and beta cells

ADCY5 immunolabeling of pancreas sections and dispersed islets in all genotypes was spread through the cytoplasm of alpha and beta cells in a speckled pattern. I expected ADCY5 to localize to the outer cell membrane because the vast majority of literature on membrane-bound ACs focuses on G protein signaling at the plasma membrane. Nevertheless, the data clearly indicates that ADCY5 attaches to a membrane of an organelle within the cell. Furthermore, data from Hodson et al. (2014) confirmed that ADCY5 immuolabeling shows a cytoplasmic distribution across alpha and beta cells in human islets. Cyclic-AMP signaling occurs in various organelles where adenylate cyclases are required to carry the signal from the membrane to the cytoplasm (DiPilato et al., 2004). Adenylate cyclases have been located in the mitochondrial membrane, nuclear membrane, and endoplasmic reticulum (Lyssand et al., 2007; Sanders, 1987; DiPilato et al, 2004). Beta-adrenergic receptors are known to be located on the nuclear membrane, but there is no direct evidence of their localization to other internal membranes (Tadevosyan et al., 2011). Association with adrenergic receptors may be a logical association considering my data from Chapter 4 that the Y1064C mutation impairs glucagon signaling.

Adenylate cyclases are also located on caveolae which endocytose to carry signals throughout the cell (Allen et al., 2009). However, if ADCY5 localized to caveolae, I would expect to see more labelling at the outer cell membrane where caveolae remain until endocytosis. Likewise, localization to the nuclear membrane would show a clear signal lining the nucleus. ADCY5 may mediate signaling at the mitochondrial or ER membranes. There is very little literature about AC signaling at these membranes leaving the specific function of ADCY5 an intriguing question. Immunogold labelling for ADCY5 and visualization under an electron microscope should resolve this question.

5.4.4 The pattern of ADCY5 distribution in Y1064C$^{-/-}$ mouse islets matches ADCY5 distribution in diabetic humans

Human pancreas sections from non-diabetic patients stained for ADCY5, insulin and glucagon showed immunolabeling of ADCY5 in beta cells but little to no labelling in alpha cells. In individuals with T2D, ADCY5 may be present in fewer beta cells than in the control patients, but it is present in most alpha cells unlike in the controls. The mouse pancreas sections show a similar pattern. The islets from the wild type mice have stronger ADCY5 labelling in beta cells while homozygotes have stronger labelling in alpha cells. This data seems to contradict the *in vivo* data presented in Chapter 4. In vivo, the data indicates that *Adcy5*^{Y1064C} homozygote mice have lower blood glucose after an oral glucose challenge and therefore may be protected against T2D. However, the pattern of localization of the ADCY5 protein in the pancreata of homozygote mice more closely matches the pattern seen in islets from human T2D patients. The shifted dispersion of ADCY5 could be an artifact of the fixation procedures of the pancreata, which involve high heat and chemical treatment and may disrupt the antigens. Only 2 pancreata from diabetic patients were tested and 3 controls. The disease progression and therefore *in vivo* environment prior to fixation may have been quite different for these specimens. The sample size is too small to determine if this effect is true for most patients.

A further experiment of interest could be to knock-in the *ADCY5*^{Y1063C} mutation in human islets and measure insulin and glucagon secretion in varying challenge conditions. The results of this experiment compared to the measurements from control human patients and type 2 diabetic subjects would clarify if the Y1064C mutation is causing the islets to behave more like diabetic islets. This *in vitro* experiment would eliminate *in vivo* factors such as blood glucose levels and endocrine hormone circulation which could have been impacting gene expression at the time of pancreas fixation.

In summary, pancreas morphometric analysis localized the ADCY5 protein to punctuate locations in the cytoplasm of both alpha and beta cells in all genotypes. Wild type mice have strong ADCY5 immunolabeling in beta cells and very little in alpha cells, but homozygous *Adcy5*^{Y1064C} mice have strong ADCY5 immunolabeling in alpha cells and very little in beta cells. In 2 T2D human subjects, ADCY5 immunolabeling more closely resembled the pattern of the homozygous *Adcy5*^{Y1064C} mice.

Chapter 6

KLF14: a promising locus for insulin resistance and a master regulator of genes in fat tissue

6.1 Introduction

6.1.1 Variants near *KLF14* are associated with multiple phenotypes in humans

A marker near *KLF14,* rs4731702, is associated with T2D with parent of origin specific effects (Kong et al., 2009; Voight et al., 2010). When inherited maternally, individuals are 1.17 times more likely to be type 2 diabetic than those with allele inherited paternally or with wild type alleles. *KLF14* is part of an imprinted region and is only expressed from the maternal allele consistent with these findings. Further findings from the data showed that possession of the G allele at the rs972283 locus upstream of *KLF14* is associated with T2D ($p=2.2x10^{-10}$ in a cohort of 42,542 cases and 98,912 controls). This variant is also associated with higher fasting insulin and reduced HOMA-IR, a measure of insulin sensitivity. The variation near *KLF14* shows maternally-restricted *cis*-regulation of *KLF14* expression levels in subcutaneous adipose tissue, strongly implicating *KLF14* as the causal transcript (Kong et al., 2009; Small et al., 2011).

In 2010, Teslovich et al., demonstrated that the same variants associated with T2D near KLF14 are also associated with HDL cholesterol levels and another variant rs1562398 in the same locus is associated with triglyceride levels in females only. This sex-specific effect has yet to be explained. Prior to the GWAS studies linking *KLF14* with T2D, Stacey et al., (2009) found variants at a different locus telomeric to *KLF14* associate with basal cell carcinoma. The variants were shown to only exert an effect when inherited paternally in contrast to the parent-of-origin effects on the T2D-associated SNPs which are only associated when inherited maternally (Kong et al., 2009; Voight et al., 2010.) The multiple associations at this locus suggest that *KLF14* is involved in several biological processes.

There is a biological pattern of genes with variation associated with metabolic phenotypes also containing other variants associated with cancer. *KLF14* is one of these genes, as is the third gene I scanned for ENU mutations, *JAZF1*. *JAZF1* locates near variation associated with T2D and other variants associated with prostate cancer. Numerous other examples exist including the gene most strongly associated with T2D to date, *TCF7L2* (variation nearby is associated with risk of breast

cancer and colon cancer [Connor et al., 2012; Slattery et al., 2008]). Another consistently strong association with T2D, *PPARG*, has a known role in cancer and an agonist of *PPARG* has been evaluated in clinical trials for liposarcoma and prostate cancer (Tachibana et al., 2008). Many hypotheses for the link between metabolic disease and cancer have been proposed. The two most likely mechanisms involve a) changes in hormone and nutrient levels causing imbalance in cell metabolism and leading to cell cycle deregulation, or b) activation of stress and inflammatory pathways leading eventually to the initiation of cancer (Joshi et al., 2015). Those genes implicated in both diseases may have a critical role to play in hormonal stabilization or inflammation.

6.1.2 Significance of the imprinting pattern of *KLF14*

Imprinting is characterized by the expression of alleles limited in a parent-of-origin manner due to epigenetic modifications. Imprinted genes are generally expressed mono-allelicly or heavily biased towards one allele. Often these imprinted genes cluster in regions in which all genes and non-coding transcripts are expressed maternally or paternally. *KLF14* falls at position 7q32, a known imprinted region. Genes in this region have been linked to disorders such as Russell-Silver Syndrome, a disorder causing growth restrictions and facial abnormalities (Gicquel C, et al., 2005). When expression was examined in various fetal and adult tissues in human and mouse, *KLF14* was shown to be expressed only from the maternal allele in every tissue examined (Parker-Katiraee et al., 2007). *KLF14* is expressed at greater levels in embryonic tissues than in adult tissues suggesting a role for the transcription factor in development (Parker-Katiraee, 2007). Many imprinted genes do have a role in development. A hypothesis has been proposed that typically, maternally-imprinted genes suppress embryonic growth to protect the health of the mother and ensure availability of resources for her, while paternally-imprinted genes encourage fetal growth to guarantee viability of the offspring (Moore and Haige, 1991). If this hypothesis applies to *KLF14*, it would predict that *KLF14* suppresses expression of genes enhancing fetal growth.

A subset of those genes implicated in both metabolic disorders and in cancer are also imprinted genes. Maternally inherited mutations near *KCNQ1* are associated with T2D, and the potassium channel is also implicated in gastrointestinal cancer (Kong et al., 2009; Than et al., 2014). Loss of imprinting at the *KCNQ1* locus and abnormal biallelic expression of *KCNQ1* gives rise to congenital growth and hyperglycemic disorder Beckwith-Wiedemann Syndrome (Lee et al., 1999). Another locus, the gene encoding the alpha subunit of the stimulatory G protein of adenylyl cyclase (*Gsα*) is imprinted in a tissue-specific manner, meaning that it is expressed primarily from the maternal allele in some tissues and not in others (Yu et al., 1998). Loss of maternal imprinting at the *Gsα* gene leads to pseudohypoparathyroidism, a condition mainly characterized by resistance to the parathyroid hormone (Sakamoto et al., 2004). While other mutations near the *Gsα* gene give rise to somatotrophic adenomas (growth hormone-secreting tumours)(Hamacher et al., 1998). Disruption of imprinting can lead to a spectrum of metabolic disorders demonstrating the importance of imprinting on growth and development. The mechanisms linking imprinting disorders to cancer are better understood as more genes are studied with implications in both diseases, but the broad principle of unchecked growth is not difficult to appreciate.

The CpG island in the promoter of a gene is a cytosine nucleotide followed by a guanine nucleotide which when located in a promoter sequence can be methylated to turn off expression of the gene. Parker-Katiraee found that the CpG island of *Klf14* is hypomethylated, but there was no difference between alleles. However, another gene in the 7q32 region *Mest* (mesoderm specific transcription homolog) had clear allele-specific histone modifications as part of a differentially methylated region (DMR). Parker-Katiraee et al. suggest that this DMR may regulate the imprinted expression of *Klf14* and the other genes in the region through long-range chromatin interactions.

6.1.3 Structure and biology of Krueppel-like transcription factors

In mammals there are 17 genes in the KLF family. The KLFs are distinguished by the three highly conserved Cys2/His2 zinc fingers located at the C-terminus. These zinc fingers are linked together by

sequences known as Krueppel-links (Turner and Crossley, 1999). *KLF*s can activate or repress genes and are involved in various processes throughout the body including haematopoiesis (Nuez et al., 1995), organ development (Wani et al., 1999), cell growth regulation (Black et al., 2001), and endothelial cell function (Lin et al., 2005).

KLF14 has been found to silence expression of the transforming growth factor-β receptor II (Truty et al, 2008), see Figure 6.3. *TGFβ* signaling is a well-known pathway for cell proliferation and differentiation. Although *TGF-β* can both promote proliferation as well as suppress it, most models of *TGFβR2* dysfunction lead to enhanced proliferation, such as in cancerous tumours (Forrester and Chytil, 2005). This would suggest a role for *KLF14* in promoting growth and proliferation since it is a suppressor of the *TGF-β* receptor II. However, this suggests the opposite of the hypothesis of maternal imprinting discussed in section 6.1.3 whereby genes imprinted maternally suppress foetal growth. Several explanations are possible: a) *KLF14* may only promote growth after foetal development, b) *KLF14* is a regulator of hundreds of different genes (Small et al., 2011) and others of these may suppress foetal growth, or c) *KLF14* may promote growth via *TGF-β* under certain conditions such as only adipose tissue but not in other tissues.

6.1.4 Krueppel-like transcription factors and adipocyte biology

KLF4, KLF6, KLF 5, KLF9, KLF15, KLF2, KLF3 and *KLF7* all have known roles in adipogenesis (Wu and Wang 2013). Adipocyte differentiation is a complex process controlled by a large network of transcription factors. At least five of the KLFs involved in adipogenesis are known to regulate *PPARγ*, a gene robustly associated with T2D susceptibility working through insulin action (Banerjee et al., 2003; Wu et al., 2005; Oishi et al., 2005; Mori et al., 2005; Pei et al., 2010; Altshuler et al., 2000). Other *KLF*s regulate *CEBPA/B/Y* or *GLUT4* at various stages during adipogenesis (Birsoy et al., 2008; Chen et al., 2005; Pei et al., 2010; Kawamura et al., 2006). Mice with 2 copies of the *Klf5* gene knocked out exhibit a deficiency of white adipose tissue (WAT), and mice with one copy knocked out had greater numbers of smaller adipocytes, suggesting a role in adipogenesis (Oishi et al., 2008). The

other 6 *KLF*s show similar phenotypes in knockout mice with either greater or less accumulation of WAT.

6.1.5 *KLF14* is a master *trans* regulator of genes in adipose

Small and colleagues from the MuTHER consortium (2011) found that the same set of SNPs located upstream of *KLF14* which were implicated in T2D and cis-expression effects were also driving the expression of over 40 transcripts in *trans* (meaning in a non-adjacent region of the genome) in subcutaneous adipose. Seven of these associations reached genome-wide significance (see Table 6.1). As an author on this paper, I provided information on the biology of *KLF14* and implicated *trans*-associated genes. The authors inferred that *KLF14* acts as a master regulator of adipocyte expression. This hypothesis is supported by the known biology of the *KLF*s as transcription regulators in various tissues (section 6.1.4). Additionally, the promoters of each of the *trans* genes contain *KLF* binding sites strongly suggesting that the T2D-risk effect is mediated through the effects of one or more of the *trans*-genes (Small et al., 2011). RNA-sequencing analyses on the same set of samples revealed 140 genes seen to be part of the trans-array for *KLF14* at 5% FDR (false discovery rate), including those seen previously by Small et al. (2011) and many more (EUROBATS et al., unpublished). Several of the *trans*-associated transcripts, including *SLC7A10*, harbor variants that are independently associated with a range of insulin-resistance traits. Small et al. hypothesized that *KLF14* and the genes it controls in *trans* are involved in communicating the abnormal function of subcutaneous adipose tissue to other organs including muscle and liver and triggering insulin resistance in the major sites of insulin action. The nature of these signals must be determined through research in cellular models, animal studies and studies in man.

Gene	Chr	*Decode Maternal* Effect, P value	*Decode Paternal* Effect, P value	*Combined MuTHER + Decode Maternal* Z score	P value	Direction
APH1B	15	0.17, 0.07	0.07, 0.44	6.2	5.5E-10	++
ARSD	X	0.51, 2.6E-8	-0.004, 0.96	8.7	2.6E-18	++
C8orf82	8	0.69, 2.1E-14	-0.09, 0.28	9.4	4.8E-21	++
GNB1	1	0.42, 1.6E-6	0.06, 0.51	7.3	3.0E-13	++
KLF13	15	0.01, 0.89	-0.02, 0.80	4.9	8.5E-7	++
MYL5	4	0.45, 1.3E-7	-0.04, 0.60	7.5	5.0E-14	++
NINJ2	12	0.24, 0.01	0.05, 0.59	6.3	2.2E-10	++
PRMT2	21	0.27, 6.7E-3	0.09, 0.33	6.4	1.2E-10	++
SLC7A10	19	-0.31, 3.3E-4	-0.11, 0.18	-7.3	2.0E-13	--
TPMT	6	-0.03, 0.78	-0.06, 0.49	6.4	1.1E-10	+-

Table 6.1: Association of gene expression levels of trans-genes with SNPs upstream of *KLF14* that reach genome-wide significance ($p<5 \times 10^{-8}$). Associations were tested in MuTHER (Multiple Tissue Human Expression Resource) data from ~850 UK twins and Decode data set containing genetic and medical information from over 500,000 individuals. The Decode data was scanned in full and then only genetic information inherited maternally and that inherited paternally to look for an imprinted pattern consistent with *KLF14*. Finally, association was tested in both data sets combined. Reproduced from Small et al. (2011) with permission from first author Dr. Kerrin Small.

The aims of this chapter are as follows:

1. To characterize the expression profile of *KLF14* in adult and fetal tissues in humans to better understand where *KLF14* may be having an impact on metabolism.

2. To profile expression of *Klf14* and associated genes in mouse adipose tissue by real-time PCR and evaluate directional consistency with *Klf14* expression to inform phenotyping of the *Klf14* R238L, *Klf14* knockout mouse lines, and trans-associated gene modeling in mice.

3. To look for evidence of enrichment of T2D-associated variants in pathways containing *KLF14* trans-associated genes to better understand the biology of how this gene network may be contributing to T2D risk and further inform phenotyping of the R238L model.

6.2 Methods

6.2.1 Samples used for RT-qPCR analysis

Human total RNA samples (testis, uterus, cerebellum, brain, skeletal muscle, adrenal gland, kidney, liver, heart, fetal liver, fetal brain, and placenta) were purchased as part of a commercially available tissue panel (Clontech, Panel #1409300, Oxford, UK), with each sample comprising a pool of between one and 64 donors (table 4.1). Omental adipose (2 donors) and subcutaneous adipose (1 donor) were obtained through Oxford University's MolSURG project, collected as part of the MolPAGE (Molecular, Phenotyping to Accelerate Genomic Epidemiology) consortium (Min et al., 2012). Tissue samples were collected with full research consent (Oxfordshire Regional Ethics Committee B) from participants undergoing elective surgery, snap frozen and stored as whole tissue in liquid nitrogen prior to extraction.

Mouse total RNA was extracted from mouse subcutaneous adipose, liver, and skeletal muscle using the protocol described in Chapter 2 section 2.3.1.

cDNA synthesis and mRNA expression quantification was completed using the procedure found in Chapter 2 sections 2.3.2 and 2.3.3. Inventoried assays were selected to cover the maximum number of transcripts and where possible, to span an exon junction (_m1 suffix). Since *KLF14, Klf14* and *Cebpa* are one exon genes, assays were ordered with the _s1 suffix. To control for genomic DNA contamination, RT-ve (lacking the RT enzyme) reactions were included in quantitative real-time PCR (qRT-PCR) experiments for *KLF14* experiments. Assay details provided in table 6.3.

6.2.2 Analyses of *KLF14 trans*-associated transcripts using MAGENTA

6.2.2 a) Analyses for enrichment of T2D risk variants in *KLF14 trans*-network gene set

The aim of this analysis was to test if there is enrichment of T2D risk variants in the gene set containing 140 transcripts associated in trans with *KLF14* from RNA-Seq data (EUROBATS, unpublished data). KLF14 was not included as part of the gene set because of its previously identified association with T2D. I will refer to this gene set as the *KLF14 trans*-network.

MAGENTA (Meta-Analysis Gene-Set enrichment of variaNT Associations) is a computational tool devised by Segre et al. (2000) to test for enrichment of genetic associations (such as associations with T2D, fasting glucose, or BMI) in sets of functionally related genes or pathways. The inputs required are a list of variant association p-values and their chromosomal positions, and pathway/s or gene set/s of interest. The output of MAGENTA is a gene set enrichment analysis p-value (GSEA p-value), and a false discovery rate (FDR) for each gene set tested.

A list of 26,914 human gene transcripts from the UCSC Genome Browser (http://genome.ucsc.edu/) in RefFlat format based on the human March 2006 (hg18) assembly was used in conjunction with the T2D GWA meta-analysis (DIAGRAMv3) (Morris et al., 2012) to test the pathway enrichment of *KLF14 trans*-associated genes with MAGENTA (Segrè et al., 2000).

Tissue	N donors	Donor age range (years)
testis	39	14-64
uterus	8	23-63
adrenal gland	62	15-61
skeletal muscle	3	43-46
heart	3	30-39
cerebellum	10	22-68
adult liver	3	29-51
fetal liver	21	26-40 (weeks)
fetal brain	59	20-33 (weeks)
placenta	15	19-33

Table 6.2: Human total RNA panel donor details.

Human Assays		
Gene ID	**Full gene name**	**Assay ID**
KLF14	krueppel-like factor 14	Hs00370951_s1
Mouse Assays		
Klf14	krueppel-llike factor 14	Mm03646643_s1
Slc7a10	solute carrier family 7 member 10	Mm00502045_m1
Tgfb1	transforming growth factor beta 1	Mm01178820_m1
Prmt2	protein arginine methyltransferase 2	Mm00459994_m1
Tgfbr2	transforming growth factor beta receptor II	Mm00436977_m1
Tpmt	thiopurine S-methyltransferase	Mm01349379_m1
Gnb1	guanine nucleotide binding protein, beta polypeptide 1	Mm00515002_m1
Cebpa	CCAAT/enhancer-binding protein alpha	Mm00514283_s1

Table 6.3: Assay details for total expression quantification. Assay IDs are TaqMan inventoried expression assays (Applied Biosystems).

The analysis was performed across 21,491 T2D cases and 55,647 controls across 25 studies from European descent populations (Morris et al., 2012) and using 2,474,735 imputed SNPs. Genes in the *KLF14 trans*-network were assigned a score that is the probability that the gene is associated with T2D. This score is a function of its regional SNP association p-values. I assigned gene boundaries to be -110kb/+40kb because these boundaries represent the 99th percentile of the distance of *cis*-eQTLs from their adjacent gene's transcript start and end sites (Segrè et al., 2010). The gene score from each of the genes in the genome were ordered from most likely to be associated with T2D to the least likely. The *KLF14 trans*-network gene set was tested for enrichment of the top 5% best-scoring genes. A p-value was calculated based on whether the gene set had more than the expected number of the top 5% genes compared to other randomly-sampled gene sets of identical size. The 95th percentile of association p-values was used to determine the enrichment cut-off because the 95th percentile has optimal power for finding genes of both smaller and larger effect size (Segrè et al., 2010). The 75th percentile enrichment cut-off was also calculated because pathways enriched at this cut-off may be a better predictor for highly polygenic diseases with many associations of weak effect (Segrè et al., 2010). This first analysis along with the second is simplified and diagrammed in Figure 6.1.

A second gene set was tested where the list of 140 *trans*-genes was condensed into 95 regions where all those within 500kb of each other were condensed into a single region. *KLF14* was not included in the gene set because of its previously discovered association with T2D. The same protocol as above was followed.

After Bonferroni correction for testing 2 gene sets (that is, the significance cut-off p = 0.05 was divided by the number of hypotheses tested), gene sets with nominal GSEA p-value of <0.025 were considered significantly enriched. However, these gene sets are highly correlated, so this cut-off may be too stringent.

6.2.2 b) Pathway analysis of *KLF14 trans*-associated transcripts in gene sets enriched for T2D variants

The strategy employed for the second analysis was to scan the MSigDB database of ~1500 gene sets for pathways enriched for T2D-associated variants and determine if the *KLF14 trans*-associated genes fall within any of the gene sets most highly enriched for T2D variants. The analysis was performed in MAGENTA using the DIAGRAMv3 T2D GWA meta-analysis as above. For all gene sets in the Gene Ontology terms section of the MSigDB database v4.0 (http://www.broadinstitute.org/gsea/index.jsp, updated May 31, 2013), I asked which gene sets were enriched for the top 5% best scoring genes near T2D associated variants (gene score calculated as above). The gene ontology set includes biological process terms, molecular function terms, and cellular localization terms. The 95^{th} percentile and 75^{th} percentile of association p-values were used to determine the enrichment cut-off as before. Gene sets with MAGENTA nominal GSEA p-value of <0.05 were considered significant. *KLF14* trans-associated genes were flagged in gene sets significantly enriched for T2D-associated SNPs. A diagram visualizing the two analyses is shown in Figure 6.1. Bonferroni correction was applied to correct for multiple hypothesis testing. In this case 1453 gene sets were tested and a single trait, so the most stringent cut-off was 3.4×10^{-5}. This may be too stringent when a large number of gene sets is tested due to overlap of genes between the different gene sets. I relaxed the correction and included the next 4 most-enriched gene sets.

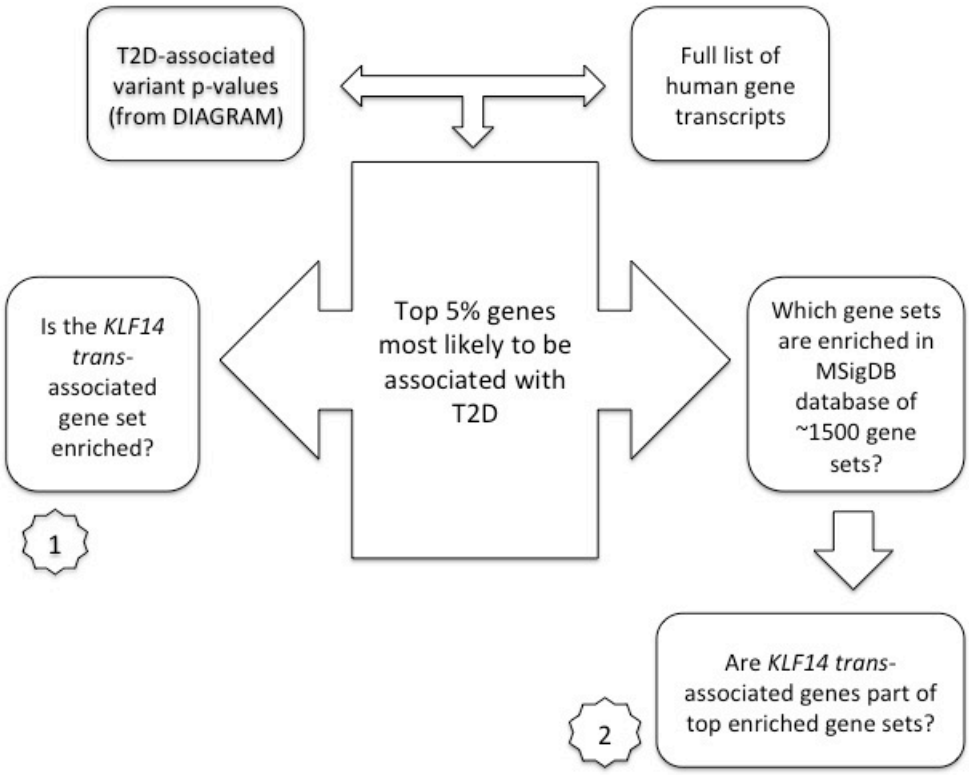


Figure 6.1:MAGENTA gene set analyses visualized. The program uses the inputs of T2D-associated variant p-values and their chromosomal locations plus a full list of human gene transcripts to rank the genes from most likely to least likely to be associated with T2D. In the first analysis described above (1), the gene set containing the *KLF14 trans*-associated genes was analysed for enrichment of the top 5% of genes most likely to be associated with T2D. In the second analysis (2), the ~1500 gene sets in the MSigDB database were queried for enrichment of the top 5% of genes most likely to be associated with T2D, and then the top enriched pathways were searched for *KLF14 trans*-associated genes.

6.3 Results

6.3.1 *KLF14* Expression Profiling in Humans

I performed qRT-PCR for *KLF14* expression over a range of human tissues to pinpoint where *KLF14* is most expressed in the body. Kong et al. (2009) found that T2D-associated signals are associated with reduced *KLF14* expression in subcutaneous adipose. Subcutaneous as well as omental fat was included in my analysis to determine if *KLF14* is only expressed in subcutaneous fat, or if it is present in omental fat as well.

KLF14 is highly expressed in testis (Figure 6.2[a]). It is also expressed in skeletal muscle and the adrenal gland but is not present in cerebellum or liver. In fetal tissues, *KLF14* was highly expressed in placenta but hardly expressed in fetal brain and fetal liver (Figure 6.2[a]). Parker-Katiraee in 2007 showed a similar pattern of expression for *KLF14,* confirming its presence in skeletal muscle, fetal liver and placenta, but finding it absent in adult liver. I also looked at *KLF14* expression in omental and subcutaneous human adipose (Figure 6.2[a]). I included two samples of omental adipose: one from an individual with low body-mass index (BMI < 21.0) (Figure 6.3, omental fat 2) and one from an individual with high BMI (BMI > 28.0, omental fat 1). *KLF14* was only present in the female with a BMI of 19.1kg/m^2. The male with a BMI of 31.1kg/m^2 did not express any *KLF14* in omental fat. *KFL14* was substantially expressed in subcutaneous fat (n=1), although confirmation in more samples is needed for *KLF14* profiling in types of adipose tissue. Small et al. (2011) performed cis-eQTL analyses of *KLF14* expression and found that variants near *KLF14* associate with levels of expression of *KLF14* in subcutaneous adipose tissue. My results also support the presence of *KLF14* in subcutaneous adipose tissue.

