Genetics of Diabetes Subtypes. Characterization of novel cluster-based diabetes subtypes.

Mansour Aly, Dina

2022

Document Version:
Publisher's PDF, also known as Version of record

Genetics of Diabetes Subtypes

Characterization of novel cluster-based diabetes subtypes

DINA MANSOUR ALY
DEPARTMENT OF CLINICAL SCIENCES MALMÖ | LUND UNIVERSITY

Genetics of Diabetes Subtypes

Characterization of novel cluster-based diabetes subtypes

Dina Mansour Aly

LUND
UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the Faculty of Medicine, Lund University, Sweden.
To be defended at Aulan, Clinical Research Center,

Jan Waldenströms gata 35, Malmö

On Thursday 20th January 2022 at 13:00.

Faculty opponent
Professor Stefan Johansson

University of Bergen

Norway

Organization LUND UNIVERSITY	Document name Doctoral Dissertation
	Date of issue 2022-01-20
Author(s) Dina Mansour Aly	Sponsoring organization

Title and subtitle

Genetics of Diabetes Subtypes

Characterization of novel cluster-based diabetes subtypes

Abstract

BACKGROUND: Type 2 diabetes (T2D) has been reproducibly clustered into five subtypes based on six-clinical variables; age at diabetes onset, body mass index (BMI), Glutamic acid decarboxylase autoantibodies (GADA), glycated hemoglobin (HbA1c) and insulin secretion and resistance estimated as HOMA2B and HOMA2IR derived from fasting glucose and C-peptide. These subtypes have different disease progression and risk of complications. The newly defined subtypes are called Severe Autoimmune Diabetes (SAID), Severe Insulin Deficient Diabetes (SIDD), Severe InsulinResistant Diabetes (SIRD), Mild Obesity-related Diabetes (MOD), and Mild Age-Related Diabetes (MARD).

AIM: The main aim of the thesis was to characterize the subtypes using genetics and biomarkers to investigate potential etiological differences, identify subtype-specific genetic associations and determine the underlying mechanisms of kidney complications in the subtypes.

METHODS: The project included individuals with diabetes (cases) from the Swedish cohort All New Diabetics In Scania (ANDIS, n=10927) and the Finnish cohort Diabetes Registry Vasa (DIREVA, n=4754) as well as diabetes-free individuals (controls) from the Swedish Malmö Diet and Cancer cohort (MDC,n=2744) and the Finnish Botnia cohort (n=1683). Clusters defined in Ahlqvist et al, 2018, were used for all analyses. The number of individuals in the subtypes were as follows: SAID (n=452, n=327), SIDD (n=1193, n=394), SIRD (n=1130, n=453), MOD (n=1374, n=596) and MARD (n=2861, n=1178), in ANDIS and DIREVA respectively. In Paper I and III, genome-wide association studies (GWAS) and genetic risk score (GRS) analyses were performed to compare underlying genetic drivers in the Swedish cohorts and replicated in the Finnish cohorts. In Paper III, the primary phenotype was estimated glomerular filtration rate (eGRF) reflecting chronic kidney disease. In Paper II, epidemiological and genetic analysis was performed using clustering, Cox regression models and GRS to compare GADA negative individuals with diabetes of Iraqi (n=286) and Swedish origin (n=10641) with respect to new diabetes subclassification and complications. In Paper IV, the proteomic profiles of the subtypes were studied using 1161 biomarkers measured on Olink panels. Machine learning algorithms were applied to prioritize biomarkers, followed by Menedelian Randomization.

RESULTS: In Paper I, the *HLA* rs9273368 variant was significantly associated with SAID (OR=2.89,P=6.5x10^{-40}), the *TCF7L2* rs7903146 variant was significantly associated with SIDD (OR=1.56, P=8.6x10^{-15}), MOD (OR=1.40, P=3.1x10^{-10}) and MARD (OR=1.42,P=6.1x10^{-16}). The rs10824307 variant near the *LRMDA* gene was uniquely associated with MOD (OR=1.35, P=1.3×10^{-09}). GRS for fasting insulin showed a unique association with SIRD (OR=1.855, P=5.91x10^{-09}). GRSs for BMI were associated with SIDD, SIRD and MOD but not MARD (OR=1.046, P=0.099). Paper II concluded thar individuals with diabetes from Iraq present with a more insulin-deficient subtype than native Swedes. They have a higher risk of coronary events but a lower risk of CKD. In Paper III, in ANDIS, eGFR was strongly associated with the A allele of rs77924615 in the well-established *PDILT-UMOD* locus (beta=0.126, p=6.61x10^{-13}) in all T2D; MARD and SIDD but not in MOD or SIRD (p>0.05). In the SIRD subtype, eGFR was associated with the C allele of rs3770382 in the *CTNNA2* gene at near genome-wide significance (beta=-0.219, p=5.5x10- 08), but was not associated in any of the other subtypes. In DIREVA, the *PDILT-UMOD* locus replicated in T2D, MARD, and SIDD, and was also associated in SIRD (beta=0.24, p=0.001) but not in MOD (beta=0.076, p=0.109). The *CTNNA2* locus did not replicate in DIREVA. Paper IV, the diabetes subtypes were shown to have different proteomic profiles and a list of prioritized biomarkers was generated for future follow-up.

CONCLUSION: The newly defined subtypes are partially distinct with genetically different backgrounds and SIRD is suggested to have more beta-cell independent pathogenesis. There is some suggestive support for different genetic backgrounds of DKD in diabetes subtypes. Biomarkers could be valuable for better discrimination of subtypes and cross-cohort comparisons in larger datasets. The diabetes subclassification approach paves the way for individualized patient management and the development of new therapeutic targets.

Keywords: Precision medicine**,** diabetes subclassification, genome-wide association studies, GWAS, genetic risk scores, GRS, chronic kidney disease, CKD, diabetic kidney disease, DKD, estimated glomerular filtration rate, eGFR.

Classification system and/or index terms (if any)

Supplementary bibliographical information	**Language** English
ISSN and key title 1652-8220 (Genetics of diabetes subtypes)	**ISBN** 978-91-8021-166-6
Recipient's notes / **Number of pages 91** / Security classification	Price

Signature Dina G Mansour Aly Date 2021-12-06

Genetics of Diabetes Subtypes

Characterization of novel cluster-based diabetes subtypes

Dina Mansour Aly

LUND
UNIVERSITY

Cover picture by Malk Amr Yehia Sakr Ahmed in Canva

Faculty of Medicine
Department of Clinical Sciences Malmö

ISSN 1652-8220
ISBN 978-91-8021-166-6
Lund University, Faculty of Medicine Doctoral Dissertation Series 2022:5

Printed in Sweden by Media-Tryck, Lund University
Lund 2021

"THE WIND MAY BLAST AND IT MAY SEEM DARK ON THE FORECAST, BUT ALL IT NEEDS IS ONE HEADSTRONG EYE TO SEE THE SOUGHT OF A FULL DREAM. THIS IS THE RIDE YOU DON'T WISH TO MISS, IT'S ONE REVOLUTION, BUT IT FOLLOWS A LOT MORE YET TO COME". Luca Falcon

Dedication to My Father; Dr. Gamaleldin Mansour Aly; Professor of Anesthesia, Faculty of Medicine, Ain Shams University, Egypt and Head of Anesthesia department Al-Razi Hospital, Kuwait; You were my inspiration and driving force to power through.

To my mother, and brothers. To my Husband; Amr Yehia Sakr Ahmed; Professor of Clinical Oncology and Radiotherapy, Faculty of Medicine, Cairo University, Egypt.

To my lovely daughters; Mahitab, Dalia, Neveen, and Malk; you were my light and strength on the journey track!

Table of Contents

Publications included in this thesis

1. **Genome-wide association analyses highlight etiological differences underlying newly defined subtypes of diabetes**

   Dina Mansour Aly, Om Prakash Dwivedi, Rashmi B. Prasad, Annemari Käräjämäki, Rebecka Hjort, Manonanthini Thangam, Mikael Åkerlund, Anubha Mahajan, Miriam S. Udler, Jose C. Florez, Mark I. McCarthy, Regeneron Genetics Center, Julia Brosnan, Olle Melander, Sofia Carlsson, Ola Hansson, Tiinamaija Tuomi1, Leif Groop and Emma Ahlqvist

   Nat Genet, 2021, PMID: 34737425

2. **Adult-onset diabetes in Middle Eastern immigrants to Sweden: Novel subgroups and diabetic complications - the ANDIS cohort diabetic complications and ethnicity**

   Louise Bennet, **Dina Mansour Aly***, Christopher Nilsson*, Anders Christensson, Leif Groop and Emma Ahlqvist

   Diabetes Metab Res Rev, 2021, PMID: 33119194

3. **Genetics of kidney complications in diabetes subtypes**

   Dina Mansour Aly, Om Prakash Dwivedi, Annemari Karajamaki, Leif Groop, Rashmi B Prasad, Ola Hansson, Tiinamaija Tuomi, and Emma Ahlqvist

   (Manuscript in preparation).

4. **Proteomic profiles of novel diabetic subtypes**

   Manonanthini Thangam, **Dina Mansour Aly**, Olof Asplund, Ola Hansson, Rashmi B Prasad, Leif Groop and Emma Ahlqvist

   (Manuscript in preparation).

   * Shared authorship position

Publications not included in this thesis

1. **Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables**

   Emma Ahlqvist, Petter Storm, Annemari Käräjämäki, Mats Martinell, Mozhgan Dorkhan, Annelie Carlsson, Petter Vikman, Rashmi B Prasad, **Dina Mansour Aly**, Peter Almgren, Ylva Wessman, Nael Shaat, Peter Spégel, Hindrik Mulder, Eero Lindholm, Olle Melander, Ola Hansson, Ulf Malmqvist, Åke Lernmark, Kaj Lahti, Tom Forsén, Tiinamaija Tuomi, Anders H Rosengren and Leif Groop

   Lancet D&E, 2018, PMID: 29503172

2. **Elevated circulating follistatin associates with an increased risk of type 2 diabetes**

   Chuanyan Wu, Yan Borné, Rui Gao, Maykel López Rodriguez, William C. Roell, Jonathan M. Wilson, Ajit Regmi, Emma Ahlqvist, Cheng Luan, Andreas Peter, Jürgen Machann, Harald Staiger, Andreas Fritsche, Rongya Tao, Robert Wagner, Mickaël Canouil, Mun-Gwan Hong, Jochen M. Schwenk, **Dina Mansour Aly**, Peter Nilsson, Angela Shore, Faisel Khan, Andrea Natali, Olle Melander, Marju Orho-Melander, Jan Nilsson, Hans-Ulrich Häring, Erik Renström, Claes B.Wollheim, Ewan R. Pearson, Paul W. Franks, Morris F. White, Kevin L. Duffin, Allan Arthur Vaag, Markku Laakso, Norbert Stefan, Leif Groop and Yang De Marinis

   Nat Commun, 2021, PMID: 34759311

3. **Genetic factors affect the susceptibility to bacterial infections in diabetes**

   Johan R Simonsen, Annemari Käräjämäki, Anni A Antikainen, Iiro Toppila, Emma Ahlqvist, Rashmi Prasad, **Dina Mansour-Aly**, Valma Harjutsalo, Asko Järvinen, Tiinamaija Tuomi, Leif Groop, Carol Forsblom, Per-Henrik Groop, Niina Sandholm and Markku Lehto

   Scientific Reports, 2021, PMID: 33947878

4. **Genetic analysis of obstructive sleep apnoea discovers obesity independent risk factors and a strong association with cardiometabolic health**

   Satu Strausz, Sanni Ruotsalainen, Hanna M Ollila, Juha Karjalainen, Tuomo Kiiskinen, Mary Reeve, Mitja Kurki, Nina Mars, Aki S Havulinna, Elina Luonsi, **Dina Mansour Aly**, Emma Ahlqvist, Maris Teder-Laving, Priit Palta, Leif Groop, Reedik Mägi, Antti Mäkitie, Veikko Salomaa, Adel Bachour, Tiinamaija Tuomi, FinnGen; Aarno Palotie, Tuula Palotie and Samuli Ripatti.

   Eur Respir J, 2021, PMID: 33243845

5. **Replication and cross-validation of type 2 diabetes subtypes based on clinical variables: an IMI-RHAPSODY study**

   Roderick C Slieker, Louise A Donnelly, Hugo Fitipaldi, Gerard A Bouland, Giuseppe N Giordano, Mikael Åkerlund, Mathias J Gerl, Emma Ahlqvist, Ashfaq Ali, Iulian Dragan, Andreas Festa, Michael K Hansen, **Dina Mansour Aly**, Min Kim, Dmitry Kuznetsov, Florence Mehl, Christian Klose, Kai Simons, Imre Pavo, Timothy J Pullen, Tommi Suvitaival, Asger Wretlind, Peter Rossing, Valeriya Lyssenko, Cristina Legido-Quigley, Leif Groop, Bernard Thorens, Paul W Franks, Mark Ibberson, Guy A Rutter, Joline W J Beulens, Leen M 't Hart and Ewan R Pearson

   Diabetologia, 2021, PMID: 34110439.

6. **Distinct molecular signatures of clinical clusters in people with type 2 diabetes: An IMI-RHAPSODY study**

   Roderick C Slieker, Louise A Donnelly, Hugo Fitipaldi, Gerard A Bouland, Giuseppe N Giordano, Mikael Åkerlund, Mathias J Gerl, Emma Ahlqvist, Ashfaq Ali, Iulian Dragan, Petra Elders, Andreas Festa, Michael K Hansen, Amber A van der Heijden, **Dina Mansour Aly**, Min Kim, Dmitry Kuznetsov, Florence Mehl, Christian Klose, Kai Simons, Imre Pavo, Timothy J Pullen, Tommi Suvitaival, Asger Wretlind, Peter Rossing, Valeriya Lyssenko, Cristina Legido Quigley, Leif Groop, Bernard Thorens, Paul W Franks, Mark Ibberson, Guy A Rutter, Joline W J Beulens, Leen M 't Hart and Ewan R Pearson

   Diabetes, 2021, PMID: 34376475

Abbreviations

Abbreviation	Term
ABCG2	ATP binding cassette subfamily G member 2 (Junior blood group)
ADCY5	adenylate cyclase 5
AFF3	AF4/FMR2 family member 3
AGER	advanced glycosylation end-product specific receptor
AGTR1	angiotensin II receptor type 1
ALMS1	ALMS1 centrosome and basal body associated protein
AMPK	5'AMP-activated protein kinase
ANDIS	All New Diabetic In Scania
ANXA9	annexin A9
APOL1	apolipoprotein L1
AUH	AU RNA binding methylglutaconyl-CoA hydratase
B37	Build 37 of the human genome
BMI	Body mass index
C10orf11	Chromosome 10 opening reading frame 11
CARS	cysteinyl-tRNA synthetase 1
CDCA7	cell division cycle associated 7
CERS2	ceramide synthase 2
CHN2	chimerin 2
CIR	Corrected Insulin response
CKD	Chronic kidney disease
CNDP1	carnosine dipeptidase 1
CNKSR3	CNKSR family member 3
CTNNA2	catenin alpha 2
DAB2	DAB adaptor protein 2
DACH1	dachshund family transcription factor 1
DKD	Diabetic kidney disease
EA	Effect allele
eGFR	Estimated glomerular filtration rate
ELMO1	engulfment and cell motility 1
eQTL	Expression quantitative locus
ERBB4	erb-b2 receptor tyrosine kinase 4
FINS	Fasting insulin
FRMD3	FERM domain containing 3
FTO	FTO alpha-ketoglutarate dependent dioxygenase
GADA	Glutamic acid decarboxylase antibodies
GCKR	glucokinase regulator

GCTA	Genome-wide Complex Trait Analysis
GLP1R	glucagon like peptide 1 receptor
GLRA3	glycine receptor alpha 3
GRM	Genetic relationship matrix
GRS	Genetic risk scores
GTEx	Genotype-Tissue Expression
GWAMA	Genome-Wide Association Meta-Analysis
GWAS	Genome wide association studies
HbA1c	Glycated haemoglobin
HDL	High density lipoprotein
HHEX	hematopoietically expressed homeobox
HLA	Human Leukocyte Antigen
HMGA2	high mobility group AT-hook 2
HOMA2B	Homeostasis Model Assessment (HOMA) Beta cell function
HOMA2IR	Homeostasis Model Assessment (HOMA) Insulin Resistance
IGV	Integrative Gene Viewer
INS	insulin
IRS1	insulin receptor substrate 1
ISI	Insulin sensitivity index
ISR	Insulin secretion rate
KCNQ1	potassium voltage-gated channel subfamily Q member 1
KEGG	Kyoto Encyclopaedia of Genes and Genomes
KNG1	kininogen 1
LD	Linkage disequilibrium
LDL	Low density lipoprotein
LIMK2	LIM domain kinase 2
LRMDA	Leucine rich melanocyte differentiation associated
MAF	Minor allele frequency
MARD	Mild Age Related Diabetes
MCTP2	multiple C2 and transmembrane domain containing 2
MDC	Malmö Diet and Cancer
MDRD	Modification of Diet in Renal Disease
MMP9	matrix metallopeptidase 9
MOD	Mild Obesity Related Diabetes
MR	Mendelian Randomization
MSRB	methionine sulfoxide reductase B2
MYH9	myosin heavy chain 9
NAT8	N-acetyltransferase 8
NEA	Reference allele
NMUR2	neuromedin U receptor 2
OR	Odds ratio
PC	Principal component
PDILT	Protein disulfide isomerase like, testis expressed
PDX1	pancreatic and duodenal homeobox 1
PPARGC1A	PPARG coactivator 1 alpha
PRKAG2	protein kinase AMP-activated non-catalytic subunit gamma 2

PRS	Polygenic risk scores
PVT1	Pvt1 oncogene
QC	Quality control
RAET1L	retinoic acid early transcript 1L
RGMA	repulsive guidance molecule BMP co-receptor a
RPS12	ribosomal protein S12
SAID	Severe Autoimmune Diabetes
SHROOM3	shroom family member 3
SIDD	Severe Insulin Deficient Diabetes
SIRD	Severe Insulin Resistant Diabetes
SLC12A3	solute carrier family 12 member 3
SLC34A1	solute carrier family 34 member 1
SNP	Single nucleotide polymorphism
SORBS1	sorbin and SH3 domain containing 1
SP3	Sp3 transcription factor
STC1	stanniocalcin 1
T1D	Type 1 diabetes
T2D	Type 2 diabetes
TC	Total Cholesterol
TCF7L2	Transcription factor 7 like 2
TG	Triglycerides
TMPO	thymopoietin
UACR	Urinary albumin creatinine ratio
UBE2Q2	ubiquitin conjugating enzyme E2 Q2
UMOD	Uromodulin
VAT	Visceral adipose tissue
VCF	Variant call file
VEP	Variant Effect Predictor
WHR	Waist hip ratio
ZMIZ1	zinc finger MIZ-type containing 1

Public Summary

Diabetes is a chronic disease defined by blood sugar levels above the normal. It is avery concerning disease to every individual at any age especially when diagnosed at a late stage and accompanied by serious complications: decrease in kidney function, eyesight changes, brain and heart conditions. Usually, patients pass with a pre-diabetic phase, characterized by an increase in body weight, water retention and in many cases increase in blood pressure, before they are diagnosed with diabetes. This clinical presentation is known as metabolic syndrome.

Blood sugar is regulated by insulin hormone secreted from the beta cells of the pancreas in response to food intake. When insulin binds to its receptors on the cells it facilitates sugar uptake by the cells which maintains blood sugar levels within the normal range. Other body systems, including the endocrine, nervous and immune systems, interact and affect the regulation of blood glucose.

Traditionally, patients are classified into two main groups: type 1 diabetes (T1D) and type 2 diabetes (T2D). T1D is usually diagnosed with diabetes at a very early age and insulin therapy is required for survival. T2D patients are diagnosed at a later age and antidiabetics other than insulin are used to control their blood sugar level. Diabetes patients, especially within the T2D group, vary in their blood sugar control, response to anti-diabetic medications and risk of diabetes complications.

In a recent study conducted in southern Sweden, 14000 adult individuals were enrolled in the All New Diabetics In Scania (ANDIS) cohort. The assessment of diabetes was done based on fasting blood glucose level conjointly with body mass index (measured by the individual's body weight and height), age, glycated haemoglobin blood level (a type of sugar-related haemoglobin), insulin resistance (the uptake of glucose by body cells in response to insulin after a meal) and insulin secretion (when and how much insulin is released after a meal). Based on the results of this study, diabetes was classified into five new subtypes[1]. These newly defined subtypes differ in their start time of diabetes, control of blood sugar levels, and in their risk to develop diabetic complications. This new subclassification has been reproduced in several ethnicities including Finnish, Indian, and Chinese diabetic populations.

In this population-based project, individuals within each diabetes subtype were compared to patients without diabetes to explore the relationship between diabetes and genetics. Family history was determined using two approaches; patient self-

reported questionnaire and genetics. The genetic association analysis was performed using the genotypes of the individuals and genetic risk scores; these are scores calculated for every individual in both cases and controls. Genetic analysis of kidney complications was performed in the subtypes using estimated glomerular filtration rate (eGFR) as an assessment of kidney disease. The results of the genetic analysis were reproduced in the Finnish cohorts and were combined in a meta-analysis. Protein profile analysis to determine the association of biomarkers (proteins) and their pathways with the subtypes was performed using the blood levels of 1161 proteins from 176 individuals of the SIDD, SIRD, MOD, and MARD subtypes.

The following is a brief description and summary of the findings in each of the newly defined subtypes.

The first subtype, Severe Autoimmune Diabetes (SAID), included 6% of the patients. The family history of T1D plays an important role in this subtype. The immune system of these patients creates antibodies against the pancreas protein GAD, leading to a decreased function in the pancreas. These patients are traditionally referred to as T1D and LADA (Latent Autoimmune Diabetes in Adults), diagnosed at relatively early age, lean body weight, their insulin blood level is low and thus they have poor control on their blood sugar level. In this project, SAID was significantly associated with *HLA* gene variants. *HLA* is known to be associated with autoimmune diseases (where the body's immune system attacks its organs). These results were expected as the immune system plays an important role in the development of this subtype. As expected, T1D genetic risk scores were strongly associated with this subtype. Insulin is often the first-line therapy to gain control of blood sugar in SAID patients.