I had the benefit of comparing my results with expression results posted in the GTex Portal (Genotype Tissue Expression Portal, GTEx Consortium, 2013). The GTex Portal showed the same pattern of expression for *KLF14* in the tissues I analyzed: testis displayed the highest expression of

KLF14 followed by adrenal gland and then subcutaneous fat (Figure 6.2[b]). Omental fat also showed substantial *KLF14* expression. The GTEx Portal did not have data for placenta or fetal tissues.

6.3.2 *Klf14* Expression Profiling in Mouse

The *Klf14* R238L model (described in Chapter 3) provided an interesting opportunity to investigate expression of *Klf14* and *trans*-associated transcripts: first to see if *Klf14* expression is altered in mice with the maternally-inherited R238L mutation, and second to observe whether expression of known *KLF14 trans*-associated transcripts in man is also changed in mice. I performed qRT-PCR in mouse tissues from 2 wild type mice, 3 mice with the *Klf14* R238L mutation inherited maternally, and 3 mice with the mutation inherited paternally. The aim was to analyze *Klf14* gene expression and expression of genes which may be regulated by *Klf14.* This was a preliminary analysis as the sample size was small.

The genes included in the analysis were: solute carrier family 7 member 10 (*Slc7a10*), protein arginine n-methyltransferase 2 (*Prmt2*), thiopurine s-methyltransferase (*Tpmt*), guanine nucleotide-binding protein beta-1 (*Gnb1*), transforming growth factor beta-1 (*Tgfb1*), transforming growth factor beta receptor 2 (*Tgfbr2*), CCAAT/enhancer-binding protein alpha (*Cebpa*), and *Klf14*. *SLC7A10, PRMT2, TPMT*, and *GNB1* have been implicated as genes under *trans*-regulation by *KLF14* (Small et al., 2011). *PRMT2* and *GNB1* also show up in the top pathways enriched for T2D variants described in section 6.3.3 (Table 6.3). *PRMT2* is an arginine methyl-transferase involved in protein metabolism, and *GNB1* is part of the G protein complex involved in extracellular signaling.

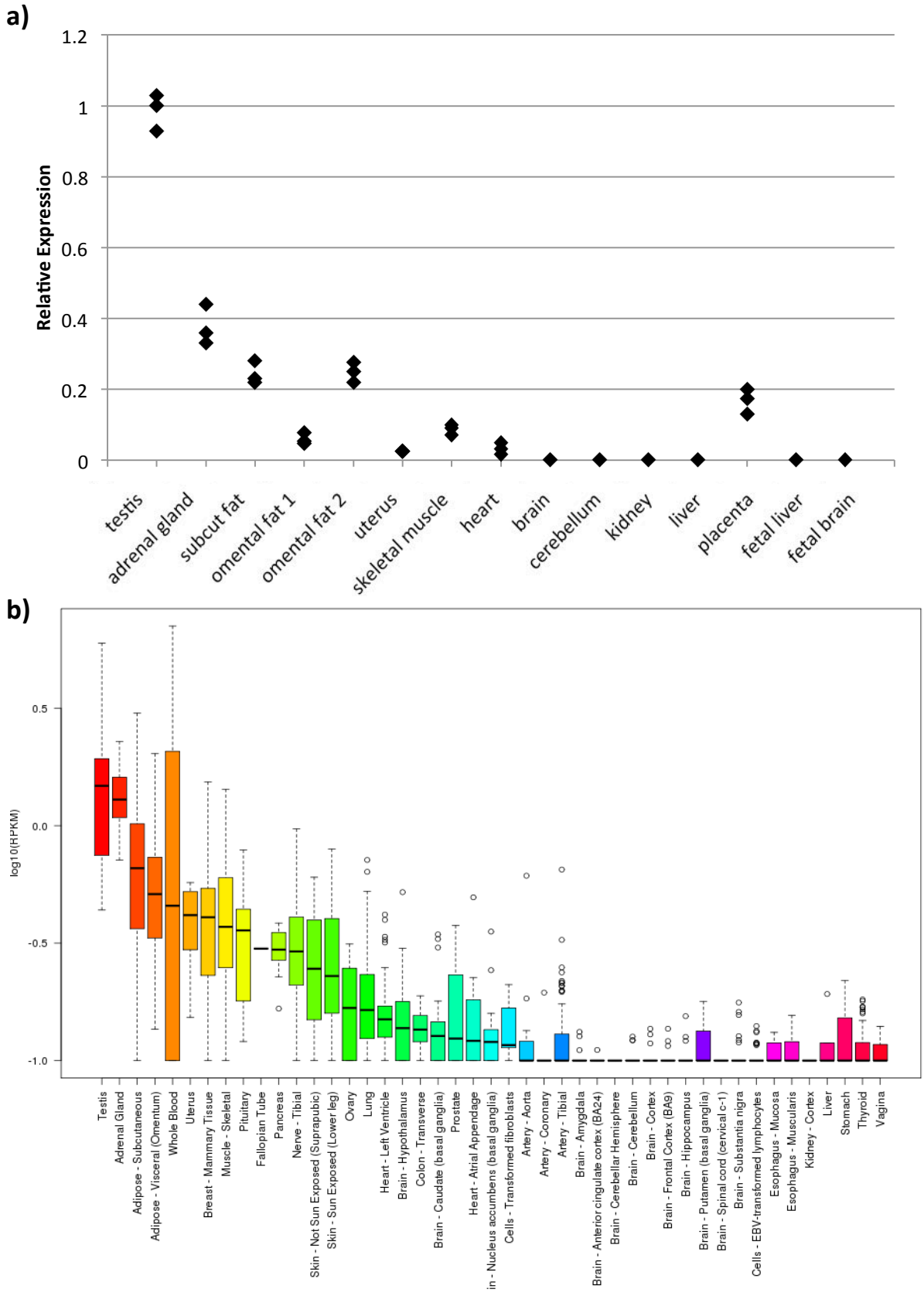


Figure 6.2: (a) Expression of *KLF14* in human adult, foetal, and adipose tissues relative to the tissue with highest level of expression, testis. *KLF14* was expressed highly in testis and adrenal gland followed closely by expression in adipose tissue. It was expressed most highly in subcutaneous fat (n=1). *KLF14* is also expressed in an omental fat sample from a female with BMI 19.1kg/m^2 (omental fat 2) and in an omental fat from a man with BMI 31.1kg/m^2 (omental fat 1). (b) Expression of KLF14 in human tissues from the GTex Portal. Gene expression patterns are similar to my analysis in common tissues.

KLF14 is implicated in TGF-beta signalling (Parker-Katiraee, 2008), so I also chose to include expression of TGF-beta protein 1 (*Tgfb1*) and TGF-beta receptor 2 (*Tgfbr2*). Finally, I included *Cebpa*. Variants near *CEBPA* have been associated with triglyceride levels in humans, and *CEBPA* is known to regulate genes involved in lipid and glucose metabolism (Olofsson et al., 2008).

Several of the krueppel-like factor family including *KLF2, KLF3,* and *KLF7* regulate expression of *CEBPA* (Banerjee et al., 2003; Sue et al., 2008; Kawamura et al., 2006). All three of these genes inhibit *CEBPA* expression which in turn inhibits adipogenesis. A correlation in expression between *Klf14* and *Cebpa* would inform understanding of the directional effect of *Klf14'*s regulation of adipogenesis.

Two controls were used in the experiment due to the imprinting pattern of *Klf14*. In addition to wild type controls, heterozygotes with the R238L allele inherited paternally were included. I expect these mice with the paternal allele not to express the mutation and therefore to act like wild types.

Klf14 was expressed highly in subcutaneous adipose tissue mice carrying the Klf14 R238L paternal allele (Figure 6.3a). Mice carrying the R238L mutation on the maternal allele expressed *Klf14* at one half of the levels expressed in mice with the mutation inherited paternally (p=0.008). *Klf14* expression in mice with 2 wild type alleles was very different between the two mice analyzed, so *Klf14* expression was not conclusive. *Slc7a10* was expressed at lower levels when the R238L allele was inherited maternally than when it was inherited paternally ($p = 2.25x10^{-6}$) or when both alleles were wild type (p = 0.01). This finding supports the *trans*-association data in humans which indicates that expression of *SLC7A10* is correlated with expression of *KLF14* in adipose (Small et al., 2011). The T2D risk allele near *KLF14* is correlated with lower expression of *KLF14* and also lower expression of *SCL7A10* in subcutaneous adipose tissue. In this analysis, the R238L allele demonstrates the same pattern of effect on expression. There was no significant difference in expression between genotypes of any of the other genes investigated in subcutaneous adipose.

In liver, *Tgf1b, Tpmt, Gnb1*, and *Cebpa* were expressed substantially while *Klf14, Prmt2* and *Slc7a10* were expressed at very low levels. *Slc7a10* expression was too low to be detected. Genotype at the *Klf14*^{R238} allele had no effect on expression of any of the genes (Figure 6.3b).

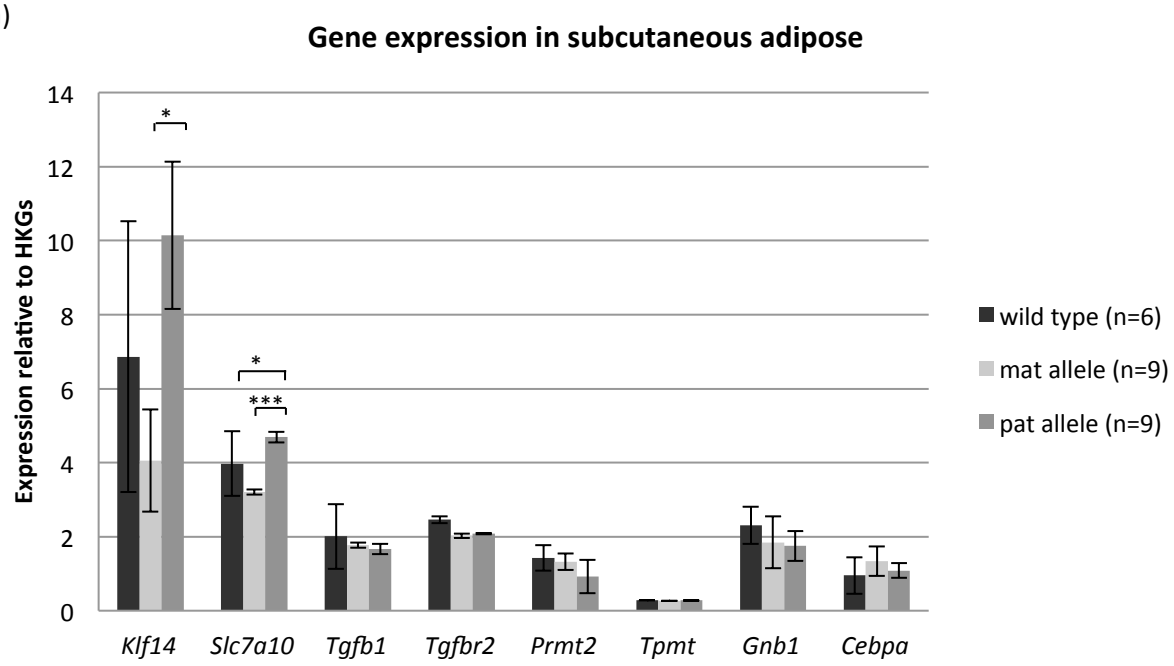
In skeletal muscle, expression of several genes was altered when the R238L allele was inherited maternally (Figure 6.3c). *Tpmt* was expressed at lower levels in muscle in mice with the maternal R238L allele than in mice with two wild type alleles (p=0.002) or the R238L paternal allele (p=0.006). *Gnb1* is expressed at lower levels in mice carrying the maternal allele than in mice carrying the paternal allele (p=0.0001), as is *Tgfb1* (p=0.006). Expression was not different for *Gnb1* and *Tgfb1* between wild type mice and those carrying the maternal allele. This may be because the number of wild type individuals was too low to detect a difference (n=2). The change in expression of these genes when the R238L mutation is inherited maternally may suggest a mechanism of regulation by *Klf14* across tissues.

6.3.3 Analysis of *KLF14 trans*-associated transcripts in gene sets enriched for T2D-associated variants

I performed a gene set analysis of the 140 transcripts associated in *trans* with *KLF14* to learn more about the pathway biology of *KLF14* and its *trans*-network. As described in the methods section of this chapter (6.2.5) I defined the *KLF14 trans*-network as a gene set and looked for evidence of enrichment of T2D risk variants. The aim was to determine if the genes under control of *KLF14* form a gene set that harbors common risk alleles for T2D. I also looked for evidence of gene set enrichment of T2D-associated variants and located *KLF14*-associated *trans* genes in these implicated gene sets.

When the *KLF14 trans*-associated genes from the RNAseq network (EUROBATS, unpublished data) are defined as a gene set, both the 140 gene set and condensed 95 region set are enriched for T2D-associated SNPs at the 75% cut-off for GSEA enrichment (p=0.016 and p=0.039) but not at the 95% cut-off (Table 6.4).

a)



b)

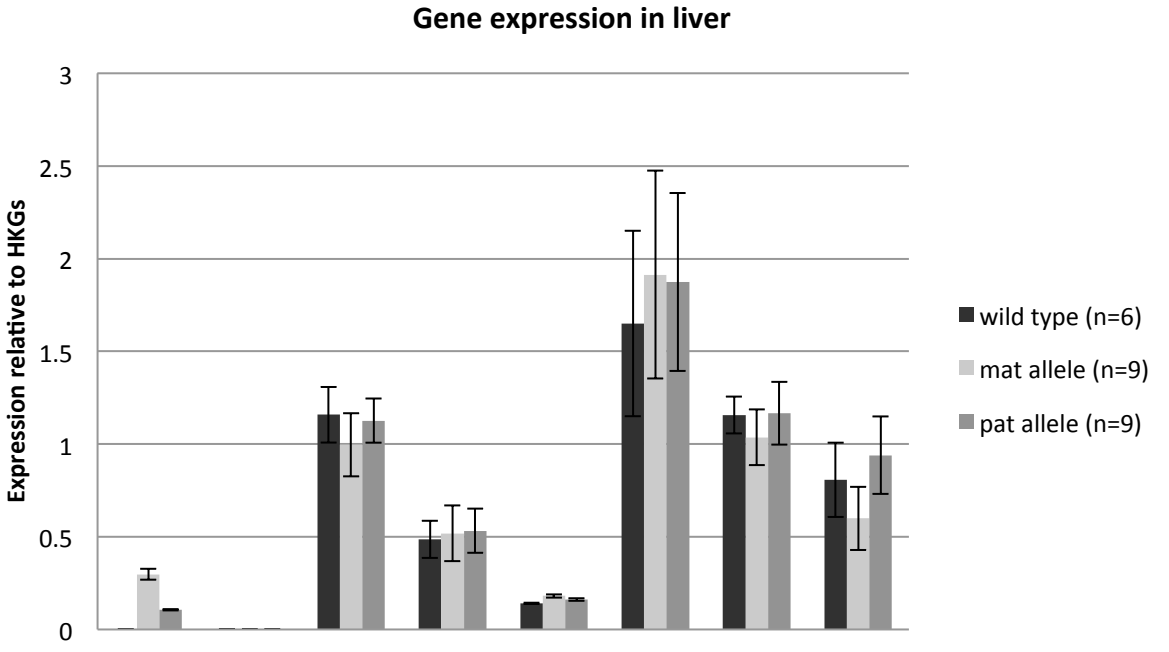


Figure 6.3(a&b): *Klf14* and related gene expression (details on following page). Sample size consists of biological replicates x technical replicates where technical replicates are equal to 3.

c)

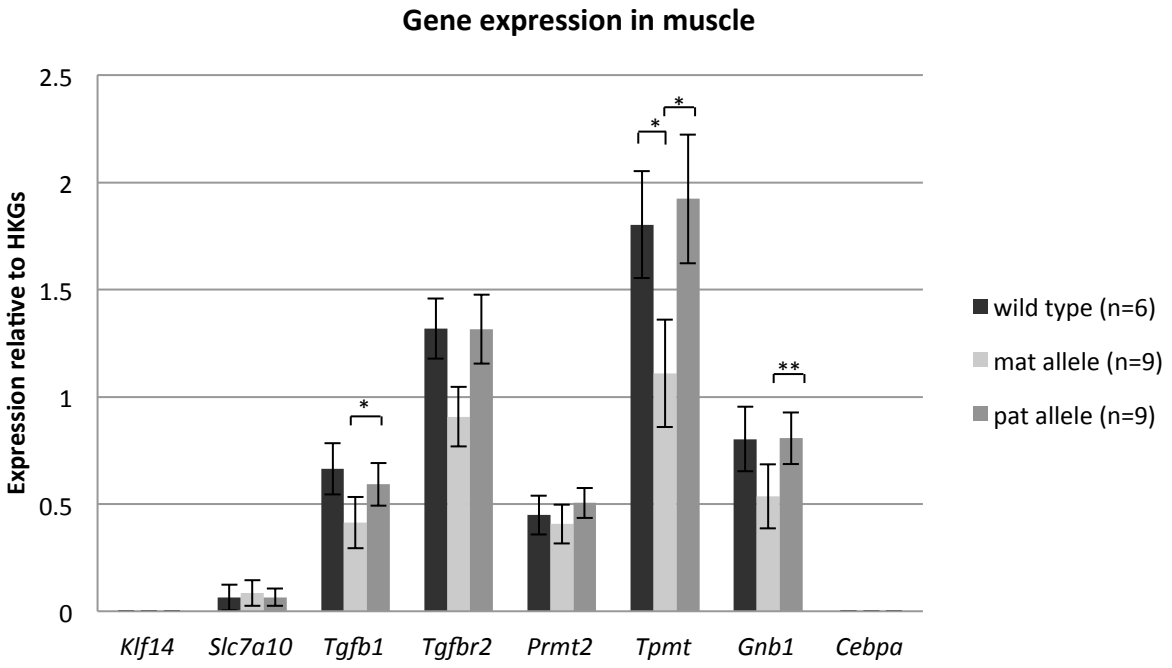


Figure 6.3: (a) *Klf14* expression in subcutaneous adipose is reduced in mice carrying the R238L mutation inherited maternally compared to mice carrying the mutation paternally (*p ≤ 0.01). *Slc7a10* expression is reduced in mice carrying the R238L mutation maternally compared to mice carrying the mutation on the paternal allele (*p≤1x10^{-5}) and compared to mice with two wild type alleles (*p≤0.01). (b) Expression of *Klf14-trans* associated genes in liver is not different between genotypes. (c) Expression of *Tgfb1,* and *Gnb1* in muscle is lower in heterozygotes with the maternally inherited allele compared to mice with the paternally-inherited allele (*p≤0.01 and **p≤0.0001 respectively), and *Tpmt* expression is lower in mice with the R238L maternal allele compared to wild type and heterozygotes with the paternal allele (*p≤0.01 in both comparisons) suggesting that these genes may be regulated by *Klf14* in muscle. Sample sizes are biological replicates x technical replicates where technical replicates are equal to 3.**

KLF14 was not included in the gene sets to eliminate the signal already discovered as a T2D-associated locus. While the 95% cut-off yields optimum power for gene enrichment in diseases with 100 or fewer associations, the 75% cut-off performed marginally better for weaker effects and may be a better predictor for highly polygenic diseases with many associations (closer to 500) of weak effect (Segrè et al., 2010). T2D is a highly polygenic disease, and thus the association at the 75% cut-off may mean that the *KLF14 trans* network is enriched for common T2D risk variants of weaker effect (Morris et al., 2012).

When all gene sets (MSigDB v4.0 , http://www.broadinstitute.org/gsea/index.jsp) separated into biological processes, molecular function, and cellular localization databases were scanned for evidence of enrichment of T2D associated variants (DIAGRAMv3: Morris et al., 2012), 3 gene sets were found to be enriched above the 95% cut-off ($p<3.4x10^{-5}$ after Bonferroni correction)(Table 6.5). I relaxed the Bonferroni correction because it is likely too stringent for this analysis as many of the gene sets are highly correlated. Four more gene sets were enriched for T2D associated variants above the 95% cut-off with $p<1.0x10^{-3}$. The associated pathways were mainly related to positive regulation of biological processes and metabolic processes. Six out of the 7 pathways with GSEA p-value less than $1x10^{-3}$ at the 95% cut-off contain genes from the *KLF14 trans* network. Thus, these genes are likely connected to processes highly associated with risk of T2D.

6.4 Conclusions

KLF14 is implicated in insulin resistance in humans and disease-associated alleles are linked to its expression in subcutaneous adipose tissue. It is a master regulator of adipose genes and plays a key part in signaling within adipocytes and to other tissues and when disrupted leads to insulin resistance.

Gene Set	# genes analyzed by GSEA	Nominal GSEA p-value 95% cut-off	FDR 95% cut-off	Nominal GSEA p-value 75% cut-off	FDR 75% cut-off
KLF14trans140	114	0.35	0.63	0.016	0.029
KLF14trans95	89	0.65	0.58	0.038	0.032

Table 6.4: The *KLF14 trans*-associated gene network of 140 genes and condensed list of 95 genes defined as pathways and analyzed for enrichment of T2D-associated SNPs. Neither pathway is significantly enriched at the 95% cut-off, but the *KLF14* 140 *trans* gene set is significantly enriched for T2D risk alleles at the 75% cut-off after Bonferroni correction for multiple hypothesis testing.

Database	Gene Set	# genes analyzed	GSEA p-value 95% cut-off	FDR 95% cut-off	GSEA p-value 75% cut-off	FDR 75% cut-off	*KLF14 trans*-associated genes near T2D validated SNPs
GO, BP	POSITIVE REGULATION OF CELLULAR PROCESS	647	3.00E-06	0.10	0.01	0.38	*ESRRG, MAP3K10, UNC13B, NDFIP2, LGALS12*
GO, BP	BIOPOLYMER METABOLIC PROCESS	1448	4.00E-06	0.097	6.64E-04	0.44	*CHD3, ESR2, ESRRG, GATA3, PRMT2, MAP3K10, NFIL3, PHKA2, POLR2L, PPP2R2A, SREBF1, RRP9, KLF13, ALG8, HDAC11, APH1B*
GO, BP	POSITIVE REGULATION OF CELLULAR PROCESS	615	5.00E-06	0.12	0.007	0.36	*ESRRG, MAP3K10, UNC13B, NDFIP2, LGALS12*
GO, BP	DNA BINDING	538	1.00E-04	0.32	0.003	0.38	*CHD3, ESR2, GATA3, NFIL3, SREBF1, TBX4, SCMH1*
GO, BP	GROWTH	74	3.00E-04	0.29	0.017	0.40	*DCBLD2*
GO, BP	TRIACYLGLYCEROL METABOLIC PROCESS	9	5.00E-04	0.10	0.009	0.37	
GO, BP	TRANSCRIPTION FACTOR ACTIVITY	320	6.00E-04	0.30	0.033	0.49	*ESR2, GATA3, NFIL3, SREBF1, TBX4, SCMH1*

Table 6.5: The 7 top gene sets for enrichment of T2D-associated risk alleles by GSEA p-value at the 95% cut-off. The top 3 meet the most stringent significance cut-off after Bonferroni correction, but this may be too strict as there is overlap between gene sets. Genes regulated in *trans* by *KLF14* are flagged and listed in column 8. Six out of the 7 top gene sets involve at least one *KLF14 trans*-associated gene.

The expression studies, pathway analysis, and mouse modelling discussed in this chapter commence the work of elucidating the mechanisms by which *KLF14* contributes to insulin resistance and the pathology of T2D. The expression work here confirming the presence of *KLF14* in adipose tissue led to the generation of mouse models of *Klf14* disruption including the *Klf14* R238L ENU model described in Chapter 3, and also the *Klf14* knockout model described below.

6.4.1 *KLF14* and *trans* gene expression in human tissue

Quantitative real-time PCR analysis showed that *KLF14* is expressed in omental and subcutaneous adipose tissue supporting its apparently crucial role in fat. Expression in adult tissues other than adipose is limited except in testis, adrenal gland, placenta, and skeletal muscle. Parker-Katiraee et al. (2006) also found high expression of *KLF14* in skeletal muscle and in placenta. They showed expression of *KLF14* in fetal heart, lung and colon which I did not test, suggesting that KLF14 may be involved in development (Parker-Katiraee et al., 2006). The TGFβ growth factors are established developmental genes providing a logical link to KLF14 in development (Leung, 1987). GENATLAS provides no expression data for *KLF14*. However GTEx (Genotype Tissue Expression Portal, GTEx Consortium, 2013) shows the same pattern of expression for *KLF14* in the tissues I analyzed: the highest expression of *KLF14* was in testis followed by adrenal gland and then subcutaneous fat. The two samples of omental fat I tested were from a woman with low BMI and a man with high BMI. The omental adipose from the woman with low BMI showed substantial levels of *KLF14* expression, but the man with high BMI had no expression of *KLF14* in omental fat at all. Further expression work in tissue samples from patients with varied BMIs is now underway by EUROBATS. There is some evidence that pathology from *KLF14* variants may only affect women. Teslovich et al., 2010, found that variants upstream of *KLF14* raise triglyceride levels in women but not in men.

6.4.2 *Klf14* R238L mouse model re-derivation and phenotyping

Re-derivation of mice carrying the *Klf14* R238L allele was problematic as described in Chapter 3, section 3.4. Sixteen live pups were eventually recovered by Dr. Naomi Nakagata and her team in

Japan at Kumamoto University. The presence of the *Klf14* R238L mutation was confirmed by genotyping by staff at the Mary Lyon Center, MRC, Harwell. The mice underwent 3 backcrosses to wild type C57BL/6J mice to breed out confounding mutations before intercrossing. All backcrosses and intercrosses were performed keeping mice carrying a maternally-inherited allele separate from those with a paternally-inherited allele. As described above, the maternal imprinting pattern of *Klf14* is such that those with the paternally-inherited allele are expected to act as controls with similar results to wild type.The quarantine and back-crossing period for these mice exceeded the time I had allotted this project in my DPhil. I chose to focus on a different project and passed the phenotyping on to another researcher.

The *in vivo* phenotyping of these mice has now been completed in Roger Cox's lab at the MRC Harwell by Alison Hough (MRC, Harwell). The mice showed no consistent differences in insulin sensitivity, glucose tolerance, triglyceride levels, HDL-C, cholesterol, or LDL-C. Body weight and fat mass data is inconclusive.

Analysis of expression of *Klf14* and selected related genes including *trans*-associated genes in the R238L model was completed by me as presented in the results section 6.3.2. The results are discussed below.

6.4.3 *Klf14* and *trans* gene expression in metabolic tissues from the *Klf14*^{R238L} mouse model

In wild type mice and those with the *Klf14* R238L paternal allele, *Klf14* is expressed highly in subcutaneous adipose tissue. It is not present in skeletal muscle or liver. The mice carrying the *Klf14* R238L maternal allele had lower expression of *Klf14* in subcutaneous adipose tissue compared to those mice carrying the paternal allele. However expression from wild type was inconclusive due to small sample size. Therefore it is difficult to say whether or not the mutation affects gene expression. It is not common that point mutations lead to lower expression of a gene, but one possible explanation is if the mutation affects RNA stability, this may lead to degradation.

I predicted R238L to disrupt binding of one of the three zinc fingers of *Klf14* to the DNA (Chapter 3). Characterization of protein folding and kinetics will be necessary to determine if the mutation is affecting DNA binding.

One of the genes associated in *trans* with Klf14, *Slc7a10,* displayed the same expression pattern as *Klf14*: it was expressed highly in subcutaneous adipose but not in liver or muscle. In adipose tissue, mice carrying the maternal R238L allele had lower expression of *Slc7a10*. The data indicates that *Klf14* may be directly regulating *Slc7a10* expression in adipose tissue, which was the expected outcome because of the *trans*-association reported by Small et al. (2011). In liver *Tpmt, Gnb1*, *Tgfb1*, *Tgfbr2, and Cebpa* were highly expressed, but there was no difference in expression between genotypes. In muscle expression of several genes was lower in mice carrying the maternal R238L allele including *Tpmt*, *Gnb1* and *Tgfb1*. *Klf14* was not expressed at all in muscle, so this implies a mechanism through which *Klf14* is signaling to these genes across tissues. The different expression profile in each tissue suggests a complex network of signaling.

6.4.4 *Klf14* knockout mouse

Work on this mouse model was not part of my project and was completed by Dr. Alison Hough and Madeleine Pope at the MRC, Harwell, but the data is relevant to report here.

At about the same time as the *Klf14*^{R238L} mouse was re-derived, a knockout model of *Klf14* became available from the Knockout Mouse Project (KOMP). This was a simple null allele model. An expression-selection cassette had been inserted into a Bac vector and through homologous recombination, the *Klf14* gene was deleted (one exon only) and the vector inserted therefore eliminating expression of *Klf14*. The knockout mice received from KOMP were crossed with wild type C57BL/6J mice and then intercrossed to generate mice with 0, 1 and 2 copies of the knockout allele by the staff at the Mary Lyon Centre, MRC Harwell, keeping the maternal heterozygotes and paternal heterozygotes separate. To verify the knockout, Madeleine Pope (University of Oxford) performed qRT-PCR and showed that *Klf14* is not expressed in mice with the maternally inherited

knockout allele. Wild type mice and those expressing the paternally inherited allele expressed *Klf14* at the same level.

In vivo phenotyping of the *Klf14* knockout model is now also complete and was done by Allison Hough of the Cox lab, MRC Harwell. Male mice with the knockout maternal allele have reduced blood insulin at most ages (p=0.01-0.05), but the insulin tolerance test data is inconclusive. There were no differences in blood glucose, glucose tolerance, or triglyceride levels. However, blood HDL-C levels were consistently reduced in male heterozygotes with the knockout maternal allele. Weights were not consistently different at all time points, but at 22 weeks, male homozygotes and heterozygotes with the knockout maternal allele had lower body weight and lower fat mass. The knockout mouse has an HDL-C phenotype consistent with the human variant discovered by Teslovich et al., (2010), but the insulin sensitivity data from the knockout mouse was inconclusive. This phenotype will have to be confirmed to determine whether the knockout allele mimics the T2D-risk allele in humans.

6.4.5 The *KLF14 trans*-network genes are involved in pathways enriched for T2D variants

Gene set analysis using MAGENTA revealed that the pathways from the MSig database which are most enriched for T2D risk variants contain many genes from the *KLF14 trans*-network. Furthermore, the *KLF14* trans-network is enriched for T2D risk variants above the 75% cut-off demonstrating that this network of genes is directly connected to processes highly correlated with T2D pathology.

Many of the *KLF* family are already known to be important genes in adipose signalling. *KLF14* is now known to also be a regulator of genes in adipose tissue and to play a role insulin resistance. My work on *KLF14* showed that it is highly expressed in adipose and not expressed in other metabolic tissues, corroborating the likelihood of *KLF14* playing a primary role in fat tissue. This chapter also presented evidence that the R238L allele, when inherited maternally, is consistent with lower expression of several genes thought to be regulated by *Klf14* in adipose and skeletal muscle. The signaling pathway that allows *KLF14* to regulate these genes in muscle may be a mechanism linking aberrant

fat metabolism to consequent insulin resistance in peripheral tissues. These data in addition to the body weight, body fat, and cholesterol phenotypes in the mouse models confirm that *KLF14* is part of a complex signaling network involved in adipose function and the pathophysiology of T2D.

Further studies investigating *KLF14* will include mouse modeling of several of the trans-associated genes. *Slc7a10* is an interesting transcript that is closely regulated by *Klf14* in adipose. Characterization of a KO mouse model of *Slc7a10* is underway in Professor Roger Cox's lab at the MRC Harwell. Phenotyping of this and other mouse models for adipose and metabolic traits will inform about the mechanisms of disrupted signaling from adipose tissue that lead to insulin resistance. Human studies will also continue on *KLF14*. One effort from the University of Oxford is to sequence *KLF14* in a database of human adipose tissue samples and correlate any mutations with metabolic phenotypes in the patients. *KLF14* is an important gene to continue investigating because it provides a rare opportunity to learn more about the poorly understood processes by which tissues become resistant to insulin.

Chapter 7

General Discussion

7.1 Achievement of Aims

Genome-wide association studies have identified >70 genetic loci associated with risk of T2D (Ng et al., 2014). A continuing challenge is to identify the causal transcript at these loci and then characterise the molecular mechanism driving the association with T2D. In this thesis several approaches are taken to select and investigate probable causal transcripts near T2D-associated loci. The broad method employed here was to generate ENU mouse models of these genes and to characterise them *in vivo* and *in vitro*. The original aims of this thesis as stated in Chapter 1 were:

a) To test candidate GWAS T2D susceptibility genes by identifying point mutations that were likely to affect function (Chapter 3).
b) To generate a mouse model of *Adcy5* dysfunction and characterise this model for glucose tolerance and birth weight phenotypes *in vivo* as well as measure insulin and glucagon secretion from isolated islets and quantify islet hormone content (Chapter 4).
c) To examine pancreas morphology of the *Adcy5* mouse model and localize the protein in islet cells to better understand an islet phenotype (Chapter 5).
d) To profile the expression of *KLF14* in humans and *Klf14* and *trans*-associated genes in mice to inform *Klf14* mouse model characterisation. Gene set analysis was completed to look for evidence of enrichment of T2D-associated variants in the *KLF14 trans*-network (Chapter 6).

During my DPhil I have addressed the above aims. Below I will summarize the work I have done toward each aim and the results which have come from my investigations as well as any questions left unanswered.

7.2 Selection of transcripts and design of ENU mouse models

JAZF1, *ADCY5*, and *KLF14* were chosen for investigation based on a set of selection criteria including but not limited to: association with T2D and other metabolic traits, location of the associated SNPs in relation to the gene, *cis*-expression data linking variants near the gene to expression levels of the transcript, and novelty of the gene. I chose to make models of the selected genes in order to test what effect mutations in these genes had on whole body phenotypes related to T2D, and further to investigate the mechanisms underlying these traits in specific tissues and cells.

There are relative merits each to investigation in man and in mouse. In this thesis loci implicated in GWA studies in man were investigated in mice, and then, where possible, results were compared with data from human cells or tissue. The mouse offers a tool for genetic manipulation, the extent of which could not be possible in humans. Lab mice are much more genetically homogeneous within a strain than humans are, so the cohort numbers can be much smaller. Gathering data from mice is also much faster due to lifespan and generation time. The major drawback to studying metabolic phenotypes in mouse is that the extent to which a phenotype in a mouse and a phenotype in a man are driven by the same mechanism cannot be known with certainty. Hence, it is imperative that wherever possible, data gathered in mice be replicated in humans. A common theme throughout this thesis is the comparison between phenotypes in man and phenotypes in mouse. For the most part, the models generated did not display phenotypes completely consistent with the human data, but the variants in mice were not the same as those discovered by GWAS in man. Different alleles were manipulated to explore the biology of the transcript through which the human variation is predicted to be working. When this thesis began, it was the first functional analysis of Adcy5 in pancreatic islets and the first functional analysis of the role of Klf14 in metabolism outside the cis-expression analyses described in Voight et al. (2010). Therefore, we did not know exactly what we would find, and we wanted to capture the full spectrum of a phenotype. The mouse offered an

opportunity to complete an extensive and efficient investigation into all aspects of a possible phenotype.

T2D is a polygenic disease with though only ~80 loci reaching genome-wide significance have yet been discovered, hundreds of variants of small effect size contribute to inherited T2D predisposition (Morris et al., 2012). Obtaining point mutations in the mouse has the advantage of generating alleles resembling naturally occurring mutations, such as those that might occur in a human and contribute to T2D susceptibility. A significant risk is that while predictions can be made about the result of a mutation, there cannot be certainty that the mutation will change protein function in the way predicted or produce a phenotype.

In this thesis, mutations were selected based on the probability of producing a phenotype after protein modelling in PyMOL and analysis by mutation prediction software Sift and Polyphen. Out of the 15 *Adcy5* mutations in the ENU archive, I selected the *Adcy5* Y1064C mutation because it was predicted to sever a polar bond intrinsic to protein folding, and therefore it was likely to result in a phenotype. I selected the *Klf14* R238L mutation because it was predicted to disrupt binding of a zinc finger to DNA and so was also likely to result in a phenotype.

7.3 Morphometrical examination and quantification of islets from *Adcy5*^{Y1064C} mice

I measured median islet size, islet number, islet density, and beta and alpha cell proportion from wild type mice and mice carrying 1 and 2 copies of the Y1064C ENU mutation at 22 and 30 weeks of age. There were no consistent differences between genotypes at either age.

Fluorescent immunolabeling of the ADCY5 protein revealed differences in alpha and beta cells between genotypes. The islets from wild type mic e had strong immunolabeling corresponding to ample presence of ADCY5 in beta cells but much less dense labelling of ADCY5 in alpha cells. In contrast, the islets from mice homozygous for the *Adcy5* Y1064C mutation had very sparse labelling

of ADCY5 in beta cells and a much stronger signal in alpha cells. Surprisingly labelling in human T2D patients demonstrated a similar immunolabeling pattern as the mice homozygous for the Y1064C mutation. The ENU mutation exhibits protective effects against T2D *in vivo*, so this result is the opposite of what might be expected. However, the mice carrying the ENU allele do have lower insulin content than wild type mice and impaired insulin secretion in response to high glucose + GLP-1 (discussed in the following section) which could be commonalities between T2D patients and Y1064C allele carriers.

Another surprising finding was that ADCY5 does not localize to the outer cell membrane. The ADCY5 protein was consistently spread throughout the cytoplasm in punctuate locations in all genotypes in mice and humans. ADCY5 is a membrane-bound adenylate cyclase, so it likely anchors to another membrane within the cell. Immunolabeling clearly did not locate ADCY5 to the nuclear membrane, so possibilities are the endoplasmic reticulum membrane, the mitochondrial membrane, or the membrane of insulin/glucagon granules. The location of ADCY5 within the cell changes the approach to investigating the molecular mechanisms involved. The stimulators and inhibitors chosen for the secretion experiments were based on the assumption that ADCY5 would be located on the outer cell membrane. However, adrenergic receptors are located both on the outer cell membrane as well as on inner cell membranes in islet cells (Boivin et al., 2006; Tadevosyan et al., 2011). Alternatively, ADCY5 could be potentiating a signal originating on the cell membrane but triggered by a secondary signaling molecule and amplifying the signal throughout the cell.

7.4 Characterisation of the *Adcy5* Y1064C mouse model

Mice carrying 0, 1, and 2 copies of the *Adcy5* Y1064C allele were characterised for metabolic phenotypes including glucose tolerance, body composition, blood insulin, and birth weight. The main findings *in vivo* were: a) Male mice carrying two copies of the *Adcy5* Y1064C allele have lower blood

glucose after an oral glucose tolerance test, and b) the *Adcy5* Y1064C allele does not change birth weight or size.

There was no difference in blood glucose between genotypes after an intraperitoneal glucose tolerance test, but after an oral glucose tolerance test, male mice homozygous for the *Adcy5* C1064 allele had lower blood glucose 30, 60, and 120 minutes after dosage than wild type mice. The reason for this difference may be that the oral route of administration results in the activation of gut hormone secretion: in particular for example, GLP-1 which is an incretin hormone that stimulates insulin secretion and inhibits glucagon secretion. To test this hypothesis I performed insulin and glucagon secretion experiments on islets isolated from the *Adcy5* Y1064C mice in a variety of conditions including in the presence of GLP-1.

Secretion experiments from isolated islets *in vitro* showed that islets from mice carrying two copies of the *Adcy5* Y1064C allele a) had impaired glucagon secretion in response to adrenaline, b) had lower insulin content, and c) had impaired insulin secretion in basal (1mM) glucose compared to wild type. The protective effect of the mutation *in vivo* could be explained by the marked impairment in glucagon secretion when in the presence of the highly stimulatory hormone adrenaline. Islets from mice homozygous for the *Adcy5* $^{Y1064C-/-}$ allele secreted 40-60% less glucagon than islets from wild type mice in 1mM glucose + 5nM adrenaline. Glucagon releases glucose stored as glycogen in the liver to raise blood glucose levels. Impairment of glucagon secretion would therefore lead to lower blood glucose levels and could explain the observed *in vivo* phenotype. However, no difference was found in blood glucagon in the allele carriers compared to wild type from a blood sample after an overnight fast. Glucagon levels were low in all genotypes and only 50% of the samples could be measured, so the experiment was underpowered.

Insulin secretion from isolated islets in basal glucose was higher from islets from cohort 1 mice carrying 1 or 2 copies of the Y1064C allele. However, this finding was not replicated in mice from cohort 2. Insulin content was lower in *Adcy5*^{Y1064C} allele carriers from cohort 1. This could be

explained by an impairment in insulin production, or it could be a result of chronic hyper secretion as was seen from the cohort 1 mice in basal glucose. Unfortunately, insulin content values could not be collected for cohort 2 mice. Hyper secretion of insulin in basal glucose could help explain the in vivo phenotype. Higher circulating levels of insulin would clear glucose from the blood faster. However, when insulin secretion measurements were collected by perfusion (work done by Nils Rorsman, OCDEM) there was no difference between genotypes at basal glucose. The perfusion method is very sensitive, but this is also a very subtle phenotype. The static incubation experiment should be repeated with a larger sample size and a longer incubation period to be sure the mice are not hyper secreting insulin in low glucose. There was some evidence from cohort 2 mice that insulin secretion was impaired in mice homozygous for the Y1064C mutation in the presence of GLP-1 and high glucose but not in low glucose + GLP-1. This finding is in contradiction with the *in vivo* phenotype because it suggests that the mice homozygous for the mutation would have lower blood insulin after an oral glucose challenge and therefore higher blood glucose, but homozygous mutants had lower blood glucose after an OGTT. One possible explanation is that impaired glucagon secretion in the homozygous mutants may still drive blood glucose down regardless of marginally lower levels of insulin secretion. However, insulin islet content could not be measured for these mice, and if insulin content is lower in the allele carriers as it is in cohort 1 mice, then insulin secretion may be lower but not as a percentage of islet content. Secretion normalized to content may be unchanged. Before further investigation it would be necessary to scale up the experiment with more mice and verify insulin secretion as well as insulin content in a variety of GLP-1 and glucose concentrations.

The shifted pattern of localization in islets observed from pancreas sections is also an interesting addition to the phenotype. The ENU allele carriers may have increased ADCY5 in alpha cells to compensate for impaired glucagon secretion and decreased ADCY5 in beta cells to reduce insulin secretion in an effort to keep blood glucose stable. For further investigation, mice fed a normal diet and also a high calorie diet could be sacrificed and pancreas sections fixed to observe localization of

ADCY5. If the islet cells are attempting to normalize blood glucose, and if ADCY5 is involved in the mechanisms suggested above, then ADCY5 localization would be expected to shift further into alpha cells the more the mice are challenged with a high calorie diet.

The risk alleles at the *ADCY5* locus associate with risk of T2D and high fasting glucose. In 2014, Hodson and colleagues found that the T2D risk allele at SNP rs11708067 is associated with lower expression of *ADCY5* in islets in man. When they silenced *ADCY5* in human islets, they found that glucose-stimulated insulin secretion was impaired but not GLP-1-stimulated insulin secretion. In contrast, the *Adcy5* Y1064C homozygote mice have lower blood glucose after an oral glucose challenge. This may suggest the Y1064C mutation is not a loss of function allele but is a gain of function allele. However, the *Adcy5* knockout mouse was reported to have lower fasting glucose than wild type (Yan et al., 2007; Ho DZX, et al., 2010; Li et al., 2006). The Y1064C mouse did not have lower fasting glucose except at 20 weeks in cohort 2 males. Therefore some conflict arises in the data as to whether *ADCY5* is performing the same role in mouse and humans, and also what directional effect the alleles are having. To determine directional effect of the Y1064C allele, the mutation could be inserted into the gene using a vector and kinetic assays performed on the protein to compare catalytic activity to activity of the normal protein. Due to time constraints on my thesis, I was not able to perform this experiment. However, the work is currently underway by Jana Rundle (OCDEM, University of Oxford).

If the mutation is a gain of function, the observation of hyper insulin secretion from isolated islets in basal glucose could be explained. Hodson and colleagues (2014) show that when *ADCY5* is silenced in human islets, insulin secretion in response to glucose is impaired, so if the Y1064C mutation was a gain of function, the islets would be more sensitive to glucose leading to hyper insulin secretion in low glucose.

7.5 Expression profiling of and gene set analysis of *KLF14* and *trans*-associated genes

The second gene investigated in this thesis was the transcription factor *KLF14*. To profile the tissues through which *KLF14* exerts its effects, expression of *KLF14* across a range of adult and foetal tissues was measured. *KLF14 i*s most highly expressed in testis, adrenal gland, subcutaneous fat, and omental fat. These results agree with the results from the GTEx portal (GTEx Consortium, 2013). The experiment confirms that *KLF14* is present in subcutaneous adipose tissue as reported by Small et al. (2013). It is not highly expressed in other metabolic tissues; it is present minimally in muscle and absent from liver in humans.

The re-derivation of the *Klf14* R238L mouse line was problematic due to unforeseen circumstances. When the mice were available, I measured the expression of *Klf14* and a selection of related genes in the *trans* network in 3 mice carrying the maternal R238L allele, 3 mice carrying the paternal allele, and 2 mice with only wild type alleles. Mice carrying the paternal allele were expected to behave as wild type because *Klf14* is a maternally expressed, imprinted gene. I looked at expression in 3 metabolic tissues from the mice: subcutaneous adipose tissue, liver, and skeletal muscle. The aims of this experiment were a) to determine if *Klf14* and related genes were present in these tissues in wild type mice, b) to determine if the R238L allele changed expression of *Klf14*, and c) to determine if the R238L allele changed expression of any of the genes predicted to be under control of *Klf14*. *Klf14* expression was lower in subcutaneous adipose tissue in mice with the maternal ENU allele compared to paternal allele carriers. There was no difference in expression between maternal allele carriers and wild type, but the experiment was underpowered. The sample size for wild type was just 2 mice due to limited availability of live mice from the re-derivation. Therefore, *Klf14* expression was inconclusive.

Expression of *Slc7a10* was lower in mice carrying the maternal ENU allele than in those carrying the paternal ENU allele or two wild type alleles in subcutaneous adipose tissue. In humans, alleles near *KLF14* associate with both lower expression of *KLF14* and lower expression of *SLC7A10* in adipose

tissue suggesting *KLF14* is regulating expression of *SLC7A10* in adipose tissue. This is consistent with the trans-association reported by Small et al. (2011). The R238L mutation may be disrupting *Klf14* regulation of *Slc7a10* leading to lower expression of *Slc7a10* in adipose tissue. No changes of expression of *Klf14* or *trans* genes were seen in liver. In skeletal muscle, several genes regulated by *Klf14* were expressed at lower levels in mice carrying the maternal allele compared to mice carrying the paternal allele including *Tgfb1*, *Gnb1*, and *Tpmt*. *TPMT* and *GNB1* were associated in *trans* with *KLF14* expression in humans in adipose tissue. My data did not show a change in expression of *Tpmt* and *Gnb1* in adipose tissue, but instead a change in skeletal muscle was observed. This could be an indicator of a more complex mechanism by which *Klf14* is regulating these transcripts across tissues. A possible mechanism for this is through regulation of *Slc7a10* in adipose tissue. *Slc7a10* (also called Asc-1) has been implicated in mediating D-serine amino acid transport in the mouse central nervous system through D-serine reuptake in synapses in the brain (Rutter et al., 2007). In adipose tissue, the protein may be mediating transport of molecules in synapses which trigger signals through the nervous system to skeletal muscle and other parts of the body about the state and health of adipose tissue.

Due to time constraints on my thesis, *in vivo* phenotyping of this model was done by Alison Hough (MRC Harwell). No insulin resistance, glucose tolerance or blood cholesterol phenotype was observed. Hence, the R238L allele does not confer the same phenotype in mice as the T2D-associated allele does in humans.

Phenotyping of a *Klf14* knockout model showed a different phenotype. Males with 1 and 2 copies of the knockout allele at 22 weeks had lower body weight and lower fat mass than wild type. HDL-C was significantly reduced at time points, and LDL was reduced at one time point in males carrying the maternal allele. The insulin sensitivity data is unclear and will need to be repeated. Lower blood HDL-C was a trait found in humans with variation at the *KLF14* locus (Teslovich et al., 2010). However, the human alleles do not have an effect on body weight. Further phenotyping of insulin

resistance in the knockout model will be necessary to determine whether the phenotype of the knockout resembles the effects of the risk alleles in man. Insulin sensitivity could be ascertained *in vivo* by administering insulin tolerance tests to *Klf14* knockout and wild type mice fed a normal diet and a high fat diet.

In Chapter 6 I describe work I did to further analyse the *KLF14 trans*-network using a computational program called MAGENTA. MAGENTA looks for enrichment of T2D-associated variation in a gene set specified by the user. I performed a gene set analysis in MAGENTA of *KLF14* and 140 trans-associated genes from humans. The aim was to determine whether the *KLF14 trans*-network, when treated as a gene set, is enriched for T2D-associated variants. I found that the network is enriched at the 75% cut-off for GSEA-enrichment, which is has been indicated as an optimal cut-off for finding variation of weaker effect for highly polygenic diseases such as T2D (Segrè et al., 2010). Further analysis revealed that 6 of the 7 of the top gene sets enriched for T2D variants contain between 1 and 16 genes associated with variation near *KLF14*. These findings indicate that the *KLF14 trans*-network is highly involved in processes which when disrupted can contribute to T2D risk. The genes implicated in top T2D gene sets would be good candidates to investigate functionally in mouse models to further understand the biological mechanisms of T2D risk. Several of the genes in the *trans*-network showed up repeatedly in the pathways most enriched for T2D association including *MAP3K10, GATA3, NFIL3* and *ESRRG.* These may be good candidates to begin functional investigation in mice or cellular models.

7.6 Future directions

Investigation into the molecular mechanism of *ADCY5* and *KLF14* continue. Adenylate cyclase kinetic assays are underway to measure the effects of the *Adcy5* Y1064C mutation on catalytic activity of the ADCY5 protein in vitro. This result will answer the question of whether the Y1064C mutation is a loss or gain of function mutation. Another experiment I would have done with more time would be

to measure cAMP in islet cells from *Adcy5* Y1064C mice. A cAMP assay would directly measure the activity of the adenylate cyclases within the cell. The cells could be assayed after exposure to glucose, adrenaline, or GLP-1.

Another important question to answer is where ADCY5 is locating within the cell. Surprisingly, it is not anchored to the outer cell membrane and is instead spread throughout the cytoplasm. Hodson et al. (2014) found ADCY5 in human islets is also cytoplasmically distributed. The protein is membrane-bound, so it is likely binding to some other membrane within the cell such as the endoplasmic reticulum, mitochondria, or granule membranes. One way to determine the localization is to perform immunogold labelling for ADCY5 and visualize the cells under an electron microscope. The gold particles will be visible under the microscope and indicate where ADCY5 is bound. The location of ADCY5 within the cell will help inform future mechanistic work by narrowing the list of possible upstream and downstream effectors of ADCY5.

The *Adcy5 in vivo* studies could also be expanded. Since my *in vitro* data did not reveal a stark GLP-1 phenotype in the mutant mice, I would have liked to repeat an IPGTT at a later age to see if intraperitoneal glucose tolerance is also improved in the homozygous mice at 20+ weeks. Also, my narrow focus on glucose tolerance phenotypes prohibited me from testing insulin tolerance in the *Adcy5* Y1064C mice. It has now been proposed that *ADCY5* may have a role in insulin sensitivity in the muscle tissue (Hodson et al., 2014). Challenging these mice with a high fat diet and insulin tolerance tests, as well as ex vivo hormone studies in the muscle cells would be extremely interesting as an entire facet of this phenotype may have been overlooked.

Further work on *KLF14* will include *in vivo* and *in vitro* functional investigation of *KLF14* trans-associated transcripts. GWA studies have definitively characterised *KLF14* as a master regulator of genes in adipose (Small et al., 2011). The work done in this thesis suggests it may control signalling to other metabolic tissues as well. Disruption of *KLF14* affects many aspects of metabolism including insulin, cholesterol, adipose cell proliferation, and growth. Further work will need to investigate the

molecular mechanisms of the genes *KLF14* is controlling to mediate these extensive effects. Generating mouse models of genes regulated by KLF14 would be one method of capturing a range of metabolic and adipose phenotypes which could be affected.

Now the CRISPR/Cas9 system offers a method of introducing targeted point mutations at any location in the genome, and this may be a better and more efficient approach in the future. While the ENU method offered a selection of mutations from which to choose, the CRISPR/Cas9 system would allow complete customization of mutations (Jinek et al., 2012; Hwang et al., 2013). Furthermore, the mutation could be introduced to a cellular system for preliminary evaluation before a mouse model was created. Creating a cell line or a mouse line with a mutation using CRISPR/Cas9 can be done quickly, so if one mutation failed preliminary characterisation in a cell line, a new mutation could be created and tested within a few weeks. The CRISPR/Cas9 technology has transformed genetic modelling.

The work presented in this thesis addresses the main challenge in understanding GWAS susceptibility data, and that is determining the biological relevance of the variation discovered. The functional analyses of *ADCY5* and *KLF14* undertaken here have contributed new knowledge of the pathophysiological mechanisms by which disruption of these genes can lead to T2D.

Appendices

References and Manuscripts

References

Abecasis, G. R., Altshuler, D., Auton, A., Brooks, L. D., Durbin, R. M., Gibbs, R. A., . . . Consortium, G. P. (2010). A map of human genome variation from population-scale sequencing. *Nature, 467*(7319), 1061-1073. doi:10.1038/nature09534

Adan, R. A., Tiesjema, B., Hillebrand, J. J., la Fleur, S. E., Kas, M. J., & de Krom, M. (2006). The MC4 receptor and control of appetite. *Br J Pharmacol, 149*(7), 815-827. doi:10.1038/sj.bjp.0706929

Allen, J. A., Yu, J. Z., Dave, R. H., Bhatnagar, A., Roth, B. L., & Rasenick, M. M. (2009). Caveolin-1 and lipid microdomains regulate Gs trafficking and attenuate Gs/adenylyl cyclase signaling. *Mol Pharmacol, 76*(5), 1082-1093. doi:10.1124/mol.109.060160

Altshuler, D., Hirschhorn, J. N., Klannemark, M., Lindgren, C. M., Vohl, M. C., Nemesh, J., . . . Lander, E. S. (2000). The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet, 26*(1), 76-80. doi:10.1038/79216

Altshuler, D. M., Gibbs, R. A., Peltonen, L., Dermitzakis, E., Schaffner, S. F., Yu, F., . . . Consortium, I. H. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature, 467*(7311), 52-58. doi:10.1038/nature09298

Ammälä, C., Ashcroft, F. M., & Rorsman, P. (1993). Calcium-independent potentiation of insulin release by cyclic AMP in single beta-cells. *Nature, 363*(6427), 356-358. doi:10.1038/363356a0

Ammälä, C., Eliasson, L., Bokvist, K., Berggren, P. O., Honkanen, R. E., Sjöholm, A., & Rorsman, P. (1994). Activation of protein kinases and inhibition of protein phosphatases play a central role in the regulation of exocytosis in mouse pancreatic beta cells. *Proc Natl Acad Sci U S A, 91*(10), 4343-4347.

Arnall, D. A., Marker, J. C., Conlee, R. K., & Winder, W. W. (1986). Effect of infusing epinephrine on liver and muscle glycogenolysis during exercise in rats. *Am J Physiol, 250*(6 Pt 1), E641-649.

Arragain, S., Handelman, S. K., Forouhar, F., Wei, F. Y., Tomizawa, K., Hunt, J. F., . . . Atta, M. (2010). Identification of eukaryotic and prokaryotic methylthiotransferase for biosynthesis of 2-methylthio-N6-threonylcarbamoyladenosine in tRNA. *J Biol Chem, 285*(37), 28425-28433. doi:10.1074/jbc.M110.106831

Ashcroft, F. M., Proks, P., Smith, P. A., Ammälä, C., Bokvist, K., & Rorsman, P. (1994). Stimulus-secretion coupling in pancreatic beta cells. *J Cell Biochem, 55 Suppl*, 54-65.

Barker, D. J., Hales, C. N., Fall, C. H., Osmond, C., Phipps, K., & Clark, P. M. (1993). Type 2 (non-insulin-dependent) diabetes mellitus, hypertension and hyperlipidaemia (syndrome X): relation to reduced fetal growth. *Diabetologia, 36*(1), 62-67.

Baroni, M. G., D'Andrea, M. P., Montali, A., Pannitteri, G., Barillà, F., Campagna, F., . . . Arca, M. (1999). A common mutation of the insulin receptor substrate-1 gene is a risk factor for coronary artery disease. *Arterioscler Thromb Vasc Biol, 19*(12), 2975-2980.