The second subtype, Severe Insulin Deficient Diabetes (SIDD), included 18% of the patients. Family history of T2D plays an important role in this subtype. Usually, patients present with very high blood sugar levels, yet they do not have GADA antibodies. These patients have poor control of their blood sugar and are at a high risk to develop diabetic eye problems, that can progress to blindness if left untreated, and nervous system problems. In this project, we could show that SIDD is genetically similar to T2D and not T1D. The protein profile for SIDD showed an association with leptin and leptin receptors. Our results suggest that patients in this subtype could benefit from this new subclassification and get insulin at an early stage to improve their blood sugar and protect them against serious eye problems.

The third subtype, Severe Insulin Resistant Diabetes (SIRD), included 15% of the patients. Family history plays a less strong role in this subtype. Usually, patients are diagnosed at a later age and suffer from increased body weight. These patients are at a high risk of developing kidney disease, fatty liver, serious heart, and brain diseases. In this project, strikingly, SIRD was not associated with the well-known T2D associated *TCF7L2* gene variant. Additionally, the T2D individual risk scores showed a smaller association compared to other subtypes. These results could

suggest that SIRD individuals are genetically different from T2D and that the pancreas has a minor role compared with other subtypes. Fasting insulin risk scores showed a unique association with SIRD, this reflects the insulin resistance in this subtype. Insulin resistance is a condition where blood sugar levels are high because of the lower response of the cells to insulin. In contrast to SIDD and MARD, kidney complication genetic analysis in SIRD did not show association with the well-known *PDILT-UMOD* gene variants instead showed unique association with variants in the *CTNNA2*. This finding suggested that the development of DKD in SIRD has a different mechanism than the other subtypes. Protein analysis showed association with proteins associated with elevated blood pressure and inflammation. The results suggest that SIRD patients are genetically different from other subtypes. Patients in this subtype, have to get their liver and kidney functions monitored to protect or delay serious diseases.

The fourth subtype, Mild Obesity-related Diabetes (MOD), included 22% of the patients. Family history of T2D is common. Patients within this group present with increased body weight and poor blood sugar control. The study revealed that this group has the least risk to develop diabetic complications (kidney, liver, and heart). In this study, genetic variants in the *LRMDA* gene specifically increased the risk of developing the MOD subtype.

The fifth subtype, Mild Age-Related Diabetes (MARD), included 39% of the patients. As its name implies, these patients are diagnosed with diabetes at a later age and have relatively good control of their blood sugar. Family history plays a minor role in this subtype and they are at low risk to develop diabetes complications.

This new subclassification of diabetes has identified three high-risk subtypes of diabetes that mandates prompt interventions by the healthcare professionals to design a personalized follow-up and treatment strategy that will improve the patient's quality of life by better control of blood sugar, earlier identification of diabetes complication and implementation of suitable therapies to delay or prevent complications. This new subclassification empowers the development of new medications for diabetes. Let us all together make the world a better place for ourselves and the new generations.

Introduction

Section I: Diabetes

Diabetes is a chronic disease characterized by elevated blood sugar above homeostatic level. The prevalence of diabetes is expected to rise from 8% in 2011 to 10% in 2030 leading to a greater burden on the healthcare budget and society[7-9].

Traditionally, the well-known diabetes types are T1D and T2D, monogenic diabetes, and gestational diabetes[10-12]. The clinical diagnosis is based on the blood level of a single variable, blood glucose. Common risk factors for developing diabetes and diabetes complications, highlighted in many research studies, include family history, body mass index (BMI), age, physical activity and stress[13-37]. Mechanisms underlying the pathogenesis of T1D include reduced insulin secretion due to autoimmune reactions against pancreatic beta cells[38-41]. On the other hand, the interplay between pancreatic insulin secretion and reduced insulin sensitivity, essentially in the liver, skeletal muscles and adipose tissue, play a central role in the pathogenesis of T2D[42-47].

The heterogeneity of T2D is reflected in its pathophysiological complexity, an orchestra of the main body systems and the pancreas, insulin system, and their effect on glucose homeostasis[15,48-52]. The following is a summary of the journey of glucose within the human body in the fed and fasting state. A fascinating tour to get an overview of diabetes.

Glucose transporters

The glucose uptake in body cells is via glucose transporters that are classified into two major families; glucose transporters (GLUTs) and sodium-glucose cotransporters (SGLTs). GLUTs are facilitated transport proteins encoded by the solute carrier family 2 (*SLC2*) genes[53]. SGLTs are encoded by solute carrier family 5 (*SLC5*) genes and are facilitated transporters of both sodium and glucose at the same time. SGLT1, SGLT2 and SGLT3 are expressed in the small intestine, renal tubular cells, and skeletal muscles respectively[54].

Pancreatic beta cells and insulin secretion

Pancreatic beta cells are the only producers of insulin hormones. In obesity, beta cell mass increases by 50%. In T1D, cytokines, and chemokines released by aggravated immune cells in the islets; macrophages, and T-cells accompanied by apoptotic signals, attack the pancreatic beta cells, decreasing the beta cell mass leading to the need for life-long insulin therapy[55-57].

Insulin secretion is an electrophysiological process initiated by beta cell sensors of blood glucose level and the cytoplasmic beta-cell metabolite concentration of glycolytic products; glyceraldehyde 3 phosphate and adenosine triphosphate to adenosine diphosphate (ATP/ADP) ratio. Glucose enters the beta cells by GLUT1 facilitated transport. In the cytoplasm, glucose is anchored by two irreversible phosphorylation steps by glucokinase enzyme (GCK), followed by phosphorylation by phosphofructokinase. These steps maintain the gradient of glucose across the pancreatic beta cell membrane and the cytoplasm enabling the entrance of glucose and, to prevent the glucose from leaving the cell, thus decreasing the blood glucose levels. Closure of the ATP potassium voltage-gated channels in response to ATP/ADP ratio initiates an action potential, where calcium influx through L-type calcium channels initiates the exocytosis of insulin from beta cells[58-62].

1. In the Fed state

Insulin receptor and glucose uptake

Once released in the blood, insulin binds to the transmembrane insulin receptors on the body cells to facilitate glucose uptake[62,63]. The cytoplasmic glucose transporters are translocated to the cell membrane to facilitate insulin-dependent glucose uptake. Net glucose uptake by the main players, liver, brain, red blood cells, skeletal muscles and fat is the rate-limiting factor in the control of the postprandial blood glucose levels. In the case of insulin resistance, the glucose uptake is reduced in the liver, skeletal muscles, and adipose tissues[62,64,65].

Insulin-dependent glucose uptake

In the liver, the net glucose uptake is via GLUT2 and is proportional to the postprandial hyperglycemia and hyperinsulinemia. The latter is greatly influenced by the portal concentration of the lipids and amino acids absorbed from the gut, the hepatic portal signal (sympathetic and parasympathetic, hypothalamic AMPK phosphorylation), and the arterial portal vein glucose gradient[64,66]. The skeletal muscle is the main blood glucose disposal organ. The glucose uptake depends on the translocation of the GLUT4 glucose transporter to the sarcolemma which is

greatly influenced mainly by insulin binding to the muscle insulin receptors at rest and the synergistic effect of muscle contraction during exercise[65,67,68]. The net glucose uptake by skeletal muscles is greatly affected by the net hepatic glucose uptake. In the adipocytes, the glucose uptake is also insulin-dependent via the GLUT4 translocation[44,65,69].

Insulin-independent glucose uptake

Basal glucose supply in most cells is maintained by insulin-dependent glucose uptake via glucose transporters GLUT1, to maintain cell functionality. The expression of GLUT1 depends on the availability of glucose; hypoglycemia increases GLUT1 expression and hyperglycemia decreases its expression. The highest expression is in erythrocytes, endothelial cells, and blood-brain barrier. GLUT1 in the skeletal muscles, maintain the basic needs of glucose during the fasting state. The phosphorylation of GLUT1-glucose by GCK causes feedback inhibition of translocation and glucose uptake by GLUT4. The latter can lead to decreased insulin sensitivity of the GLUT4 glucose uptake pathway[70].

Adipose tissue

The quality and quantity of adipose tissue have an influential effect on energy expenditure and glucose homeostasis. Adipose tissue has three phenotypes: brown, beige and white. The brown (healthy) phenotype is a heat generator characterized by high mitochondrial content and is involved in thermogenesis[71]. On the other hand, the white phenotype is the energy storage (toxic monster) and has differential deposition, subcutaneous under the skin (WAT) and visceral (VAT), and is strongly associated with the development of the metabolic syndrome. The beige phenotype is formed by the browning of white adipose tissue and is embedded in the white adipose tissue[71-74]. White adipose tissue releases adipokines (leptin, adiponectin, resistin), tumor necrosis factor-alpha and vasoactive substance, that mediates the neural, endocrine, vascular and immune effects of the adipocytes. Collectively they play an important role in the development of obesity and diabetes-related micro and macro-vascular complications[73,75-77].

2. In the fasting state

During fasting, endogenous glucose is produced by glycogenolysis and gluconeogenesis from metabolic precursors; deamination of amino acids, glycerol, and Kreb's cycle intermediates in response to glucagon hormone secreted from pancreatic alpha cells in response to low blood glucose levels. Gluconeogenesis takes place in the liver, intestine, kidneys and skeletal muscles[78-80].

The balance between fed and fasting states is crucial in controlling insulin secretion and insulin sensitivity and hence prevention and management of T2D.

Cell bioenergetics

Once glucose enters the cell, glucose metabolism and cell bioenergetics are vital limiting steps in glucose homeostasis. The balance between glycolysis (anaerobic respiration) and oxidative phosphorylation, Kreb's cycle (aerobic respiration), plays an important role in the energy production and output[81,82].

Renal glucose reabsorption

Renal glucose reabsorption (GR) is the final destination of glucose. In healthy individuals, GR is 100% and no glucose is passed in the urine. SGLT2 and SGLT1 transporters expressed in the early and distal renal proximal tubules contribute to 90% and 10% glucose reabsorption respectively and have a significant impact on blood glucose levels[83].

Additional factors

Brain, hormones, and blood composition have a critical effect on insulin secretion and glucose homeostasis include circadian clock and inflammatory responses[52,84-86]. Melatonin hormone secreted from the pineal gland in the brain to regulate the sleep-wake cycle reduces insulin secretion. Nocturnal meal intake is unhealthy due to the inhibition of insulin secretion by elevated melatonin[87,88]. The peripheral clocks of skeletal muscles are greatly affected by physical activity and play an important role in the regulation of blood insulin levels[89-92]. Hypothalamic Pituitary Adrenal (HPA) axis dysregulation and metabolic memory are vital modulators of glucose homeostasis. Adrenal hormones, epinephrine, and cortisol are key players in the HPA[93-95]. Iron deficiency, so-called anemia, decreases the oxygen-carrying capacity of the red blood cells by downregulating the synthesis of hemoglobin (the red pigment in RBCs that combines with oxygen molecules at the alveolar gas exchange surface to form oxy-hemoglobin) leading to a state of hypoxia (decreased oxygen levels in the cells). Moreover, anemia has a negative effect on Kreb's cycle enzymes and the mitochondrial energy production leading to a shift in the balance toward the glycolytic pathway, thus affecting glucose metabolism and homeostasis[96-98].

Insulin resistance, obesity, & metabolic syndrome

Insulin resistance is a progressive condition and the severity depends on the duration of the condition[45]. In the beginning, the glucose uptake decreases due to decreased

insulin sensitivity in the main organs; liver, skeletal muscle, and adipose tissues. This results in elevated post-prandial blood glucose levels, leading to an increase in pancreatic insulin secretion, hyperinsulinemia, and insulin-mediated lipogenesis. Later on, the pancreatic insulin secretion decreases and cannot compensate for the reduced glucose uptake and hyperglycemia leading to decreased insulin-mediated gluconeogenesis suppression, and impaired free fatty acid oxidation in the liver. As a consequence, there will be an increase in the overproduction of hepatic endogenous glucose as well as an increase in ectopic hepatic fat deposition (visceral fat) and development of non-alcoholic fatty liver disease (NAFLD), de novo lipogenesis and dyslipidaemia, leading to the increase in subcutaneous white adipose tissue and development of obesity[4,99]. At this prediabetic stage, metabolic syndrome develops characterized by visceral adipose tissue, insulin resistance, dyslipidemia, and hypertension induce inflammatory responses that counteract the anti-inflammatory role of insulin leading to elevated cytokine secretion and chronic low-grade inflammation[100]. Finally, the pancreatic insulin secretion decreases to an extent where there is persistent hyperglycemia and the development of diabetes. Diabetes accompanied by metabolic syndrome increases the risk for atherosclerosis, macrovascular and microvascular complications of diabetes[6,101,102]. The prevalence of NAFLD is 50-75% in T2D diagnosed by ultrasound[4]. In the case of insulin resistance, bioenergetics' balance is shifted toward the glycolytic pathways leading to less energy production and reactive oxygen species formation (ROS) in macrophages and monocytes promoting inflammation and vascular complications[103-105]. Insulin resistance in skeletal muscle leads to mitochondrial dysfunction and decreased mitochondrial capacity, which increases the glycolytic to the oxidative capacity of the cell and the net energy production is eventually diminished[106,107]. Additionally, insulin resistance causes an increase in the proinflammatory cytokines and their anti-erythropoietic effect leading to apoptosis of immature RBCs and hence decreasing the hemoglobin, causing anaemia[30,108] Hyperinsulinemia increases the expression of SGLT2 in response to the glucose load in the kidneys[83,109,110]. Despite this, in the case of diabetes, glucosuria is manifested.[111] Diabetes and diabetic complications are the major consequences of poor glycemic control.

Glycation

Glycation is the non-enzymatic glycosylation of proteins due to persistent high glucose levels in plasma and in the tissues. The degree and extent of damage differs between cells and proteins. Plasma protein (albumin, globulin, and fibrinogen) modification can have serious effects on platelet aggregation, immune system, and drug plasma protein binding. Advanced glycation end products (AGEs) affect many body parts, including the heart, eye, lungs, brain, kidney, liver, vascular tissue and bones, leading to the development of complications through the cross-linking of the extracellular matrix and by binding to the receptor for AGE (RAGE)[112-114].

Antidiabetic agents

Insulin is a peptide hormone secreted by pancreatic beta cells. Insulin replacement therapy is the first line treatment in T1D, where the pancreatic beta cells are attacked by the immune system. Insulin is dispensed in many formulations that differ in their drug delivery system and duration of action[115].

Sulfonylureas stimulate insulin secretion by blocking the voltage-gated potassium ATP sensitive channels in the pancreatic beta cells, reducing potassium efflux and depolarizing the cell membrane along with an increase in calcium influx, the net of which is an increase in insulin secretion. Sulfonylureas differ in their onset of action and their duration and elimination. An important consideration during sulfonylurea therapy is their elimination route (liver, kidney function) and their increased risk of life-threatening hypoglycaemia[116].

Meglitinides act in the same way as sulfonylureas. They bind to a different site on the ATP-voltage gated potassium channel and enhance insulin secretion but has a weaker effect, hence they do not cause life-threatening hypoglycaemia[117].

Biguanides, available as Metformin, is a non-sulfonylurea and is the most common first-line treatment in T2D alone or in combination with other antidiabetic agents. Metformin does not act on the pancreatic beta cell's insulin secretion, instead, it acts at the insulin binding receptors and thus increases insulin-dependent glucose uptake in the tissues. In the liver, it decreases gluconeogenesis. The dual effect leads to a gradual decrease in blood glucose levels without causing life-threatening hypoglycaemia[118-121].

Alpha-glucosidase inhibitors are competitive and reversible inhibitors of the intestinal alpha-glucosidase enzyme. The alpha-glucosidase enzyme is responsible for carbohydrate digestion in the intestine, so the inhibitors slow the digestion and delay glucose absorption resulting in a gradual rise in postprandial blood levels[122].

Incretin mimetics are incretin hormones, receptor agonists. Incretins are metabolic hormones secreted from the gut in response to glucose. The antihyperglycemic effect of glucagon-like peptide-1 (GLP-1) and gastric inhibitory peptide (GIP) is by their collective effect on pancreatic beta-cell insulin secretion, pancreatic alpha cell glucagon secretion, slowing of gastric emptying rate, and appetite suppression. They are used to treat T2D alone or in combination with other antidiabetic agents[123-129].

Dipeptidyl peptidase 4 (DPP-4) inhibitors are competitive reversible inhibitors of dipeptidyl peptidase enzymes that are responsible for incretin degradation. As a result, they increase incretin levels accompanied by an enhanced antihyperglycemic effect. Food & Drug Administration (FDA) warning of severe joint pain and risk of heart failure are reported side effects[130].

SGLT-2 inhibitors inhibit the renal SGLT-2 proteins responsible for glucose reabsorption, thus decreasing blood glucose levels, increasing glucose elimination

in the urine. They are beta-cell independent. The major side effect is the increased risk of urinary tract infections, fungal infections, and genital infections due to the high glucose content[131].

Thiazolidinediones bind to the adipocyte receptor called the peroxisome proliferator-activated receptor-gamma (PPAR$^{\gamma}$) and promote fat cell maturation and fat deposition into peripheral tissues. They increase insulin sensitivity of the peripheral tissues by decreasing circulating fat. They have protective cardiovascular effects. Their major side effect is a fat deposition and cosmetic weight gain[132,133].

Vitamins are food supplements. Their additional benefit to T2D is controversial, yet some benefits likely exist[134,135].

The potential protective role of Vitamin D in the prediabetic and diabetic phase is published in many research studies[136-149].

Vitamin B6 (pyridoxine) is a co-factor in glucose, lipid, and amino acid metabolism. It has antioxidant effects against ROS and AGEs, and protective effects against progressive diabetic retinopathy.[121,135,150]

Antihypertensives are blood pressure-lowering agents, which vary in their mechanism of action, usually used in combination with antidiabetics in order to get better control over insulin resistance, glucose, and lipid metabolism and manage diabetic complications. ACE-inhibitors are antihypertensive agents used to control blood pressure in patients suffering from diabetic kidney disease[151-155]

Subclassification of diabetes

Diabetes, especially T2D, is a very heterogeneous disease. Diabetes occurs when insulin secretion can no longer compensate for increased insulin requirements due to obesity or insulin resistance. Several pathways can lead to this situation.

Cluster analysis aims to group individuals with similar characteristics into one group using a set of variables. Connectivity models (hierarchical clustering) and centroid models (k-mean) are the most common methods used for cluster analysis[156,157].

In order to identify more homogeneous diabetes groups, Ahlqvist et al. used k-means clustering in the Swedish All New Diabetics In Scania (ANDIS) cohort using six clinical variables: age at diabetes onset, BMI, GADA, HbA1c, insulin secretion estimated as HOMA2B, and insulin sensitivity as HOMA2IR, derived from fasting glucose and C-peptide. The clustering identified five clusters of patients with different clinical characteristics, disease progression and outcomes[1]. These clusters were replicated in three other Swedish cohorts and a Finnish cohort. The clustering has also been replicated in numerous cohorts since.

A cross-validation of ANDIS-subclassification was done by Zaharia et al, in the German Diabetes Study (GDS) to test comprehensive phenotyping at diagnosis and determine the differences in the diabetes-related complications among the clusters during 5-years of follow-up[158]. In an IMI-RHAPSODY study, clustering using age, HbA1c, HDL-cholesterol, BMI, and C-peptide, showed stability of clusters between the two cohorts with the identification of an additional cluster characterized by slow-glycemic deterioration[159]. The ANDIS subclassification was also replicated in different ethnicities; Chinese, Indian, and Mexican[160,161]. More studies are ongoing in different populations.

The response to medication within the clusters was tested in the ADOPT trial. Dennis et al. showed that clusters differed in their glycaemic response and that patients in SIRD could benefit from thiazolidinediones and MARD could benefit from sulfonylureas[162].

Subclassification of diabetes using clustering serves as a valuable approach considering the complex heterogeneity. Clustering diabetes to more homogenous subpopulations that differ in their disease profiles has several advantages. First, to identify patients at high risk of developing diabetes complications. Second, to help optimize the cost-effective allocation of clinical resources to these patients, thus improving patients' health and quality of life as well as complying with the constrained healthcare budget. Another advantage of the subclassification is the increased statistical power for clinical, genetic, and experimental analysis. Clinical parameters that are used for the clustering are easily available and reflect pathogenesis underlying the disease as well as the genetic, environmental exposures, and response to antidiabetics[163].

The following is a brief description of the newly defined diabetes subtypes based on Ahlqvist et al. 2018 (**Figure 1**).

Severe Autoimmune Diabetes (SAID), 6% of adult individuals in ANDIS, includes GADA antibody-positive individuals. Since this project only included adult individuals older than 18 years, juvenile T1D diabetes was not captured. SAID includes both T1D (20%) and LADA (80%). Clinically, individuals are characterized by early disease onset, low insulin secretion, relatively low BMI, and elevated HbA1c levels.

Severe Insulin Deficient Diabetes (SIDD), 18% of adult individuals in ANDIS. Clinically, these individuals do not have GADA antibody, yet are diagnosed by relatively early disease onset, low insulin secretion, relatively low BMI, and elevated HbA1c levels.

Severe Insulin Resistance Diabetes (SIRD), in my opinion the most interesting subtype, included 15% of adult individuals in ANDIS. Clinically, these individuals are characterized by late disease onset, high BMI, relatively low HbA1c levels and insulin resistance.

Mild Obesity related Diabetes (MOD), 22% of adult individuals in ANDIS, are characterized by early disease onset and obesity (BMI >33).

Mild Age-Related Diabetes (MARD), included 39% of adult individuals in ANDIS. Clinically, these individuals present with late-onset diabetes, intermediate insulin secretion, relatively low BMI, and relatively low HbA1c levels.

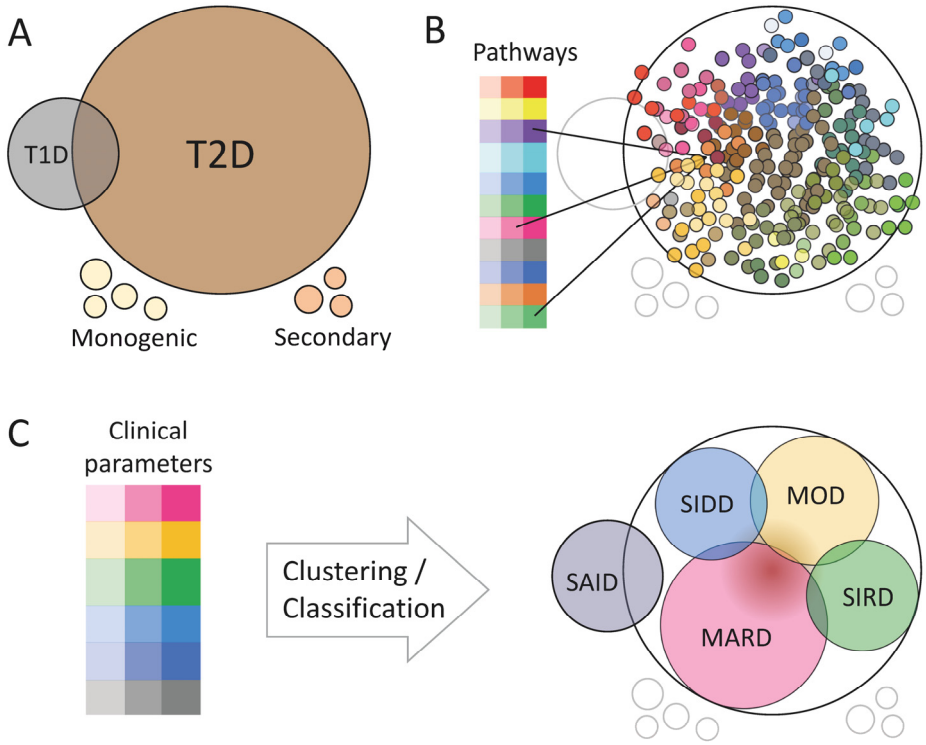

Figure 1. New diabetes subclassification. Panel A shows the classical diabetes types as homogeneous diseases. Panel B shows heterogeneity of T2D with each patient being represented by the most dominant pathways. Panel C shows how clustering using clinical parameters can be used to group patients into groups with different dominant pathways. Modified from Ahlqvist et al, Diabetes, 2020.

Clustering using genetic approaches

Clustering has also been used in other approaches. Udler et al used soft clustering to group known diabetes risk variants by their association with diabetes-related traits. They identified five SNP clusters related to different pathways; beta-cell function, proinsulin, obesity, lipodystrophy and liver-lipid. Genetic risk scores based on these clusters were used to identify patient groups with extreme phenotypes[164]. These clusters were examined for their association with ANDIS subclassification in Paper I of this thesis.

Mahajan et al. also clustered known diabetes loci by function, identifying SNP clusters associated with adiposity, insulin secretion, insulin action, insulin secretion – action combined, and impaired lipids. These clusters were also used to construct risk scores in Paper I to test their association with the ANDIS subtypes[165].

Section II: Genetics of diabetes

Monogenic and complex genetic traits

Monogenic diseases, like MODY and neonatal diabetes, follow traditional Mendelian inheritance; one gene is responsible for a certain trait and is passed to the next generation in accordance with the Mendelian law of inheritance. Complex diseases and traits are caused by interaction between genetic variation in multiple genetic loci and the environment. Diabetes, cardiovascular disease, cancer, and neurological disorders are examples of complex genetic traits[166-168].

The complex heterogeneity of diabetes is attributed to both genetic and environmental influencers. Implementation of multi-omics methods; Genome-wide association studies, polygenic risk scores, epigenomics, candidate gene analysis, single-cell, and tissue transcriptomics as well as proteomics provide a valuable understanding of the heritability, the genetic make-up, and the pathways underlying the pathogenesis of diabetes[169-171].

Genetic variation

Genetic variation is a permanent change in the genetic make-up due to mutations and recombination[172]. Deoxyribonucleic acid (DNA) mutations include point mutations usually referred to as single nucleotide polymorphisms (SNPs), insertions, deletions, duplications, translocations, and inversions. This genetic variation is passed to daughter cells during cell division, mitosis, in case of autosomal genetic variation (somatic mutations) or to the gametes during meiosis, in the case of germline genetic variation. Acquired mutations are caused by environmental factors and are not inherited from the parents[173,174]. Since the release

of the first draft of the human genome sequence in 2003, by the Human Genome Project Consortium, this influential project has opened the gate for the sequencing of the individual's DNA, RNA, and proteins. High through-put sequencing technologies have been implemented that enables the sequencing of many individuals in parallel in a timely, convenient and affordable manner. In the omics era, genomics, epigenomics, and transcriptomics are different methods used to determine the role of genetics in complex genetic traits[172,175].

Genotyping methods

Genotyping is a technique to determine the sequence variation at specific positions of the genome. Genotyping by sequencing technique is commonly used for genome-wide association studies (GWAS) and depends on the hybridization with complementary oligonucleotides. The quality control of the raw genotyped data is a crucial step to ensure the validity of the GWAS results[176].

Heritability

The narrow-sense heritability commonly referred to as just heritability, is used to estimate the proportion of genetic variation that contributes to the phenotype. The missing heritability usually refers to the portion of heritability that could not be captured by the implemented methods, for example, SNP heritability is an estimation of heritability using SNPs. Usually, the set of SNPs is not fully representative of the genetic variation, so the estimated heritability will represent only part of the true genetic variation[177,178].

In the early phase of genetics, the heritability of a given trait was determined using studies conducted on monozygotic identical twins (100% similar genetic-make-up) or dizygotic non-identical twins (50% similar genetic make-up) to study the influence of environment. The concordance is the probability that a pair of twins will have the trait given one has the trait. The discordance means one of the twins does not have the trait, this gives valuable information about the influence of the environment[179,180].

Heritability of T1D is estimated to be >80%, most of which is attributed to the *HLA* locus, while in T2D it is 25-80%[181]. In twin studies, the proband-wise concordance rate for T1D was 23-61% while for T2D it was 17.5% in monozygous twins[182-184].

Linkage analysis

Linkage analysis is based on the principle that genetic sequences that are located together on a chromosome tend to be inherited together and not separated during the homologous recombination in meiosis. If so, these sequences have genetic linkage,

which violates the random assortment assumption for Mendelian law of inheritance. In family pedigrees, when markers co-segregate in a certain phenotype, the markers, and the phenotype are linked or associated[185].

Linkage studies have been successfully used to identify genetic variants causing monogenic diseases, like MODY, but were also used for complex traits before the sequencing of the human genome enabled more powerful methods[168]. In 1996, linkage analysis of T1D showed that major histocompatibility complex loci (*HLA*), located on chromosome 6, are genetic susceptibility loci for T1D[186-188]. For T2D, the calpain-10 gene (*CAPN10*) located on chromosome 2 was identified by genome-wide screening and positional cloning in 2004[189,190]. The now well-known T2D locus transcription factor 7 like 2 (*TCF7L2*), was mapped to chromosome 10 in a Mexican-American population in 1999, fine mapped in the Icelandic population in 2006, and has been replicated numerous times in GWAS of T2D[191,192]. *TCF7L2* is highly conserved among species, plays a fundamental role in the Wnt/β-catenin signaling pathway, and regulates the expression of genes involved in lipid metabolism in adipocytes and glucose-induced insulin exocytosis[193].

Candidate genes studies

Candidate studies are performed for selected genes based on prior information about the gene's effect on the trait of interest. These studies have more statistical power to detect the differences between cases and controls because they only test a small set of genetic variants but are prone to false-positive findings. Another major limitation is the selection bias of genes and pathways of interest[194-196].

Candidate genes studies in diabetes were conducted for beta-cell function affecting genes that had substitutions in the protein-coding regions. Insulin receptor substrate 1 *(IRS1),* Peroxisome proliferator-activated receptor gamma *(PPARG),* and insulin receptor substrate 2 (*IRS2*)*,* Wolfram syndrome 1 *(wolframin) (WFS1),* potassium inwardly-rectifying channel, subfamily J, member 11 *(KCNJ11),* HNF1 homeobox A *(HNF1A),* HNF1 homeobox B *(HNF1B)* and *HNF4A,* their association with T2D was identified by candidate gene studies[194,197].

Genome-wide association studies (GWAS)

GWASs determine the genetic variation across the entire genome in the form of SNPs and compare the frequency of variants e.g. in cases and controls. Usually, GWAS is performed using large sample sizes. Successful GWAS depends on cohort selection, quality of genotyping, and imputation of the genetic data.

In 2007, GWAS caused a revolution in the field of complex genetics. The first GWAS studies of T2D identified haematopoietically expressed homeobox (*HHEX*)*,* solute carrier family 30 member 8 (*SLC30A8*)*,* cyclin-dependent kinase inhibitor

2A/2B (*CDKN2A/2B*), insulin-like growth factor 2 mRNA binding protein 2 (*IGF2BP2*), CDK5 regulatory subunit associated protein 1 like 1 (*CDKAL1*) and FTO alpha-ketoglutarate dependent dioxygenase (*FTO*)[198-202].

Following the GWAS revolution, the genetic architecture of T1D and T2D was determined in many populations with different ethinicities[203-212]. Large consortia were constructed to increase sample size and perform meta-analysis; DIAbetes Genetics Replication And Meta-analysis (DIAGRAM), Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC), SUrrogate Markers for Micro- and macro-vascular hard endpoints for Innovative diabetes Tools (SUMMIT) and DIAbetes Meta-ANalysis of Trans-Ethnic association studies (DIAMANTE) are some important examples[165,213-215]. GWAS of T1D has identified >50 loci associated with T1D, revealed pathways underlying disease, and overlap of pathways with autoimmune diseases[181]. GWAS in T2D are of low predictive value yet give valuable information about the mechanisms underlying the diseases.

One of the major meta-analysis studies was published by Mahajan et al. in 2018, in Nature genetics[165]. The study identified 403 T2D associated loci that have been used and investigated in this thesis.

Open access GWAS results repositories are available to get information about variants and their associations with T2D, e.g. T2D knowledge portal, GWAS catalog, PhenScanner, and open Target Genetics[216,217].

One of the challenges of GWAS is that identified variants that are associated with the trait, may not be the causal variants but rather in linkage disequilibrium with the causal variants. Further, it is often difficult to determine how the variant affects the trait as the variant is often outside coding regions and could affect genes and regulatory elements at a distance. Expression data, eQTLs, and Mendelian randomization are tools for inference of SNP function.

In the omics era, there are many facilities to implement in the research of diabetes. The incorporation of genetics and proteomics facilitated the discovery of new biomarkers to aid the screening and diagnosis of T2D, follistatin, and osteopontin[218,219]. Recently, the development of new antidiabetic agents is based on the findings of the omics research[220].

Epigenetics

Initial epigenetic studies of human pancreatic islet candidate genes for T2D, such as *INS, PDX1, PPARGC1A*, and *GLP1R,* found that hyperglycaemic induced DNA methylation decreased the differential expression of these genes leading to impaired insulin secretion. Recently, findings in the islet's epigenome revealed differential DNA methylation of CpG active sites in GWAS identified genes associated with T2D: *ADCY5, FTO, HHEX, IRS1, KCNQ1, PPARG, and TCF7L2*[221,222]. Genome-

wide epigenetic studies in T2D revealed the modifications of genes affected by HbA1c, BMI, age, adipose tissue, and exercise[223-236]. Transcriptomics and metabolomics studies in diabetes bridge the gap between GWAS and clinical studies. Using RNAseq data from diabetes donors and non-diabetic controls, revealed differential expression of transcription regulators of genes affecting adipogenesis and some underlying diabetes pathways[237].

Genetics of novel diabetes subtypes

In the original paper by Ahlqvist et al., I performed a genetic association analysis s of ANDIS individuals clustered into the newly defined diabetes subtypes for a designed panel of 172 candidate SNPs. The results illustrated in the Venn diagram represent the nearest gene of top associated variant's with each diabetes subtypes ($p<0.01$). Strikingly, there were no genes associated in all the subtypes (**Figure 2**).

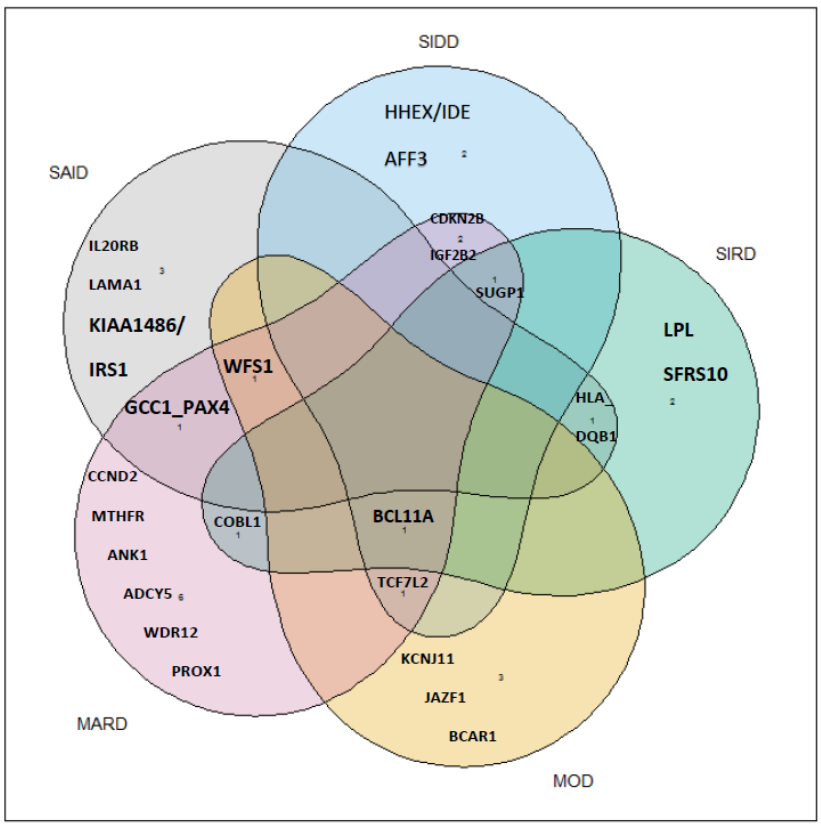

Figure 2. Venn diagram illustrating the T2D associated genes and the top associations with the newly defined diabetes subtypes.

Section III: Diabetic kidney disease

Diabetic kidney disease (DKD), presenting as chronic kidney disease (CKD) often accompanied by albuminuria, is a major complication in both T1D and T2D. Even though DKD in T1D differs from T2D in the onset and prevalence, they share some common clinical presentations[238].

Kidney and nephron

Nephrons form the foundation of the kidney, each consisting of the bowman capsule, proximal tubules, loop of Henle, distal convoluted tubules, and the collecting duct[239]. Glomerular filtration takes place in the renal cortical nephron through the glomerular filtration barrier (GFB)[240,241]. Healthy GFB consists of podocytes; intact glomerular filtration membrane, endothelium, and capillary lumen. Podocytes are responsible for the formation of the glomerular basement membrane (GBM) components (**Figure 3**). They are fully differentiated visceral epithelial cells with special cytoplasm foot processes (pedicels). Adjacent pedicels form the filtration slits. Nephrin, an immunoglobulin cell adhesion molecule in cooperation with podocyte actin forms the backbone of the slit diaphragm and determines the slit size[241]. An intact GFB allows the blood from the renal artery to pass through the bowman's capsule, filtering large protein molecules (mainly albumin). Then the filtrate passes to proximal tubules, where reabsorption of glucose and electrolytes takes place. In the case of diabetes, GFB changes including thickening of the basement membrane, podocyte effacement, and apoptosis, lead to the albumin leak seen in patients with albuminuria[242].

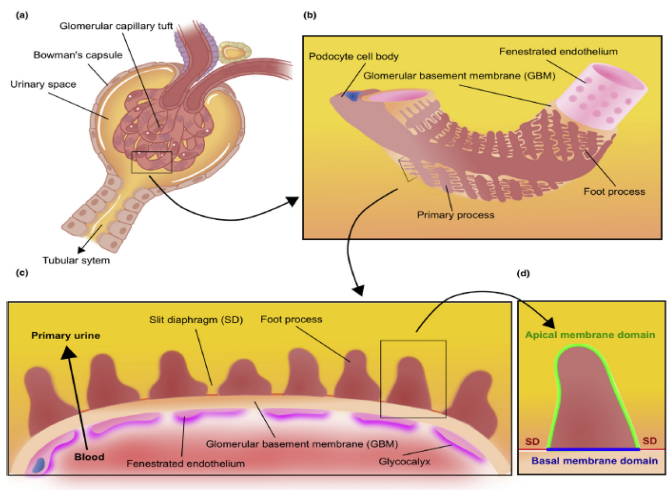

Figure 3. Glomerular filtration barrier (adapted from Trimarchi et al.[243].

Stages of DKD

Stage I starts around 5 years from the onset of diabetes and is characterized by normal or increased eGFR, no albuminuria, normal blood pressure, 20% increase in the size of the kidney (nephromegaly), and a 10-15% increase in the renal plasma flow. Stage II is characterized by the thickening of basement membrane and mesangial proliferation, eGFR is normalized and no clinical signs of DKD. However, stage III usually occurs 5-10 years after the DKD onset and is characterized by a decline in eGFR accompanied by microalbuminuria (albumin 30-300 mg/day) with or without hypertension. Stage IV is characterized by irreversible proteinuria (>300 mg/day), decreased eGFR <60 mL/min/1.73 m^2, and sustained hypertension. Stage V, end-stage renal disease (ESRD), is characterized by eGFR <15 mL/min/1.73 m^2. At this stage, patients require renal replacement therapy (peritoneal dialysis, hemodialysis, or kidney transplantation)[238,244-246].

Histopathology of DKD in T1D and T2D

Up to date, there is no recommendation of kidney biopsies for patients with DKD. Strong recommendations are that the biopsies if taken, should be taken with care at professional centers. Renal biopsies show that in T1D, thickening of the GBM occurs in the early stages followed by podocyte decrease in number in later stages. On the other hand, in T2D, a decrease in podocyte number takes place earlier than albuminuria, and the histological lesions are highly heterogenous[247].

Genetics of diabetic kidney disease

DKD is a multi-phenotypic trait based on phenome-wide association studies and augmented by epigenetic studies[248]. Genetic studies of CKD in diabetes-free individuals have identified 150 genes. GWAS of kidney disease; CKD was performed in many populations. *PRKAG2, ANXA9, DAB2, SHROOM3, DACH1, STC1, SLC34A1, ALMS1/NAT8, UBE2Q2, GCKR, AGTR1, CNDP1* have published loci that are associated with kidney function[249,250].

GWAS has over the past 10 years identified 33 DKD associated genes: *ABCG2, AFF3, AGER, APOL1, AUH, CARS, CERS2, CDCA7/SP3, CHN2, CNDP1, ELMO1, ERBB4, FRMD3, GCKR, GLRA3, KNG1, LIMK2, MMP9, NMUR2, MSRB3/HMGA2, MYH9, PVT1, RAET1L, RGMA/MCTP2, RPS12, SASH1, SCAF8/CNKSR3, SHROOM3, SLC12A3, SORBS1, TMPO, UMOD,* and *ZMIZ1*[246,251-261]. Yet these genes need to be further replicated. *UMOD* is a well-established GWAS significant loci associated with DKD[249,262,263].

Aim of the studies

Paper I

To elucidate to what extent the new subclassification of diabetes results in etiologically distinct subtypes with different genetic risk profiles by performing GWAS and restricted genetic risk scores analysis**.**

To identify new loci affecting subtype-specific disease pathways based on the new subclassification of T2D.

Paper II

To study the prevalence of new subtypes of diabetes within the ANDIS framework and assess the risk of diabetic macro-and microvascular complications in Iraqi immigrants and native Swedes.

Paper III

To identify genetic variants associated with kidney complications and to compare genetic associations in the new subtypes of diabetes to determine if the underlying mechanisms differ.

Paper IV

To analyse the proteomic profiles of the GADA negative subtypes (SIDD, SIRD, MOD and MARD) to investigate to what extent these four groups can be described by their proteomic profile and to understand the underlying pathogenesis.

Study populations

The following is a brief description of the four study cohorts included in this project.

All New Diabetics In Scania (ANDIS)

The ANDIS project was initiated by Professor Leif Groop in 2008. This project aims to recruit all incident cases of diabetes within the Scania (Skåne) County in southern Sweden (1,200,000 inhabitants). In this thesis, we have used individuals recruited from the period 2008 until 2016. In this time, 177 clinics registered 14,625 patients aged 0-96 years, within a median of 40 days (IQR 12-99) after diagnosis. Individuals were assigned to clusters/subtypes as previously described[1]. Patient characteristics are in **Table 1**.

Table S1. Characteristics of patients in ANDIS included in the genetic studies

ANDIS	SAID	SIDD	SIRD	MOD	MARD	T2D
N total	452	1193	1130	1374	2861	9486
N (Males/fFemales)	285/167	951/242	732/398	917/457	1959/902	7534/1952
Age, diagnosis (years)	51.2(18.0)	57.5(11.0)	66.0(9.4)	49.8(9.5)	68.0(8.6)	61.1(13.3)
Age, onset of CKD*	57.6(18.1)	63.5(11.2)	71.4(9.4)	54.8(9.7)	73.8(8.6)	68.4(11.8)
Mean Duration*	6.58 (2.63)	6.2(2.64)	5.79(2.71)	6.17(2.4)	6.16(2.55)	6.11(2.53)
BMI (Kg/m^2)*	27.4(6.4)	28.9(4.8)	33.8(5.3)	35.7(5.4)	28.0(1.4)	30.5(5.9)
HOMA2B*	55.74(42.1)	47.7(28.9)	150.6(47.6)	94.92(32.6)	86.7(26.5)	87.9(49.7)
HOMA2IR*	2.13(1.49)	3.18(1.72)	5.52(2.71)	3.34(1.18)	2.55(0.84)	3.43(4.76)
HbA1c (mmol/l)	79.7(31.1)	101.8(19.3)	54.0(15.3)	57.8(15.9)	50.1(9.9)	64.0(25.1)
eGFR>60	396 (87.80%)	1002 (84.27%)	796 (70.44%)	1290 (93.95%)	2285 (79.87%)	7841 (82.66%)
eGFR<60	50 (11.09%)	178 (14.97%)	320 (28.32%)	72 (5.24%)	545 (19.05%)	1645 (17.34%)
eGFR<45	24 (5.32%)	81 (6.81%)	146 (12.92%)	26 (1.89%)	219 (7.65%)	710 (7.48%)
eGFR<15	6 (1.33%)	32 (2.69%)	46 (4.07%)	10 (0.73%)	42 (1.47%)	209 (2.20%)
Micro-albuminuria	36 (7.96%)	111 (9.30%)	131 (11.59%)	109 (70.93%)	229 (8.00%)	847 (8.92%)
Macro-albuminuria	13 (2.88%)	41 (3.44%)	52 (4.60%)	29 (2.11%)	46 (1.61%)	249 (2.62%)

*All reported values mean(SD).