Bernstein, B. E., Birney, E., Dunham, I., Green, E. D., Gunter, C., Snyder, M., & Consortium, E. P. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature, 489*(7414), 57-74. doi:10.1038/nature11247

Björklund, A., & Grill, V. E. (2000). Relief from glucose-induced over-stimulation sensitizes the adenylate cyclase-cAMP system of rat pancreatic islets. *J Endocrinol, 166*(3), 537-544.

Björklund, A., Lansner, A., & Grill, V. E. (2000). Glucose-induced [Ca2+]i abnormalities in human pancreatic islets: important role of overstimulation. *Diabetes, 49*(11), 1840-1848.

Black, A. R., Black, J. D., & Azizkhan-Clifford, J. (2001). Sp1 and krüppel-like factor family of transcription factors in cell growth regulation and cancer. *J Cell Physiol, 188*(2), 143-160. doi:10.1002/jcp.1111

Bock, T., Pakkenberg, B., & Buschard, K. (2003). Increased islet volume but unchanged islet number in ob/ob mice. *Diabetes, 52*(7), 1716-1722.

Boivin, B., Lavoie, C., Vaniotis, G., Baragli, A., Villeneuve, L. R., Ethier, N., . . . Hébert, T. E. (2006). Functional beta-adrenergic receptor signalling on nuclear membranes in adult rat and mouse ventricular cardiomyocytes. *Cardiovasc Res, 71*(1), 69-78. doi:10.1016/j.cardiores.2006.03.015

Bonnefond, A., Clément, N., Fawcett, K., Yengo, L., Vaillant, E., Guillaume, J. L., . . . (MAGIC), M.-A. o. G. a. I.-R. T. C. (2012). Rare MTNR1B variants impairing melatonin receptor 1B function contribute to type 2 diabetes. *Nat Genet, 44*(3), 297-301. doi:10.1038/ng.1053

Bouatia-Naji, N., Bonnefond, A., Cavalcanti-Proença, C., Sparsø, T., Holmkvist, J., Marchand, M., . . . Froguel, P. (2009). A variant near MTNR1B is associated with increased fasting plasma glucose levels and type 2 diabetes risk. *Nat Genet, 41*(1), 89-94. doi:10.1038/ng.277

Braun, S., Bitton-Worms, K., & LeRoith, D. (2011). The link between the metabolic syndrome and cancer. *Int J Biol Sci, 7*(7), 1003-1015.

Brereton, M. F., Iberl, M., Shimomura, K., Zhang, Q., Adriaenssens, A. E., Proks, P., . . . Ashcroft, F. M. (2014a). Reversible changes in pancreatic islet structure and function produced by elevated blood glucose. *Nat Commun, 5*, 4639. doi:10.1038/ncomms5639

Brereton, M. F., Iberl, M., Shimomura, K., Zhang, Q., Adriaenssens, A. E., Proks, P., . . . Ashcroft, F. M. (2014b). Reversible changes in pancreatic islet structure and function produced by elevated blood glucose. *Nat Commun, 5*, 4639. doi:10.1038/ncomms5639

Butler, A. E., Janson, J., Bonner-Weir, S., Ritzel, R., Rizza, R. A., & Butler, P. C. (2003). Beta-cell deficit and increased beta-cell apoptosis in humans with type 2 diabetes. *Diabetes, 52*(1), 102-110.

Buyske, S., Wu, Y., Carty, C. L., Cheng, I., Assimes, T. L., Dumitrescu, L., . . . North, K. E. (2012). Evaluation of the metabochip genotyping array in African Americans and implications for fine mapping of GWAS-identified loci: the PAGE study. *PLoS One, 7*(4), e35651. doi:10.1371/journal.pone.0035651

Chambers, J. C., Elliott, P., Zabaneh, D., Zhang, W., Li, Y., Froguel, P., . . . Kooner, J. S. (2008). Common genetic variation near MC4R is associated with waist circumference and insulin resistance. *Nat Genet, 40*(6), 716-718. doi:10.1038/ng.156

Chan, C. B., MacPhail, R. M., Sheu, L., Wheeler, M. B., & Gaisano, H. Y. (1999). Beta-cell hypertrophy in fa/fa rats is associated with basal glucose hypersensitivity and reduced SNARE protein expression. *Diabetes, 48*(5), 997-1005.

Chan, C. B., Saleh, M. C., Purje, A., & MacPhail, R. M. (2002). Glucose-inducible hypertrophy and suppression of anion efflux in rat beta cells. *J Endocrinol, 173*(1), 45-52.

Charton, M., & Charton, B. I. (1982). The structural dependence of amino acid hydrophobicity parameters. *J Theor Biol, 99*(4), 629-644.

Chasman, D. I., Paré, G., Mora, S., Hopewell, J. C., Peloso, G., Clarke, R., . . . Ridker, P. M. (2009). Forty-three loci associated with plasma lipoprotein size, concentration, and cholesterol content in genome-wide analysis. *PLoS Genet, 5*(11), e1000730. doi:10.1371/journal.pgen.1000730

Chen, G., Bentley, A., Adeyemo, A., Shriner, D., Zhou, J., Doumatey, A., . . . Rotimi, C. (2012). Genome-wide association study identifies novel loci association with fasting insulin and insulin resistance in African Americans. *Hum Mol Genet, 21*(20), 4530-4536. doi:10.1093/hmg/dds282

Chen, X., Li, S., Yang, Y., Yang, X., Liu, Y., Hu, W., . . . Wang, X. (2012). Genome-wide association study validation identifies novel loci for atherosclerotic cardiovascular disease. *J Thromb Haemost, 10*(8), 1508-1514. doi:10.1111/j.1538-7836.2012.04815.x

Cheng, H., & Farquhar, M. G. (1976). Presence of adenylate cyclase activity in Golgi and other fractions from rat liver. II. Cytochemical localization within Golgi and ER membranes. *J Cell Biol, 70*(3), 671-684.

Cho, Y. S., Chen, C. H., Hu, C., Long, J., Ong, R. T., Sim, X., . . . Consortium, M. (2012). Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. *Nat Genet, 44*(1), 67-72. doi:10.1038/ng.1019

Church, C., Lee, S., Bagg, E. A., McTaggart, J. S., Deacon, R., Gerken, T., . . . Cox, R. D. (2009). A mouse model for the metabolic effects of the human fat mass and obesity associated FTO gene. *PLoS Genet, 5*(8), e1000599. doi:10.1371/journal.pgen.1000599

Church, C., Moir, L., McMurray, F., Girard, C., Banks, G. T., Teboul, L., . . . Cox, R. D. (2010). Overexpression of Fto leads to increased food intake and results in obesity. *Nat Genet, 42*(12), 1086-1092. doi:10.1038/ng.713

Chèvre, J. C., Hani, E. H., Boutin, P., Vaxillaire, M., Blanché, H., Vionnet, N., . . . Froguel, P. (1998). Mutation screening in 18 Caucasian families suggest the existence of other MODY genes. *Diabetologia, 41*(9), 1017-1023. doi:10.1007/s001250051025

Coghill, E. L., Hugill, A., Parkinson, N., Davison, C., Glenister, P., Clements, S., . . . Brown, S. D. (2002). A gene-driven approach to the identification of ENU mutants in the mouse. *Nat Genet, 30*(3), 255-256. doi:10.1038/ng847

Collins, L. L., Lee, Y. F., Heinlein, C. A., Liu, N. C., Chen, Y. T., Shyr, C. R., . . . Chang, C. (2004). Growth retardation and abnormal maternal behavior in mice lacking testicular orphan nuclear receptor 4. *Proc Natl Acad Sci U S A, 101*(42), 15058-15063. doi:10.1073/pnas.0405700101

Connor, A. E., Baumgartner, R. N., Baumgartner, K. B., Kerber, R. A., Pinkston, C., John, E. M., . . . Slattery, M. L. (2012). Associations between TCF7L2 polymorphisms and risk of breast cancer among Hispanic and non-Hispanic white women: the Breast Cancer Health Disparities Study. *Breast Cancer Res Treat, 136*(2), 593-602. doi:10.1007/s10549-012-2299-7

Consortium, D. G. R. A. M.-a. D. (2014). Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet*. doi:10.1038/ng.2897

Consortium, I. H. (2005). A haplotype map of the human genome. *Nature, 437*(7063), 1299-1320. doi:10.1038/nature04226

Cooper, D. M. (2003). Regulation and organization of adenylyl cyclases and cAMP. *Biochem J, 375*(Pt 3), 517-529. doi:10.1042/BJ20031061

Cooper, D. M., & Crossthwaite, A. J. (2006a). Higher-order organization and regulation of adenylyl cyclases. *Trends Pharmacol Sci, 27*(8), 426-431. doi:10.1016/j.tips.2006.06.002

Cooper, D. M., & Crossthwaite, A. J. (2006b). Higher-order organization and regulation of adenylyl cyclases. *Trends Pharmacol Sci, 27*(8), 426-431. doi:10.1016/j.tips.2006.06.002

Cooper, D. M., Mons, N., & Fagan, K. (1994). Ca(2+)-sensitive adenylyl cyclases. *Cell Signal, 6*(8), 823-840.

Cooper, D. M., & Tabbasum, V. G. (2014). Adenylate cyclase-centred microdomains. *Biochem J, 462*(2), 199-213. doi:10.1042/BJ20140560

Cristancho, A. G., & Lazar, M. A. (2011). Forming functional fat: a growing understanding of adipocyte differentiation. *Nat Rev Mol Cell Biol, 12*(11), 722-734. doi:10.1038/nrm3198

Crossthwaite, A. J., Seebacher, T., Masada, N., Ciruela, A., Dufraux, K., Schultz, J. E., & Cooper, D. M. (2005a). The cytosolic domains of Ca2+-sensitive adenylyl cyclases dictate their targeting to plasma membrane lipid rafts. *J Biol Chem, 280*(8), 6380-6391. doi:10.1074/jbc.M411987200

Crossthwaite, A. J., Seebacher, T., Masada, N., Ciruela, A., Dufraux, K., Schultz, J. E., & Cooper, D. M. (2005b). The cytosolic domains of Ca2+-sensitive adenylyl cyclases dictate their targeting to plasma membrane lipid rafts. *J Biol Chem, 280*(8), 6380-6391. doi:10.1074/jbc.M411987200

Dayhoff, M. O., Schwartz, R. D., & Or.

Dayhoff, M. O., Schwartz, R. M., & Orcutt, B. C. (1978). A Model of Evolutionary Change in Proteins. In M. O. Dayhoff (Ed.), *Atlas of Protein Sequence and Structure.* (Vol. 5, pp. 345-352). Washington D.C.: National Biochemical Research Foundation.

De Marinis, Y. Z., Salehi, A., Ward, C. E., Zhang, Q., Abdulkader, F., Bengtsson, M., . . . Rorsman, P. (2010a). GLP-1 inhibits and adrenaline stimulates glucagon release by differential modulation of N- and L-type Ca2+ channel-dependent exocytosis. *Cell Metab, 11*(6), 543-553. doi:10.1016/j.cmet.2010.04.007

De Marinis, Y. Z., Salehi, A., Ward, C. E., Zhang, Q., Abdulkader, F., Bengtsson, M., . . . Rorsman, P.

(2010b). GLP-1 inhibits and adrenaline stimulates glucagon release by differential modulation of N- and L-type Ca2+ channel-dependent exocytosis. *Cell Metab, 11*(6), 543-553. doi:10.1016/j.cmet.2010.04.007

Debuyser, A., Drews, G., & Henquin, J. C. (1991). Adrenaline inhibition of insulin release: role of cyclic AMP. *Mol Cell Endocrinol, 78*(3), 179-186.

Deeb, S. S., Fajas, L., Nemoto, M., Pihlajamäki, J., Mykkänen, L., Kuusisto, J., . . . Auwerx, J. (1998). A Pro12Ala substitution in PPARgamma2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity. *Nat Genet, 20*(3), 284-287. doi:10.1038/3099

Dimas, A. S., Lagou, V., Barker, A., Knowles, J. W., Mägi, R., Hivert, M. F., . . . Investigators, M. (2014). Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes, 63*(6), 2158-2171. doi:10.2337/db13-0949

Dimas, A. S., Lagou, V., Barker, A., Knowles, J. W., Mägi, R., Hivert, M. F., . . . investigators, o. b. o. t. M. (2013). Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes*. doi:10.2337/db13-0949

DiPilato, L. M., Cheng, X., & Zhang, J. (2004). Fluorescent indicators of cAMP and Epac activation reveal differential dynamics of cAMP signaling within discrete subcellular compartments. *Proc Natl Acad Sci U S A, 101*(47), 16513-16518. doi:10.1073/pnas.0405973101

Dupuis, J., Langenberg, C., Prokopenko, I., Saxena, R., Soranzo, N., Jackson, A. U., . . . investigators, M. (2010). New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet, 42*(2), 105-116. doi:10.1038/ng.520

Dyachok, O., & Gylfe, E. (2004). Ca(2+)-induced Ca(2+) release via inositol 1,4,5-trisphosphate receptors is amplified by protein kinase A and triggers exocytosis in pancreatic beta-cells. *J Biol Chem, 279*(44), 45455-45461. doi:10.1074/jbc.M407673200

Dyachok, O., Idevall-Hagren, O., Sågetorp, J., Tian, G., Wuttke, A., Arrieumerlou, C., . . . Tengholm, A. (2008). Glucose-induced cyclic AMP oscillations regulate pulsatile insulin secretion. *Cell Metab, 8*(1), 26-37. doi:10.1016/j.cmet.2008.06.003

Dyachok, O., Isakov, Y., Sågetorp, J., & Tengholm, A. (2006). Oscillations of cyclic AMP in hormone-stimulated insulin-secreting beta-cells. *Nature, 439*(7074), 349-352. doi:10.1038/nature04410

Eberle, M. A., Ng, P. C., Kuhn, K., Zhou, L., Peiffer, D. A., Galver, L., . . . Murray, S. S. (2007). Power to detect risk alleles using genome-wide tag SNP panels. *PLoS Genet, 3*(10), 1827-1837. doi:10.1371/journal.pgen.0030170

Eliasson, L., Ma, X., Renström, E., Barg, S., Berggren, P. O., Galvanovskis, J., . . . Rorsman, P. (2003a). SUR1 regulates PKA-independent cAMP-induced granule priming in mouse pancreatic B-cells. *J Gen Physiol, 121*(3), 181-197.

Eliasson, L., Ma, X., Renström, E., Barg, S., Berggren, P. O., Galvanovskis, J., . . . Rorsman, P. (2003b). SUR1 regulates PKA-independent cAMP-induced granule priming in mouse pancreatic B-cells. *J Gen Physiol, 121*(3), 181-197.

Eliasson, L., Ma, X., Renström, E., Barg, S., Berggren, P. O., Galvanovskis, J., . . . Rorsman, P. (2003c). SUR1 regulates PKA-independent cAMP-induced granule priming in mouse pancreatic B-cells. *J Gen Physiol, 121*(3), 181-197.

Fadista, J., Vikman, P., Laakso, E. O., Mollet, I. G., Esguerra, J. L., Taneera, J., . . . Groop, L. (2014). Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. *Proc Natl Acad Sci U S A, 111*(38), 13924-13929. doi:10.1073/pnas.1402665111

Fagan, K. A., Mahey, R., & Cooper, D. M. (1996). Functional co-localization of transfected Ca(2+)-stimulable adenylyl cyclases with capacitative Ca2+ entry sites. *J Biol Chem, 271*(21), 12438-12444.

Fagan, K. A., Smith, K. E., & Cooper, D. M. (2000). Regulation of the Ca2+-inhibitable adenylyl cyclase type VI by capacitative Ca2+ entry requires localization in cholesterol-rich domains. *J Biol*

Chem, 275(34), 26530-26537. doi:10.1074/jbc.M001369200
Florez, J. C., Hirschhorn, J., & Altshuler, D. (2003). The inherited basis of diabetes mellitus: implications for the genetic analysis of complex traits. *Annu Rev Genomics Hum Genet, 4*, 257-291. doi:10.1146/annurev.genom.4.070802.110436
Forrester, E., Chytil, A., Bierie, B., Aakre, M., Gorska, A. E., Sharif-Afshar, A. R., . . . Moses, H. L. (2005). Effect of conditional knockout of the type II TGF-beta receptor gene in mammary epithelia on mammary gland development and polyomavirus middle T antigen induced tumor formation and metastasis. *Cancer Res, 65*(6), 2296-2302. doi:10.1158/0008-5472.CAN-04-3272
Frayling, T. M., Colhoun, H., & Florez, J. C. (2008). A genetic link between type 2 diabetes and prostate cancer. *Diabetologia, 51*(10), 1757-1760. doi:10.1007/s00125-008-1114-9
Frayling, T. M., Timpson, N. J., Weedon, M. N., Zeggini, E., Freathy, R. M., Lindgren, C. M., . . . McCarthy, M. I. (2007). A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science, 316*(5826), 889-894. doi:10.1126/science.1141634
Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., . . . Consortium, I. H. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature, 449*(7164), 851-861. doi:10.1038/nature06258
Freathy, R. M., Bennett, A. J., Ring, S. M., Shields, B., Groves, C. J., Timpson, N. J., . . . Hattersley, A. T. (2009). Type 2 diabetes risk alleles are associated with reduced size at birth. *Diabetes, 58*(6), 1428-1433. doi:10.2337/db08-1739
Fredriksson, R., Hägglund, M., Olszewski, P. K., Stephansson, O., Jacobsson, J. A., Olszewska, A. M., . . . Schiöth, H. B. (2008). The obesity gene, FTO, is of ancient origin, up-regulated during food deprivation and expressed in neurons of feeding-related nuclei of the brain. *Endocrinology, 149*(5), 2062-2071. doi:10.1210/en.2007-1457
Furuta, H., Iwasaki, N., Oda, N., Hinokio, Y., Horikawa, Y., Yamagata, K., . . . Bell, G. I. (1997). Organization and partial sequence of the hepatocyte nuclear factor-4 alpha/MODY1 gene and identification of a missense mutation, R127W, in a Japanese family with MODY. *Diabetes, 46*(10), 1652-1657.
Galassetti, P., & Davis, S. N. (2000). Effects of insulin per se on neuroendocrine and metabolic counter-regulatory responses to hypoglycaemia. *Clin Sci (Lond), 99*(5), 351-362.
Gaulton, K. J., Nammo, T., Pasquali, L., Simon, J. M., Giresi, P. G., Fogarty, M. P., . . . Ferrer, J. (2010). A map of open chromatin in human pancreatic islets. *Nat Genet, 42*(3), 255-259. doi:10.1038/ng.530
Gicquel, C., Rossignol, S., Cabrol, S., Houang, M., Steunou, V., Barbu, V., . . . Le Bouc, Y. (2005). Epimutation of the telomeric imprinting center region on chromosome 11p15 in Silver-Russell syndrome. *Nat Genet, 37*(9), 1003-1007. doi:10.1038/ng1629
Gloyn, A. L., Weedon, M. N., Owen, K. R., Turner, M. J., Knight, B. A., Hitman, G., . . . Frayling, T. M. (2003). Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. *Diabetes, 52*(2), 568-572.
Gragnoli, C., Lindner, T., Cockburn, B. N., Kaisaki, P. J., Gragnoli, F., Marozzi, G., & Bell, G. I. (1997). Maturity-onset diabetes of the young due to a mutation in the hepatocyte nuclear factor-4 alpha binding site in the promoter of the hepatocyte nuclear factor-1 alpha gene. *Diabetes, 46*(10), 1648-1651.
Gómez Dumm, C. L., Cónsole, G. M., Luna, G. C., Dardenne, M., & Goya, R. G. (1995). Quantitative immunohistochemical changes in the endocrine pancreas of nonobese diabetic (NOD) mice. *Pancreas, 11*(4), 396-401.
Haiman, C. A., Fesinmeyer, M. D., Spencer, K. L., Buzková, P., Voruganti, V. S., Wan, P., . . . Pankow, J. S. (2012). Consistent directions of effect for established type 2 diabetes risk variants across populations: the population architecture using Genomics and Epidemiology (PAGE)

Consortium. *Diabetes, 61*(6), 1642-1647. doi:10.2337/db11-1296
Hamacher, C., Bröcker, M., Adams, E. F., Lei, T., Fahlbusch, R., Buchfelder, M., & Derwahl, M. (1998). Overexpression of stimulatory G protein alpha-subunit is a hallmark of most human somatotrophic pituitary tumours and is associated with resistance to GH-releasing hormone. *Pituitary, 1*(1), 13-23.
Hamed, M., Ismael, S., Paulsen, M., & Helms, V. (2012). Cellular functions of genetically imprinted genes in human and mouse as annotated in the gene ontology. *PLoS One, 7*(11), e50285. doi:10.1371/journal.pone.0050285
Hare, K. J. (2010). Role of GLP-1 induced glucagon suppression in type 2 diabetes mellitus. *Dan Med Bull, 57*(9), B4181.
Hare, K. J., Vilsbøll, T., Asmar, M., Deacon, C. F., Knop, F. K., & Holst, J. J. (2010). The glucagonostatic and insulinotropic effects of glucagon-like peptide 1 contribute equally to its glucose-lowering action. *Diabetes, 59*(7), 1765-1770. doi:10.2337/db09-1414
Heid, I. M., Jackson, A. U., Randall, J. C., Winkler, T. W., Qi, L., Steinthorsdottir, V., . . . MAGIC. (2010). Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat Genet, 42*(11), 949-960. doi:10.1038/ng.685
Hellman, B., Idahl, L. A., Lernmark, A., & Täljedal, I. B. (1974). The pancreatic beta-cell recognition of insulin secretagogues: does cyclic AMP mediate the effect of glucose? *Proc Natl Acad Sci U S A, 71*(9), 3405-3409.
Hennige, A. M., Ozcan, U., Okada, T., Jhala, U. S., Schubert, M., White, M. F., & Kulkarni, R. N. (2005). Alterations in growth and apoptosis of insulin receptor substrate-1-deficient beta-cells. *Am J Physiol Endocrinol Metab, 289*(2), E337-346. doi:10.1152/ajpendo.00032.2004
Henquin, J. C., Nenquin, M., Stiernet, P., & Ahren, B. (2006). In vivo and in vitro glucose-induced biphasic insulin secretion in the mouse: pattern and role of cytoplasmic Ca2+ and amplification signals in beta-cells. *Diabetes, 55*(2), 441-451.
Hitotsumachi, S., Carpenter, D. A., & Russell, W. L. (1985). Dose-repetition increases the mutagenic effectiveness of N-ethyl-N-nitrosourea in mouse spermatogonia. *Proc Natl Acad Sci U S A, 82*(19), 6619-6621.
Ho, D., Yan, L., Iwatsubo, K., Vatner, D. E., & Vatner, S. F. (2010a). Modulation of beta-adrenergic receptor signaling in heart failure and longevity: targeting adenylyl cyclase type 5. *Heart Fail Rev, 15*(5), 495-512. doi:10.1007/s10741-010-9183-5
Ho, D., Yan, L., Iwatsubo, K., Vatner, D. E., & Vatner, S. F. (2010b). Modulation of beta-adrenergic receptor signaling in heart failure and longevity: targeting adenylyl cyclase type 5. *Heart Fail Rev, 15*(5), 495-512. doi:10.1007/s10741-010-9183-5
Hodson, D. J., Mitchell, R. K., Marselli, L., Pullen, T. J., Brias, S. G., Semplici, F., . . . Rutter, G. A. (2014). ADCY5 couples glucose to insulin secretion in human islets. *Diabetes*. doi:10.2337/db13-1607
Hodson, D. J., Mitchell, R. K., Marselli, L., Pullen, T. J., Gimeno Brias, S., Semplici, F., . . . Rutter, G. A. (2014). ADCY5 Couples Glucose to Insulin Secretion in Human Islets. *Diabetes, 63*(9), 3009-3021. doi:10.2337/db13-1607
Holt, L. J., Lyons, R. J., Ryan, A. S., Beale, S. M., Ward, A., Cooney, G. J., & Daly, R. J. (2009). Dual ablation of Grb10 and Grb14 in mice reveals their combined role in regulation of insulin signaling and glucose homeostasis. *Mol Endocrinol, 23*(9), 1406-1414. doi:10.1210/me.2008-0386
Holz, G. G. (2004). Epac: A new cAMP-binding protein in support of glucagon-like peptide-1 receptor-mediated signal transduction in the pancreatic beta-cell. *Diabetes, 53*(1), 5-13.
Holz, G. G., Kang, G., Harbeck, M., Roe, M. W., & Chepurny, O. G. (2006). Cell physiology of cAMP sensor Epac. *J Physiol, 577*(Pt 1), 5-15. doi:10.1113/jphysiol.2006.119644
Holz, G. H., Kuhtreiber, W. M., & Habener, J. F. (1992). Induction of glucose competence in pancreatic beta cells by glucagon-like peptide-1(7-37). *Trans Assoc Am Physicians, 105*, 260-

267.
Horikoshi, M., Yaghootkar, H., Mook-Kanamori, D. O., Sovio, U., Taal, H. R., Hennig, B. J., . . . Consortium, E. G. G. E. (2013). New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism. *Nat Genet, 45*(1), 76-82. doi:10.1038/ng.2477
Hu, C. L., Chandra, R., Ge, H., Pain, J., Yan, L., Babu, G., . . . Vatner, D. E. (2009). Adenylyl cyclase type 5 protein expression during cardiac development and stress. *Am J Physiol Heart Circ Physiol, 297*(5), H1776-1782. doi:10.1152/ajpheart.00050.2009
Huyghe, J. R., Jackson, A. U., Fogarty, M. P., Buchkovich, M. L., Stančáková, A., Stringham, H. M., . . . Mohlke, K. L. (2013). Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. *Nat Genet, 45*(2), 197-201. doi:10.1038/ng.2507
Jiang, W., Bikard, D., Cox, D., Zhang, F., & Marraffini, L. A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol, 31*(3), 233-239. doi:10.1038/nbt.2508
Johansson, A., Marroni, F., Hayward, C., Franklin, C. S., Kirichenko, A. V., Jonasson, I., . . . Consortium, E. (2009). Common variants in the JAZF1 gene associated with height identified by linkage and genome-wide association analysis. *Hum Mol Genet, 18*(2), 373-380. doi:10.1093/hmg/ddn350
Joshi, S., Liu, M., & Turner, N. (2015). Diabetes and Its Link with Cancer: Providing the Fuel and Spark to Launch an Aggressive Growth Regime. *Biomed Res Int, 2015*, 390863. doi:10.1155/2015/390863
Joung, J. K., & Sander, J. D. (2013). TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol, 14*(1), 49-55. doi:10.1038/nrm3486
Justice, M. J., Noveroske, J. K., Weber, J. S., Zheng, B., & Bradley, A. (1999). Mouse ENU mutagenesis. *Hum Mol Genet, 8*(10), 1955-1963.
Kadowaki, T., Tamemoto, H., Tobe, K., Terauchi, Y., Ueki, K., Kaburagi, Y., . . . Yazaki, Y. (1996). Insulin resistance and growth retardation in mice lacking insulin receptor substrate-1 and identification of insulin receptor substrate-2. *Diabet Med, 13*(9 Suppl 6), S103-108.
Kang, G., Chepurny, O. G., & Holz, G. G. (2001). cAMP-regulated guanine nucleotide exchange factor II (Epac2) mediates Ca2+-induced Ca2+ release in INS-1 pancreatic beta-cells. *J Physiol, 536*(Pt 2), 375-385.
Kang, G., Chepurny, O. G., Malester, B., Rindler, M. J., Rehmann, H., Bos, J. L., . . . Holz, G. G. (2006). cAMP sensor Epac as a determinant of ATP-sensitive potassium channel activity in human pancreatic beta cells and rat INS-1 cells. *J Physiol, 573*(Pt 3), 595-609. doi:10.1113/jphysiol.2006.107391
Kang, G., Leech, C. A., Chepurny, O. G., Coetzee, W. A., & Holz, G. G. (2008). Role of the cAMP sensor Epac as a determinant of KATP channel ATP sensitivity in human pancreatic beta-cells and rat INS-1 cells. *J Physiol, 586*(5), 1307-1319. doi:10.1113/jphysiol.2007.143818
Katsanis, N., Yaspo, M. L., & Fisher, E. M. (1997). Identification and mapping of a novel human gene, HRMT1L1, homologous to the rat protein arginine N-methyltransferase 1 (PRMT1) gene. *Mamm Genome, 8*(7), 526-529.
Kaufman, R. J. (2011). Beta-cell failure, stress, and type 2 diabetes. *N Engl J Med, 365*(20), 1931-1933. doi:10.1056/NEJMcibr1109442
Keays, D. A., Clark, T. G., & Flint, J. (2006). Estimating the number of coding mutations in genotypic- and phenotypic-driven N-ethyl-N-nitrosourea (ENU) screens. *Mamm Genome, 17*(3), 230-238. doi:10.1007/s00335-005-0101-4
Keinan, A., & Clark, A. G. (2012). Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science, 336*(6082), 740-743. doi:10.1126/science.1217283
Kellis, M., Wold, B., Snyder, M. P., Bernstein, B. E., Kundaje, A., Marinov, G. K., . . . Hardison, R. C. (2014). Defining functional DNA elements in the human genome. *Proc Natl Acad Sci U S A, 111*(17), 6131-6138. doi:10.1073/pnas.1318948111
Kheirbek, M. A., Britt, J. P., Beeler, J. A., Ishikawa, Y., McGehee, D. S., & Zhuang, X. (2009). Adenylyl

cyclase type 5 contributes to corticostriatal plasticity and striatum-dependent learning. *J Neurosci, 29*(39), 12115-12124. doi:10.1523/JNEUROSCI.3343-09.2009

Kilpeläinen, T. O., Zillikens, M. C., Stančákova, A., Finucane, F. M., Ried, J. S., Langenberg, C., . . . Loos, R. J. (2011). Genetic variation near IRS1 associates with reduced adiposity and an impaired metabolic profile. *Nat Genet, 43*(8), 753-760. doi:10.1038/ng.866

Kim, K. S., Kim, H., Baek, I. S., Lee, K. W., & Han, P. L. (2011a). Mice lacking adenylyl cyclase type 5 (AC5) show increased ethanol consumption and reduced ethanol sensitivity. *Psychopharmacology (Berl), 215*(2), 391-398. doi:10.1007/s00213-010-2143-x

Kim, K. S., Kim, H., Baek, I. S., Lee, K. W., & Han, P. L. (2011b). Mice lacking adenylyl cyclase type 5 (AC5) show increased ethanol consumption and reduced ethanol sensitivity. *Psychopharmacology (Berl), 215*(2), 391-398. doi:10.1007/s00213-010-2143-x

Kim, K. S., Lee, K. W., Baek, I. S., Lim, C. M., Krishnan, V., Lee, J. K., . . . Han, P. L. (2008). Adenylyl cyclase-5 activity in the nucleus accumbens regulates anxiety-related behavior. *J Neurochem, 107*(1), 105-115. doi:10.1111/j.1471-4159.2008.05592.x

Kim, Y. J., Go, M. J., Hu, C., Hong, C. B., Kim, Y. K., Lee, J. Y., . . . consortium, M. (2011). Large-scale genome-wide association studies in East Asians identify new genetic loci influencing metabolic traits. *Nat Genet, 43*(10), 990-995. doi:10.1038/ng.939

Knoll, N., Jarick, I., Volckmar, A. L., Klingenspor, M., Illig, T., Grallert, H., . . . Hinney, A. (2013). Gene set of nuclear-encoded mitochondrial regulators is enriched for common inherited variation in obesity. *PLoS One, 8*(2), e55884. doi:10.1371/journal.pone.0055884

Kong, A., Steinthorsdottir, V., Masson, G., Thorleifsson, G., Sulem, P., Besenbacher, S., . . . Consortium, D. (2009). Parental origin of sequence variants associated with complex diseases. *Nature, 462*(7275), 868-874. doi:10.1038/nature08625

.