Diabetes Registry Vasa (DIREVA)

The DIREVA is a Finnish cohort from Western Finland (~170,000 inhabitants) that includes 5107 individuals with diabetes recruited 2009-2014 in the Vaasa hospital district. Patients' characteristics are in **Table 2.**

Table 2. Characteristics of patients in DIREVA included in the genetic studies

DIREVA	SAID	SIDD	SIRD	MOD	MARD	T2D
N	327	394	453	596	1178	3453
Frequency (%)	11.09	13.37	15.34	20.22	39.96	
N Men	175	265	235	300	703	1949
Men (%)	53.50	67.26	51.87	50.33	59.67	56.44
HBA1C (mmol/l)*	60.31 (18.52)	76.55 (19.11)	46.77 (8.57)	49.74 (10.42)	45.30 (6.42)	52.93 (16.13)
BMI (Kg/m2)*	28.52 (5.43)	28.8 (4.66)	32.47 (4.80)	35.77 (5.52)	27.86 (3.36)	30.17 (5.48)
Age, diagnosis (years)*	45.69 (15.65)	48.48 (13.42)	61.98 (8.78)	47.57 (9.58)	63.48 (8.66)	53.87 (17.39)
HOMA2B*	38.21 (42.53)	29.79 (21.90)	120.18 (39.10)	62.74 (25.50)	61.76 (22.50)	64.12 (39.63)
HOMA2IR*	1.25 (1.69)	1.49 (1.22)	4.17 (2.19)	2.02 (0.98)	1.62 (0.70)	2.09 (2.86)
eGFR > 60ml/min	255 (77.98%)	279 (70.81%)	251 (55.40%)	488 (81.88%)	801 (67.89%)	2524 (73.10%)
eGFR <60ml/min	72 (22.02%)	115 (29.18%)	202 (44.59%)	107 (17.95%)	376 (31.92%)	922 (26.70%)
eGFR < 15ml/min (ESRD)	17 (5.19%)	21 (5.33%)	35 (7.73%)	20 (3.36%)	20 (1.69%)	132 (3.82%)

*All values are represented as mean(SD).

Malmö Diet and Cancer (MDC)

The MDC cohort started in the early 1990s and aimed to screen middle-aged individuals (born between 1923 and 1950) from Malmö, Sweden, to examine the effect of diet on cancer incidence. Cardiovascular risk factors were measured in about 6000 individuals. Diabetes-free individuals (n=2744) from the MDC cardiovascular arm (MDC-CVA) re-examination cohort (age 61-85) were used as controls in genetic analyses in this thesis. Patient characteristics are in **Table 3**.

Table S3. Patient characteristics of Malmö Diet and Cancer (MDC), diabetes-free controls for ANDIS

MDC	Diabetes free individuals
N (Male/Female)	2744 (997/1746)
Age (years)	71.69 (5.24)
BMI (Kg/m^2)	26.91 (4.47)
Fasting Glucose (mmol/l)	6.16(2.022)

Botnia

The Botnia study has recruited patients with T2D and their family members in the area of five primary health care centers in Western Finland since 1990[36]. Unrelated (based on estimated genetic relationships) diabetes-free individuals were used as controls for DIREVA. Patient characteristics are in **Table 4**.

Table S4. Patient characteristics Botnia, diabetes-free controls for DIREVA

Botnia	Diabetes free individuals
N (Male/Female)	1683 (941/742)
Age (years)	55.95 (10.49)
BMI (Kg/m^2)	26.62 (4.05)
Fasting Glucose (mmol/l)	5.57 (0.58)

Methodology

Genotyping

ANDIS and DIREVA were genotyped with InfiniumCoreExome-24v1-1 BeadChip arrays (Illumina, San Diego, CA, USA), at Lund University Diabetes Centre, Malmö, Sweden. MDC was genotyped at the Broad genotyping facility using Infinium OmniExpressExome-8 version 1.0 BeadChip arrays (Illumina, San Diego, CA, USA). Botnia were genotyped using Illumina Global Screening array-24v1 at Regeneron Pharmaceuticals Inc, NY, US.

Power Calculations

The power to detect genetic associations depends on the risk allele frequency, the magnitude of the genetic risk (i.e. effect size), the type 1 error rate, and imputation quality, and the sample size. In Paper I, non-centrality parameter calculations was based on double genomic controlled standard error estimates from the additive model meta-analysis; these estimates integrate information on allele frequency, imputation quality, and sample size, which typically vary across studies. The type 1 error was set at 5×10^{-08} and an additive risk model was assumed.

Quality control of the genetic data

Quality control for the individuals and the markers ensures the validity of the data analysis results. The quality control is performed using a generalized approved protocol [264]. The protocol adjusts for individuals (sample) and markers (SNPs). For individuals, the adjustment is based on sex, relatedness, and population stratification, and for markers based on genotyping call rate, minor allele frequency, and Hardy-Weinberg equilibrium. The QC protocol uses the PLINK platform and is run in the form of a pipeline with adjusted parameters. Samples were excluded if ambiguous gender, call rate < 95%, and any duplicate or related individuals (pi_hat ≥ 0.2). SNPs were excluded if monomorphic SNPs, SNPs with MAF < 0.05, and SNPs with missingness rate > 0.05.

Since the genotyping of ANDIS and MDC was done at a different time and using different arrays, genotypes from the ANDIS (12770 individuals) and MDC (3344 individuals) cohorts genetic were merged using PLINK,[265] including only SNPs present on both genotyping arrays. After the ANDIS-MDC merge, 16804 individuals and 324063 SNPs passed quality control (QC).

Imputation of the genetic data

Imputation is the inference of the haplotype of non-genotyped SNPs based on alignment with a reference genome. One of the most important GWAS limitations is the sample size. The imputation method increases the number of SNPs beyond those that were genotyped, thus increasing the resolution and power of GWAS in detecting true associations. The genotype imputation quality depends on the software used, reference genome selection, SNP density of the dataset, number of samples, and the sequencing coverage[266]. In ANDIS-MDC merged dataset, autosomal chromosomes, data files were submitted to the Haplotype Reference Consortium (HRC) Michigan server in the form of variant call files (VCF) after passing QC[267].

X-chromosome imputation

The X-chromosome imputation was problematic as the HRC imputation server imputed only the autosomal chromosomes in 2017. MINIMAC3 implementation based on the protocol stated by the software was used for imputation[267,268].

Study design and phenotypes

Paper I

In Paper I, we used a case-control study design where the newly defined subtypes for ANDIS and DIREVA were used as cases for the genetic analysis:

Severe Autoimmune Diabetes (SAID, N_{ANDIS}=452, N_{DIREVA}=327),

Severe Insulin Deficient Diabetes (SIDD, N_{ANDIS}=1193, N_{DIREVA}=394),

Severe Insulin Resistant Diabetes (SIRD, N_{ANDIS}=1130, N_{DIREVA}=453),

Mild Obesity-Related Diabetes (MOD, N_{ANDIS}=1374, N_{DIREVA}=596),

Mild Age-Related Diabetes (MARD, N_{ANDIS}=2861, N_{DIREVA}=1178).

The controls were non-diabetic individuals from MDC (N=2744) and Botnia (N=1683) for ANDIS and DIREVA, respectively.

The analysis included SNP heritability, GWAS, GRS, and LD score regression. Discovery analysis was done in ANDIS-MDC followed by replication in DIREVA-Botnia and meta-analysis.

Heritability was studied using questionnaire data about family history of diabetes from ANDIS and ESTRID, a substudy of ANDIS, as well as SNP-based heritability.

For the GRS analysis, trait GRS was created using published genome-wide significant SNPs.

A flow chart of study design and methods is found in **Figure 4.**

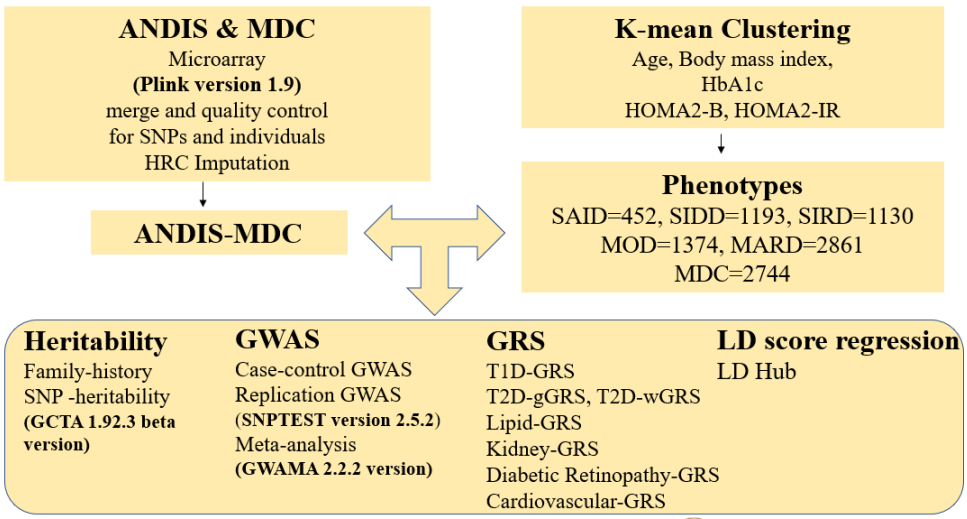

Figure 4. Paper I flow chart.

Paper II

In Paper II, Iraqi individuals from ANDIS (N=286) were used as cases, and Swedish individuals (N=10641) were used as controls. Only GADA-negative individuals were included. Prevalence of subtypes and diabetic complications defined by ICD10 codes were compared between the cases and controls.

The analysis included cluster analysis, risk of complications, and genetic risk scores. For the clustering analysis, Ahlqvist clustering was applied to 286 Iraqi individuals, and the distribution within the subtypes was reported. The risk for developing complications, including coronary events, CKD, stroke, and diabetic retinopathy

were compared using Cox-regression. GRS was designed as in Paper I for traits; T2D, insulin sensitivity, insulin secretion, and BMI.

A flow chart of study design and methods is found in **Figure 5.**

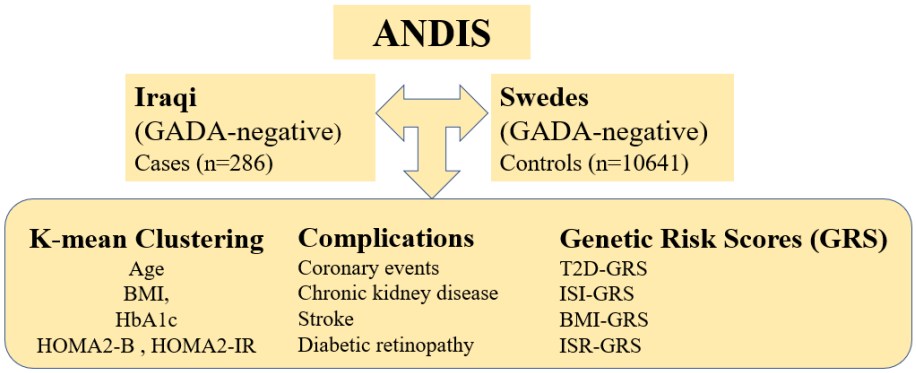

Figure 5. Paper II flow chart.

Paper III

In Paper III, the geometric mean of eGFR in the last year of follow-up was used as the primary phenotype (quantitative trait) for GWAS analysis due to the relatively low number of individuals with CKD and albuminuria. The secondary phenotypes were CKD60 (eGFR less than 60 ml/min/1.73m^2), CKD45 (<45ml/min/1.73m^2) and ESRD as binary traits. Measures of eGFR and albuminuria were calculated from data collected from the Skåne Clinical Chemistry database. The eGFR was calculated using serum creatinine (in mg/dL) as an input for the MDRD formula, GFR = 186 × Serum Cr$^{-1.154}$ × age$^{-0.203}$ × 0.742 (if female).

The analysis included the prevalence of CKD, GWAS of eGFR, association analysis with CKD and GRS. The GWAS was performed using primary phenotype for each diabetes subtype in ANDIS and DIREVA followed by meta-analysis. Association analysis with secondary phenotypes swas performed for significant variants in the subtypes in ANDIS. The GRS analysis for the traits; kidney (CKD, eGFR, and UACR), T1D-DKD, T2D-DKD were designed from the published genome-wide SNPs associated with the traits.

A flow chart of study design and methods is found in **Figure 6.**

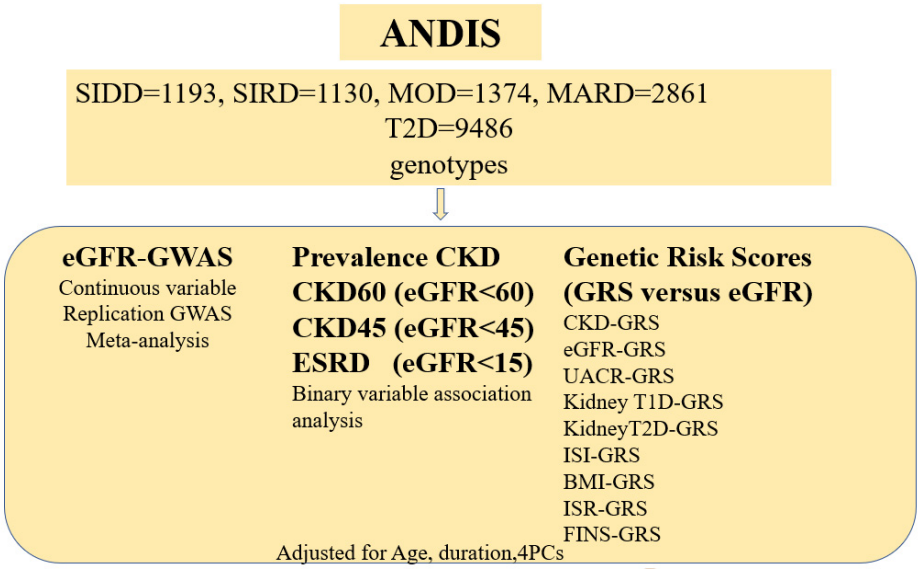

Figure 6. Paper III flow chart.

Paper IV

In Paper IV, 176 individuals were selected from SIDD, SIRD, MOD, and MARD to represent their subtypes based on Euclidian distance to their cluster center. Equal sample sizes were selected from males and females. All selected individuals were of European origin and were born in Sweden. The concentration of 1161 protein biomarkers was measured in the blood samples using Olink panels.
Generalized linear models adjusted for covariates were performed for biomarkers in each subtype to identify differential biomarkers. Tree-based machine learning algorithms were applied to identify biomarkers that can fine-tune the clustering analysis. A flow chart of study design and methods is found in **Figure 7.**

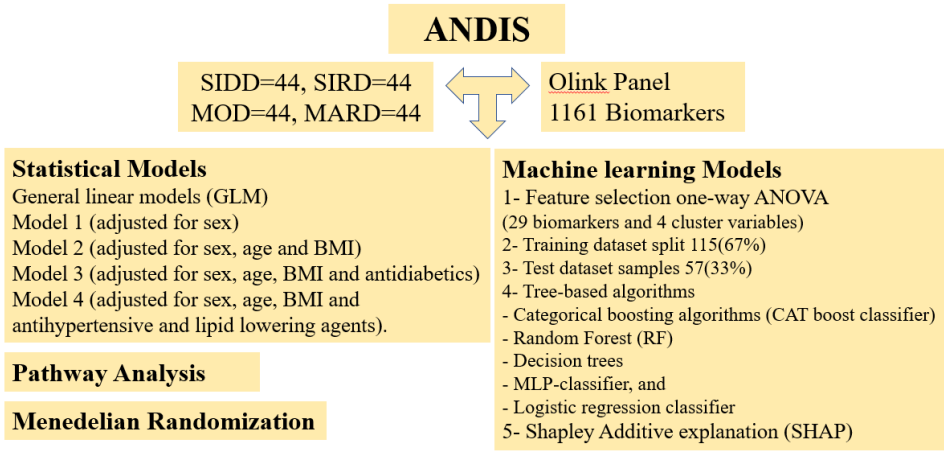

Figure 7. Paper IV flow chart.

Analysis of genetic data

Genetic datasets are a rich source of information that empowers the development of new approaches for screening, diagnosis, and management of complex genetic traits. There is an array of methods used to analyze genetic data depending on the research questions and they output valuable information about the complex genetic trait. This section describes in brief the principles, software, and output of the methods used in this thesis.

Heritability

In Paper I, heritability was performed to determine to what extent the newly defined diabetes subtypes were inherited, the percentage of the genetic component that affects the complex genetic trait. The heritability analysis was done by two approaches; self-reported family history of diabetes from questionnaires and the SNP heritability calculated using GCTA 1.92.3 beta version software[269].

The principle of the GCTA software is to estimate the proportion of genetic variation of the complex genetic traits by using the restricted maximum likelihood method based on the genetic relation matrix (GRM) calculated using the SNPs of all individuals included in the analysis.

Genetic risk scores

GRS also called polygenic risk scores, is a method to predict each individual's risk for a certain trait. GRS was estimated based on GWAS significant trait-associated SNPs weighted by their effect size published in the GWAS catalog[217]. The analysis aimed to determine the genetic risk of individuals in each subtype for the trait of interest. The GRSs were estimated for individuals within each diabetes subtype and non-diabetic controls. Logistic regression was used to determine the association of the trait-GRS and each diabetes subtype versus non-diabetic controls. GRS was calculated using PLINK[270] software. Association with the traits was performed using logistic regression in R. In Paper I, logistic regression was run for traits-GRS for each diabetes subtype versus non-diabetic controls in both ANDIS and DIREVA. In Paper II, the logistic regression was performed for the traits-GRS of Iraqi cases versus Swedish controls. In Paper III, the logistic regression was performed for traits-GRS versus eGFR within each diabetes subtype.

Genome wide association analysis

In this thesis, SNPTEST version 2.5.2[170] was used for the GWAS analysis assuming an additive model and the method used was the "score"; likelihood score test. This estimates the likelihood function for each subtype under the null hypothesis[271].

The effect size is estimated by the beta; an estimate of the increase in log-odds that can be attributed to each copy of the effect allele (allele B). The association of the variant is described as genome-wide significant if the association p-value is below $5x10^{-08}$ and suggestive if the p-value is $<10^{-05}$.

In Paper I, the case-control GWAS model was adjusted for sex and the first four principal components (PCs) calculated in the QC.

In Paper III, GWAS was performed for eGFR as a quantitative trait in all GADA-negative individuals and the newly defined subtypes separately. The model was adjusted for sex, age at onset of diabetes diagnosis, duration, and the first four PCs. Adjusting for BMI, HbA1c and HOMA2IR were done to test the effect of these covariates on eGFR. In ANDIS, association analysis for the binary traits; CKD60, CKD45 were performed for the findings of the former GWAS of CKD (quantitative trait) in T2D, SIDD, MOD, and MARD and ESRD in SIRD only.

Replication of the GWAS results

A major step in GWAS is the reproducibility of the genotype-phenotype association results. Replication using different study designs and populations ensures that the association is true and not by chance or artefact in the genetic data. Optimal replication cohorts have the minimal heterogeneity tested by Cochran's Q test of

homogeneity. Being of Scandinavian European origin, DIREVA and Botnia were the convenient replication cohorts[272].

In Paper I, the phenotypes used were the newly defined subtypes as in ANDIS. Case-control GWAS was performed using DIREVA-Botnia individuals

In Paper III, in DIREVA, the geometric mean of eGFR in the last year's follow-up was used as a quantitative trait as in ANDIS. In DIREVA, the geometric mean was normally distributed so was used without transformation. GWAS was done in T2D and the newly defined diabetes subtypes.

A flowchart of replication is in **Figure 8.**

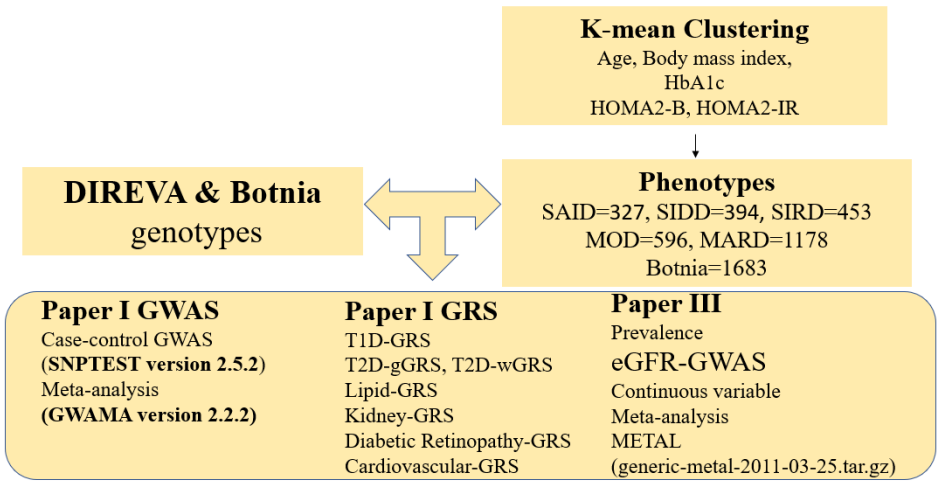

Figure 8. Replication cohort analysis flow chart.