Kubota, N., Tobe, K., Terauchi, Y., Eto, K., Yamauchi, T., Suzuki, R., . . . Kadowaki, T. (2000). Disruption of insulin receptor substrate 2 causes type 2 diabetes because of liver insulin resistance and lack of compensatory beta-cell hyperplasia. *Diabetes, 49*(11), 1880-1889.

Kulkarni, R. N. (2002). Receptors for insulin and insulin-like growth factor-1 and insulin receptor substrate-1 mediate pathways that regulate islet function. *Biochem Soc Trans, 30*(2), 317-322. doi:10.1042/

Kulkarni, R. N., Mizrachi, E. B., Ocana, A. G., & Stewart, A. F. (2012). Human β-cell proliferation and intracellular signaling: driving in the dark without a road map. *Diabetes, 61*(9), 2205-2213. doi:10.2337/db12-0018

Kulkarni, R. N., Winnay, J. N., Daniels, M., Brüning, J. C., Flier, S. N., Hanahan, D., & Kahn, C. R. (1999). Altered function of insulin receptor substrate-1-deficient mouse islets and cultured beta-cell lines. *J Clin Invest, 104*(12), R69-75. doi:10.1172/JCI8339

Lee, M. P., DeBaun, M. R., Mitsuya, K., Galonek, H. L., Brandenburg, S., Oshimura, M., & Feinberg, A. P. (1999). Loss of imprinting of a paternally expressed transcript, with antisense orientation to KVLQT1, occurs frequently in Beckwith-Wiedemann syndrome and is independent of insulin-like growth factor II imprinting. *Proc Natl Acad Sci U S A, 96*(9), 5203-5208.

Lemaire, K., Ravier, M. A., Schraenen, A., Creemers, J. W., Van de Plas, R., Granvik, M., . . . Schuit, F. C. (2009). Insulin crystallization depends on zinc transporter ZnT8 expression, but is not required for normal glucose homeostasis in mice. *Proc Natl Acad Sci U S A, 106*(35), 14872-14877. doi:10.1073/pnas.0906587106

Levey, A. I. (1993). Immunological localization of m1-m5 muscarinic acetylcholine receptors in peripheral tissues and brain. *Life Sci, 52*(5-6), 441-448.

Levey, A. I., Hersch, S. M., Rye, D. B., Sunahara, R. K., Niznik, H. B., Kitt, C. A., . . . Ciliax, B. J. (1993). Localization of D1 and D2 dopamine receptors in brain with subtype-specific antibodies. *Proc Natl Acad Sci U S A, 90*(19), 8861-8865.

Li, M., Chiu, J. F., Mossman, B. T., & Fukagawa, N. K. (2006). Down-regulation of manganese-superoxide dismutase through phosphorylation of FOXO3a by Akt in explanted vascular

smooth muscle cells from old rats. *J Biol Chem, 281*(52), 40429-40439. doi:10.1074/jbc.M606596200

Lin, Z., Kumar, A., SenBanerjee, S., Staniszewski, K., Parmar, K., Vaughan, D. E., . . . Jain, M. K. (2005). Kruppel-like factor 2 (KLF2) regulates endothelial thrombotic function. *Circ Res, 96*(5), e48-57. doi:10.1161/01.RES.0000159707.05637.a1

Lindner, T., Gragnoli, C., Furuta, H., Cockburn, B. N., Petzold, C., Rietzsch, H., . . . Bell, G. I. (1997). Hepatic function in a family with a nonsense mutation (R154X) in the hepatocyte nuclear factor-4alpha/MODY1 gene. *J Clin Invest, 100*(6), 1400-1405. doi:10.1172/JCI119660

Liu, T., Johnson, J. A., Casella, G., & Wu, R. (2004). Sequencing complex diseases With HapMap. *Genetics, 168*(1), 503-511. doi:10.1534/genetics.104.029603

Loos, R. J., & Bouchard, C. (2008). FTO: the first gene contributing to common forms of human obesity. *Obes Rev, 9*(3), 246-250. doi:10.1111/j.1467-789X.2008.00481.x

Lupski, J. R., Belmont, J. W., Boerwinkle, E., & Gibbs, R. A. (2011). Clan genomics and the complex architecture of human disease. *Cell, 147*(1), 32-43. doi:10.1016/j.cell.2011.09.008

Lyssand, J. S., & Bajjalieh, S. M. (2007). The heterotrimeric [corrected] G protein subunit G alpha i is present on mitochondria. *FEBS Lett, 581*(30), 5765-5768. doi:10.1016/j.febslet.2007.11.044

Lyssenko, V., Nagorny, C. L., Erdos, M. R., Wierup, N., Jonsson, A., Spégel, P., . . . Groop, L. (2009). Common variant in MTNR1B associated with increased risk of type 2 diabetes and impaired early insulin secretion. *Nat Genet, 41*(1), 82-88. doi:10.1038/ng.288

Mahajan, A., Go, M. J., Zhang, W., Below, J. E., Gaulton, K. J., Ferreira, T., . . . Consortium, T. D. G. E. b. N.-g. s. i. m.-E. S. T. D.-G. (2014). Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet, 46*(3), 234-244. doi:10.1038/ng.2897

Manolio, T. A., Brooks, L. D., & Collins, F. S. (2008). A HapMap harvest of insights into the genetics of common disease. *J Clin Invest, 118*(5), 1590-1605. doi:10.1172/JCI34772

Matsuzaki, H., Dong, S., Loi, H., Di, X., Liu, G., Hubbell, E., . . . Mei, R. (2004). Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nat Methods, 1*(2), 109-111. doi:10.1038/nmeth718

McClellan, J., & King, M. C. (2010). Genetic heterogeneity in human disease. *Cell, 141*(2), 210-217. doi:10.1016/j.cell.2010.03.032

McMurray, F., & Cox, R. D. (2011). Mouse models and type 2 diabetes: translational opportunities. *Mamm Genome, 22*(7-8), 390-400. doi:10.1007/s00335-011-9345-3

McVean, G., Spencer, C. C., & Chaix, R. (2005). Perspectives on human genetic variation from the HapMap Project. *PLoS Genet, 1*(4), e54. doi:10.1371/journal.pgen.0010054

Miki, H., Yamauchi, T., Suzuki, R., Komeda, K., Tsuchida, A., Kubota, N., . . . Kadowaki, T. (2001). Essential role of insulin receptor substrate 1 (IRS-1) and IRS-2 in adipocyte differentiation. *Mol Cell Biol, 21*(7), 2521-2532. doi:10.1128/MCB.21.7.2521-2532.2001

Min, J. L., Nicholson, G., Halgrimsdottir, I., Almstrup, K., Petri, A., Barrett, A., . . . Consortium, M. (2012). Coexpression network analysis in abdominal and gluteal adipose tissue reveals regulatory genetic loci for metabolic syndrome and related phenotypes. *PLoS Genet, 8*(2), e1002505. doi:10.1371/journal.pgen.1002505

Moore, T., & Haig, D. (1991). Genomic imprinting in mammalian development: a parental tug-of-war. *Trends Genet, 7*(2), 45-49. doi:10.1016/0168-9525(91)90230-N

Morris, A. P., Voight, B. F., Teslovich, T. M., Ferreira, T., Segrè, A. V., Steinthorsdottir, V., . . . Consortium, D. G. R. A. M.-a. D. (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet, 44*(9), 981-990. doi:10.1038/ng.2383

Nakajima, T., Fujino, S., Nakanishi, G., Kim, Y. S., & Jetten, A. M. (2004). TIP27: a novel repressor of the nuclear orphan receptor TAK1/TR4. *Nucleic Acids Res, 32*(14), 4194-4204. doi:10.1093/nar/gkh741

Nejentsev, S., Walker, N., Riches, D., Egholm, M., & Todd, J. A. (2009). Rare variants of IFIH1, a gene

implicated in antiviral responses, protect against type 1 diabetes. *Science, 324*(5925), 387-389. doi:10.1126/science.1167728

Nelson, M. R., Wegmann, D., Ehm, M. G., Kessner, D., St Jean, P., Verzilli, C., . . . Mooser, V. (2012). An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science, 337*(6090), 100-104. doi:10.1126/science.1217876

Ng, M. C., Shriner, D., Chen, B. H., Li, J., Chen, W. M., Guo, X., . . . Consortium, M.-a. o. t. D. i. A. A. (2014). Meta-analysis of genome-wide association studies in African Americans provides insights into the genetic architecture of type 2 diabetes. *PLoS Genet, 10*(8), e1004517. doi:10.1371/journal.pgen.1004517

Nicolson, T. J., Bellomo, E. A., Wijesekara, N., Loder, M. K., Baldwin, J. M., Gyulkhandanyan, A. V., . . . Rutter, G. A. (2009). Insulin storage and glucose homeostasis in mice null for the granule zinc transporter ZnT8 and studies of the type 2 diabetes-associated variants. *Diabetes, 58*(9), 2070-2083. doi:10.2337/db09-0551

Nolan, P. M., Peters, J., Strivens, M., Rogers, D., Hagan, J., Spurr, N., . . . Hunter, J. (2000). A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. *Nat Genet, 25*(4), 440-443. doi:10.1038/78140

Nuez, B., Michalovich, D., Bygrave, A., Ploemacher, R., & Grosveld, F. (1995). Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature, 375*(6529), 316-318. doi:10.1038/375316a0

Ohara-Imaizumi, M., Yoshida, M., Aoyagi, K., Saito, T., Okamura, T., Takenaka, H., . . . Nagamatsu, S. (2010). Deletion of CDKAL1 affects mitochondrial ATP generation and first-phase insulin exocytosis. *PLoS One, 5*(12), e15553. doi:10.1371/journal.pone.0015553

Ohshige, T., Iwata, M., Omori, S., Tanaka, Y., Hirose, H., Kaku, K., . . . Maeda, S. (2011). Association of new loci identified in European genome-wide association studies with susceptibility to type 2 diabetes in the Japanese. *PLoS One, 6*(10), e26911. doi:10.1371/journal.pone.0026911

Oishi, Y., Manabe, I., Tobe, K., Ohsugi, M., Kubota, T., Fujiu, K., . . . Nagai, R. (2008). SUMOylation of Krüppel-like transcription factor 5 acts as a molecular switch in transcriptional programs of lipid metabolism involving PPAR-delta. *Nat Med, 14*(6), 656-666. doi:10.1038/nm1756

Okumura, S., Takagi, G., Kawabe, J., Yang, G., Lee, M. C., Hong, C., . . . Ishikawa, Y. (2003). Disruption of type 5 adenylyl cyclase gene preserves cardiac function against pressure overload. *Proc Natl Acad Sci U S A, 100*(17), 9986-9990. doi:10.1073/pnas.1733772100

Olofsson, L. E., Orho-Melander, M., William-Olsson, L., Sjöholm, K., Sjöström, L., Groop, L., . . . Olsson, B. (2008). CCAAT/enhancer binding protein alpha (C/EBPalpha) in adipose tissue regulates genes in lipid and glucose metabolism and a genetic variation in C/EBPalpha is associated with serum levels of triglycerides. *J Clin Endocrinol Metab, 93*(12), 4880-4886. doi:10.1210/jc.2008-0574

Orho-Melander, M., Melander, O., Guiducci, C., Perez-Martinez, P., Corella, D., Roos, C., . . . Kathiresan, S. (2008). Common missense variant in the glucokinase regulatory protein gene is associated with increased plasma triglyceride and C-reactive protein but lower fasting glucose concentrations. *Diabetes, 57*(11), 3112-3121. doi:10.2337/db08-0516

Ostenson, C. G., & Efendic, S. (2007). Islet gene expression and function in type 2 diabetes; studies in the Goto-Kakizaki rat and humans. *Diabetes Obes Metab, 9 Suppl 2*, 180-186. doi:10.1111/j.1463-1326.2007.00787.x

Ostrom, R. S. (2002). New determinants of receptor-effector coupling: trafficking and compartmentation in membrane microdomains. *Mol Pharmacol, 61*(3), 473-476.

Ostrom, R. S., Gregorian, C., Drenan, R. M., Xiang, Y., Regan, J. W., & Insel, P. A. (2001). Receptor number and caveolar co-localization determine receptor coupling efficiency to adenylyl cyclase. *J Biol Chem, 276*(45), 42063-42069. doi:10.1074/jbc.M105348200

Ostrom, R. S., & Insel, P. A. (2004). The evolving role of lipid rafts and caveolae in G protein-coupled receptor signaling: implications for molecular pharmacology. *Br J Pharmacol, 143*(2), 235-245. doi:10.1038/sj.bjp.0705930

Owen, K., & Hattersley, A. T. (2001). Maturity-onset diabetes of the young: from clinical description to molecular genetic characterization. *Best Pract Res Clin Endocrinol Metab, 15*(3), 309-323. doi:10.1053/beem.2001.0148

Palmer, N. D., McDonough, C. W., Hicks, P. J., Roh, B. H., Wing, M. R., An, S. S., . . . Investigators, M. (2012). A genome-wide association search for type 2 diabetes genes in African Americans. *PLoS One, 7*(1), e29202. doi:10.1371/journal.pone.0029202

Parker-Katiraee, L., Carson, A. R., Yamada, T., Arnaud, P., Feil, R., Abu-Amero, S. N., . . . Scherer, S. W. (2007). Identification of the imprinted KLF14 transcription factor undergoing human-specific accelerated evolution. *PLoS Genet, 3*(5), e65. doi:10.1371/journal.pgen.0030065

Pattanayak, V., Ramirez, C. L., Joung, J. K., & Liu, D. R. (2011). Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat Methods, 8*(9), 765-770. doi:10.1038/nmeth.1670

Potter, K. J., Westwell-Roper, C. Y., Klimek-Abercrombie, A. M., Warnock, G. L., & Verchere, C. B. (2014). Death and dysfunction of transplanted β-cells: lessons learned from type 2 diabetes? *Diabetes, 63*(1), 12-19. doi:10.2337/db12-0364

Poulsen, P., Kyvik, K. O., Vaag, A., & Beck-Nielsen, H. (1999). Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance--a population-based twin study. *Diabetologia, 42*(2), 139-145.

Pound, L. D., Sarkar, S. A., Ustione, A., Dadi, P. K., Shadoan, M. K., Lee, C. E., . . . O'Brien, R. M. (2012). The physiological effects of deleting the mouse SLC30A8 gene encoding zinc transporter-8 are influenced by gender and genetic background. *PLoS One, 7*(7), e40972. doi:10.1371/journal.pone.0040972

Prokopenko, I., Langenberg, C., Florez, J. C., Saxena, R., Soranzo, N., Thorleifsson, G., . . . Abecasis, G. R. (2009). Variants in MTNR1B influence fasting glucose levels. *Nat Genet, 41*(1), 77-81. doi:10.1038/ng.290

Pørksen, N., Grøfte, T., Greisen, J., Mengel, A., Juhl, C., Veldhuis, J. D., . . . Vilstrup, H. (2002). Human insulin release processes measured by intraportal sampling. *Am J Physiol Endocrinol Metab, 282*(3), E695-702. doi:10.1152/ajpendo.00516.2000

Quwailid, M. M., Hugill, A., Dear, N., Vizor, L., Wells, S., Horner, E., . . . Cox, R. D. (2004). A gene-driven ENU-based approach to generating an allelic series in any gene. *Mamm Genome, 15*(8), 585-591. doi:10.1007/s00335-004-2379-z

Rajala, R. V., Chan, M. D., & Rajala, A. (2005). Lipid-protein interactions of growth factor receptor-bound protein 14 in insulin receptor signaling. *Biochemistry, 44*(47), 15461-15471. doi:10.1021/bi0513148

Rees, M. G., Wincovitch, S., Schultz, J., Waterstradt, R., Beer, N. L., Baltrusch, S., . . . Gloyn, A. L. (2012). Cellular characterisation of the GCKR P446L variant associated with type 2 diabetes risk. *Diabetologia, 55*(1), 114-122. doi:10.1007/s00125-011-2348-5

Rees, S. D., Hydrie, M. Z., Shera, A. S., Kumar, S., O'Hare, J. P., Barnett, A. H., . . . Kelly, M. A. (2011). Replication of 13 genome-wide association (GWA)-validated risk variants for type 2 diabetes in Pakistani populations. *Diabetologia, 54*(6), 1368-1374. doi:10.1007/s00125-011-2063-2

Renström, E., Eliasson, L., & Rorsman, P. (1997a). Protein kinase A-dependent and -independent stimulation of exocytosis by cAMP in mouse pancreatic B-cells. *J Physiol, 502 (Pt 1)*, 105-118.

Renström, E., Eliasson, L., & Rorsman, P. (1997b). Protein kinase A-dependent and -independent stimulation of exocytosis by cAMP in mouse pancreatic B-cells. *J Physiol, 502 (Pt 1)*, 105-118.

Rich, T. C., Fagan, K. A., Nakata, H., Schaack, J., Cooper, D. M., & Karpen, J. W. (2000). Cyclic nucleotide-gated channels colocalize with adenylyl cyclase in regions of restricted cAMP diffusion. *J Gen Physiol, 116*(2), 147-161.

Roger, B., Papin, J., Vacher, P., Raoux, M., Mulot, A., Dubois, M., . . . Lang, J. (2011a). Adenylyl cyclase 8 is central to glucagon-like peptide 1 signalling and effects of chronically elevated glucose in

rat and human pancreatic beta cells. *Diabetologia, 54*(2), 390-402. doi:10.1007/s00125-010-1955-x

Roger, B., Papin, J., Vacher, P., Raoux, M., Mulot, A., Dubois, M., . . . Lang, J. (2011b). Adenylyl cyclase 8 is central to glucagon-like peptide 1 signalling and effects of chronically elevated glucose in rat and human pancreatic beta cells. *Diabetologia, 54*(2), 390-402. doi:10.1007/s00125-010-1955-x

Rorsman, P., Eliasson, L., Kanno, T., Zhang, Q., & Gopel, S. (2011). Electrophysiology of pancreatic β-cells in intact mouse islets of Langerhans. *Prog Biophys Mol Biol, 107*(2), 224-235. doi:10.1016/j.pbiomolbio.2011.06.009

Rorsman, P., Ramracheya, R., Rorsman, N. J., & Zhang, Q. (2014). ATP-regulated potassium channels and voltage-gated calcium channels in pancreatic alpha and beta cells: similar functions but reciprocal effects on secretion. *Diabetologia, 57*(9), 1749-1761. doi:10.1007/s00125-014-3279-8

Rorsman, P., Salehi, S. A., Abdulkader, F., Braun, M., & MacDonald, P. E. (2008). K(ATP)-channels and glucose-regulated glucagon secretion. *Trends Endocrinol Metab, 19*(8), 277-284. doi:10.1016/j.tem.2008.07.003

Rosengren, A. H., Braun, M., Mahdi, T., Andersson, S. A., Travers, M. E., Shigeto, M., . . . Eliasson, L. (2012). Reduced insulin exocytosis in human pancreatic β-cells with gene variants linked to type 2 diabetes. *Diabetes, 61*(7), 1726-1733. doi:10.2337/db11-1516

Rung, J., Cauchi, S., Albrechtsen, A., Shen, L., Rocheleau, G., Cavalcanti-Proença, C., . . . Sladek, R. (2009). Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. *Nat Genet, 41*(10), 1110-1115. doi:10.1038/ng.443

Russell, L. B., Bangham, J. W., Stelzner, K. F., & Hunsicker, P. R. (1988). High frequency of mosaic mutants produced by N-ethyl-N-nitrosourea exposure of mouse zygotes. *Proc Natl Acad Sci U S A, 85*(23), 9167-9170.

Russell, L. B., & Montgomery, C. S. (1982). Supermutagenicity of ethylnitrosourea in the mouse spot test: comparisons with methylnitrosourea and ethylnitrosourethane. *Mutat Res, 92*(1-2), 193-204.

Russell, W. L., Carpenter, D. A., & Hitotsumachi, S. (1988). Effect of X-ray and ethylnitrosourea exposures separated by 24 h on specific-locus mutation frequency in mouse stem-cell spermatogonia. *Mutat Res, 198*(2), 303-307.

Rutter, A. R., Fradley, R. L., Garrett, E. M., Chapman, K. L., Lawrence, J. M., Rosahl, T. W., & Patel, S. (2007). Evidence from gene knockout studies implicates Asc-1 as the primary transporter mediating d-serine reuptake in the mouse CNS. *Eur J Neurosci, 25*(6), 1757-1766. doi:10.1111/j.1460-9568.2007.05446.x

Sabbatini, M. E., D'Alecy, L., Lentz, S. I., Tang, T., & Williams, J. A. (2013). Adenylyl cyclase 6 mediates the action of cyclic AMP-dependent secretagogues in mouse pancreatic exocrine cells via protein kinase A pathway activation. *J Physiol, 591*(Pt 15), 3693-3707. doi:10.1113/jphysiol.2012.249698

Sakamoto, A., Liu, J., Greene, A., Chen, M., & Weinstein, L. S. (2004). Tissue-specific imprinting of the G protein Gsalpha is associated with tissue-specific differences in histone methylation. *Hum Mol Genet, 13*(8), 819-828. doi:10.1093/hmg/ddh098

Samani, N. J., Erdmann, J., Hall, A. S., Hengstenberg, C., Mangino, M., Mayer, B., . . . Consortium, W. a. t. C. (2007). Genomewide association analysis of coronary artery disease. *N Engl J Med, 357*(5), 443-453. doi:10.1056/NEJMoa072366

Sanders, E. J. (1987). Ultrastructural cytochemical localization of adenylate cyclase in the early chick embryo. *Cell Tissue Res, 247*(2), 465-468.

Saxena, R., Elbers, C. C., Guo, Y., Peter, I., Gaunt, T. R., Mega, J. L., . . . consortium, D. (2012). Large-scale gene-centric meta-analysis across 39 studies identifies type 2 diabetes loci. *Am J Hum Genet, 90*(3), 410-425. doi:10.1016/j.ajhg.2011.12.022

Saxena, R., Voight, B. F., Lyssenko, V., Burtt, N. P., de Bakker, P. I., Chen, H., . . . Diabetes Genetics

Initiative of Broad Institute of Harvard and MIT, L. n. U., and Novartis Institutes of BioMedical Research. (2007). Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science, 316*(5829), 1331-1336. doi:10.1126/science.1142358

Scott, L. J., Mohlke, K. L., Bonnycastle, L. L., Willer, C. J., Li, Y., Duren, W. L., . . . Boehnke, M. (2007). A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science, 316*(5829), 1341-1345. doi:10.1126/science.1142382

Scott, R. A., Lagou, V., Welch, R. P., Wheeler, E., Montasser, M. E., Luan, J., . . . Consortium, D. G. R. a. M.-a. D. (2012). Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet, 44*(9), 991-1005. doi:10.1038/ng.2385

Scuteri, A., Sanna, S., Chen, W. M., Uda, M., Albai, G., Strait, J., . . . Abecasis, G. R. (2007). Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. *PLoS Genet, 3*(7), e115. doi:10.1371/journal.pgen.0030115

Seed Ahmed, M., Kovoor, A., Nordman, S., Abu Seman, N., Gu, T., Efendic, S., . . . Gu, H. F. (2012). Increased expression of adenylyl cyclase 3 in pancreatic islets and central nervous system of diabetic Goto-Kakizaki rats: a possible regulatory role in glucose homeostasis. *Islets, 4*(5), 343-348. doi:10.4161/isl.22283

Seng, K. C., & Seng, C. K. (2008). The success of the genome-wide association approach: a brief story of a long struggle. *Eur J Hum Genet, 16*(5), 554-564. doi:10.1038/ejhg.2008.12

Shin, S. Y., Fauman, E. B., Petersen, A. K., Krumsiek, J., Santos, R., Huang, J., . . . Consortium, M. T. H. E. R. M. (2014). An atlas of genetic influences on human blood metabolites. *Nat Genet, 46*(6), 543-550. doi:10.1038/ng.2982

Slattery, M. L., Folsom, A. R., Wolff, R., Herrick, J., Caan, B. J., & Potter, J. D. (2008). Transcription factor 7-like 2 polymorphism and colon cancer. *Cancer Epidemiol Biomarkers Prev, 17*(4), 978-982. doi:10.1158/1055-9965.EPI-07-2687

Slavin, B. G., Zarow, C., Warden, C. H., & Fisler, J. S. (2010). Histological, immunocytochemical, and morphometrical analyses of pancreatic islets in the BSB mouse model of obesity. *Anat Rec (Hoboken), 293*(1), 108-116. doi:10.1002/ar.21019

Small, K. S., Hedman, A. K., Grundberg, E., Nica, A. C., Thorleifsson, G., Kong, A., . . . Consortium, M. (2011). Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. *Nat Genet, 43*(6), 561-564. doi:10.1038/ng.833

Smith, K. E., Gu, C., Fagan, K. A., Hu, B., & Cooper, D. M. (2002). Residence of adenylyl cyclase type 8 in caveolae is necessary but not sufficient for regulation by capacitative Ca(2+) entry. *J Biol Chem, 277*(8), 6025-6031. doi:10.1074/jbc.M109615200

Soranzo, N., Rivadeneira, F., Chinappen-Horsley, U., Malkina, I., Richards, J. B., Hammond, N., . . . Deloukas, P. (2009). Meta-analysis of genome-wide scans for human adult stature identifies novel Loci and associations with measures of skeletal frame size. *PLoS Genet, 5*(4), e1000445. doi:10.1371/journal.pgen.1000445

Soumaya, K. (2012). Molecular mechanisms of insulin resistance in diabetes. *Adv Exp Med Biol, 771*, 240-251.

Stacey, S. N., Sulem, P., Masson, G., Gudjonsson, S. A., Thorleifsson, G., Jakobsdottir, M., . . . Stefansson, K. (2009). New common variants affecting susceptibility to basal cell carcinoma. *Nat Genet, 41*(8), 909-914. doi:10.1038/ng.412

Steinthorsdottir, V., Thorleifsson, G., Reynisdottir, I., Benediktsson, R., Jonsdottir, T., Walters, G. B., . . . Stefansson, K. (2007). A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat Genet, 39*(6), 770-775. doi:10.1038/ng2043

Steinthorsdottir, V., Thorleifsson, G., Sulem, P., Helgason, H., Grarup, N., Sigurdsson, A., . . . Stefansson, K. (2014). Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. *Nat Genet*. doi:10.1038/ng.2882

Sunahara, R. K., Dessauer, C. W., Whisnant, R. E., Kleuss, C., & Gilman, A. G. (1997). Interaction of

Gsalpha with the cytosolic domains of mammalian adenylyl cyclase. *J Biol Chem, 272*(35), 22265-22271.

Tachibana, K., Yamasaki, D., Ishimoto, K., & Doi, T. (2008). The Role of PPARs in Cancer. *PPAR Res, 2008*, 102737. doi:10.1155/2008/102737

Tadevosyan, A., Vaniotis, G., Allen, B. G., Hébert, T. E., & Nattel, S. (2012a). G protein-coupled receptor signalling in the cardiac nuclear membrane: evidence and possible roles in physiological and pathophysiological function. *J Physiol, 590*(Pt 6), 1313-1330. doi:10.1113/jphysiol.2011.222794

Tadevosyan, A., Vaniotis, G., Allen, B. G., Hébert, T. E., & Nattel, S. (2012b). G protein-coupled receptor signalling in the cardiac nuclear membrane: evidence and possible roles in physiological and pathophysiological function. *J Physiol, 590*(Pt 6), 1313-1330. doi:10.1113/jphysiol.2011.222794

Tai, H. L., Krynetski, E. Y., Yates, C. R., Loennechen, T., Fessing, M. Y., Krynetskaia, N. F., & Evans, W. E. (1996). Thiopurine S-methyltransferase deficiency: two nucleotide transitions define the most prevalent mutant allele associated with loss of catalytic activity in Caucasians. *Am J Hum Genet, 58*(4), 694-702.

Tamemoto, H., Kadowaki, T., Tobe, K., Yagi, T., Sakura, H., Hayakawa, T., . . . Satoh, S. (1994). Insulin resistance and growth retardation in mice lacking insulin receptor substrate-1. *Nature, 372*(6502), 182-186. doi:10.1038/372182a0

Tengholm, A. (2012). Cyclic AMP dynamics in the pancreatic β-cell. *Ups J Med Sci, 117*(4), 355-369. doi:10.3109/03009734.2012.724732

Tepe, N. M., & Liggett, S. B. (1999). Transgenic replacement of type V adenylyl cyclase identifies a critical mechanism of beta-adrenergic receptor dysfunction in the G alpha q overexpressing mouse. *FEBS Lett, 458*(2), 236-240.

Tepe, N. M., Lorenz, J. N., Yatani, A., Dash, R., Kranias, E. G., Dorn, G. W., & Liggett, S. B. (1999). Altering the receptor-effector ratio by transgenic overexpression of type V adenylyl cyclase: enhanced basal catalytic activity and function without increased cardiomyocyte beta-adrenergic signalling. *Biochemistry, 38*(50), 16706-16713.

Teslovich, T. M., Musunuru, K., Smith, A. V., Edmondson, A. C., Stylianou, I. M., Koseki, M., . . . Kathiresan, S. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. *Nature, 466*(7307), 707-713. doi:10.1038/nature09270

Than, B. L., Goos, J. A., Sarver, A. L., O'Sullivan, M. G., Rod, A., Starr, T. K., . . . Cormier, R. T. (2014). The role of KCNQ1 in mouse and human gastrointestinal cancers. *Oncogene, 33*(29), 3861-3868. doi:10.1038/onc.2013.350

Thaung, C., Hough, T., Hunter, A. J., Hardisty, R., & Nolan, P. M. (2002). In search of new disease models in the mouse using ENU mutagenesis. *Ernst Schering Res Found Workshop*(36), 109-134.

Thaung, C., West, K., Clark, B. J., McKie, L., Morgan, J. E., Arnold, K., . . . Cross, S. H. (2002). Novel ENU-induced eye mutations in the mouse: models for human eye disease. *Hum Mol Genet, 11*(7), 755-767.

Thomas, G., Jacobs, K. B., Yeager, M., Kraft, P., Wacholder, S., Orr, N., . . . Chanock, S. J. (2008). Multiple loci identified in a genome-wide association study of prostate cancer. *Nat Genet, 40*(3), 310-315. doi:10.1038/ng.91

Tomita, T., Doull, V., Pollock, H. G., & Krizsan, D. (1992). Pancreatic islets of obese hyperglycemic mice (ob/ob). *Pancreas, 7*(3), 367-375.

Truty, M. J., Lomberk, G., Fernandez-Zapico, M. E., & Urrutia, R. (2009). Silencing of the transforming growth factor-beta (TGFbeta) receptor II by Kruppel-like factor 14 underscores the importance of a negative feedback mechanism in TGFbeta signaling. *J Biol Chem, 284*(10), 6291-6300. doi:10.1074/jbc.M807791200

Turner, J., & Crossley, M. (1999). Basic Krüppel-like factor functions within a network of interacting haematopoietic transcription factors. *Int J Biochem Cell Biol, 31*(10), 1169-1174.

Voight, B. F., Kang, H. M., Ding, J., Palmer, C. D., Sidore, C., Chines, P. S., . . . Boehnke, M. (2012). The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet, 8*(8), e1002793. doi:10.1371/journal.pgen.1002793

Voight, B. F., Scott, L. J., Steinthorsdottir, V., Morris, A. P., Dina, C., Welch, R. P., . . . Consortium, G. (2010). Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet, 42*(7), 579-589. doi:10.1038/ng.609

Wang, X., Zhou, J., Doyle, M. E., & Egan, J. M. (2001a). Glucagon-like peptide-1 causes pancreatic duodenal homeobox-1 protein translocation from the cytoplasm to the nucleus of pancreatic beta-cells by a cyclic adenosine monophosphate/protein kinase A-dependent mechanism. *Endocrinology, 142*(5), 1820-1827. doi:10.1210/endo.142.5.8128

Wang, X., Zhou, J., Doyle, M. E., & Egan, J. M. (2001b). Glucagon-like peptide-1 causes pancreatic duodenal homeobox-1 protein translocation from the cytoplasm to the nucleus of pancreatic beta-cells by a cyclic adenosine monophosphate/protein kinase A-dependent mechanism. *Endocrinology, 142*(5), 1820-1827. doi:10.1210/endo.142.5.8128

Wani, M. A., Wert, S. E., & Lingrel, J. B. (1999). Lung Kruppel-like factor, a zinc finger transcription factor, is essential for normal lung development. *J Biol Chem, 274*(30), 21180-21185.

Wei, F. Y., Suzuki, T., Watanabe, S., Kimura, S., Kaitsuka, T., Fujimura, A., . . . Tomizawa, K. (2011). Deficit of tRNA(Lys) modification by Cdkal1 causes the development of type 2 diabetes in mice. *J Clin Invest, 121*(9), 3598-3608. doi:10.1172/JCI58056

Wei, F. Y., & Tomizawa, K. (2011). Functional loss of Cdkal1, a novel tRNA modification enzyme, causes the development of type 2 diabetes. *Endocr J, 58*(10), 819-825.

Weir, G. C., & Bonner-Weir, S. (2013). Islet β cell mass in diabetes and how it relates to function, birth, and death. *Ann N Y Acad Sci, 1281*, 92-105. doi:10.1111/nyas.12031

Wijesekara, N., Chimienti, F., & Wheeler, M. B. (2009). Zinc, a regulator of islet function and glucose homeostasis. *Diabetes Obes Metab, 11 Suppl 4*, 202-214. doi:10.1111/j.1463-1326.2009.01110.x

Wijesekara, N., Dai, F. F., Hardy, A. B., Giglou, P. R., Bhattacharjee, A., Koshkin, V., . . . Wheeler, M. B. (2010). Beta cell-specific Znt8 deletion in mice causes marked defects in insulin processing, crystallisation and secretion. *Diabetologia, 53*(8), 1656-1668. doi:10.1007/s00125-010-1733-9

Wilcox, G. (2005). Insulin and insulin resistance. *Clin Biochem Rev, 26*(2), 19-39.

Willoughby, D., Masada, N., Wachten, S., Pagano, M., Halls, M. L., Everett, K. L., . . . Cooper, D. M. (2010). AKAP79/150 interacts with AC8 and regulates Ca2+-dependent cAMP synthesis in pancreatic and neuronal systems. *J Biol Chem, 285*(26), 20328-20342. doi:10.1074/jbc.M110.120725

Willoughby, D., Wachten, S., Masada, N., & Cooper, D. M. (2010). Direct demonstration of discrete Ca2+ microdomains associated with different isoforms of adenylyl cyclase. *J Cell Sci, 123*(Pt 1), 107-117. doi:10.1242/jcs.062067

Withers, D. J., Gutierrez, J. S., Towery, H., Burks, D. J., Ren, J. M., Previs, S., . . . White, M. F. (1998). Disruption of IRS-2 causes type 2 diabetes in mice. *Nature, 391*(6670), 900-904. doi:10.1038/36116

Wu, Z., & Wang, S. (2013). Role of kruppel-like transcription factors in adipogenesis. *Dev Biol, 373*(2), 235-243. doi:10.1016/j.ydbio.2012.10.031

Yang, X., Horn, K., & Wand, G. S. (1998). Chronic ethanol exposure impairs phosphorylation of CREB and CRE-binding activity in rat striatum. *Alcohol Clin Exp Res, 22*(2), 382-390.

Yoshimura, M., & Tabakoff, B. (1995). Selective effects of ethanol on the generation of cAMP by particular members of the adenylyl cyclase family. *Alcohol Clin Exp Res, 19*(6), 1435-1440.

Yu, S., Yu, D., Lee, E., Eckhaus, M., Lee, R., Corria, Z., . . . Weinstein, L. S. (1998). Variable and tissue-specific hormone resistance in heterotrimeric Gs protein alpha-subunit (Gsalpha) knockout mice is due to tissue-specific imprinting of the gsalpha gene. *Proc Natl Acad Sci U S A, 95*(15), 8715-8720.

Zambrowicz, B. P., & Sands, A. T. (2003). Knockouts model the 100 best-selling drugs--will they model the next 100? *Nat Rev Drug Discov, 2*(1), 38-51. doi:10.1038/nrd987

.

Zeggini, E., Scott, L. J., Saxena, R., Voight, B. F., Marchini, J. L., Hu, T., . . . Consortium, W. T. C. C. (2008). Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet, 40*(5), 638-645. doi:10.1038/ng.120

Zeggini, E., Weedon, M. N., Lindgren, C. M., Frayling, T. M., Elliott, K. S., Lango, H., . . . (WTCCC), W. T. C. C. C. (2007). Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science, 316*(5829), 1336-1341. doi:10.1126/science.1142364

Zhang, J., Bakheet, R., Parhar, R. S., Huang, C. H., Hussain, M. M., Pan, X., . . . Hashmi, S. (2011). Regulation of fat storage and reproduction by Krüppel-like transcription factor KLF3 and fat-associated genes in Caenorhabditis elegans. *J Mol Biol, 411*(3), 537-553. doi:10.1016/j.jmb.2011.06.011

Zhang, Q., Ramracheya, R., Lahmann, C., Tarasov, A., Bengtsson, M., Braha, O., . . . Rorsman, P. (2013). Role of KATP channels in glucose-regulated glucagon secretion and impaired counterregulation in type 2 diabetes. *Cell Metab, 18*(6), 871-882. doi:10.1016/j.cmet.2013.10.014

Zhong, H., Beaulaurier, J., Lum, P. Y., Molony, C., Yang, X., Macneil, D. J., . . . Schadt, E. E. (2010). Liver and adipose expression associated SNPs are enriched for association to type 2 diabetes. *PLoS Genet, 6*(5), e1000932. doi:10.1371/journal.pgen.1000932

Published Manuscripts

nature genetics

Identification of an imprinted master *trans* regulator at the *KLF14* locus related to multiple metabolic phenotypes

Kerrin S Small[1,2,10], Åsa K Hedman[3,10], Elin Grundberg[1,2,10], Alexandra C Nica[4], Gudmar Thorleifsson[5], Augustine Kong[5], Unnur Thorsteindottir[5,6], So-Youn Shin[2], Hannah B Richards[7], the GIANT Consortium[8], the MAGIC Investigators[8], the DIAGRAM Consortium[8], Nicole Soranzo[1,2], Kourosh R Ahmadi[1], Cecilia M Lindgren[3], Kari Stefansson[5,6,10], Emmanouil T Dermitzakis[4,10], Panos Deloukas[2,10], Timothy D Spector[1,10] & Mark I McCarthy[3,7,9,10] for the MuTHER Consortium[8]

© 2011 Nature America, Inc. All rights reserved.

Genome-wide association studies have identified many genetic variants associated with complex traits. However, at only a minority of loci have the molecular mechanisms mediating these associations been characterized. In parallel, whereas *cis* regulatory patterns of gene expression have been extensively explored, the identification of *trans* regulatory effects in humans has attracted less attention. Here we show that the type 2 diabetes and high-density lipoprotein cholesterol–associated *cis*-acting expression quantitative trait locus (eQTL) of the maternally expressed transcription factor *KLF14* acts as a master *trans* regulator of adipose gene expression. Expression levels of genes regulated by this *trans*-eQTL are highly correlated with concurrently measured metabolic traits, and a subset of the *trans*-regulated genes harbor variants directly associated with metabolic phenotypes. This *trans*-eQTL network provides a mechanistic understanding of the effect of the *KLF14* locus on metabolic disease risk and offers a potential model for other complex traits.

Variants near the maternally expressed transcription factor *KLF14* (encoding Kruppel-like factor 14) are robustly associated with both type 2 diabetes (T2D) and high-density lipoprotein (HDL) cholesterol levels in large-scale genome-wide association studies (GWAS)[1,2]. These studies have implicated a group of highly correlated SNPs, including rs4731702 and rs972283 located ~14 kb upstream of *KLF14*. *KLF14* is the gene in the region likely to be mediating these effects, as the same SNPs show adipose-specific, maternally restricted *cis*-regulatory associations with *KLF14* expression levels, a pattern that mirrors the parent-of-origin effects for T2D susceptibility at this locus[3].

Because transcription factors such as *KLF14* typically modulate expression of other genes in *trans*, we tested for association between rs4731702 and expression levels of ~24,000 probes (16,663 genes) on the Illumina Human HT12 array in subcutaneous adipose tissue biopsies from a cohort of 776 healthy female twins[4]. The enrichment of rs4731702 *trans* associations for low *P* values (**Fig. 1** and **Supplementary Fig. 1**) suggests that *KLF14* is a master regulator of gene expression in adipose tissue. The pattern of *trans* associations at *KLF14* mirrors the GWAS associations (**Fig. 2** and **Supplementary**

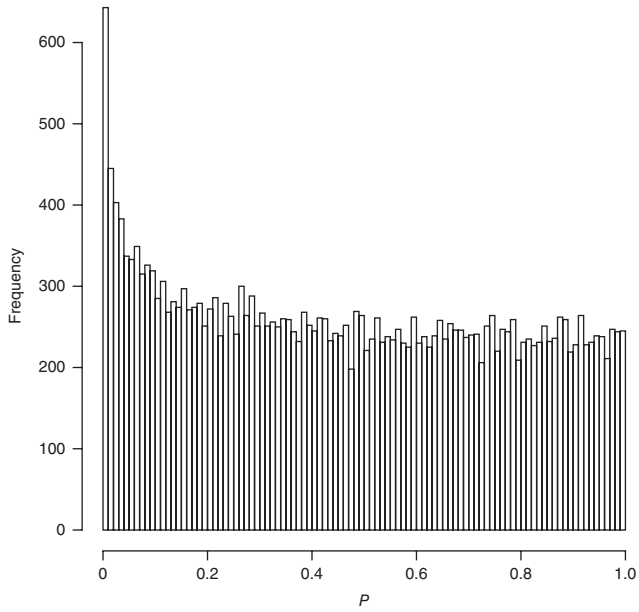


Figure 1 *KLF14* is a master regulator of gene expression in adipose tissue. Shown is the *P* value distribution of association between the *KLF14 cis*-eQTL rs4731702 and expression levels of ~24,000 probes in adipose tissue.

[1]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. [2]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK. [3]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [4]Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland. [5]deCODE genetics, Reykjavik, Iceland. [6]Faculty of Medicine, University of Iceland, Reykjavík, Iceland. [7]Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, Churchill Hospital, Oxford, UK. [8]A full list of members is provided in the **Supplementary Note**. [9]Oxford National Institute for Health Research Biomedical Research Centre, Churchill Hospital, Oxford, UK. [10]These authors contributed equally to this work. Correspondence should be addressed to T.D.S. (tim.spector@kcl.ac.uk) or M.I.M. (mark.mccarthy@drl.ox.ac.uk).

Received 17 December 2010; accepted 14 April 2011; published online 15 May 2011; doi:10.1038/ng.833

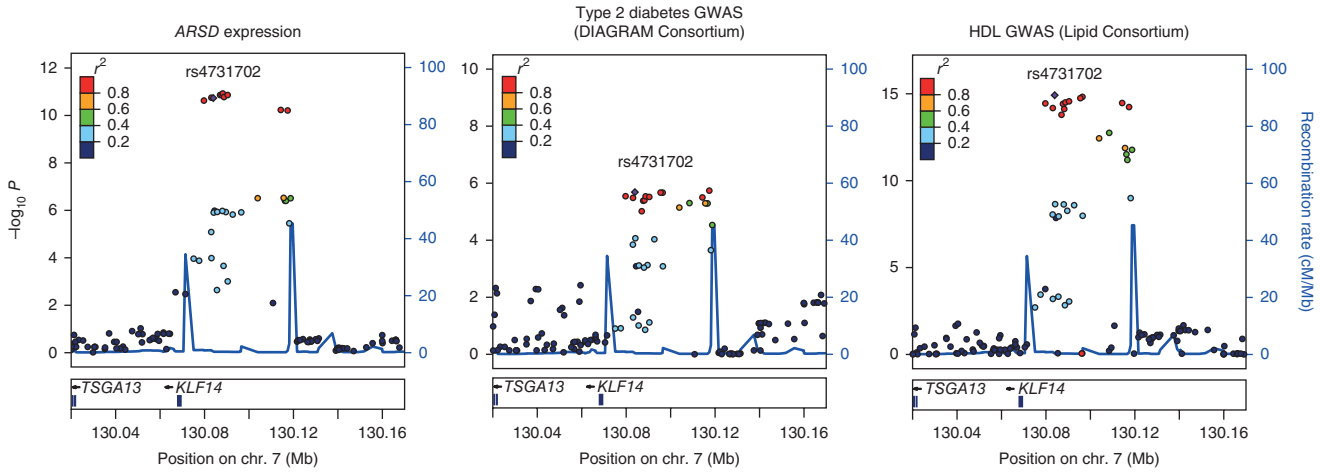


Figure 2 Regional signal plots at the *KLF14* locus. The left panel shows a signal plot of a representative *trans*-regulated gene in the MuTHER samples ($n = 776$). The middle panel shows a signal plot for the type 2 diabetes GWAS meta-analysis performed by the DIAGRAM consortium[1] (stage 1 data only, effective sample size of 22,044; rs972283 reached genome-wide significance after further replication). The right panel shows a signal plot for the HDL cholesterol GWAS meta-analysis performed by the Lipid Consortium[2] ($n = 99,900$). Signal plots of all ten genome-wide significant *trans* regulated genes are included in **Supplementary Figure 2**.

© 2011 Nature America, Inc. All rights reserved.

Fig. 2), and conditioning the *trans* associations on rs4731702 abolished the signal at all other SNPs. These findings indicate that the same set of SNPs (and presumably the same causal variant) underlies the *cis*, *trans* and metabolic trait associations at this locus.

We focused on the ten genes (*TPMT*, *ARSD*, *SLC7A10*, *C8orf82*, *APH1B*, *PRMT2*, *NINJ2*, *KLF13*, *GNB1* and *MYL5*) showing genome-wide significant *trans* (GWST) associations ($P < 5 \times 10^{-8}$) driven by rs4731702 (**Table 1**). First, we sought replication of the *trans* associations in an independent set of adipose tissue samples (deCODE genetics; $n = 589$)[5]. As previously reported[3], the deCODE data showed a strong maternally specific *cis* association between rs4731702 and *KLF14* expression in adipose tissue ($P = 1 \times 10^{-19}$). We did not detect this *cis* effect in the MuTHER data because of apparent problems with the *KLF14* probe represented on the Illumina HT12 array used for the MuTHER data (Online Methods). Seven of the GWST genes from the MuTHER analysis had a directionally consistent *trans* association with $P < 0.05$ in the deCODE replication set (**Table 1**), and we were able to show parent-of-origin effects for the *trans* associations consistent with the maternally specific *cis* effects for *KLF14* expression and T2D risk[3]. In the deCODE replication data, maternally inherited *trans* associations were markedly more significant than general analyses, and we saw no paternally inherited *trans* associations (**Table 1**).

The *trans* effects explain a substantial portion of the genetically regulated variation in GWST expression levels. Our heritability estimates of GWST gene expression levels ranged from 0.13 to 0.79; the rs4731702 *trans*-eQTL explained between 3% and 7.8% of the variance in expression, corresponding to 6% to 25% of the heritability (**Table 2**). Expression levels of the ten GWST genes were moderately correlated in adipose tissue, with a mean |pairwise ρ| of 0.29 (standard deviation = 0.15). *SLC7A10* is the only GWST gene downregulated by the T2D risk allele (and hence the only transcript showing anti-correlated expression levels within the GWST genes). This pattern is consistent with the known ability of the KLF family of transcription factors to act as both transcription activators and repressors[6].

We obtained further support for the hypothesis that the *trans* effects are mediated by *KLF14* expression from analysis of transcription factor binding sites in *trans*-associated genes using PSCAN[7] with the JASPAR database[8]. KLF14 itself is not represented in JASPAR, but other KLF family members have closely related binding sites (and in some cases have been shown to compete for the same binding site)[9], and KLF4 (the only KLF family member in the JASPAR database) and KLF14 share highly similar DNA-binding C-terminal regions[10]. Though we found no evidence for enrichment after correction for multiple testing when examining the ten GWST genes alone, inclusion

Table 1 Genome-wide significant ($P < 5 \times 10^{-8}$) associations of gene expression levels with rs4731702

		MuTHER		deCODE all		deCODE maternal		deCODE paternal		Combined MuTHER + deCODE maternal		
Gene	Chr.	Effect (s.e.)	*P*	Effect (s.e.)	*P*	Effect (s.e.)	*P*	Effect (s.e.)	*P*	Z score	*P*	Direction
APH1B	15	0.08 (0.013)	1.2×10^{-9}	0.11 (0.059)	0.08	0.17 (0.085)	0.07	0.07 (0.083)	0.44	6.1	9.7×10^{-10}	+ +
ARSD	X	0.08 (0.012)	1.9×10^{-11}	0.24 (0.059)	2.2×10^{-4}	0.51 (0.083)	2.6×10^{-8}	−0.004 (0.083)	0.96	8.6	5.4×10^{-18}	+ +
C8orf82	8	0.09 (0.014)	4.8×10^{-10}	0.28 (0.058)	8.9×10^{-6}	0.69 (0.080)	2.1×10^{-14}	−0.09 (0.082)	0.28	9.3	1.1×10^{-20}	+ +
GNB1	1	0.05 (0.009)	4.0×10^{-8}	0.23 (0.059)	1.8×10^{-4}	0.42 (0.085)	1.6×10^{-6}	0.06 (0.084)	0.51	7.2	6.1×10^{-13}	+ +
KLF13	15	0.10 (0.017)	2.2×10^{-8}	−0.01 (0.060)	0.94	0.01 (0.086)	0.89	−0.02 (0.084)	0.80	4.8	1.4×10^{-6}	+ +
MYL5	4	0.09 (0.017)	4.5×10^{-8}	0.20 (0.059)	1.3×10^{-3}	0.45 (0.083)	1.3×10^{-7}	−0.04 (0.083)	0.60	7.4	1.1×10^{-13}	+ +
NINJ2	12	0.08 (0.013)	8.4×10^{-9}	0.14 (0.060)	0.03	0.24 (0.087)	0.01	0.05 (0.085)	0.59	6.3	4.1×10^{-10}	+ +
PRMT2	21	0.06 (0.010)	6.9×10^{-9}	0.18 (0.060)	0.01	0.27 (0.087)	6.7×10^{-3}	0.09 (0.085)	0.33	6.4	2.1×10^{-10}	+ +
SLC7A10	19	−0.27 (0.042)	2.7×10^{-10}	−0.21 (0.057)	7.4×10^{-4}	−0.31 (0.082)	3.3×10^{-4}	−0.11 (0.081)	0.18	−7.3	3.8×10^{-13}	− −
TPMT	6	0.10 (0.013)	1.6×10^{-14}	−0.04 (0.060)	0.49	−0.03 (0.087)	0.78	−0.06 (0.084)	0.49	6.4	1.8×10^{-10}	+ −

The effect allele is the type 2 diabetes risk allele C, which has a frequency of 55% in the HapMap CEU population. Chr., chromosome; s.e., standard error.

Table 2 Heritability and *trans*-eQTL variance of GWST gene expression

Probe ID	Gene	Chr.	Transcription start site (build 36)	Percent variance explained in *trans*	h^2	Percent h^2 explained in *trans*
ILMN_1767816	*APH1B*	15	61,356,801	3.1	0.13	23.2
ILMN_1684873	*ARSD*	X	2,832,010	6.6	0.52	12.6
ILMN_1693862	*C8orf82*	8	145,722,410	6.9	0.27	25.4
ILMN_1760320	*GNB1*	1	1,706,588	4.1	0.19	21.3
ILMN_1679929	*KLF13*	15	29,406,374	4.3	0.70	6.2
ILMN_1746948	*MYL5*	4	661,710	4.0	0.40	10.0
ILMN_1731745	*NINJ2*	12	543,722	4.7	0.48	9.8
ILMN_1675038	*PRMT2*	21	46,879,954	4.9	0.55	8.9
ILMN_1681087	*SLC7A10*	19	38,391,409	5.4	0.79	6.8
ILMN_1740185	*TPMT*	6	18,236,523	7.8	0.48	16.4

Chr., chromosome.

of a larger number of *trans*-associated genes (46 with *trans* $P < 10^{-4}$ or 121 with *trans* $P < 10^{-3}$) showed strong evidence of enrichment for KLF4 binding sites. KLF4 was the most over-represented binding site in the former set (Bonferroni corrected $P = 0.01$) and the second most over-represented site in the latter set (Bonferroni corrected $P = 1.3 \times 10^{-7}$) after EGR1. These data indicate that one feature of the transcripts showing *trans* associations with the *KLF14* SNPs is enrichment for KLF binding sites.

Having shown that the same set of SNPs influences *cis* expression of *KLF14, trans* expression of members of the GWST gene network and a variety of metabolic traits including T2D and HDL cholesterol, we sought to clarify the causal connections between these effects and, in particular, to establish whether or not the *trans* effects were likely to be mediating the metabolic associations at *KLF14*. First, we examined the correlations between *trans* gene expression and concurrently measured metabolic phenotypes. At an array-wide Bonferroni threshold of $P < 1.9 \times 10^{-6}$, expression levels of six of the ten GWST genes were associated with body mass index (BMI) and HDL cholesterol, five with triglycerides and fasting insulin levels, four with HOMA-IR (an index of insulin sensitivity) and two with fasting glucose and adiponectin levels (**Table 3**). Compared to all genes on the array, this represents an enrichment for expression and metabolic phenotype associations, with significance ranging from $P = 0.001$ to $P = 3.3 \times 10^{-5}$. The strength of these associations is consistent with a causal link between *trans* gene expression and metabolic phenotypes and provides clues to the biological processes in which these genes may participate.