Meta-analysis

For GWAS to detect common variant associations with small effect sizes, large sample sizes are required and this aim is often not attainable in a single cohort. Instead, meta-analysis using the results of more than one GWAS is performed to increase the sample size and hence the power to detect true association[273].

In Paper I, meta-analysis was performed for pairwise GWAS of both ANDIS and DIREVA respectively. Meta-analysis was performed based on effect sizes, using genome-wide association meta-analysis (GWAMA version 2.2.2)[274] software program. In Paper III, since the phenotype eGFR was on different scales; log-transformed in ANDIS and not transformed in DIREVA, meta-analysis was performed using weighted p-value based analysis for the GWAS results for eGFR as a quantitative trait) using METAL (generic-metal-2011-03-25.tar.gz) [275-277].

Post-GWAS quality control and result visualization

In Paper I and Paper III, the GWAS results were quality controlled based on the imputation score, Hardy-Weinberg, and allele frequency using PERL scripts. Genomic control was performed by calculating the inflation factor; lambda for the p-values of each subtype using the "GenABEL" package in R[278]. Manhattan plots and Q-Q plots were generated in R using the "QQMAN" package to illustrate the results of the twenty-two autosomal chromosomes and highlight the GWAS significant SNPs and suggestive SNPs. Locus zoom plots for the chromosome region +/- 400KB, were done for the GWAS significant SNPs in the open web interface for "LocusZoom" [279].

Post-GWAS variant interpretation

In Paper I and III, variants that reached genome-wide significance level and suggestive level were further interpreted. To get information on the variant gene and nearby gene, many available open access databases can be used for the variant interpretation. Locus zoom plots give information about the genes within the region. The Variant Effect Predictor (VEP)[280], the open web interface implemented by Ensembl gives information about the variant (SNP), type of variant either upstream, downstream, intron variant, type of the mutation; missense variants, stop variants, mapped gene, and the variant deleteriousness. The Integrative Genome Viewer (IGV)[281,282], is an open web interface that gives information about the mapped gene, and by scrolling the slider you can also identify the nearest gene. Following variant annotation, using the National Center of Biotechnology Information (NCBI), a thorough study of the genes in the region, their coding protein function, tissue expression levels, and publications about the genes. The Genotype-Tissue Expression (GTEx) portal gives the expression quantitative loci (eQTL) for the variants of interest[283]. T2D knowledge portal[284], GWAS catalog[217], and UK Biobank[285] give information about the association of the variant in published GWAS studies.

Linkage Disequilibrium Score Regression analysis (LDscore)

In GWAS, the test statistic could be inflated either due to polygenicity (true finding) or population stratification (False positive due to confounding). LD score regression method is used to differentiate between true findings and false positive, by regressing the standardized test statistic versus the LD score, the greater the correlation between test statistics and LD score the more likely it is true finding. For trait association analysis, two traits are used; trait 1 (newly defined diabetes subtype) and trait 2 (LDHub trait), the SNPs selected are known to be associated with trait 2,

so if the trait1 test statistics showed correlation with the LD score of the trait-associated SNPs, then the two traits are correlated[178,286].

In Paper I, this method was used to determining the association between newly defined diabetes subtypes and traits of interest. The analysis was performed in the open web interface LDHub[287] using LDHub SNPs and traits. LDHub has selected trait-associated SNPs for which the LD score was calculated. The main assumption is that the population used to estimate LD scores for the SNPs should be matched with the population of the GWAS. The European population was used for the analysis.

Machine learning Models (ML)

Machine learning models are algorithms built to make predictions or decisions. Unsupervised machine learning makes predictions without prior information about the targets. Supervised machine learning (SML) uses information given about features in the training dataset and predicts outcomes about the targets for a test dataset. Support vector machines (SVM), logistic regression, naïve Bayes, K-nearest neighbors, decision trees, random forests are examples of SML that are used to analyze data for classification and regression analysis. A linear classifier performs linear combination analysis using the features to make classification decisions. Gradient boosting is the implemented technique in decision trees. Usually, combinations of classifiers are used to increase the accuracy of the prediction[288-291].

In Paper IV, the input dataset included both clinical observations and NPX data that was evaluated for outliers using an unsupervised clustering algorithm, One-Class Support Vector Machine (OC-SVM). Recursive feature selection for the classification model was performed by applying a one-way ANOVA F-test (Bonferroni corrected; $p<4.191e-05$) to the standardized data. Features selected for classification were 33 biomarkers and 4 minimal cluster variables (age at diagnosis, BMI, HbA1c, and HOMA2-IR). The main dataset (176 samples and 1161 biomarkers) was split into a training dataset (67% of the samples) and a cross-validation test dataset (33% of the samples), these were used to train and evaluate the performance of the tree-based algorithms. The Shapley Additive explanation package (SHAP) was used to interpret the models. The CAT boost classifier performed best, so the output was used for downstream analysis[292]. The entire analysis was performed using packages Catboost, shap, pandas, sklearn, seaborn, and matplotlib in Python 3.6.9 along with the OlinkAnalyze package R version 4.0.2 for data pre-processing and normalization.

Pathway analysis

Pathway analysis gives information about the pathways genes and biomarkers are involved in and thus reflects the importance of the pathogenesis of the complex genetic trait. The gene and tissue enrichment analysis of the variants of interest is performed before the pathway analysis. There are open-source pathway databases: Gene Ontology[293], KEGG[294], STRING[295], PANTHER[169], REACTOME[296], Elsevier database[297,298], and BIOCARTA[299]. Pathway scoring algorithm (Pascal)[300], computes gene and pathway scores from summary statistics of the SNP-phenotype association. In Paper I, the DEPICT[301] software was used to annotating the SNPs of interest. Pathway analysis was performed in REACTOME[302]. In Paper IV, differentially expressed biomarkers were used for the protein-protein interactions and pathway enrichment analysis using STRING, KEGG, and REACTOME. Elsevier database was used to identify related phenotypes and diseases.

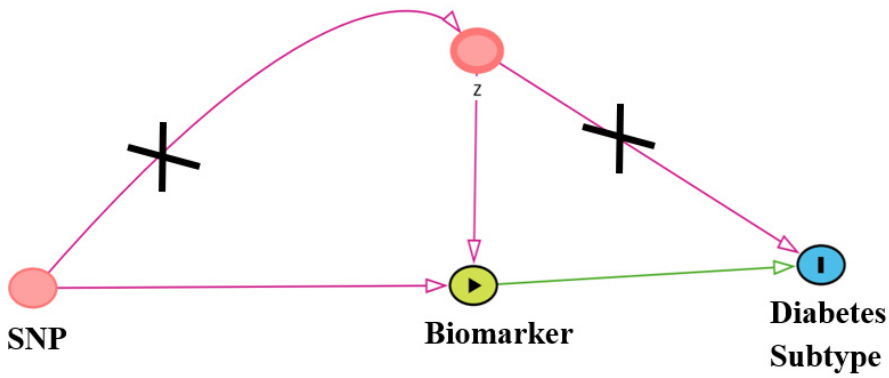

Figure 9. Mendelian randomization principle (developed by DAGitty[303])

Mendelian Randomization (MR)

MR is a method that determines the causal relationship between genetic variation and modifiable exposure on the outcome using SNPs as instruments. GWAS Catalog published trait-associated SNPs (GWAS significant) are used as

instrumental variables to estimate the causality between two traits[304]. Natural randomization of alleles at the gene level takes place during crossover and random assortment processes in the meiosis and gamete formation thus MR is a randomized controlled study. As a major assumption for MR, the instrumental variables (SNPs) should be associated with the exposure and independent from the outcome, to be valid instruments **(Figure 9)**. Pleiotropy happens when the SNPs have an indirect effect on the outcome, and this should be taken into consideration. In Paper IV, MR was performed using the TwoSampleMR package implemented in R using SNPs associated with the blood level of the biomarker as the exposure dataset and the results of the case-control GWAS for the same set of SNPs as the outcome. Five models are implemented in TwoSampleMR: the Egger regression model, inverse variance model, weighted-median model, and the two mode-estimate models; simple-mode model and weighted-mode model. The weighted median model is considered the optimal model to account for pleiotropy. The results of TwoSampleMR represent the five models and the magnitude of the causal effect i.e. the effect of the SNP on the outcome when the exposure is changed by one unit provided no confounding.

Results

Paper I. Genome-wide association analyses highlight etiological differences underlying newly defined subtypes of diabetes.

Heritability

The family history of diabetes was different between subtypes. As expected, SAID showed association with a family history of T1D. SIDD and MOD showed the strongest association with a family history of T2D.

Genome wide association study (GWAS)

Three genome-wide significant associations were identified (**Figure 10**). The *HLA* gene variant rs9273368 was significantly associated with SAID. The well-known T2D associated variant rs7903146 in the *TCF7L2* gene was associated with SIDD, MOD, and MARD. The variant rs10824307 near the *LRMDA* gene was uniquely associated with MOD. The look-ups for the variant rs10824307 in the AGEN and DIAMANTE study[212], revealed a significant association with T2D supporting that it is a true finding. In the UK biobank[285], variant rs10824307 was associated with a higher basal metabolic rate and higher whole-body fat-free mass. In the GTEx database, the same variant was an eQTL for the *LRMDA* gene in adipose tissue and pancreas[283].

Genetic risk scores (GRS)

The GRSs analysis was performed using logistic regression of GRSs of each diabetes subtype versus non-diabetic controls to determine the association of each subtype and the traits.

T1D GRS and T2D GRS

Two GRS for T1D were used: a T1D GRS calculated using all variants associated with T1D at genome-wide significant levels in the largest European T1D fine-

mapping study [211] and the T1D GRS2 score developed by Sharp et al that also takes interactions between SNPs into account.[39]. Both scores were only significantly associated with SAID in ANDIS and DIREVA. T2D GRSs were constructed from SNPs from Mahajan et al, the largest European T2D meta-analysis study published SNPs at the time of analysis[165]. A global score (T2D-gPRS) including all SNPs in the genome showed a significant association for all the diabetes subtypes with the greatest effect size for SIDD and MOD and smaller risk for SAID, SIRD, and MARD. When using T2D restricted GRSs (T2D-GRS) calculated by including only GWAS significant SNPs (n=304 SNPs) from Mahajan et al[165], the difference between subtypes was even larger. In ANDIS, T2D GRS was strongly associated with the greatest effect size in SIDD with much smaller effect sizes in SIRD. T2D-gPRS and T2D-GRS were replicated in DIREVA.

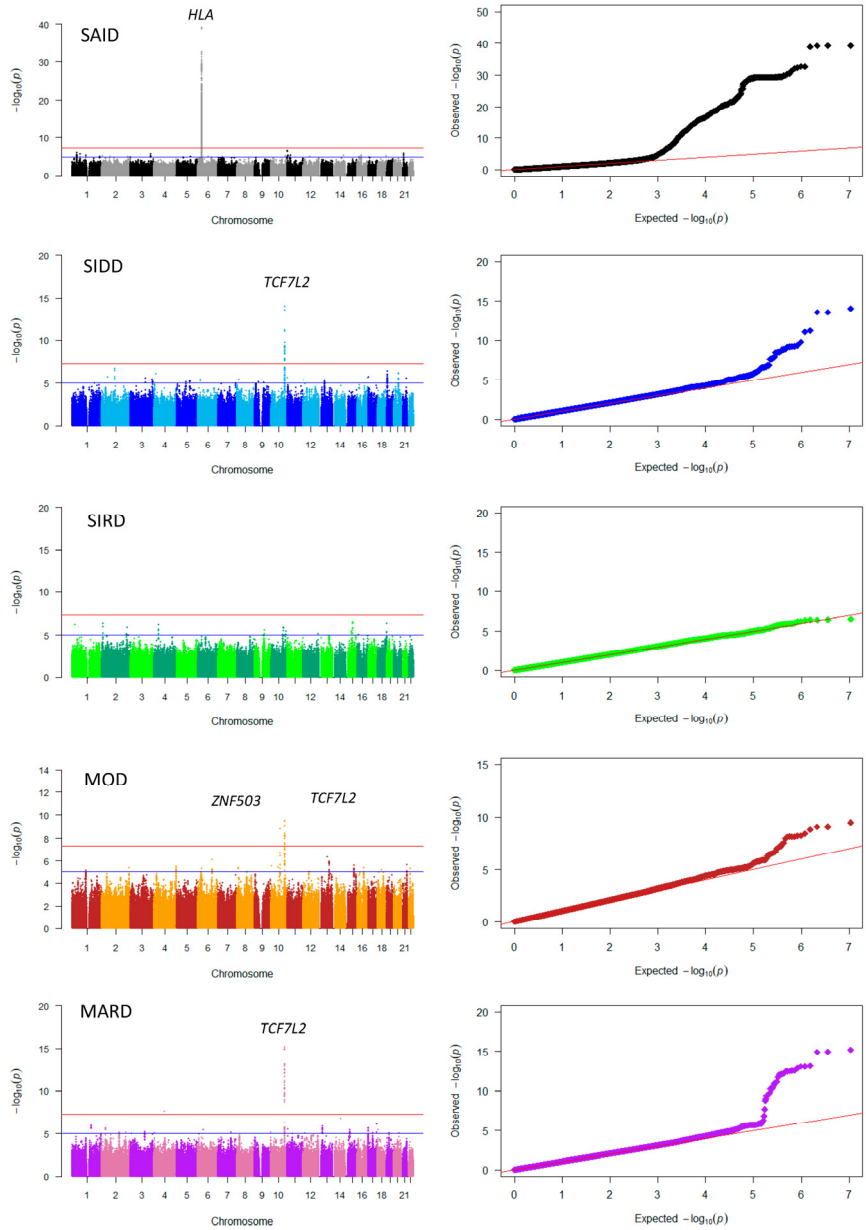

Figure 10. Manhattan plots for case-control GWAS in ANDIS. SAID-MDC (panel A), SIDD-MDC (panel B), SIRD-MDC (panel C), MOD-MDC (panel D), and MARD-MDC (panel E). The red line indicates the genome-wide significance threshold (p < 5 ×10-08) and the blue line suggestive association (P< 10-05).

GRS analysis for insulin secretion and sensitivity measures

T2D weighted GRS (T2D-wGRS) were estimated from known T2D associated SNPs using their effect sizes on the first-phase insulin response calculated in non-diabetic individuals. T2D-wGRSs included corrected insulin rate (CIR), insulin secretion rate (ISR), fasting insulin (FINS), insulin sensitivity index (ISI), fasting glucose, 2h-glucose, and fasting proinsulin. CIR and ISR were strongly associated with SIDD and not with SIRD. On the other hand, FINS showed a unique association with SIRD and not with the other subtypes. ISI was strongly associated with SIRD and MOD and nominally associated with SIDD and MARD.

GRS analysis for weight-related phenotypes.

The Body mass index (BMI), waist circumference (WC), and visceral adipose tissue (VAT) GRSs were strongly associated with MOD and not MARD. The waist-hip ratio adjusted for BMI GRS had a strong association with SIRD and not MOD.

SNP-cluster GRS

Two previous publications have clustered diabetes-associated SNPs by their associations with diabetes-related traits[164,165,305]

GRS based on clustered SNPs from Mahajan et al. included adiposity that had an association with MOD and not in MARD. Impaired lipids GRS was nominally associated with SIRD and not MOD. The insulin secretion GRSs (independent GRSs) had no association with SIRD ($P>0.2$) yet were associated with SIDD, MOD, and MARD, and the insulin action GRS were associated with all the diabetes subtypes. GRS based on clustered SNPs from Udler et al. included beta cell function which was strongly associated with SIDD and not SIRD. Proinsulin-GRS had an association with increased risk of SIDD and MARD but was protective in SIRD. The liver dystrophy GRS was associated with increased risk of SIDD, SIRD, MOD, and MARD subtypes and a weak association with SAID and the liver GRS had a strong association with SAID and not the other subtypes. The obesity GRS had an association with SIRD and MOD.

LD Score regression analysis

The case-control GWAS results for LDHub specified trait-associated SNPs were used to estimate the association between the subtypes (Trait 1) and LDHub traits (Trait2). Obesity-associated traits (fat body mass, waist circumference, hip circumference, basal metabolic rate, cholesterol-related traits, hypertension, were strongly associated with MOD. Maternal and paternal diabetes, cholesterol levels were associated with SIDD, while fasting glucose, birth weight aswas associated with MARD. SIRD had no association with the LDHub traits. The results of this analysis were not included in the published version of Paper I.

Paper II. Adult-Onset Diabetes in Middle Eastern Immigrants to Sweden: Novel subgroups and diabetic complications

Comparison of Iraqi and Swedish individuals with diabetes

The onset of diabetes was a decade earlier in Iraqi patients than in Swedes. The frequency of male individuals was higher in Iraqi (71%) than in Swedes (59.9%). Iraqi individuals were younger than Swedes individuals. The HbA1c in Iraqi (66.9 mmol/ml) was greater than in Swedes (62.7 mmol/ml). The prevalence of CKD stage 3A (eGFR <60 mL/min/1.73m2) in Iraqi (0.05%) was lower than in Swedes (64%) at baseline and no Iraqi had the CKD stage 3B (eGFR<45 mL/min/1.73m2). Fewer Iraqi (45.4%) had hypertension than Swedes (72.9%).

Prevalence of diabetes subtypes

In Iraqi immigrants, the MOD subtype was the most prevalent (39.3 vs 19.1% in native Swedes) followed by the SIDD subtype (27.9 vs 16.2%). On the other hand, in native Swedes, the MARD subtype, the SIRD subtype, and the SAID subtype, respectively, were 2-3 times as prevalent as in Iraqi immigrants (MARD 41.3 vs 25.1%; SIRD 16.3 vs 5.5%; SAID 7.0 vs 2.2%).

Risk of diabetic complications

During the 8-year follow-up, Iraqi patients had a higher risk of coronary events than native Swedes, females being at considerably lower risk than males, and higher BMI patients at baseline had a slightly lower risk of coronary events during follow-up. Iraqi patients had a considerably lower risk to develop CKD as compared to native Swedes adjusted for sex, age at diabetes onset, baseline BMI and HbA1c that was no longer significant after adjusting the model for baseline eGFR. The risk of developing CKD in females was considerably higher than in males. There were no significant differences in the incidence of stroke between individuals from Iraqi and Swedes, during follow-up. Early-onset of diabetes onset predicted increased risk for stroke, whereas HbA1c and BMI did not affect the risk. The risk of developing stroke in females was less than males. Fundus photography showed that the prevalence of patients displaying at least moderate retinopathy was almost twice as high in the Iraqi individuals compared to Swedish individuals, but the difference was not statistically significant, possibly due to the small sample size.

Genetic risk score analysis

GRS analysis for T2D, BMI, insulin secretion rate (ISR), and insulin sensitivity (ISI) traits using Iraqi individuals (cases) versus Swedish individuals (controls). T2D-GRS and ISI-GRS showed a greater association with Iraqi individuals than Swedes. On the other hand, BMI-GRS and ISR-GRS showed lower association in Iraqi individuals compared to Swedish individuals (**Figure 13**).

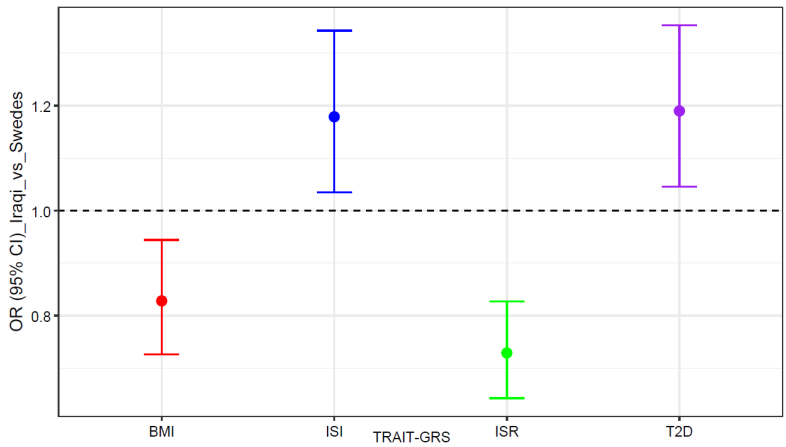

Figure 13. Genetic Risk Scores (GRS) in Iraqi (cases) versus Swedes (controls).

Paper III. Genetics of kidney complications in diabetes subtypes

Frequency of diabetic kidney disease

In the ANDIS cohort, the mean prevalence of CKD, the highest frequency was seen in the SIRD subtype with 28.32% eGFR<60 and 12.92% eGFR<45. Micro and macroalbuminuria were observed in 8.92% and 2.62% of T2D respectively, with the highest prevalence of microalbuminuria in SIDD 9.30% and of macroalbuminuria in SIRD 4.60%. In DIREVA, the prevalence of CKD was similar to ANDIS, SIRD had higher frequencies for eGFR<60 and eGFR <15 (ESRD) (**Tables 1, 3**).

Genome-wide association analysis (GWAS)

In ANDIS eGFR-GWAS, the variant rs77924615, A allele, in the Protein Disulfide Isomerase Like Testis expressed (*PDILT*)-Uromodulin (*UMOD*) locus, reached genome-wide significance in T2D. This association was stable after adjusting for BMI and HbA1c but was weakened by adjusting for HOMA2IR. The variant was also associated with eGFR in MARD and SIDD but not SIRD or MOD. In SIRD, the variant rs377038, C allele in the Catenin alpha 2 (*CTNNA2*) locus reached near genome-wide significance. This association was strengthened on adjusting for HOMA2IR. This variant was not associated with eGFR in T2D or other subtypes.

In DIREVA, the *PDILT-UMOD* association was replicated for T2D, MARD, and SIDD and was also associated with SIRD but there was no association in MOD. Unfortunately, the *CTNNA2* variant did not replicate in SIRD.

Only *PDILT-UMOD* loci reached genome-wide significance in meta-analysis in T2D and MARD.

Genetic risk scores analysis

GRS were created for three kidney traits; CKD (N_{SNPs}=34), eGFR (N_{SNPs}=625), and UACR (N_{SNPs}=94). For CKD-GRS, the strongest association was seen in the MARD subtype. Kidney-TlD GRS and kidney-T2D GRS were created using GWAS significant SNPs for kidney disease in TlD and T2D respectively. Both scores showed no association with eGFR in T2D or the subtypes after removal of the *PDILT-UMOD* SNP.