Next, we examined large-scale association data made available by trait-specific GWAS meta-analysis consortia, focusing on SNPs in the 250-kb region surrounding each GWST gene. The rs4731702 T2D risk allele is associated with higher fasting insulin[1], indicating that the primary effect on diabetes risk is mediated by decreased peripheral insulin sensitivity. Accordingly, we focused on a set of insulin-resistance–related traits, including fasting insulin[11], fasting glucose[11], HOMA-IR[11], T2D[1], lipids (HDL, low-density lipoprotein (LDL) and triglycerides)[2], body fat distribution (BMI-adjusted waist-hip ratio (WHR))[12] and BMI[13]. In GWAS datasets ranging in size from 22,044 to 123,865 individuals, we found eight associations in five genes at a study-wide significance threshold of $P < 1.03 \times 10^{-4}$ (**Table 4**; see Online Methods for the threshold determination). For example, SNPs near *APH1B* are associated with HDL (rs2729787; $P = 9.8 \times 10^{-9}$) and triglycerides (rs17184382; $P = 1.5 \times 10^{-5}$), and SNPs near *KLF13* are associated with BMI-adjusted WHR (rs4779526; $P = 1.8 \times 10^{-5}$) and LDL (rs8034505; $P = 5.8 \times 10^{-5}$). In addition, SNPs in *MSRA* (expression levels of which marginally failed to reach genome-wide significance; *trans* association $P = 9.8 \times 10^{-8}$) have been previously associated with waist circumference[10] and are here associated with triglycerides (rs615171; $P = 7.5 \times 10^{-7}$). This pattern of association signals shows that variation involving GWST genes has the potential to affect insulin-resistance–related traits and thereby supports the notion that a subset of these genes are directly implicated in mediating the effects of *KLF14* variation on disease susceptibility.

One of the more notable transcripts identified by these analyses is *SLC7A10*, a member of the solute carrier family that mediates transport of neutral amino acids. Adipose expression of *SLC7A10* is highly

© 2011 Nature America, Inc. All rights reserved.

npg

Table 3 Association between expression of GWST genes and concurrently measured metabolic phenotypes

Gene	Adiponectin	HDL	LDL	Triglycerides	BMI	Fasting insulin	Fasting glucose	HOMA-IR
APH1B	7.7×10^{-7} (–0.02)	1.9×10^{-8} (–0.11)	0.01 (0.02)	5.2×10^{-17} (0.13)	5.2×10^{-15} (0.01)	6.7×10^{-18} (0.002)	0.001 (0.05)	1.2×10^{-12} (0.05)
ARSD	0.02 (–0.008)	6.8×10^{-9} (–0.10)	0.02 (0.02)	3.6×10^{-11} (0.09)	5.3×10^{-11} (0.01)	4.3×10^{-16} (0.002)	4.2×10^{-10} (0.07)	2.1×10^{-11} (0.04)
C8orf82	2.8×10^{-5} (0.02)	4.4×10^{-8} (0.11)	0.008 (–0.03)	4.1×10^{-5} (–0.07)	2.4×10^{-14} (–0.01)	6.0×10^{-5} (–0.001)	0.52 (–0.009)	0.0003 (–0.02)
GNB1	0.004 (–0.008)	6.1×10^{-10} (–0.08)	0.03 (0.01)	2.1×10^{-12} (0.08)	2.9×10^{-21} (0.01)	8.4×10^{-8} (0.001)	0.00053 (0.03)	9.8×10^{-6} (0.02)
KLF13	0.17 (0.005)	0.11 (–0.04)	0.86 (–0.002)	0.33 (0.02)	0.93 (0)	0.87 (0)	0.02 (0.04)	0.48 (–0.0002)
MYL5	0.00075 (0.02)	0.00053 (0.10)	0.19 (–0.02)	0.8 (–0.005)	8.4×10^{-5} (–0.009)	0.007 (–0.001)	0.47 (–0.01)	0.001 (–0.02)
NINJ2	0.85 (–0.002)	0.03 (–0.04)	0.03 (0.02)	0.00012 (0.06)	0.003 (0.005)	0.0006 (0.001)	0.51 (0.009)	0.05 (0.02)
PRMT2	0.04 (0.005)	0.06 (0.03)	0.56 (–0.004)	0.21 (–0.015)	1.1×10^{-5} (–0.006)	0.008 (–0.0004)	0.3 (–0.01)	0.002 (–0.01)
SLC7A10	1.6×10^{-14} (0.07)	7.8×10^{-30} (0.65)	2.7×10^{-5} (–0.12)	1.7×10^{-34} (–0.55)	3.6×10^{-48} (–0.08)	1.2×10^{-51} (–0.01)	6.7×10^{-7} (–0.21)	5.3×10^{-45} (–0.27)
TPMT	0.0003 (–0.01)	2.3×10^{-9} (–0.11)	0.01 (0.023)	4.9×10^{-16} (0.12)	7.5×10^{-24} (0.02)	7.3×10^{-21} (0.002)	0.00057 (0.05)	6.7×10^{-15} (0.06)

Values in each cell represent *P* values, with β values in parentheses. HDL, high density lipoprotein; LDL, low-density lipoprotein; BMI, body mass index.

Table 4 Genome-wide association meta-analysis signals within 250 kb of GWST genes

Gene	Trait	SNP	Effect allele	Z score	*P*
APH1B	HDL	rs2729787	T	5.73	9.81×10^{-9}
APH1B	Triglycerides	rs17184382	A	4.32	1.58×10^{-5}
C8orf82	Type 2 diabetes	rs2294120	A	1.14 (OR)	8.43×10^{-5}
NINJ2	LDL	rs2302408	T	3.91	9.06×10^{-5}
SLC7A10	HDL	rs8182584	T	–5.11	3.19×10^{-7}
SLC7A10	Waist-hip ratio	rs7251505	A	–4.85	3.21×10^{-6}
KLF13	LDL	rs8034505	A	4.02	5.85×10^{-5}
KLF13	Waist-hip ratio	rs4779526	A	4.50	1.79×10^{-5}
MSRA	Triglycerides	rs615171	T	–4.95	7.50×10^{-7}
MSRA	Waist circumference	rs7826222	G	5.75	8.89×10^{-9}

This table also includes the results for *MSRA*, for which the *trans* association marginally failed to reach genome-wide significance ($P = 9.8 \times 10^{-8}$). OR, odds ratio.

heritable ($h^2 = 0.79$) and is downregulated by the *KLF14* T2D risk allele. *SLC7A10* expression is strongly associated with diverse metabolic phenotypes, is negatively correlated with BMI ($P = 3 \times 10^{-48}$), insulin ($P = 1.1 \times 10^{-51}$), HOMA-IR ($P = 7 \times 10^{-48}$), glucose ($P = 6 \times 10^{-7}$) and triglycerides ($P = 1 \times 10^{-34}$) and is positively correlated with HDL ($P = 7 \times 10^{-30}$) and adiponectin ($P = 1 \times 10^{-12}$) levels (**Table 3**). The *SLC7A10* locus contains independent ($r^2 = 0.03$) SNPs associated with HDL (rs8182584; $P = 3.2 \times 10^{-7}$) and BMI-adjusted WHR (rs7251505; $P = 3.2 \times 10^{-6}$). The former SNP (rs8182584) is weakly associated with insulin levels ($P = 0.002$) and BMI ($P = 1.4 \times 10^{-3}$), suggesting that this gene has a wide-ranging role in metabolism.

Our data provide convincing evidence of a *bona fide* adipose *trans*-eQTL and implicate this *trans* expression network in the link between *KLF14* variation and risk of metabolic disease. This *trans* regulation uncovers new biological links between previously identified genome-wide significant associations at *KLF14* (with HDL and T2D), *APH1B* (with HDL) and *MSRA* (with waist circumference) and additional signals where metabolic trait associations have not yet been established to genome-wide significance (*SLC7A10*, *KLF13*, *C8orf82* and *NINJ2*). These links provide a framework for hypothesis-directed investigation of genetic interactions among GWAS loci and provide an example of the power of 'integrative genomics' to leverage 'omics' data from multiple sources to discover new biological and functional insights.

URLs. R, http://www.r-project.org/.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Accession numbers. Expression data has been deposited in Array Express under accession number E-TABM-1140.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

The MuTHER study was funded by the Wellcome Trust Program grant # 081917. Genotyping of TwinsUK samples was provided by the Wellcome Trust Sanger Institute and the National Eye Institute via a US National Institutes of Health (NIH)/Center for Inherited Disease Research (CIDR) genotyping project. TwinsUK also receives support from the ENGAGE project grant agreement HEALTH-F4-2007-201413 and from the Department of Health via the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's & St. Thomas' National Health Service Foundation Trust in partnership with King's College London. T.D.S. is an NIHR senior investigator and European Research Council (ERC) senior investigator. M.I.M. is supported by the Oxford NIHR Biomedical Research Centre. Additional support was provided by the Louis-Jeantet Foundation to E.T.D. and A.C.N. and via NIH-NIMH grant R01 MH090941 to E.T.D. and M.I.M.

AUTHOR CONTRIBUTIONS

K.S.S., Å.K.H., E.G., G.T. and A.C.N. analyzed data. G.T., A.K., S.-Y.S., H.B.R., N.S. and C.M.L. contributed reagents, materials and analysis tools. U.T., K.R.A., K.S., E.T.D., P.D., M.I.M. and T.D.S. conceived and designed the experiments. K.S.S. and M.I.M. wrote the paper with contributions from Å.K.H. and E.G. All authors read and approved the manuscript before submission.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Voight, B.F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.* **42**, 579–589 (2010).
2. Teslovich, T.M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
3. Kong, A. *et al.* Parental origin of sequence variants associated with complex diseases. *Nature* **462**, 868–874 (2009).
4. Nica, A.C. *et al.* The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet.* **7**, e1002003 (2011).
5. Emilsson, V. *et al.* Genetics of gene expression and its effect on disease. *Nature* **452**, 423–428 (2008).
6. Dang, D.T., Pevsner, J. & Yang, V.W. The biology of the mammalian Kruppel-like family of transcription factors. *Int. J. Biochem. Cell Biol.* **32**, 1103–1121 (2000).
7. Zambelli, F., Pesole, G. & Pavesi, G. Pscan: finding over-represented transcription factor binding site motifs in sequences from co-regulated or co-expressed genes. *Nucleic Acids Res.* **37**, W247–W252 (2009).
8. Portales-Casamar, E. *et al.* JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **38**, D105–D110 (2010).
9. Kaczynski, J., Cook, T. & Urrutia, R. Sp1- and Kruppel-like transcription factors. *Genome Biol.* **4**, 206 (2003).
10. McConnell, B.B. & Yang, V.W. Mammalian Kruppel-like factors in health and diseases. *Physiol. Rev.* **90**, 1337–1381 (2010).
11. Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* **42**, 105–116 (2010).
12. Heid, I.M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat. Genet.* **42**, 949–960 (2010).
13. Speliotes, E.K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).

© 2011 Nature America, Inc. All rights reserved.

ONLINE METHODS

Sample collection. The MuTHER (Multiple Tissue Human Expression Resource) study includes 856 female individuals of European descent (336 monozygotic and 520 dizygotic twins) recruited from the TwinsUK Adult twin registry[14] (776 MuTHER participants had both adipose tissue expression profiles and genome-wide genotypes available). The age at inclusion ranged from 40 to 87 years, with a median age of 62 years. Punch biopsies (8 mm) were taken from a relatively photoprotected area adjacent and inferior to the umbilicus. Subcutaneous fat tissue was carefully dissected from each biopsy using forceps and a scalpel, weighed and immediately stored in liquid nitrogen. The project was approved by the local ethics committees of all institutions involved, and all samples were collected after obtaining written and signed informed consent. Metabolic phenotypes were measured at the same time point as the biopsies and were collected as previously described, including anthropometric traits (height, weight and BMI)[15], blood lipids[16], adiponectin[17], fasting glucose[18] and fasting insulin[19]. The deCODE replication samples were collected as previously described[5].

RNA extraction. RNA was extracted using TRIzol Reagent (Invitrogen) according to protocol provided by the manufacturer. Briefly, tissue samples were homogenized, and pelleted cells were lysed by repetitive pipetting. RNA quality was confirmed using the Agilent 2100 BioAnalyzer (Agilent Technologies), and the concentrations were determined using NanoDrop ND-1000 (NanoDrop Technologies).

Expression profiling. Expression profiling of the samples, each with either two or three technical replicates, were performed using the Illumina Human HT-12 V3 BeadChips (Illumina Inc.) including more than 48,000 probes, where 200 ng of total RNA was processed according to the protocol supplied by Illumina. All samples were randomized before array hybridization, and the technical replicates were always hybridized on different BeadChips. Log2-transformed expression signals were normalized with quantile normalization of the replicates of each individual followed by quantile normalization across all individuals. Post–quality-control expression profiles were obtained for 825 individuals. The Illumina probe annotations were cross-checked by mapping the probe sequence to the NCBI build 36 genome with MAQ[20]. Only the uniquely mapping probes with no mismatches and either an Ensembl or RefSeq ID were kept for analysis. Probes mapping to genes of uncertain function (LOC symbols) and those encompassing a common SNP were further excluded, leaving 23,552 probes to be used in the analysis.

Genetic association. All MuTHER eQTL associations were performed with the GenABEL and ProbABEL packages[21,22] using both twins from each pair in a polygenic model incorporating a kinship matrix in GenABEL followed by the ProbABEL mmscore option. Age and experimental batch were included as cofactors in the analysis. Seven hundred seventy six samples had both expression profiles and genotypes. Analysis of the deCODE samples, including parent-of-origin analyses, were performed as previously described[3].

Lead SNP selection. The *trans* associations did not have a single lead SNP. Each probe had barely distinguishable *P* values at five SNPs in high linkage disequilibrium ($r^2 > 0.96$, $D' = 1$). Conditioning the *trans* associations on any of the five lead SNPs (rs13234269, rs4731702, rs972283, rs11979110 and rs13233731) completely removed the signal at the other SNPs. rs13234269 had slightly lower *P* values for most probes but was not present in the deCODE replication data. The lead SNP in the T2D study[1] was rs972283, and the parent-of-origin study[3] and the HDL study[2] had rs4731702 as the lead SNP. As rs4731702 and rs972283 are in perfect linkage disequilibrium ($r^2 = 1$) and rs4731702 was directly genotyped in MuTHER and deCODE, whereas rs972283 was imputed in only some of the samples, we chose to report results with respect to rs4731702.

Isoform and probe differences. The Illumina HT12 array contains three probes for *ARSD* and *PRMT2* that are designed to assay different isoforms. Expression levels measured by the two alternate probes for *ARSD* were not associated in *trans* with rs4731702 (ILMN_1684956/NM_001669, $P = 0.53$; ILMN_1661624/NM_009589, $P = 0.13$; deCODE $P = 0.45$ and $P = 0.83$, respectively), whereas one of the two *PRMT2* alternate probes was associated in *trans* with rs4731702 (ILMN_2393544/NM_001535, $P = 6.5 \times 10^{-4}$; ILMN_2259119/NM_206962, $P = 0.59$; no alternate probes in deCODE).

The Illumina HT12 probe targeting *KLF14* (ILMN_1681168) did not detect any expression in the MuTHER samples. As expression of *KLF14* in adipose tissue has been previously reported in the deCODE data as well as three independent studies of adipose tissue in the GEO database[23], we believe the apparent lack of expression in the MuTHER data is because of a technical failure of the Illumina probe. The Illumina *KLF14* probe and the Agilent *KLF14* probe used in the deCODE data target nonoverlapping locations in the single exon of *KLF14*.

Heritability. Expression heritability was calculated with a linear mixed effects model as described[24], including age and experimental batch as covariates. All available complete twin pairs were included, corresponding to 143 monozygotic pairs and 214 dizygotic pairs.

Phenotypic association. Associations between mRNA expression and phenotypes were modeled using a linear mixed effects model, using R (see URLs) and the *lmer()* function in the lme4 package, fitted by maximum likelihood. The linear mixed effects model was adjusted for age and experimental batch (fixed effects) and family relationship (twin pairing) and zygosity (random effects). A likelihood ratio test was used to assess the significance of the phenotype effect. The *P* value of the phenotype effect in each model was calculated from the χ^2 distribution with 1 degree of freedom using −2log (likelihood ratio) as the test statistic. A Fisher exact test was used to assess enrichment of phenotype associations in the GWST gene set versus the full set of genes in adipose tissue.

GWAS lookups. The ten genome-wide significant *trans* genes and their ~250-kb flanking regions were interrogated for SNPs associated with metabolic phenotypes in large GWAS meta-analyses[1,2,11–13] (the GIANT, MAGIC, DIAGRAM and Lipids consortia). A study-wide significance of $P = 1.03 \times 10^{-4}$ was calculated by Bonferroni correcting for the 487 independents tests in the ten regions. The number of independent tests was approximated by the number of tag SNPs needed to tag all CEU HapMap SNPs with minor allele frequency (MAF) > 0.05 in the ten regions with an r^2 of 0.5 as calculated by the Tagger algorithm[25]. Alternatively, requiring an r^2 of 0.8 resulted in 880 tests for a threshold of $P = 5.7 \times 10^{-5}$. The region around *ARSD* was included despite the lack of X chromosome results in most studies. All genes within the ~250-kb flanking regions are listed in **Supplementary Table 1**.

Transcription factor binding sites. We searched 450 bp upstream and 50 bp downstream of the transcription start site of the GWST genes for overrepresented JASPAR binding sites using PSCAN[7].

Genotyping. Genotyping of the TwinsUK dataset (n = ~6,000) was done with a combination of Illumina arrays (HumanHap300, HumanHap610Q, 1M-Duo and 1.2MDuo 1M). Intensity data for each of the three arrays were pooled separately (with 1M-Duo and 1.2MDuo 1M pooled together), and genotypes were called with the Illuminus[26] calling algorithm, thresholding on a maximum posterior probability of 0.95. Similar exclusion criteria were applied to each of the three datasets separately. Exclusion criteria for samples were: (i) sample call rate <98%; (ii) heterozygosity across all SNPs ≥2 standard deviations from the sample mean; (iii) evidence of non-European ancestry as assessed by principal component analysis comparison with HapMap3 populations; and (iv) observed pairwise identity by descent probabilities suggestive of sample identity errors. Exclusion criteria for SNPs were: (i) Hardy-Weinberg $P < 10^{-6}$, assessed in a set of unrelated samples; (ii) MAF < 1%, assessed in a set of unrelated samples; or (iii) SNP call rate <97% (SNPs with MAF ≥ 5%) or <99% (for 1% ≤ MAF < 5%).

Prior to merging the three datasets, we performed pairwise comparison among the three datasets and further excluded SNPs and samples as follows: (i) concordance at duplicate samples <1%; (ii) concordance at duplicate SNPs <1%; (iii) visual inspection of quantile-quantile plots for logistic regression applied to all pairwise dataset comparisons; (iv) Hardy-Weinberg $P < 10^{-6}$,

© 2011 Nature America, Inc. All rights reserved.

assessed in a set of unrelated samples; or (v) observed pairwise identity by descent probabilities suggestive of sample identity errors.

Genotype imputation. Imputation was performed using the IMPUTE software package (v2)[27] using two reference panels, P0 (HapMap2, rel 22, combined CEU, YRI and, ASN panels) and P1 (610K+, including the combined HumanHap610K and 1M array). After imputation, SNPs were filtered at a MAF > 5% and an IMPUTE info value of >0.8.

14. Spector, T.D. & Williams, F.M. The UK Adult Twin Registry (TwinsUK). *Twin Res. Hum. Genet.* **9**, 899–906 (2006).
15. Skidmore, P.M. *et al.* Relation of birth weight, body mass index, and change in size from birth to adulthood to insulin resistance in a female twin cohort. *J. Clin. Endocrinol. Metab.* **93**, 516–520 (2008).
16. Aulchenko, Y.S. *et al.* Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat. Genet.* **41**, 47–55 (2009).
17. Richards, J.B., Valdes, A.M., Burling, K., Perks, U.C. & Spector, T.D. Serum adiponectin and bone mineral density in women. *J. Clin. Endocrinol. Metab.* **92**, 1517–1523 (2007).
18. Prokopenko, I. *et al.* Variants in *MTNR1B* influence fasting glucose levels. *Nat. Genet.* **41**, 77–81 (2009).
19. Falchi, M., Wilson, S.G., Paximadas, D., Swaminathan, R. & Spector, T.D. Quantitative linkage analysis for pancreatic B-cell function and insulin resistance in a large twin cohort. *Diabetes* **57**, 1120–1124 (2008).
20. Li, H., Ruan, J. & Durbin, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* **18**, 1851–1858 (2008).
21. Aulchenko, Y.S., Ripke, S., Isaacs, A. & van Duijn, C.M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294–1296 (2007).
22. Aulchenko, Y.S., Struchalin, M.V. & van Duijn, C.M. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics* **11**, 134 (2010).
23. Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets–10 years on. *Nucleic Acids Res.* **39**, D1005–D1010 (2011).
24. Visscher, P.M., Benyamin, B. & White, I. The use of linear mixed models to estimate variance components from data on twin pairs by maximum likelihood. *Twin Res.* **7**, 670–674 (2004).
25. de Bakker, P.I. *et al.* Efficiency and power in genetic association studies. *Nat. Genet.* **37**, 1217–1223 (2005).
26. Teo, Y.Y. *et al.* A genotype calling algorithm for the Illumina BeadArray platform. *Bioinformatics* **23**, 2741–2746 (2007).
27. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).

© 2011 Nature America, Inc. All rights reserved.

Author's personal copy

Curr Cardiovasc Risk Rep (2013) 7:66–72
DOI 10.1007/s12170-012-0281-x

DIABETES AND INSULIN RESISTANCE (M RUTTER, SECTION EDITOR)

Recent Developments in the Genetic and Genomic Basis of Type 2 Diabetes

Hannah B. Richards · Mark I. McCarthy

Published online: 16 November 2012
© Springer Science+Business Media New York 2012

Abstract Genome wide association studies (GWAS) have transformed the study of heritable factors influencing complex diseases such as type 2 diabetes (T2D), with the current tally of established risk loci approaching 70. Each of these loci has the potential to offer novel insights into the biology of this disease, and opportunities for clinical exploitation. However, the complexity of this condition has often frustrated efforts to achieve these functional and translational advances. This review describes progress made over the past year to expand genome wide association studies, to characterize the mechanisms through which diabetes risk loci operate, and to define the processes involved in diabetes predisposition.

Keywords Type 2 diabetes · Genetic association · Genome-wide association studies · Heritability · Hyperglycaemia · Cardiovascular disease · Complex traits

H. B. Richards · M. I. McCarthy (✉)
Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Old Road, Headington, Oxford OX3 7LJ, UK
e-mail: mark.mccarthy@drl.ox.ac.uk

H. B. Richards
MRC Harwell, Mammalian Genetics Unit, Harwell Science and Innovation Campus, Oxfordshire OX11 0RD, UK

M. I. McCarthy
Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK

M. I. McCarthy
Oxford NIHR Biomedical Research Centre, Churchill Hospital, Old Road, Headington, Oxford OX3 7LJ, UK

Introduction

The massive and growing impact of type 2 diabetes (T2D) on global public health has made a priority of efforts to define the pathogenesis of this disease, and explore the mechanistic basis of individual predisposition. The current count of ~350 million people with T2D worldwide is expected to rise to over 500 million by 2030 (World Health Organization, 2011; International Diabetes Federation, 2011). This rapid rise in overall prevalence is presumed to reflect dramatic shifts in lifestyle and diet, but predisposition is also driven by individual genetic risk. Estimates of the heritability of T2D vary between studies, but figures of 25 % [1] or more [2] are typical.

The search for the DNA sequence variants that underlie the risk of non-autoimmune forms of diabetes (of which T2D is by far the most prevalent) started almost 2 decades ago, and initial efforts were mostly based around family-based linkage analyses. These studies identified several genes (such as those encoding the beta-cell glucose sensor, glucokinase, and key transcription factors such as hepatocyte nuclear factors 1A and 4A) that contained mutations responsible for monogenic forms of early-onset diabetes [3]. However, linkage studies conducted for more common forms of T2D had little impact, and the same was true (with a few notable exceptions [4, 5]) for the candidate-gene based association studies that flourished in their place. The advent of genome wide association studies (GWAS) in the mid-2000s, enabled by completion of the HapMap project and developments in genotyping technology, has had a profound effect on discovery efforts, and as of late 2012, over 65 genetic loci have been shown to be associated with T2D [6••]. While the identification of so many loci offers many potential insights into disease biology, collectively these loci explain only a minor part (around 10 %) of the variation in individual predisposition to T2D, and therefore only a minority of that part which is heritable [6••]. In parallel with efforts to explore the biology of the loci already discovered (see below), there is continued interest in uncovering new loci, and in identifying the variants that are driving the GWAS loci, in the expectation that this will augment the proportion of variance

Author's personal copy

explained. This review catalogues recent developments in these areas. First, we describe recent progress in expanding the number of T2D susceptibility loci and in defining the variants that may be responsible for them. Second, we focus on efforts to define the biology behind these association signals.

Recent Advances in Identifying the Genetic Basis of T2D

While some might question the rationale for ever larger GWAS studies that seem destined to collect growing lists of variants of ever smaller effects [7], expansion of the GWAS approach to access greater sample size and a broader ethnic range has a number of powerful motivations. These include: (a) a more complete definition of the overall contribution of common variant signals to T2D pathogenesis; (b) greater power, usually in combination with imputation to sequence-based reference sets, such as 1000 Genomes, to detect associations to low frequency variants, some of which may have larger effects than the common variants detected by "traditional" GWAS; (c) recognition that the modest effect sizes of common risk-alleles that are tolerated to segregate at high frequency over many generations have little bearing on the potential for the products of the same genes to be valuable therapeutic targets (as common T2D-risk variants in *KCNJ11* and *PPARG* clearly illustrate); (d) the expectation that an expanded list of associated loci will power analyses that look for evidence that these loci are enriched for genes influencing particular networks or pathways; (e) the need to establish the extent to which the variants and loci involved in predisposition are shared between major ancestral and continental populations; and (f) attempts to identify the variants that are causal for genome wide association signals through fine-mapping.

Several of these features emerge from the most recent T2D GWA meta-analysis performed by the DIAbetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium [6••]. Using genotype data from almost 150,000 individuals, predominantly of European descent, the consortium was able to define 10 novel T2D-susceptibility loci to genome-wide significance, and to highlight several hundreds more that, whilst failing to reach the stringent criteria typically regarded as proof, are nonetheless highly likely to reflect genuine associations. By combining variation at both the "proven" and "probable" loci, the authors were able to explain ~10.7 % of overall variance in risk of T2D (and therefore as much as 40 % of the inherited component). Further, using approaches that take advantage of the complete set of GWAS results, these data were consistent with the view that a substantial component (as much as half) of the variation in individual risk of diabetes can be defined by common sequence variants. As described below, the expanded list of loci also provided a powerful tool for revealing networks and pathways implicated in T2D pathogenesis.

Whilst the activities of the DIAGRAM consortium have focused on samples of European descent, the past year has seen considerable expansion of efforts to identify common variants influencing T2D-risk in other populations, including those of South and East Asian origin, and African-Americans. For example, Kooner and colleagues [8] completed a GWA meta-analysis in over 5500 T2D cases and 14,400 controls from the UK, Singapore, and Pakistan, all with origins in the Indian subcontinent. This analysis identified 6 novel association signals, including variants near the genes encoding the GRB14 adaptor protein and hepatocyte nuclear factor-4A, the latter already implicated in monogenic forms of diabetes [8]. Equivalent efforts in East Asian subjects have been similarly productive, adding a further 8 loci to the global tally [9, 10]. Studies in individuals of African descent are of particular interest given their extensive genetic diversity, and during the past year the first large-scale association studies from African-American subjects have emerged [11, 12]. These studies have highlighted some of the particular challenges associated with genetic studies in African-descent populations (such as limited linkage disequilibrium, and genetic admixture) but did reveal a number of novel genome-wide significant signals, including those mapping near *RND3* and *BCL2*.