GRS for BMI and fasting insulin showed nominal association with eGFR in T2D, the latter mostly driven by the association in SIRD. GRS of fasting glucose showed association with eGFR only in SIDD.

Paper IV. Proteomic profiles of novel diabetic subtypes

Protein biomarkers whose concentrations differed between subtypes

In one-way ANOVA Model 1, adjusted for sex 168 biomarkers were significant in at least two pairwise comparisons; 12 with SIDD, 88 with SIRD, 16 with MOD, 29 with MARD, 17 others in more than one subtypes and 6 in all subtypes. Leptin and GDF-15 differed significantly in all subtype comparisons. PLC, TIMP4, FSTL3, Gal-4, EDA2R, and TRAIL-R2 were associated with the SIRD subtype. In Model 2, adjusting for age and BMI; LEP, Gal-4, GDF-15, PLXNB2, C1QTNF1, ACE2, KIM1, GUSB, and CDHR2, were the differential significant biomarkers among the subtypes. The biomarkers that had the strongest association after adjustment for sex, age, BMI, and medications were leptin, leptin receptor, SELE, GUSB, and Gal4.

Pathway analysis

In SIDD, the overrepresented pathways were mostly related to leptin signaling and incretin hormone biology. Associated phenotypes were polycystic ovary syndrome and arterial hypertension. Cytokine signaling, metabolism of angiotensinogen, and TNF pathways were overrepresented in SIRD along with myocardial ischemia. Atherosclerosis was associated with both SIRD and MOD. In MOD, hormones that underlie human gastrointestinal functions and eating behavior were the second top association. The cytokine-cytokine receptor signaling pathway was overrepresented in MOD compared to MARD.

Accuracy of classification with and without original variables

Biomarker prioritization was based on the model development and performance using different clinical variables and biomarkers that were significant for each subtype. The CAT boosting algorithm developed with the free variables age, sex, and BMI along with biomarkers, classified the patient samples into precise subtypes with no confusion between subtypes using only 33 variables.

Mendelian randomization

The causality between the most important biomarkers and both subtypes and T2D GADA negative individuals was performed using Mendelian Randomization. The strongest association was observed for Selenocysteine lyase (SCLY) in T2D and all GADA-negative subtypes in ANDIS. The biomarkers showed no significant association after adjustment for multiple testing.

Discussion

Ahlqvist et al. clutered diabetes into five subtypes based on six clinical variables. They reflect the heterogeneity of diabetes by the differences in prevalence, age at onset, glycemic control and risk to develop diabetic complications. In this thesis we characterized these subtypes using GWAS and biomarkers.

For the SAID subtype, findings were by as expectatied considering that this subtype is made up of T1D and LADA patients. Family history of T1D, defined as a diagnosis at age below 40 years and insulin treatment, was strongly associated with SAID. SAID also showed strong association with *HLA* variants, a well-known locus for its association with autoimmune disease[39,306,307]. As expected, SAID had the strongest association with the T1D-GRS and T1D-GRS2 (HLA variants excluded)[39]. In accordance with the large proportion of LADA in this subtype, SAID also showed some association with T2D-GRS and family history of T2D.

For the SIDD subtype, family history of T2D was common and this finding was supported by the results of SNP heritability. SIDD showed association with *TCF7L2* variants, a well-known T2D locus, and had the strongest association with the T2D-GRS and insulin secretion GRSs. The results from Paper I show that SIDD belongs to T2D rather than T1D and highlights the role of pancreatic beta-cell function in this subtype. The non-autoimmune nature of the subtype is also supported by the lack of autoantibodies.

In Paper IV, the biomarker analysis indicated an important role for leptin and the leptin receptor in SIDD, which was supported by a few nominal associations in MR. Leptin has a major role in energy regulation, food intake, obesity, inflammation, metabolic syndrome, diabetes, and diabetes-related cardiac dysfunction[101,308-311].

For the SIRD subtype, the family history of T2D had a smaller effect compared to SIDD and MOD and this finding was supported by the results of SNP heritability. SIRD showed no association with the *TCF7L2* variants or insulin secretion GRS suggesting a mostly beta-cell-independent pathway in this subtype. SIRD showed a unique association with fasting insulin GRS versus non-diabetic controls and versus eGFR, reflecting the insulin resistance status in these individuals[312]. In Paper III, SIRD showed no association with *PDILT-UMOD* but instead had a unique association with *CTNNA2* locus. The *UMOD* gene encodes uromodulin, the most abundant protein in the urine of healthy adults which is produced in the loop of

Henle following the proteolytic cleavage of the luminal cell surface ectodomain protein. Uromodulin is expressed only in the kidneys and is protective against renal calcium crystallization and bacterial urinary tract infections. The lack of association in individuals with high BMI supports a finding by Cornelia et al. that the presence of metabolic syndrome decreases serum uromodulin levels[313]. The catenin alpha 2 protein encoded by the *CTNNA2* locus, belongs to the cell adhesion protein family, which plays an important role in connecting cadherins located on the plasma membrane to the actin filaments inside the cell[255]. The main units in GFB are podocytes; fully differentiated kidney cells where the vital processes of podocyte differentiation and effacement are highly regulated by the cytoskeleton mechanisms, actin filament, and foot mobility. The association of *CTNNA2* with SIRD could suggest podocyte malfunction and disintegration of the glomerular basement membrane of the nephron and this explains the accompanying macroalbuminuria in this subtype[257,314-316]. Unfortunately, in DIREVA, the *CTNNA2* locus did not replicate, which, excluding the possibility of the original finding being a false positive, could be due to the low power to detect the association and the great random variation due to small sample size or other reasons; population differences (Swedish and Finnish), follow-up time differences (6 years in ANDIS versus 9.7 years in DIREVA), and sampling routines.

In Paper IV, for the biomarker pathway analysis, cytokine pathway including numerous TNF Receptor Superfamily members and an angiotensin related pathway including ACE2, CPB1, REN, and CTSD, showed a strong association with SIRD. Cytokines are potent immunomodulating proteins that play a vital role in cell signaling, endocrine function, and inflammatory processes. Many studies published information about the intimate relationship of metabolic syndrome and inflammation, indicating an important role of the immune system in SIRD[51,77,105,317-319]. Renin-angiotensin System (RAS) pathway in the kidney, is known to play an important role in blood pressure control, electrolyte homeostasis and influences different processes like immune response, inflammation, and ageing[106,320-322].

For the MOD subtype, the family history of T2D was common, which was supported by SNP heritability. MOD showed association with the *TCF7L2* variants, T2D-GRS, insulin secretion GRS and no association with fasting insulin, in contrast to SIRD, suggesting that insulin resistance plays a smaller role in this subtype. One of the main aims of this project was to identify subtype-specific locus, which was accomplished in MOD. The *LRMDA* loci showed association in the recent studies, AGEN and DIAMANTE, suggesting it is a true diabetes locus[212,323] . In GTEx, the associated SNP was an eQTL for *LRMDA* in adipose tissue and pancreas, supporting that this is the functional gene but this remains to be proven[323]. The *LRMDA* is highly expressed in many tissues and the main known function is melanocyte differentiation, however, knock out of the gene in mice show a muscle-related function suggested by elevated circulation creatinine and increased grip strength.

These lookups collectively support the role of obesity in the MOD subtype. For the lipid-related GRS, MOD showed a greater association with BMI and VAT GRS and smaller for WHR adjusted for BMI in contrast to SIRD, reflecting the dominant role of obesity and to a less extent the role of metabolic syndrome in this subtype. A recent study conducted by Isidor et al. to compare the transcriptomics of WAT and VAT, revealed that in the visceral adipocyte, systemic insulin resistance leads to gene expression dysfunction in adipose tissue that is not similar to that caused by increased BMI. This could explain the healthier profile seen in the MOD subtype compared to the SIRD subtype[324].

In Paper II, MOD was the most prevalent subtype in Middle Eastern; Iraqi individuals compare to European; Swedish individuals.

In Paper III, MOD showed no association with *PDILT-UMOD* as in SIRD yet also no association with *CTNNA2* in contrast to SIRD. The lack of association with *PDILT-UMOD* could be due to the clinical presentation of obesity and *PDILT-UMOD* being an obesity-dependent locus, but could also be due to the low prevalence of CKD cases in MOD, especially in ANDIS (5.26%)[313]. Replication in other cohorts with a longer duration time is strongly recommended to make definitive conclusions about the interaction between clinical variables and the *PDILT-UMOD* locus. On the other hand, the lack of association with *CTNNA2* could be due to a true difference in the underlying pathways for DKD in both subtypes. In Paper IV, the biomarker interaction GH1-LEP-PPY-CCL11 suggested mechanisms related to obesity, including the synthesis of Ghrelin and appetite regulation by leptin[17,325-329]. The concentration of GH1 and PPY aswas lower in MOD compared to all other subtypes including SIRD, while leptin levels were higher in SIRD and MOD subtypes.

For the MARD subtype, the family history of T2D was less common compared to SIDD and MOD and this finding was supported by the results of SNP heritability. MARD showed association with the *TCF7L2* variant, T2D-GRS, and insulin secretion GRSs suggesting an important role of the pancreas. The most striking genetic finding for MARD was a lack of association with GRS for BMI and obesity, suggesting this is of less importance in the development of this subtype.

In Paper IV, the combination of biomarkers EGFR-CDH1-SELE-NOS3 and LEP-IL1RN-IL1R2 showed a unique association with MARD except for LEP[249,330,331]. The related diseases include several age-related complications (presented by cardiovascular and kidney complications) and the finding supports the incidence of these complications in MARD.

The overall findings suggest etiological differences between the newly defined subtypes in the development of diabetes and diabetic complications. A hypothesis suggestion of the interaction of different body organs and the newly defined diabetes subtypes is illustrated in **Figure 14**.

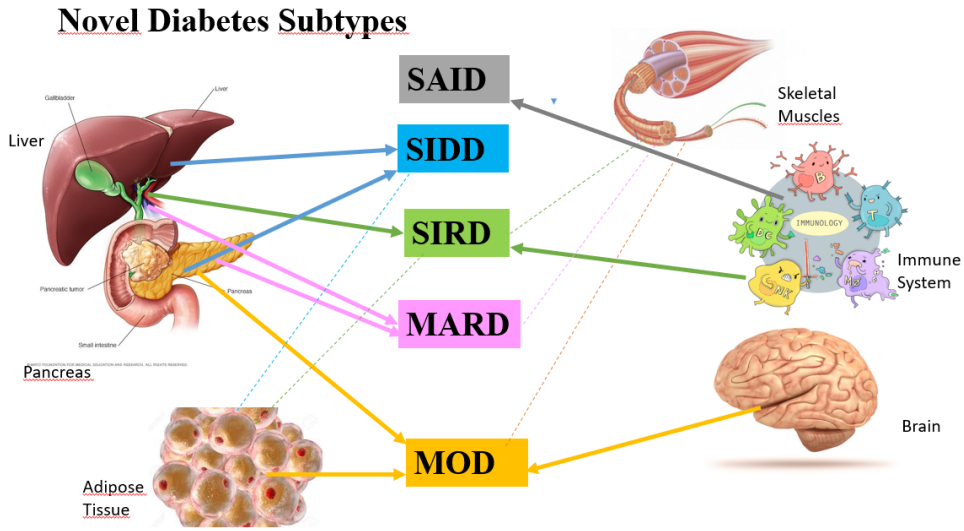

Figure 14. Hypothesis suggesting organ-subtype interactions for the new defined diabetes subtypes.

Summary and conclusions

A new subclassification is a promising approach in the future of diabetes. Clinical variables used for the clustering are valuable parameters to diagnose diabetes. A subclassification application can be implemented in the healthcare clinics to get information about the patient's diabetes subtype. SIDD and SIRD are high-risk subtypes and recommendations are to initiate early management and monitor for complications to prevent or delay the onset and severity of diabetic complications.

In Paper I, the genetic analysis provides strong evidence for distinct genetic backgrounds of the new diabetes subtypes. Strikingly, SIRD stands out in contrast to the other subtypes showing the genetic evidence for beta-cell independent pathogenesis and the unique association with fasting insulin GRS, reflecting liver insulin resistance. In Paper II, Middle Eastern individuals with diabetes differed in the distribution of the diabetes subtypes but also showed different risks of cardiovascular and kidney complications compared to native Swedes. This illustrates the importance of considering ethnicity in the study of diabetes and the subtypes. In Paper III, there is some suggestive support for different genetic backgrounds of DKD in diabetes subtypes. In Paper IV, the different proteomic profiles indicate variability in the underlying pathways of the subtypes. Differentially expressed biomarkers could be used to fine-tune the clustering.

A new diabetes subclassification is a promising approach however application in the healthcare system could be challenging. The results of this project shed highlight the genetic background of the newly defined diabetes subtypes. The main limitation of this project is the low statistical power due to the small samples size. Further analysis in larger populations will increase the power, hopefully, support the findings and maybe enable new findings that were not captured in this project.

The initiation of awareness programs to inform people about the clinical presentation of diabetes and its complications, who is at high risk, and how prediabetic individuals can benefit from early detection and the new subclassification of diabetes would be valuable. This could help in diabetes management and prevention or delay of complications.

New approaches in managing chronic disease especially diabetes by screening, early detection, new subclassification and genetics is a game changer and can decrease diabetes incidence and diabetes impact on healthcare budget in the future.

Future prospectives

Diabetes is one of the most challenging chronic diseases. As a complex outcome of a cascade of events, it requires a deep understanding of the molecular mechanisms and the behavior of the pivotal molecules and their interactions within the cell and the neighboring cells, using genetic data analysis, transcriptomics, single-cell biology technologies, proteomics, and metabolomics.

The new diabetes subclassification allows the focus of the Omics technologies on specific patient subgroups with certain features. This focus makes it easier to explore how biological molecules interact and affect the specific subtype.

I think from now and on, diabetes will not be seen as two major subtypes T1D and T2D, instead, as different five subtypes (SAID, SIDD, SIRD, MOD, and MARD) based on the clinical and genetic data. Larger cohorts are required for the replication of the results presented in this thesis. Following these replications, downstream functional studies for the top genome-wide significant variants will determine the metabolic pathways underlying the diabetes subtype and be valuable for new drug development.

Until the time, when diabetes is officially announced as five subtypes and the new subclassification becomes the main guideline for diagnosis and treatment of diabetic patients, the researchers and healthcare professionals should work hand in hand to facilitate the application of the new subclassification, measure the patient outcomes and develop new medications to serve the new approach in the future.

Acknowledgment

I am grateful to everyone who inspired and support me on my successful research journey, in chronological order.

Cairo University Professors:

Professor Dr. Mohamed F. El-Miligi (Professor of Pharmaceutics & Industrial Pharmacy, Cairo University, Advisor to Chancellor of Misr University of Science and Technology (MUST), Ex-Dean College of Pharmacy), my main supervisor during my first Master degree in pharmaceutical sciences.

Professor Dr. / Soad Ali Yehia Ali Amin (Pharmaceutics & Industrial Pharmacy, Cairo University**).**

Professor Dr. Amro Hanora (Professor and head of Microbiology & Immunology Department, Faculty of Pharmacy, Suez Canal University, Egypt) conducted the first course of Bioinformatics at the faculty of pharmacy, Cairo University. He recommended Lund University for my Master's degree in Bioinformatics.

Broad Institute:

Professor Eric Lander (president and founding director of the Broad Institute of MIT and Harvard University, USA), for his interesting and highly informative online course "Secret of Life" on EDX, I was honored to meet him in person at the Third Swedish Diabetes Summit in 2018, Karolinska Institute, Sweden.

Lund University Master program of Bioinformatics team:

Professor Björn Canbäck, (Senior Lecturer, Department of Biology, Lund University), the coordinator of the Master program in Bioinformatics at Lund University, and my main teacher for coding syntax, programming, software implementation, and genetic data analysis.

Professor Jessica Abbott (Senior Lecturer, Department of Biology, Lund University), the course conductor in R, and the fantastic assistance in making R magic code run.

Professor Mattias Ohlsson (Professor of Computational Biology and Biological Physics, Lund University), the conductor of "Introduction to programming course in PERL".

Professor Torbjörn Säll (Professor of molecular cell biology, Lund University), first to introduce me to population genetics and association analysis.

Genomics, Diabetes and Endocrinology group at Lund University Diabetes Centre (LUDC): ***The perfect research team and unreplaceable in my heart.***

Professor Leif Groop (Senior Professor, Genomics, Diabetes and Endocrinology, LUDC, Institute for Molecular Medicine Finland, Centre of Excellence in Complex Disease Genetics). Thank you for believing in me.

Emma Ahlqvist (Associate professor, Genomics, Diabetes, and Endocrinology). Thank you for your professional scientific guidance and for making this happen.

Ola Hansson (Associate Professor, Group leader, Genomics, Diabetes, and Endocrinology). Thank you for giving me the opportunity to join your research group.

Rashmi Prasad (Associate Professor, Genomics, Diabetes, and Endocrinology)**.** Thank you for your positivity, and for always giving me hope**.**

Yang De Marinis (Associate Professor, Genomics, Diabetes, and Endocrinology). It was an honor to be part of the discovery of FST**.**

Professor Valeryia Lyssenko (Professor, MD, University of Bergen, LUDC). Even though we have not worked together, yet your presence was always inspiring.

Gad Hatem (PhD student) Excellent programmer and loyal colleague**.**

Manonanthini Thagman (Post-doctorate fellow, Genomics, Diabetes and Endocrinology), for your help with my thesis.

Pradeep Bompada (PhD) Thank you for your advice and tips about migration issues**.**

Mattias Borell (Information Technology coordinator). Thank you so much for all the timely support.

Johan Hultman (Systems Engineer, Information Technology coordinator). Thank you for your support and assistance.

Bushra Shadid (PhD) Nice beautiful sister during my lonely stay in Sweden.

Jacquline Postma (Grant manager LUDC, Project manager IMI-BEAt-DKD) for her great support with communication with the migration agency and Ph.D. committee, and many things esp. “Noble breakfast”.

Ulrika Blom-Nilsson (Research administrator LUDC)**.** Lovely Ulrika you were the driving force behind every successful step.

Mikael Åkerlund (System engineer at LUDC), for his valuable assistance in the quality control of genetic data; pipeline.

Peter Almgren (Senior Biostatistician at LEO Pharma). Thank you for your time and effort on the ANDIS project.

University of Helsinki Team:

Professor Tiinamaija Tuomi (Professor, MD, Research Director, Institute for Molecular Medicine Finland, Centre of Excellence in Complex Disease Genetics). Thanks for the guidance, time, and contribution to the Nature Genetics paper.

Om Prakash Dwivedi (PhD, Senior Researcher, Institute for Molecular Medicine Finland). Thank you for your effort and contribution to the polygenic risk scores.

Annemari Käräjämäki (MD, Specialist in general medicine, Chief Physician). Thanks for your time and effort and assistance in the DIREVA dataset.

References

1. Ahlqvist E, Storm P, Karajamaki A, et al. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol* 2018; **6**(5): 361-9.
2. Kelly GS. Insulin resistance: lifestyle and nutritional interventions. *Alternative medicine review : a journal of clinical therapeutic* 2000; **5**(2): 109-32.
3. McAuley KA, Williams SM, Mann JI, et al. Diagnosing Insulin Resistance in the General Population. *Diabetes Care* 2001; **24**(3): 460-4.
4. Petrovic G, Bjelakovic G, Benedeto-Stojanov D, Nagorni A, Brzacki V, Markovic-Zivkovic B. Obesity and metabolic syndrome as risk factors for the development of non-alcoholic fatty liver disease as diagnosed by ultrasound. *Vojnosanit Pregl* 2016; **73**(10): 910-20.
5. Isokuortti E, Zhou Y, Peltonen M, et al. Use of HOMA-IR to diagnose non-alcoholic fatty liver disease: a population-based and inter-laboratory study. *Diabetologia* 2017; **60**(10): 1873-82.
6. Jelenik T, Kaul K, Séquaris G, et al. Mechanisms of Insulin Resistance in Primary and Secondary Nonalcoholic Fatty Liver. *Diabetes* 2017; **66**(8): 2241-53.
7. WHO Diabetes keyfacts sheet. 2021.
8. Glovaci D, Fan W, Wong ND. Epidemiology of Diabetes Mellitus and Cardiovascular Disease. *Current cardiology reports* 2019; **21**(4): 21.
9. Andersson E, Persson S, Hallén N, et al. Costs of diabetes complications: hospital-based care and absence from work for 392,200 people with type 2 diabetes and matched control participants in Sweden. *Diabetologia* 2020; **63**(12): 2582-94.
10. Li Y, Zhang H, Tu Y, et al. Monogenic Obesity Mutations Lead to Less Weight Loss After Bariatric Surgery: a 6-Year Follow-Up Study. *Obes Surg* 2019.
11. Hinney A, Vogel CI, Hebebrand J. From monogenic to polygenic obesity: recent advances. *Eur Child Adolesc Psychiatry* 2010; **19**(3): 297-310.
12. Senat MV, Deruelle P. [Gestational diabetes mellitus]. *Gynecologie, obstetrique & fertilite* 2016; **44**(4): 244-7.
13. DeFronzo RA. Pathogenesis of type 2 diabetes mellitus. *The Medical clinics of North America* 2004; **88**(4): 787-835, ix.
14. Kim J-a, Wei Y, Sowers JR. Role of Mitochondrial Dysfunction in Insulin Resistance. *Circulation Research* 2008; **102**(4): 401-14.
15. Huang PL. A comprehensive definition for metabolic syndrome. *Dis Model Mech* 2009; **2**(5-6): 231-7.

16. Aboul Ella NA, Shehab DI, Ismail MA, Maksoud AA. Prevalence of metabolic syndrome and insulin resistance among Egyptian adolescents 10 to 18 years of age. *J Clin Lipidol* 2010; **4**(3): 185-95.
17. Yadav A, Jyoti P, Jain SK, Bhattacharjee J. Correlation of adiponectin and leptin with insulin resistance: a pilot study in healthy north Indian population. *Indian J Clin Biochem* 2011; **26**(2): 193-6.
18. Borgundvaag E, Mak J, Kramer CK. Metabolic Impact of Intermittent Fasting in Patients With Type 2 Diabetes Mellitus: A Systematic Review and Meta-analysis of Interventional Studies. *The Journal of Clinical Endocrinology & Metabolism* 2020; **106**(3): 902-11.
19. Gowey MA, Khodneva Y, Tison SE, et al. Depressive symptoms, perceived stress, and metabolic health: The REGARDS study. *Int J Obes (Lond)* 2019; **43**(3): 615-32.
20. Gupta S, Bansal S. Does a rise in BMI cause an increased risk of diabetes?: Evidence from India. *PLOS ONE* 2020; **15**(4): e0229716.
21. Jovanovic J, Sarac I, Jovanovic S, Sokolovic D, Govedarovic N, Jovanovic J. The relationship between occupational stress and health status, temporary and permanent work disability among security guards in Serbia. *Int J Occup Saf Ergon* 2019: 1-50.
22. Ehrmann D, Kulzer B, Roos T, Haak T, Al-Khatib M, Hermanns N. Risk factors and prevention strategies for diabetic ketoacidosis in people with established type 1 diabetes. *Lancet Diabetes Endocrinol* 2020; **8**(5): 436-46.
23. Lindgren CM, McCarthy MI. Mechanisms of disease: genetic insights into the etiology of type 2 diabetes and obesity. *Nat Clin Pract Endocrinol Metab* 2008; **4**(3): 156-63.
24. Andrade-Silva SG, Caranti DA, Sallet JA, Leal LP, Leal AJ, Damaso AR. Age and gender may influence the results of Roux-en-Y gastric bypass? Metabolic syndrome parameters. *Arq Gastroenterol* 2014; **51**(3): 171-9.
25. Ahlgren U, Jonsson J, Jonsson L, Simu K, Edlund H. beta-cell-specific inactivation of the mouse Ipf1/Pdx1 gene results in loss of the beta-cell phenotype and maturity onset diabetes. *Genes Dev* 1998; **12**.
26. Sicard DA, Taylor JR. Comparison of point-of-care HbA1c test versus standardized laboratory testing. *Ann Pharmacother* 2005; **39**(6): 1024-8.
27. Nitin S. HbA1c and factors other than diabetes mellitus affecting it. *Singapore medical journal* 2010; **51**(8): 616-22.
28. Al-Rubeaan K, Siddiqui K, Abu Risheh K, et al. Correlation between serum electrolytes and fasting glucose and Hb1Ac in Saudi diabetic patients. *Biol Trace Elem Res* 2011; **144**(1-3): 463-8.
29. Tanaka M, Yoshida T, Bin W, Fukuo K, Kazumi T. FTO, abdominal adiposity, fasting hyperglycemia associated with elevated HbA1c in Japanese middle-aged women. *J Atheroscler Thromb* 2012; **19**(7): 633-42.
30. Christy AL, Manjrekar PA, Babu RP, Hegde A, Rukmini MS. Influence of iron deficiency anemia on hemoglobin A1c levels in diabetic individuals with controlled plasma glucose levels. *Iran Biomed J* 2014; **18**(2): 88-93.

31. Pivovarov R, Albers DJ, Hripcsak G, Sepulveda JL, Elhadad N. Temporal trends of hemoglobin A1c testing. *J Am Med Inform Assoc* 2014; **21**(6): 1038-44.
32. Caumo A, Perseghin G, Brunani A, Luzi L. New Insights on the Simultaneous Assessment of Insulin Sensitivity and β-Cell Function With the HOMA2 Method. *Diabetes Care* 2006; **29**(12): 2733-4.
33. Gautier A, Roussel R, Lange C, et al. Effects of genetic susceptibility for type 2 diabetes on the evolution of glucose homeostasis traits before and after diabetes diagnosis: data from the D.E.S.I.R. Study. *Diabetes* 2011; **60**(10): 2654-63.
34. Song YS, Hwang YC, Ahn HY, Park CY. Comparison of the Usefulness of the Updated Homeostasis Model Assessment (HOMA2) with the Original HOMA1 in the Prediction of Type 2 Diabetes Mellitus in Koreans. *Diabetes & metabolism journal* 2016; **40**(4): 318-25.
35. Hariri S, Yoon PW, Qureshi N, Valdez R, Scheuner MT, Khoury MJ. Family history of type 2 diabetes: A population-based screening tool for prevention? *Genetics in Medicine* 2006; **8**(2): 102-8.
36. Groop L, Forsblom C, Lehtovirta M, et al. Metabolic consequences of a family history of NIDDM (the Botnia study): evidence for sex-specific parental effects. *Diabetes* 1996; **45**(11): 1585-93.
37. Neumann A, Norberg M, Schoffer O, et al. Risk equations for the development of worsened glucose status and type 2 diabetes mellitus in a Swedish intervention program. *BMC Public Health* 2013; **13**: 1014.
38. Rutter GA, Pullen TJ, Hodson DJ, Martinez-Sanchez A. Pancreatic β-cell identity, glucose sensing and the control of insulin secretion. *The Biochemical journal* 2015; **466**(2): 203-18.
39. Sharp SA, Rich SS, Wood AR, et al. Development and Standardization of an Improved Type 1 Diabetes Genetic Risk Score for Use in Newborn Screening and Incident Diagnosis. *Diabetes Care* 2019; **42**(2): 200-7.
40. Wedrychowicz A, Stec M, Sztefko K, Starzyk JB. Associations between bone, fat tissue and metabolic control in children and adolescents with type 1 diabetes mellitus. *Exp Clin Endocrinol Diabetes* 2014; **122**(8): 491-5.
41. Roep BO, Thomaidou S, van Tienhoven R, Zaldumbide A. Type 1 diabetes mellitus as a disease of the β-cell (do not blame the immune system?). *Nat Rev Endocrinol* 2021; **17**(3): 150-61.
42. Blair M. Diabetes Mellitus Review. *Urologic nursing* 2016; **36**(1): 27-36.
43. Ottosson-Laakso E, Krus U, Storm P, et al. Glucose-Induced Changes in Gene Expression in Human Pancreatic Islets: Causes or Consequences of Chronic Hyperglycemia. *Diabetes* 2017; **66**(12): 3013-28.
44. Stolic M, Russell A, Hutley L, et al. Glucose uptake and insulin action in human adipose tissue—influence of BMI, anatomical depot and body fat distribution. *International Journal of Obesity* 2002; **26**(1): 17-23.
45. Wilcox G. Insulin and insulin resistance. *Clin Biochem Rev* 2005; **26**(2): 19-39.

46. Abdul-Ghani MA, Tripathy D, DeFronzo RA. Contributions of beta-cell dysfunction and insulin resistance to the pathogenesis of impaired glucose tolerance and impaired fasting glucose. *Diabetes Care* 2006; **29**(5): 1130-9.
47. Maahs DM, Nadeau K, Snell-Bergeon JK, et al. Association of insulin sensitivity to lipids across the lifespan in people with Type 1 diabetes. *Diabet Med* 2011; **28**(2): 148-55.
48. Gurung M, Li Z, You H, et al. Role of gut microbiota in type 2 diabetes pathophysiology. *EBioMedicine* 2020; **51**: 102590.
49. Umpierre D, Ribeiro PA, Kramer CK, et al. Physical activity advice only or structured exercise training and association with HbA1c levels in type 2 diabetes: a systematic review and meta-analysis. *JAMA* 2011; **305**(17): 1790-9.
50. Brunton S. Pathophysiology of Type 2 Diabetes: The Evolution of Our Understanding. *The Journal of family practice* 2016; **65**(4 Suppl).
51. Escobar-Morreale HF, Martinez-Garcia MA, Montes-Nieto R, Fernandez-Duran E, Temprano-Carazo S, Luque-Ramirez M. Effects of glucose ingestion on circulating inflammatory mediators: Influence of sex and weight excess. *Clin Nutr* 2017; **36**(2): 522-9.
52. Poggiogalle E, Jamshed H, Peterson CM. Circadian regulation of glucose, lipid, and energy metabolism in humans. *Metabolism* 2018; **84**: 11-27.
53. Mueckler M, Thorens B. The SLC2 (GLUT) family of membrane transporters. *Mol Aspects Med* 2013; **34**(2-3): 121-38.
54. Harada N, Inagaki N. Role of sodium-glucose transporters in glucose uptake of the intestine and kidney. *Journal of diabetes investigation* 2012; **3**(4): 352-3.
55. Tuomi T, Groop LC, Zimmet PZ, Rowley MJ, Knowles W, Mackay IR. Antibodies to glutamic acid decarboxylase reveal latent autoimmune diabetes mellitus in adults with a non-insulin-dependent onset of disease. *Diabetes* 1993; **42**(2): 359-62.
56. Li Z, Wu TM, Ming WJ, Chen X, Xiao XM. The influence of maternal islet beta-cell autoantibodies in conjunction with gestational hyperglycemia on neonatal outcomes. *PLoS One* 2015; **10**(3): e0120414.
57. Hjort R, Ahlqvist E, Carlsson PO, et al. Overweight, obesity and the risk of LADA: results from a Swedish case-control study and the Norwegian HUNT Study. *Diabetologia* 2018; **61**(6): 1333-43.
58. Braun M, Ramracheya R, Bengtsson M, et al. Voltage-gated ion channels in human pancreatic beta-cells: electrophysiological characterization and role in insulin secretion. *Diabetes* 2008; **57**(6): 1618-28.
59. German MS. Glucose sensing in pancreatic islet beta cells: the key role of glucokinase and the glycolytic intermediates. *Proceedings of the National Academy of Sciences* 1993; **90**(5): 1781-5.
60. Navale AM, Paranjape AN. Glucose transporters: physiological and pathological roles. *Biophysical reviews* 2016; **8**(1): 5-9.
61. Cho NH, Jang HC, Choi SH, et al. Abnormal Liver Function Test Predicts Type 2 Diabetes. *Diabetes Care* 2007; **30**(10): 2566.

62. Grunberger G, Taylor SI, Dons RF, Gorden P. Insulin receptors in normal and disease states. *Clinics in endocrinology and metabolism* 1983; **12**(1): 191-219.
63. Boucher J, Kleinridders A, Kahn CR. Insulin receptor signaling in normal and insulin-resistant states. *Cold Spring Harb Perspect Biol* 2014; **6**(1): a009191.
64. Moore MC, Coate KC, Winnick JJ, An Z, Cherrington AD. Regulation of hepatic glucose uptake and storage in vivo. *Adv Nutr* 2012; **3**(3): 286-94.
65. Chadt A, Al-Hasani H. Glucose transporters in adipose tissue, liver, and skeletal muscle in metabolic health and disease. *Pflugers Arch* 2020; **472**(9): 1273-98.
66. Sherwin RS. Role of Liver in Glucose Homeostasis. *Diabetes Care* 1980; **3**(2): 261-5.
67. Richter EA, Hargreaves M. Exercise, GLUT4, and Skeletal Muscle Glucose Uptake. *Physiological reviews* 2013; **93**(3): 993-1017.
68. Richter EA. Is GLUT4 translocation the answer to exercise-stimulated muscle glucose uptake? *American Journal of Physiology-Endocrinology and Metabolism* 2021; **320**(2): E240-E3.
69. Belman JP, Habtemichael EN, Bogan JS. A proteolytic pathway that controls glucose uptake in fat and muscle. *Rev Endocr Metab Disord* 2014; **15**(1): 55-66.
70. Ebeling P, Koistinen HA, Koivisto VA. Insulin-independent glucose transport regulates insulin sensitivity. *FEBS Lett* 1998; **436**(3): 301-3.
71. Whitehead A, Krause FN, Moran A, et al. Brown and beige adipose tissue regulate systemic metabolism through a metabolite interorgan signaling axis. *Nature Communications* 2021; **12**(1): 1905.
72. Hulgan T, Ramsey BS, Koethe JR, et al. Relationships Between Adipose Mitochondrial Function, Serum Adiponectin, and Insulin Resistance in Persons With HIV After 96 Weeks of Antiretroviral Therapy. *J Acquir Immune Defic Syndr* 2019; **80**(3): 358-66.
73. Scheja L, Heeren J. The endocrine function of adipose tissues in health and cardiometabolic disease. *Nature Reviews Endocrinology* 2019; **15**(9): 507-24.
74. Kaisanlahti A, Glumoff T. Browning of white fat: agents and implications for beige adipose tissue to type 2 diabetes. *Journal of physiology and biochemistry* 2019; **75**(1): 1-10.
75. Xu A, Wang Y, Lam KS, Vanhoutte PM. Vascular actions of adipokines molecular mechanisms and therapeutic implications. *Advances in pharmacology (San Diego, Calif)* 2010; **60**: 229-55.
76. Straub RH. Chapter II - Pathogenesis and Neuroendocrine Immunology. In: Straub RH, ed. The Origin of Chronic Inflammatory Systemic Diseases and their Sequelae: Academic Press; 2015: 59-129.
77. Shi J, Fan J, Su Q, Yang Z. Cytokines and Abnormal Glucose and Lipid Metabolism. *Frontiers in Endocrinology* 2019; **10**(703).
78. Jiang G, Zhang BB. Glucagon and regulation of glucose metabolism. *American Journal of Physiology-Endocrinology and Metabolism* 2003; **284**(4): E671-E8.
79. Woitiski CB, Neufeld RJ, Soares AF, Figueiredo IV, Veiga FJ, Carvalho RA. Evaluation of hepatic glucose metabolism via gluconeogenesis and glycogenolysis

after oral administration of insulin nanoparticles. *Drug Dev Ind Pharm* 2012; **38**(12): 1441-50.

80. Hussain Z, Khan JA. Food intake regulation by leptin: Mechanisms mediating gluconeogenesis and energy expenditure. *Asian Pacific Journal of Tropical Medicine* 2017; **10**(10): 940-4.
81. Fernie AR, Carrari F, Sweetlove LJ. Respiratory metabolism: glycolysis, the TCA cycle and mitochondrial electron transport. *Current opinion in plant biology* 2004; **7**(3): 254-61.
82. Keane KN, Calton EK, Carlessi R, Hart PH, Newsholme P. The bioenergetics of inflammation: insights into obesity and type 2 diabetes. *European Journal of Clinical Nutrition* 2017; **71**(7): 904-12.
83. Di Vincenzo A, Bettini S, Russo L, Mazzocut S, Mauer M, Fioretto P. Renal structure in type 2 diabetes: facts and misconceptions. *Journal of nephrology* 2020; **33**(5): 901-7.
84. Nienaber-Rousseau C, Swanepoel B, Dolman RC, Pieters M, Conradie KR, Towers GW. Interactions between C-reactive protein genotypes with markers of nutritional status in relation to inflammation. *Nutrients* 2014; **6**(11): 5034-50.
85. Asare-Anane H, Botchey CPK, Ofori EK, Boamah I, Crabbe S, Asamoah-Kusi K. Altered immunoglobulins (A and G) in Ghanaian patients with type 2 diabetes. *SAGE Open Med* 2018; **6**: 2050312118762042.
86. Hirotsu C, Albuquerque RG, Nogueira H, et al. The relationship between sleep apnea, metabolic dysfunction and inflammation: The gender influence. *Brain Behav Immun* 2017; **59**: 211-8.
87. Morris CJ, Aeschbach D, Scheer FA. Circadian system, sleep and endocrinology. *Mol Cell Endocrinol* 2012; **349**(1): 91-104.
88. Javeed N, Matveyenko AV. Circadian Etiology of Type 2 Diabetes Mellitus. *Physiology (Bethesda)* 2018; **33**(2): 138-50.
89. Yin H, Li W, Chatterjee S, et al. Metabolic-sensing of the skeletal muscle clock coordinates fuel oxidation. *The FASEB Journal* 2020; **34**(5): 6613-27.
90. Nakao R, Nikawa T, Oishi K. The skeletal muscle circadian clock: current insights. *ChronoPhysiology and Therapy* 2017; **7**: 47-57.
91. Aoyama S, Shibata S. The Role of Circadian Rhythms in Muscular and Osseous Physiology and Their Regulation by Nutrition and Exercise. *Frontiers in Neuroscience* 2017; **11**(63).
92. Schiaffino S, Blaauw B, Dyar KA. The functional significance of the skeletal muscle clock: lessons from Bmal1 knockout models. *Skeletal Muscle* 2016; **6**(1): 33.
93. Raff H, Magill SB. Is the hypothalamic–pituitary–adrenal axis disrupted in type 2 diabetes mellitus? *Endocrine* 2016; **54**(2): 273-5.
94. Jax TW. Metabolic memory: a vascular perspective. *Cardiovascular Diabetology* 2010; **9**(1): 51.
95. Felšöci M, Schroner Z, Petrovičová J, Lazúrová I. Relationship between type 2 diabetes mellitus and hypothalamic-pituitary-adrenal axis. *Wien Klin Wochenschr* 2011; **123**(1-2): 28-33.

96. Benaiges D, Chillaron JJ, Pedro-Botet J, et al. Role of A1c in the postpartum screening of women with gestational diabetes. *Gynecol Endocrinol* 2013; **29**(7): 687-90.
97. Rajagopal L, Ganapathy S, Arunachalam S, Raja V, Ramraj B. Does Iron Deficiency Anaemia and its Severity Influence HbA1C Level in Non Diabetics? An Analysis of 150 Cases. *J Clin Diagn Res* 2017; **11**(2): EC13-EC5.
98. Cetinkalp S, Simsir IY, Ertek S. Insulin resistance in brain and possible therapeutic approaches. *Curr Vasc Pharmacol* 2014; **12**(4): 553-64.
99. Khan RS, Bril F, Cusi K, Newsome PN. Modulation of Insulin Resistance in Nonalcoholic Fatty Liver Disease. *Hepatology* 2019; **70**(2): 711-24.
100. Chen Z, Yu R, Xiong Y, Du F, Zhu S. A vicious circle between insulin resistance and inflammation in nonalcoholic fatty liver disease. *Lipids in Health and Disease* 2017; **16**(1): 203.
101. Chuang SM, Shih HM, Chien MN, Liu SC, Wang CH, Lee CC. Risk factors in metabolic syndrome predict the progression of diabetic nephropathy in patients with type 2 diabetes. *Diabetes Res Clin Pract* 2019; **153**: 6-13.
102. Daenen K, Andries A, Mekahli D, Van Schepdael A, Jouret F, Bammens B. Oxidative stress in chronic kidney disease. *Pediatric nephrology (Berlin, Germany)* 2019; **34**(6): 975-91.
103. Blackstock JC. CHAPTER 13 - Mitochondrial oxidative phosphorylation. In: Blackstock JC, ed. Guide to Biochemistry: Butterworth-Heinemann; 1989: 160-8.
104. Blackstock JC. CHAPTER 12 - The tricarboxylate cycle. In: Blackstock JC, ed. Guide to Biochemistry: Butterworth-Heinemann; 1989: 149-59.
105. Rendra E, Riabov V, Mossel DM, Sevastyanova T, Harmsen MC, Kzhyshkowska J. Reactive oxygen species (ROS) in macrophage activation and function in diabetes. *Immunobiology* 2019; **224**(2): 242-53.
106. Shou J, Chen P-J, Xiao W-H. Mechanism of increased risk of insulin resistance in aging skeletal muscle. *Diabetology & Metabolic Syndrome* 2020; **12**(1): 14.
107. Townsend LK, Brunetta HS, Mori MAS. Mitochondria-associated ER membranes in glucose homeostasis and insulin resistance. *American Journal of Physiology-Endocrinology and Metabolism* 2020; **319**(6): E1053-E60.
108. Barbieri J, Fontela PC, Winkelmann ER, et al. Anemia in Patients with Type 2 Diabetes Mellitus. *Anemia* 2015; **2015**: 354737-.
109. Vallon V. Molecular determinants of renal glucose reabsorption. Focus on "Glucose transport by human renal Na+/D-glucose cotransporters SGLT1 and SGLT2". *Am J Physiol Cell Physiol* 2011; **300**(1): C6-C8.
110. Nakamura N, Matsui T, Ishibashi Y, Yamagishi S-I. Insulin stimulates SGLT2-mediated tubular glucose absorption via oxidative stress generation. *Diabetology & metabolic syndrome* 2015; **7**: 48-.
111. Chao EC. SGLT-2 Inhibitors: A New Mechanism for Glycemic Control. *Clinical Diabetes* 2014; **32**(1): 4-11.
112. Rodriguez-Segade S, Rodriguez J, Garcia-Lopez JM, et al. Influence of the glycation gap on the diagnosis of type 2 diabetes. *Acta Diabetol* 2015; **52**(3): 453-9.