The detection of so many novel signals in non-European samples might appear to indicate a considerable degree of ethnic specificity in the genetic basis of T2D. However, most of the top association signals detected in one ethnic group have been visible (not always at genome-wide significant levels) in others, and the likelihood that a given signal will be detected for the first time in a particular population often has more to do with chance, and/or with differences in allele frequency and linkage disequilibrium that impact on study power. In the most complete study of transethnic associations to date, Haiman and colleagues [13•] tested 20 of the known T2D risk variants across 6 major racial and ethnic groups from the US (including African Americans, Hispanics, American Indians, and Native Hawaiians) and found marked overlap in the patterns of association, with almost complete directional consistency of effects. These data provide compelling evidence that common variant associations for T2D are shared between global populations, and in turn, that these associations likely reflect the action of causal variants that are themselves common and widely-distributed. (The logic runs as follows: if the causal alleles that underlie the common variant signals detected by GWAS were rare, they would almost certainly have arisen after the human diaspora out of Africa, and would not, therefore, result in broadly consistent patterns of association across diverse ethnic groups).

Efforts to identify the specific causal alleles through fine-mapping approaches have also gained momentum over the past year, building on the more complete surveys of regional sequence variation accessible through targeted custom arrays, such as the metabochip [14•], and through imputation

Author's personal copy

from sequence based reference panels, such as those generated by the 1000 Genomes project. It has long been hypothesized that transethnic analyses, which might allow shared causal variants to be distinguished from their partially-correlated neighbors, would be particularly valuable in this respect, a proposition recently tested by Buyske and colleagues [15•] in African Americans. They were able to show, at the *CELSR2/PSRC1/SORT1* locus known to influence lipid levels, that rs12740374 stood out as the strongest associated SNP for LDL-cholesterol in African-descent subjects, whilst, in European ancestry samples, this SNP was amongst a group of highly-correlated variants which could not be separated by association analyses. The availability of large multiethnic GWAS and the growing size and diversity of sequence-based reference panels is now bolstering efforts to fine-map the variants driving the T2D risk loci uncovered by GWAS.

GWAS methodology has, to date, largely focused on the contribution of common variants to disease susceptibility. However, deep sequencing studies have revealed the extent to which the rapid, recent, population growth of mankind has enriched human sequence variation for rare, even private, mutations [16•, 17]. This has generated additional support for the view that low frequency and rare variants, many acting in the face of strong adverse selection, are likely to play an important role in defining individual predisposition to common disease [18].

Rapid improvements in the throughput and accuracy of next-generation sequencing mean that systematic association analysis of variants of all frequencies and all types is now possible, and several large-scale efforts are underway to identify and characterize the contribution of low frequency and rare variants in T2D. These studies, which involve whole genome and whole exome sequencing in many thousands of individuals, are likely to report in the coming year. Over the past year, published research in this area has been restricted to single gene studies, but these have served to illustrate some of the challenges associated with the generation and interpretation of rare variant associations.

Bonnefond and colleagues [19•], for example, focused on resequencing of the *MTNR1B* gene, encoding one of the receptors for the hormone melatonin. Several GWAS have shown that common variants near *MTNR1B* influence fasting glucose levels in non-diabetic individuals, as well as having an impact on T2D risk [20–22] and there is compelling biological evidence that these are likely to be acting through altered expression of the receptor [21]. Deep resequencing of *MTNR1B* coding sequence, in more than 7500 European subjects, revealed a total of 40 nonsynonymous variants, 36 of them with minor allele frequency below 0.1 %. When considered collectively, these rare variants were seen to be associated with T2D, with estimated mean allelic odds ratios of ~3, far exceeding the equivalent figure for the common variants identified by GWAS (~1.10). To determine the impact of each of these variants on receptor function, the authors functionally characterized each one, distinguishing those with molecular effects from those with none. Reanalysis of the genetic data demonstrated that the T2D association signal was restricted to individuals carrying those functionally-defective alleles. These findings confirm *MTNR1B* as the gene mediating the GWAS signal, but also illustrate that prediction of the functional and physiological impact of rare novel alleles, such as those which might be revealed by medical resequencing, is difficult in the absence of direct empirical examination. Rees and colleagues [23] reached similar conclusions in their analysis of the effects of rare variants in the *GCKR* gene (encoding the glucokinase regulatory protein) on lipid profiles. Studies such as these have important implications for both the discovery of rare variant signals, and the clinical use of sequence data for diagnostic and prognostic purposes in common disease.

The Next Step: Elucidating the Biological Mechanisms Driving Association

Once a robust association signal has been identified, the challenge shifts to converting these discoveries into meaningful biological insights. For GWAS signals, this has generally proved a greater challenge than might have been foreseen. There are a number of reasons for this: (1) most of the signals map to non-coding areas of the genome, such that the transcript mediating the signal is typically not obvious; (2) extensive linkage disequilibrium means that the variant(s) responsible for each signal is poorly-specified; and (3) individual effect sizes are modest. The sequence-based studies now underway are expected to start to address these issues by enabling more systematic fine-mapping of the causal alleles and by defining additional risk-alleles at these loci, some of them of lower frequency and larger effect, and some of them mapping within coding sequence [24]. In addition, the improving annotation of regulatory elements, through the ENCODE project [25] and related efforts in disease relevant tissues [26], will accelerate the interpretation of signals in the non-coding genome.

In the meantime, investigators have been making steady progress in using these genetic data to drive biological inference. Broadly-speaking, these efforts have followed one of two approaches. In the first, the emphasis has been on exploiting the collective properties of the variants and regions discovered (in the absence of precise information regarding the causal alleles and transcripts) to define the broader phenotypic and genomic attributes of the common variant contribution to T2D. In the second, the focus has been on characterizing the mechanisms operating at individual loci.

Author's personal copy

The studies from the DIAbetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium provide examples of the former. In the most recent of these, described earlier, DIAGRAM investigators [6••] characterized the common variant architecture of T2D through a case–control study of ~150,000 European subjects, identifying hundreds of loci that were highly likely (but not absolutely certain) to house further T2D-risk variants. They used this enlarged set of loci to perform a series of "enrichment" studies, asking the following question: is there anything about the transcripts that map near to T2D GWAS signals that is different (from the perspective of functional annotation, protein-protein interactions and the like) from transcripts that do not? These analyses revealed that these T2D loci were collectively enriched for genes involved in: (1) cell cycle regulation (speaking to the importance of beta-cell mass as a determinant of insulin secretion); (2) adipocytokine signalling (reflecting tight links between adipose tissue and glucose metabolism); and (3) CREBBP-interacting transcription factor activity.

Using the same data, the DIAGRAM investigators were also able to extend previous analyses which derive biological insights from the association effects of T2D-risk variants on related traits, such as body mass index, fasting glucose (in non-diabetic individuals), and indices of beta-cell function and insulin action [6••, 27, 28•]. They were able to confirm: (1) partial, but not complete, overlap between variants that influence individual risk of T2D, and those that modulate physiological variation in fasting glucose amongst healthy individuals; (2) that the only signals which are driven by a primary effect on obesity are those at *FTO* and *MC4R*; and (3) that, whilst most risk loci operate via beta-cell dysfunction, a growing number (see Table 1) exert their T2D-risk effects through an obesity-independent deterioration in insulin sensitivity. This list of "insulin resistance" loci offers interesting insights into key players mediating the actions of insulin in peripheral tissues. In the case of the *GRB14* locus for example, which emerged from GWAS in South Asians as well as Europeans [8], RNA expression data from fat confirms *GRB14* as the strongest candidate transcript at the locus: its product is an adaptor protein that binds to the insulin receptor to inhibit tyrosine kinase signaling [29].

The second type of approach seeks to use a wide range of approaches, including association fine-mapping and resequencing in man, and functional examination of variants and candidate transcripts in cellular and animal models, to define the mechanisms whereby the T2D-risk variants at each locus exert their effects on T2D predisposition. These studies aim both to identify specific alleles that drive association (which can deliver insights into the biology upstream of the locus), and the transcripts through which they act (which feeds into downstream mechanisms). We briefly mention data for a couple of T2D loci that illustrate progress in these endeavors.

One of the first novel T2D-risk signals to emerge from GWAS mapped to the region around *CDKAL1* [30–33]. It was clear that this effect was mediated via a defect in beta-cell function, and that the risk variants mapped within an intron of *CDKAL1*, a gene of unknown function, its official name (CDK5 regulatory subunit associated protein 1-like 1) having been assigned on the basis of sequence homology alone. However, it was not clear how *CDKAL1* might be mediating T2D risk, or, indeed whether *CDKAL1* was actually the culprit. Recent

Table 1 T2D loci acting via insulin resistance

Gene locus	Associated phenotypes	eQTL data	Phenotypes Influenced
Obesity-independent			
PPARG	T2D[4], BMI[42], CAD[43]	None	Adipogenesis, lipid uptake in adipocytes[44]
IRS1	T2D[45], body fat %[46], CAD[47], blood lipids[39], hyperinsulinemia[45]	Cis: adipose[39]	Fat distribution[39]; adipocyte biology[39, 48]
GCKR	T2D[49], fasting glucose[50], fasting insulin[50], triglycerides[31]	None	Glucose sensing, signal transduction[23]
KLF14	T2D[27], HDL[39]	Cis: adipose[27]; Trans: >40 transcripts[41••]	Adipocyte growth and signaling[41••]
ADAMTS9	T2D[51], waist-hip ratio[51]	Cis: omental adipose[52]	Unknown
GRB14	T2D[6••], waist-hip ratio (in females) [51], HDL[39]	Cis: omental adipose[6••]	Insulin receptor signaling, regulates insulin action[53]
Primary effect on obesity			
FTO	T2D[54], BMI[55], obesity[55], body fat %[46], waist circum[54], CAD[43]	None	Food intake[56]; lypolytic activity in adipose[57]
MC4R	T2D[6], waist circum[58], obesity, weight, BMI, body fat %[59], CAD[43]	None	Food intake[60]

Examples of T2D loci considered to be acting primarily through defects in insulin action.

Author's personal copy

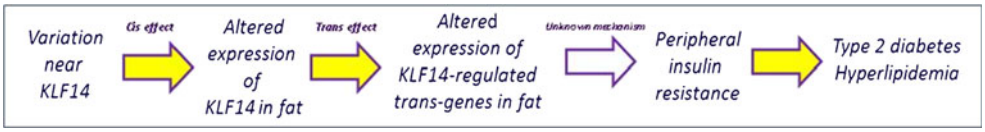


Fig. 1 Model for *KLF14* effects. SNPs upstream of *KLF14* influence *KLF14* expression in adipose tissue. Since *KLF14* is a master regulator of dozens of transcripts in adipose tissue, these same SNPs influence expression of these trans-regulated genes. In turn, altered expression of certain of these transcripts promotes insulin resistance in other key sites of insulin action (through mechanisms yet to be defined) and, thereby, increases risk of T2D and hyperlipidemia

work has established that the beta-cell specific *Cdkal1* knockout mouse displays decreased insulin secretion and impaired glucose tolerance, strongly supporting the candidacy of *CDKAL1* in man [34–36]. At the same time, it has emerged that CDKAL1 is not a modulator of cyclin dependent kinase activity, as homology analyses had intimated: instead, it acts as a methylthiotransferase with a crucial role in stabilizing the codon-anticodon interaction involving the transfer RNA for lysine in particular, and is consequently involved in the correct incorporation of that amino acid into protein [34, 37]. One possible mechanism for the effect on T2D risk might be that lack of CDKAL1 in the beta cell causes mistranslation in proinsulin synthesis resulting in defective insulin processing and endoplasmic reticulum stress [38].

In contrast, T2D-risk at the *KLF14* locus is clearly driven by obesity-independent defects in insulin action [27, 39]. Analysis of RNA expression data in adipose tissue samples has demonstrated that the T2D-associated alleles are also associated with reduced *KLF14* expression [27, 40, 41••], and that both the T2D and the transcript expression association signals are imprinted (that is, subject to epigenetic effects that limit expression to one of the 2 inherited copies of the gene), with effects restricted to the maternal allele [40]. These data clearly implicate *KLF14*, which encodes the Krueppel-like factor 14 transcription factor, as the causal transcript at this locus. More recently, Small and colleagues from the MuTHER consortium, again using expression data from subcutaneous fat, were able to show that the same set of SNPs upstream of *KLF14* influencing T2D-risk and modifying the expression of *KLF14* itself, was also driving expression of a suite of several dozen transcripts elsewhere in the genome (that is, in trans). The inference that *KLF14* acts as a master regulator of adipocyte expression is supported by the known biology of KLFs, and by the observation that the promoter regions of the trans genes are enriched for KLF binding sites [41••]. On the basis of these data, the authors hypothesized that the *KLF14*-associated T2D-risk effect was mediated through the effects of one or more of these trans-genes (see Fig. 1), and were able to find support for this in the evidence that several of the trans-genes, including *SLC7A10*, themselves harbored variants that were independently associated with a range of insulin-resistant traits [41••]. These data suggest that *KLF14* and its set of downstream effectors are intimately involved in the generation of the signals whereby the abnormal function of subcutaneous adipose tissue is communicated to other organs (notably muscle and liver) which are the major sites of insulin action. Further research, involving cellular and animal models, as well as physiological, genomic, and biomarker studies in man, is required to establish what these signals might be.

Conclusion

The studies highlighted in this review demonstrate how increasingly powerful genetic and genomic approaches are being used to harvest and then characterize genetic variation that influences individual predisposition to T2D. These studies are motivated by a desire to improve understanding of the pathogenesis of this common, yet mysterious, disease, and to exploit the particular value of human genetics to provide causal inference. Though the pace of biological inference continues to lag behind that of genetic discovery, this has quite a lot to do with the types of alleles that have emerged from genome wide association studies, and the expectation is that the shift towards sequence-based discovery will bring greater access to pathogenic variation which has larger effects and is more often located within coding sequence. From a clinical perspective, the value of this research lies not so much in the expectation that genetic variation alone will provide clinically useful prediction of disease risk, but that an enhanced understanding of disease biology will drive efforts to define novel therapeutic agents acting on well-validated molecular targets, and non-genetic biomarkers that can be used to monitor disease state.

Acknowledgements HR is a DPhil student funded by the Nuffield Department of Medicine and the Clarendon Fund. MMcC is an NIHR Senior Investigator.

Author's personal copy

Disclosure No potential conflicts of interest relevant to this article were reported.

References

Papers of particular interest, published recently, have been highlighted as:

- • Of importance
- •• Of major importance

1. Poulsen P, Ohm-Kyvik K, Vaag A, Beck-Nielsen H. Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance–a population-based twin study. Diabetologia. 1999;42(2):139–45.
2. Florez J, Hirschhorn J, Altshuler D. The inherited basis of diabetes mellitus: implications for the genetic analysis of complex traits. Annu Rev Genomics Hum Genet. 2003;4:257–91.
3. Owen K, Hattersley A. Maturity-onset diabetes of the young: from clinical description to molecular genetic characterization. Best Pract Res Clin Endocrinol Metab. 2001;15(3):309–23.
4. Altshuler D, Hirschhorn J, Klannemark M, Lindgren C, Vohl M, Nemesh J, et al. The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. Nat Genet. 2000;26(1):76–80.
5. Gloyn A, Weedon M, Owen K, Turner M, Knight B, Hitman G, et al. Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (*KCNJ11*) and SUR1 (*ABCC8*) confirm that the *KCNJ11* E23K variant is associated with type 2 diabetes. Diabetes. 2003;52(2):568–72.
6. •• Morris A, Voight B, Teslovich T, Ferreira T, Segre A, Steinthorsdottir V, Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet. 2012;44(9):981–90. *Ten new T2D-risk loci were identified as well as hundreds of association signals which did not reach genome-side significance but are likely to contain more T2D-risk variants. Enrichment analyses revealed associations between T2D loci and genes involved in three main biological pathways. Other analyses further extended previous biological insights rom association effects of T2D-risk variants on related traits*.
7. McClellan J, King M. Genetic heterogeneity in human disease. Cell. 2010;141(2):210–7.
8. Kooner JS, Saleheen D, Sim X, Sehmi J, Zhang W, Frossard P, et al. Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. Nat Genet. 2011;43(10):984–9.
9. Cho Y, Chen C, Hu C, Long J, Ong RT, Sim X, et al. Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. Nat Genet. 2012;44(1):67–72.
10. Kim Y, Go M, Hu C, Hong C, Kim Y, Lee J, et al. Large-scale genome-wide association studies in East Asians identify new genetic loci influencing metabolic traits. Nat Genet. 2011;43(10):990–5.
11. Palmer N, McDonough C, Hicks P, Roh B, Wing M, An S, et al. A genome-wide association search for type 2 diabetes genes in African Americans. PLoS One. 2012;7(1):e29202.
12. Saxena R, Elbers C, Guo Y, Peter I, Gaunt T, Mega J, et al. Large-scale gene-centric meta-analysis across 39 studies identifies type 2 diabetes loci. Am J Hum Genet. 2012;90(3):410–25.
13. • Haiman C, Fesinmeyer M, Spencer K, Buzkova P, Voruganti V, Wan P, et al. Consistent directions ofeffect for established type 2 diabetes risk variants across populations: the Population Architecture using Genomics and Epidemiology (PAGE) Consortium. Diabetes. 2012;61(6):1642–7. *In the most complete trans-ethnic T2D GWAS study to date, Haiman et al. show directional consistency for almost all variants studied across a number of populations, providing strong evidence that common variant associations are the same between populations.*
14. • Voight B, Kang H, Ding J, Palmer C, Sidore C, Chines P, et al. The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. PLoS Genet. 2012;8(8):e1002793. *A custom genotyping array which targets nearly 200,000 SNP markers called the Metabochip is described and evaluated for identification of new associated loci and characterisation of genome significant loci by fine-mapping.*
15. • Buyske S, Wu Y, Carty CL, Cheng I, Assimes T, Dumitrescu L, et al. Evaluation of the metabochip genotyping array in African Americans and implications for fine mapping of GWAS-identified loci: the PAGE study. PLoS One. 2012;7(4):e35651. *This study demonstrates how large multiethnic GWAS and expanding sequence-based reference panels is improving efforts to fine-map the variants driving the risk loci uncovered by GWAS.*
16. • Keinan A, Clark A. Recent explosive human population growth has resulted in an excess of rare genetic variants. Science. 2012;336(6082):740–3. *Keinan and Clark's observation that the exponential human population growth over the last 400 generations has resulted in an abundance of rare variants could have ground-breaking results for the methods of studying complex diseases and traits.*
17. Nelson MR, Wegmann D, Ehm M, Kessner D, Jean St P, Verzilli C, et al. An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. Science. 2012;337 (6090):100–4.
18. Lupski J, Belmont J, Boerwinkle E, Gibbs R. Clan genomics and the complex architecture of human disease. Cell. 2011;147(1):32–43.
19. • Bonnefond A, Clement N, Fawcett K, Yengo L, Vaillant E, Guillaume J, et al. Rare *MTNR1B* variants impairing melatonin receptor 1B function contribute to type 2 diabetes. Nat Genet. 2012;44(3):297–301. *Exon re-sequencing yields 40 non-synonymous rare variants in the MTNR1B gene which are submitted to rigorous functional investigation in one of the most complete rare variant functional follow-ups of a GWAS candidate to date.*
20. Prokopenko I, Langenberg C, Florez JC, Saxena R, Soranzo N, Thorleifsson G, et al. Variants in *MTNR1B* influence fasting glucose levels. Nat Genet. 2009;41(1):77–81.
21. Lyssenko V, Nagorny CL, Erdos M, Wierup N, Jonsson A, Spegel P, et al. Common variant in *MTNR1B* associated with increased risk of type 2 diabetes and impaired early insulin secretion. Nat Genet. 2009;41(1):82–8.
22. Bouatia-Naji N, Bonnefond A, Cavalcanti-Proenca C, Sparso T, Holmkvist J, Marchand M, et al. A variant near *MTNR1B* is associated with increased fasting plasma glucose levels and type 2 diabetes risk. Nat Genet. 2009;41(1):89–94.
23. Rees M, Wincovitch S, Schultz J, Waterstradt R, Beer N, Baltrusch S, et al. Cellular characterisation of the *GCKR* P446L variant associated with type 2 diabetes risk. Diabetologia. 2012;55 (1):114–22.
24. Nejentsev S, Walker N, Riches D, Egholm M, Todd J, et al. Rare variants of *IFIH1*, a gene implicated in antiviral responses, protect against type 1 diabetes. Science. 2009;324(5925):387–9.
25. ENCODE Project Consortium, Bernstein B, Birney E, Dunham I, Green E, Gunter C, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57–74.
26. Gaulton K, Nammo T, Pasquali L, Simon J, Giresi P, Fogarty M, et al. A map of open chromatin in human pancreatic islets. Nat Genet. 2010;42(3):255–9.
27. Voight B, Scott L, Steinthorsdottir V, Morris A, Dina C, Welch R, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet. 2010;42(7):579–89.

Author's personal copy

28. • Scott R, Lagou V, Welch R, Wheeler E, Montasser M, Luan J, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nat Genet. 2012;44(9):991–1005. *A thorough genome-wide meta-analysis study in Europeans increases the list of confirmed loci influencing glycaemic traits to 53 and provides strong evidence of association at other biologically relevant loci*.
29. Holt L, Lyons R, Ryan A, Beale S, Ward A, Cooney G, et al. Dual ablation of *Grb10* and *Grb14* in mice reveals their combined role in regulation of insulin signaling and glucose homeostasis. Mol Endocrinol. 2009;23(9):1406–14.
30. Zeggini E, Weedon M, Lindgren C, Frayling T, Elliott K, Lango H, et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science. 2007;316 (5829):1336–41.
31. Manning A, Hivert M, Scott R, Grimsby J, Bouatia-Naji N, Chen H, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science. 2007;316 (5829):1331–6.
32. Scott L, Mohlke K, Bonnycastle L, Willer C, Li Y, Duren W, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science. 2007;316 (5829):1341–5.
33. Steinthorsdottir V, Thorleifsson G, Reynisdottir I, Benediktsson R, Jonsdottir T, Walters G, et al. A variant in *CDKAL1* influences insulin response and risk of type 2 diabetes. Nat Genet. 2007;39 (6):770–5.
34. Wei F, Suzuki T, Watanabe S, Kimura S, Kaitsuka T, Fujimura A, et al. Deficit of tRNA(Lys) modification by *Cdkal1* causes the development of type 2 diabetes in mice. J Clin Invest. 2011;121 (9):3598–608.
35. Wei F, Tomizawa K. Functional loss of *Cdkal1*, a novel tRNA modification enzyme, causes the development of type 2 diabetes. Endocr J. 2011;58(10):819–25.
36. Ohara-Imaizumi M, Yoshida M, Aoyagi K, Saito T, Okamura T, Takenaka H, et al. Deletion of *CDKAL1* affects mitochondrial ATP generation and first-phase insulin exocytosis. PLoS One. 2010;5 (12):e15553.
37. Arragain S, Handelman S, Forouhar F, Wei F, Tomizawa K, Hunt J, et al. Identification of eukaryotic and prokaryotic methylthiotransferase for biosynthesis of 2-methylthio-N6-threonylcarbamoyladenosine in tRNA. J Biol Chem. 2010;285(37):28425–33.
38. Kaufman RJ. Beta-cell failure, stress, and type 2 diabetes. N Engl J Med. 2011;365(20):1931–3.
39. Teslovich T, Musunuru K, Smith A, Edmondson A, Stylianou I, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature. 2010;466(7307):707–13.
40. Kong A, Steinthorsdottir V, Masson G, Thorleifsson G, Sulem P, Besenbacher S, et al. Parental origin of sequence variants associated with complex diseases. Nature. 2009;462(7275):868–74.
41. •• Small K, Hedman A, Grundberg E, Nica A, Thorleifsson G, Kong A, et al. Identification of an imprinted master trans regulator at the *KLF14* locus related to multiple metabolic phenotypes. Nat Genet. 2011;43(6):561–4. *The authors find that variants near KLF14, previously associated with T2D, insulin resistance, HDL cholesterol, and expression of KLF14 in fat, are associated in trans with over 40 transcripts in the genome. These data provide strong evidence that KLF14 is a master regulator of genes in fat and suggest a mechanism through which adipose signalling may lead to peripheral insulin resistance.*
42. Deeb S, Fajas L, Nemoto M, Pihlajamaki J, Mykkanen L, Kuusisto J, et al. A Pro12Ala substitution in PPARgamma2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity. Nat Genet. 1998;20(3):284–7.
43. Samani N, Erdmann J, Hall A, Hengstenberg C, Mangino M, Mayer B, et al. Genomewide association analysis of coronary artery disease. N Engl J Med. 2007;357(5):443–53.
44. Cristancho A, Lazar M. Forming functional fat: a growing understanding of adipocyte differentiation. Nat Rev Mol Cell Biol. 2011;12(11):722–34.
45. Rung J, Cauchi S, Albrechtsen A, Shen L, Rocheleau G, Cavalcanti-Proenca C, et al. Genetic variant near *IRS1* is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. Nat Genet. 2009;41(10):1110–5.
46. Kilpelainen TO, Zillikens MC, Stančakova A, Finucane FM, Ried JS, Langenberg C, et al. Genetic variation near *IRS1* associates with reduced adiposity and an impaired metabolic profile. Nat Genet. 2011;43(8):753–60.
47. Baroni MG, D'Andrea MP, Montali A, Pannitteri G, Barilla F, Campagna F, et al. A common mutation of the insulin receptor substrate-1 gene is a risk factor for coronary artery disease. Arterioscler Thromb Vasc Biol. 1999;19(12):2975–80.
48. Miki H, Yamauchi T, Suzuki R, Komeda K, Tsuchida A, Kubota N, et al. Essential role of insulin receptor substrate 1 (IRS-1) and IRS-2 in adipocyte differentiation. Mol Cell Biol. 2001;21(7):2521–32.
49. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet. 2010;42(2):105–16.
50. Orho-Melander M, Melander O, Guiducci C, Perez-Martinez P, Corella D, Roos C, et al. Common missense variant in the glucokinase regulatory protein gene is associated with increased plasma triglyceride and C-reactive protein but lower fasting glucose concentrations. Diabetes. 2008;57(11):3112–21.
51. Heid IM, Jackson AU, Randall JC, Winkler TW, Qi L, Steinthorsdottir V, et al. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. Nat Genet. 2010;42(11):949–60.
52. Zhong H, Beaulaurier J, Lum PY, Molony C, Yang X, Macneil DJ, et al. Liver and adipose expression associated SNPs are enriched for association to type 2 diabetes. PLoS Genet. 2010;6(5):e1000932.
53. Rajala RV, Chan MD, Rajala A. Lipid-protein interactions of growth factor receptor-bound protein 14 in insulin receptor signaling. Biochemistry. 2005;44(47):15461–71.
54. Scuteri A, Sanna S, Chen WM, Uda M, Albai G, Strait J, et al. Genome-wide association scan shows genetic variants in the *FTO* gene are associated with obesity-related traits. PLoS Genet. 2007;3 (7):e115.
55. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, et al. A common variant in the *FTO* gene is associated with body mass index and predisposes to childhood and adult obesity. Science. 2007;316(5826):889–94.
56. Fredriksson R, Hägglund M, Olszewski PK, Stephansson O, Jacobsson JA, Olszewska AM, et al. The obesity gene, *FTO*, is of ancient origin, up-regulated during food deprivation and expressed in neurons of feeding-related nuclei of the brain. Endocrinology. 2008;149(5):2062–71.
57. Loos RJ, Bouchard C. *FTO*: the first gene contributing to common forms of human obesity. Obes Rev. 2008;9(3):246–50.
58. Chambers JC, Elliott P, Zabaneh D, Zhang W, Li Y, Froguel P, et al. Common genetic variation near *MC4R* is associated with waist circumference and insulin resistance. Nat Genet. 2008;40(6):716–8.
59. Loos RJ, LIndgren CM, Li S, Wheeler E, Zhao JH, Prokopenko I, et al. Common variants near *MC4R* are associated with fat mass, weight and risk of obesity. Nat Genet. 2008;40(6):768–75.
60. Adan RA, Tiesjema B, Hillebrand JJ, la Fleur SE, Kas MJ, de Krom M. The MC4 receptor and control of appetite. Br J Pharmacol. 2006;149(7):815–27.