113. Rodrigues KL, Borges JP, Lopes GO, et al. Influence of Physical Exercise on Advanced Glycation End Products Levels in Patients Living With the Human Immunodeficiency Virus. *Front Physiol* 2018; **9**: 1641.
114. Zhu J-l, Cai Y-q, Long S-l, Chen Z, Mo Z-c. The role of advanced glycation end products in human infertility. *Life Sciences* 2020; **255**: 117830.
115. Janež A, Guja C, Mitrakou A, et al. Insulin Therapy in Adults with Type 1 Diabetes Mellitus: a Narrative Review. *Diabetes Ther* 2020; **11**(2): 387-409.
116. Ghosh S, Collier A. Section 3 - Management of diabetes. In: Ghosh S, Collier A, eds. Churchill's Pocketbook of Diabetes (Second Edition) (Second Edition). Oxford: Churchill Livingstone; 2012: 83-125.
117. Meglitinides. In: Aronson JK, ed. Meyler's Side Effects of Drugs (Sixteenth Edition). Oxford: Elsevier; 2016: 800-7.
118. Zhou K, Donnelly L, Yang J, et al. Heritability of variation in glycaemic response to metformin: a genome-wide complex trait analysis. *Lancet Diabetes Endocrinol* 2014; **2**(6): 481-7.
119. Pernicova I, Korbonits M. Metformin—mode of action and clinical implications for diabetes and cancer. *Nature Reviews Endocrinology* 2014; **10**(3): 143-56.
120. Minamii T, Nogami M, Ogawa W. Mechanisms of metformin action: In and out of the gut. *Journal of Diabetes Investigation* 2018; **9**(4): 701-3.
121. Porter KM, Ward M, Hughes CF, et al. Hyperglycemia and Metformin Use Are Associated With B Vitamin Deficiency and Cognitive Dysfunction in Older Adults. *J Clin Endocrinol Metab* 2019; **104**(10): 4837-47.
122. Abbas G, Al-Harrasi AS, Hussain H. Chapter 9 - α-Glucosidase Enzyme Inhibitors From Natural Products. In: Brahmachari G, ed. Discovery and Development of Antidiabetic Agents from Natural Products: Elsevier; 2017: 251-69.
123. Li WX, Gou JF, Tian JH, Yan X, Yang L. Glucagon-like peptide-1 receptor agonists versus insulin glargine for type 2 diabetes mellitus: A systematic review and meta-analysis of randomized controlled trials. *Curr Ther Res Clin Exp* 2010; **71**(4): 211-38.
124. Hlebowicz J, Lindstedt S, Bjorgell O, Dencker M. The effect of endogenously released glucose, insulin, glucagon-like peptide 1, ghrelin on cardiac output, heart rate, stroke volume, and blood pressure. *Cardiovasc Ultrasound* 2011; **9**: 43.
125. Breitman I, Saraf N, Kakade M, et al. The effects of an amino acid supplement on glucose homeostasis, inflammatory markers, and incretins after laparoscopic gastric bypass. *J Am Coll Surg* 2011; **212**(4): 617-25; discussion 25-7.
126. Gjesing AP, Ekstrom CT, Eiberg H, et al. Fasting and oral glucose-stimulated levels of glucose-dependent insulinotropic polypeptide (GIP) and glucagon-like peptide-1 (GLP-1) are highly familial traits. *Diabetologia* 2012; **55**(5): 1338-45.
127. Vardarli I, Arndt E, Deacon CF, Holst JJ, Nauck MA. Effects of sitagliptin and metformin treatment on incretin hormone and insulin secretory responses to oral and "isoglycemic" intravenous glucose. *Diabetes* 2014; **63**(2): 663-74.
128. Larsen MP, Torekov SS. Glucagon-Like Peptide 1: A Predictor of Type 2 Diabetes? *J Diabetes Res* 2017; **2017**: 7583506.

129. Preiss D, Dawed A, Welsh P, et al. Sustained influence of metformin therapy on circulating glucagon-like peptide-1 levels in individuals with and without type 2 diabetes. *Diabetes Obes Metab* 2017; **19**(3): 356-63.

130. information-dipeptidyl-peptidase-4-dpp-4-inhibitors. 2016.

131. Hsia DS, Grove O, Cefalu WT. An update on sodium-glucose co-transporter-2 inhibitors for the treatment of diabetes mellitus. *Current opinion in endocrinology, diabetes, and obesity* 2017; **24**(1): 73-9.

132. Pastromas SC, Koulouris SN. Thiazolidinediones: antidiabetic drugs with cardiovascular effects. *Hellenic journal of cardiology : HJC = Hellenike kardiologike epitheorese* 2006; **47 6**: 352-60.

133. Genovese S, De Berardis G, Nicolucci A, et al. Effect of pioglitazone versus metformin on cardiovascular risk markers in type 2 diabetes. *Adv Ther* 2013; **30**(2): 190-202.

134. Valdés-Ramos R, Guadarrama-López AL, Martínez-Carrillo BE, Benítez-Arciniega AD. Vitamins and type 2 diabetes mellitus. *Endocrine, metabolic & immune disorders drug targets* 2015; **15**(1): 54-63.

135. Mascolo E, Vernì F. Vitamin B6 and Diabetes: Relationship and Molecular Mechanisms. *Int J Mol Sci* 2020; **21**(10).

136. Cutillas-Marco E, Prosper AF, Grant WB, Morales-Suarez-Varela MM. Vitamin D status and hypercholesterolemia in Spanish general population. *Dermatoendocrinol* 2013; **5**(3): 358-62.

137. Meerza D, Naseem I, Ahmed J. Effect of 1, 25(OH)(2) vitamin D(3) on glucose homeostasis and DNA damage in type 2 diabetic mice. *J Diabetes Complications* 2012; **26**(5): 363-8.

138. Lee JI, Oh SJ, Ha WC, et al. Serum 25-hydroxyvitamin D concentration and arterial stiffness among type 2 diabetes. *Diabetes Res Clin Pract* 2012; **95**(1): 42-7.

139. Nimitphong H, Chailurkit LO, Chanprasertyothin S, Sritara P, Ongphiphadhanakul B. The Association of vitamin D status and fasting glucose according to body fat mass in young healthy Thais. *BMC Endocr Disord* 2013; **13**: 60.

140. Al-Shoumer KA, Al-Asoosi AA, Ali AH, Nair VS. Does insulin resistance in type 2 diabetes alter vitamin D status? *Prim Care Diabetes* 2013; **7**(4): 283-7.

141. Shore-Lorenti C, Brennan SL, Sanders KM, Neale RE, Lucas RM, Ebeling PR. Shining the light on Sunshine: a systematic review of the influence of sun exposure on type 2 diabetes mellitus-related outcomes. *Clin Endocrinol (Oxf)* 2014; **81**(6): 799-811.

142. Fu J, Hou C, Li L, et al. Vitamin D modifies the associations between circulating betatrophin and cardiometabolic risk factors among youths at risk for metabolic syndrome. *Cardiovasc Diabetol* 2016; **15**(1): 142.

143. Strobel F, Reusch J, Penna-Martinez M, et al. Effect of a randomised controlled vitamin D trial on insulin resistance and glucose metabolism in patients with type 2 diabetes mellitus. *Horm Metab Res* 2014; **46**(1): 54-8.

144. Leong A, Rehman W, Dastani Z, et al. The causal effect of vitamin D binding protein (DBP) levels on calcemic and cardiometabolic diseases: a Mendelian randomization study. *PLoS Med* 2014; **11**(10): e1001751.
145. Fagundes GE, Macan TP, Rohr P, et al. Vitamin D3 as adjuvant in the treatment of type 2 diabetes mellitus: modulation of genomic and biochemical instability. *Mutagenesis* 2019.
146. Jiang X, Kiel DP, Kraft P. The genetics of vitamin D. *Bone* 2018.
147. Kim H, Lee H, Yim HW, Kim HS. Association of serum 25-hydroxyvitamin D and diabetes-related factors in Korean adults without diabetes: The Fifth Korea National Health and Nutrition Examination Survey 2010-2012. *Prim Care Diabetes* 2018; **12**(1): 59-65.
148. Zhang J, Song W, Sun Y, Cheng B, Shan A. Changes in glucose metabolism and mRNA expression of IRS-2 in rats exposed to phoxim and the protective effects of vitamin E. *Toxicol Res (Camb)* 2018; **7**(2): 201-10.
149. Mihrshahi S, Gow ML, Baur LA. Contemporary approaches to the prevention and management of paediatric obesity: an Australian focus. *Med J Aust* 2018; **209**(6): 267-74.
150. HORIKAWA C, AIDA R, TANAKA S, et al. Impact of Vitamin B6 Intake on the Risk of Diabetic Retinopathy—Analysis from Multicenter Prospective Study of Japanese Patients with Type 2 Diabetes. *Diabetes* 2018; **67**(Supplement 1): 597-P.
151. Lithell HOL. Effect of Antihypertensive Drugs on Insulin, Glucose, and Lipid Metabolism. *Diabetes Care* 1991; **14**(3): 203-9.
152. Asterios K, Konstantinos T, Panagiotis A, et al. The Effect of Antihypertensive Agents on Insulin Sensitivity, Lipids and Haemostasis. *Current Vascular Pharmacology* 2010; **8**(6): 792-803.
153. Liu Y-x. A new antihypertensive drug ameliorate insulin resistance. *Acta Pharmacologica Sinica* 2012; **33**(4): 429-30.
154. Xiao W-y, Ning N, Tan M-H, et al. Effects of antihypertensive drugs losartan and levamlodipine besylate on insulin resistance in patients with essential hypertension combined with isolated impaired fasting glucose. *Hypertension Research* 2016; **39**(5): 321-6.
155. Mikhailova OO, Litvin AY, Rogoza AN. [Influence of modifiable cardiovascular risk factors on antihypertensive therapy efficiency escape]. *Ter Arkh* 2017; **89**(9): 10-4.
156. Islam MZ, Estivill-Castro V, Rahman MA, Bossomaier T. Combining K-Means and a genetic algorithm through a novel arrangement of genetic operators for high quality clustering. *Expert Systems with Applications* 2018; **91**: 402-17.
157. Pagnuco IA, Pastore JI, Abras G, Brun M, Ballarin VL. Analysis of genetic association using hierarchical clustering and cluster validation indices. *Genomics* 2017; **109**(5): 438-45.
158. Zaharia OP, Strassburger K, Strom A, et al. Risk of diabetes-associated diseases in subgroups of patients with recent-onset diabetes: a 5-year follow-up study. *Lancet Diabetes Endocrinol* 2019; **7**(9): 684-94.

159. Slieker RC, Donnelly LA, Fitipaldi H, et al. Replication and cross-validation of type 2 diabetes subtypes based on clinical variables: an IMI-RHAPSODY study. *Diabetologia* 2021; **64**(9): 1982-9.
160. Anjana RM, Baskar V, Nair ATN, et al. Novel subgroups of type 2 diabetes and their association with microvascular outcomes in an Asian Indian population: a data-driven cluster analysis: the INSPIRED study. *BMJ Open Diabetes Research & Care* 2020; **8**(1): e001506.
161. Bello-Chavolla OY, Bahena-López JP, Vargas-Vázquez A, et al. Clinical characterization of data-driven diabetes subgroups in Mexicans using a reproducible machine learning approach. *BMJ Open Diabetes Research & Care* 2020; **8**(1): e001550.
162. Dennis JM, Shields BM, Henley WE, Jones AG, Hattersley AT. Disease progression and treatment response in data-driven subgroups of type 2 diabetes compared with models based on simple clinical features: an analysis using clinical trial data. *Lancet Diabetes Endocrinol* 2019; **7**(6): 442-51.
163. Ahlqvist E, Prasad RB, Groop L. Subtypes of Type 2 Diabetes Determined From Clinical Parameters. *Diabetes* 2020; **69**(10): 2086-93.
164. Udler MS, Kim J, von Grotthuss M, et al. Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: A soft clustering analysis. *PLoS Med* 2018; **15**(9): e1002654.
165. Mahajan A, Taliun D, Thurner M, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nature Genetics* 2018; **50**(11): 1505-13.
166. Zeggini E. Next-generation association studies for complex traits. *Nat Genet* 2011; **43**(4): 287-8.
167. Rao DC. An overview of the genetic dissection of complex traits. *Adv Genet* 2008; **60**: 3-34.
168. Bowden DW, Akots G, Rothschild CB, et al. Linkage analysis of maturity-onset diabetes of the young (MODY): genetic heterogeneity and nonpenetrance. *Am J Hum Genet* 1992; **50**(3): 607-18.
169. Thomas PD, Campbell MJ, Kejariwal A, et al. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res* 2003; **13**(9): 2129-41.
170. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007; **39**(7): 906-13.
171. Wu Y, Zeng J, Zhang F, et al. Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nature Communications* 2018; **9**(1): 918.
172. Ewens WJ. Genetic Variation. In: Maloy S, Hughes K, eds. Brenner's Encyclopedia of Genetics (Second Edition). San Diego: Academic Press; 2013: 290-1.
173. Nedoszytko B, Arock M, Lyons JJ, et al. Clinical Impact of Inherited and Acquired Genetic Variants in Mastocytosis. *International journal of molecular sciences* 2021; **22**(1): 411.

174. Zhang N, Gong Y. Clinical Significance of Germ Line and Acquired Somatic Mutations of 30 High Frequency Genes in Hematologic Malignancies. *Blood* 2018; **132**(Supplement 1): 5126-.
175. Siu H, Zhu Y, Jin L, Xiong M. Implication of next-generation sequencing on association studies. *BMC Genomics* 2011; **12**: 322.
176. Spencer CC, Su Z, Donnelly P, Marchini J. Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. *PLoS Genet* 2009; **5**(5): e1000477.
177. Wray NV, P. . Estimating trait heritability. *Nature Education* 2008; **1(1):29**.
178. Ni G, Moser G, Schizophrenia Working Group of the Psychiatric Genomics C, Wray NR, Lee SH. Estimation of Genetic Correlation via Linkage Disequilibrium Score Regression and Genomic Restricted Maximum Likelihood. *American journal of human genetics* 2018; **102**(6): 1185-94.
179. Sahu M, Prasuna JG. Twin Studies: A Unique Epidemiological Tool. *Indian J Community Med* 2016; **41**(3): 177-82.
180. Boomsma D, Busjahn A, Peltonen L. Classical twin studies and beyond. *Nature Reviews Genetics* 2002; **3**(11): 872-82.
181. Cerolsaletti K, Hao W, Greenbaum CJ. Genetics Coming of Age in Type 1 Diabetes. *Diabetes Care* 2019; **42**(2): 189-91.
182. Hyttinen V, Kaprio J, Kinnunen L, Koskenvuo M, Tuomilehto J. Genetic Liability of Type 1 Diabetes and the Onset Age Among 22,650 Young Finnish Twin Pairs. *A Nationwide Follow-Up Study* 2003; **52**(4): 1052-5.
183. Kyvik KO, Green A, Beck-Nielsen H. Concordance rates of insulin dependent diabetes mellitus: a population based study of young Danish twins. *BMJ (Clinical research ed)* 1995; **311**(7010): 913-7.
184. Poulsen P, Grunnet LG, Pilgaard K, et al. Increased Risk of Type 2 Diabetes in Elderly Twins. *Diabetes* 2009; **58**(6): 1350-5.
185. Yu Z, Garner C, Ziogas A, Anton-Culver H, Schaid DJ. Genotype determination for polymorphisms in linkage disequilibrium. *BMC Bioinformatics* 2009; **10**: 63.
186. Todd JA, Farrall M. Panning for gold: genome-wide scanning for linkage in type 1 diabetes. *Hum Mol Genet* 1996; **5 Spec No**: 1443-8.
187. She J-X, Marron MP. Genetic susceptibility factors in type 1 diabetes: linkage, disequilibrium and functional analyses. *Current Opinion in Immunology* 1998; **10**(6): 682-9.
188. Todd JA. Genetics of type 1 diabetes. *Pathologie-biologie* 1997; **45**(3): 219-27.
189. Tripathy D, Eriksson KF, Orho-Melander M, Fredriksson J, Ahlqvist G, Groop L. Parallel manifestation of insulin resistance and beta cell decompensation is compatible with a common defect in Type 2 diabetes. *Diabetologia* 2004; **47**(5): 782-93.
190. Song Y, Niu T, Manson JE, Kwiatkowski DJ, Liu S. Are variants in the CAPN10 gene related to risk of type 2 diabetes? A quantitative assessment of population and family-based association studies. *American journal of human genetics* 2004; **74**(2): 208-22.

191. Duggirala R, Blangero J, Almasy L, et al. Linkage of type 2 diabetes mellitus and of age at onset to a genetic location on chromosome 10q in Mexican Americans. *Am J Hum Genet* 1999; **64**(4): 1127-40.
192. Grant SF, Thorleifsson G, Reynisdottir I, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet* 2006; **38**(3): 320-3.
193. del Bosque-Plata L, Martínez-Martínez E, Espinoza-Camacho MÁ, Gragnoli C. The Role of TCF7L2 in Type 2 Diabetes. *Diabetes* 2021; **70**(6): 1220.
194. Barroso I, Luan Ja, Middelberg RPS, et al. Candidate gene association study in type 2 diabetes indicates a role for genes involved in beta-cell function as well as insulin action. *PLoS biology* 2003; **1**(1): E20-E.
195. Lucas G, Lluis-Ganella C, Subirana I, et al. Post-genomic update on a classical candidate gene for coronary artery disease: ESR1. *Circ Cardiovasc Genet* 2011; **4**(6): 647-54.
196. Kahler AK, Djurovic S, Rimol LM, et al. Candidate gene analysis of the human natural killer-1 carbohydrate pathway and perineuronal nets in schizophrenia: B3GAT2 is associated with disease risk and cortical surface area. *Biol Psychiatry* 2011; **69**(1): 90-6.
197. Minton JAL, Hattersley AT, Owen K, et al. Association Studies of Genetic Variation in the WFS1 Gene and Type 2 Diabetes in U.K. Populations. *Diabetes* 2002; **51**(4): 1287-90.
198. Frayling TM, Timpson NJ, Weedon MN, 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.
199. Dina C, Meyre D, Gallina S, et al. Variation in FTO contributes to childhood obesity and severe adult obesity. *Nat Genet* 2007; **39**(6): 724-6.
200. Diabetes Genetics Initiative of Broad Institute of H, Mit LU, Novartis Institutes of BioMedical R, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007; **316**(5829): 1331-6.
201. Sanghera DK, Ortega L, Han S, et al. Impact of nine common type 2 diabetes risk polymorphisms in Asian Indian Sikhs: PPARG2 (Pro12Ala), IGF2BP2, TCF7L2 and FTO variants confer a significant risk. *BMC Med Genet* 2008; **9**: 59.
202. Zeggini E, Weedon MN, Lindgren CM, et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 2007; **316**(5829): 1336-41.
203. Strawbridge RJ, Dupuis J, Prokopenko I, et al. Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. *Diabetes* 2011; **60**(10): 2624-34.
204. Ahlqvist E, Ahluwalia TS, Groop L. Genetics of type 2 diabetes. *Clinical chemistry* 2011; **57**(2): 241-54.
205. Tabassum R, Chauhan G, Dwivedi OP, et al. Genome-wide association study for type 2 diabetes in Indians identifies a new susceptibility locus at 2q21. *Diabetes* 2013; **62**(3): 977-86.

206. Steinthorsdottir V, Thorleifsson G, Sulem P, et al. Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. *Nat Genet* 2014; **46**(3): 294-8.
207. Go MJ, Lee Y, Park S, Kwak SH, Kim BJ, Lee J. Genetic-risk assessment of GWAS-derived susceptibility loci for type 2 diabetes in a 10 year follow-up of a population-based cohort study. *J Hum Genet* 2016; **61**(12): 1009-12.
208. Fuchsberger C, Flannick J, Teslovich TM, et al. The genetic architecture of type 2 diabetes. *Nature* 2016; **536**(7614): 41-7.
209. Wang X, Strizich G, Hu Y, Wang T, Kaplan RC, Qi Q. Genetic markers of type 2 diabetes: Progress in genome-wide association studies and clinical application for risk prediction. *J Diabetes* 2016; **8**(1): 24-35.
210. Scott RA, Scott LJ, Magi R, et al. An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. *Diabetes* 2017; **66**(11): 2888-902.
211. Onengut-Gumuscu S, Chen W-M, Burren O, et al. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nature genetics* 2015; **47**(4): 381-6.
212. Spracklen CN, Horikoshi M, Kim YJ, et al. Identification of type 2 diabetes loci in 433,540 East Asian individuals. *Nature* 2020; **582**(7811): 240-5.
213. Morris AP, Voight BF, Teslovich TM, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* 2012; **44**(9): 981-90.
214. Chen J, Spracklen CN, Marenne G, et al. The trans-ancestral genomic architecture of glycemic traits. *Nat Genet* 2021; **53**(6): 840-60.
215. Karras S, Goulis DG, Mintziori G, Katsiki N, Tzotzas T. The effects of incretins on energy homeostasis: physiology and implications for the treatment of type 2 diabetes mellitus and obesity. *Current vascular pharmacology* 2012; **10**(6): 781-91.
216. Mountjoy E, Schmidt EM, Carmona M, et al. An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nature Genetics* 2021.
217. Buniello A, Jacqueline A, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research* 2019; **47**(D1): D1005-D12.
218. Cai M, Bompada P, Salehi A, et al. Role of osteopontin and its regulation in pancreatic islet. *Biochem Biophys Res Commun* 2018; **495**(1): 1426-31.
219. Wu C, Borné Y, Gao R, et al. Elevated circulating follistatin associates with an increased risk of type 2 diabetes. *Nature Communications* 2021; **12**(1): 6486.
220. Paananen J, Fortino V. An omics perspective on drug target discovery platforms. *Briefings in Bioinformatics* 2020; **21**(6): 1937-53.
221. Gillberg L, Jacobsen SC, Ribel-Madsen R, et al. Does DNA methylation of PPARGC1A influence insulin action in first degree relatives of patients with type 2 diabetes? *PLoS One* 2013; **8**(3): e58384.
222. Ling C, Rönn T. Epigenetics in Human Obesity and Type 2 Diabetes. *Cell metabolism* 2019; **29**(5): 1028-44.

223. Bysani M, Agren R, Davegårdh C, et al. ATAC-seq reveals alterations in open chromatin in pancreatic islets from subjects with type 2 diabetes. *Sci Rep* 2019; **9**(1): 7785.
224. Hall E, Jönsson J, Ofori JK, et al. Glucolipotoxicity Alters Insulin Secretion via Epigenetic Changes in Human Islets. *Diabetes* 2019; **68**(10): 1965-74.
225. Garcia-Calzon S, Perfilyev A, Martinell M, et al. Epigenetic markers associated with metformin response and intolerance in drug-naive patients with type 2 diabetes. *Science translational medicine* 2020; **12**(561).
226. Broholm C, Ribel-Madsen R, Hjort L, et al. Epigenome- and Transcriptome-wide Changes in Muscle Stem Cells from Low Birth Weight Men. *Endocr Res* 2020; **45**(1): 58-71.
227. Bysani M, Agren R, Davegårdh C, et al. Author Correction: ATAC-seq reveals alterations in open chromatin in pancreatic islets from subjects with type 2 diabetes. *Sci Rep* 2020; **10**(1): 1744.
228. Davegårdh C, Säll J, Benrick A, et al. VPS39-deficiency observed in type 2 diabetes impairs muscle stem cell differentiation via altered autophagy and epigenetics. *Nat Commun* 2021; **12**(1): 2431.
229. Perfilyev A, Dahlman I, Gillberg L, et al. Impact of polyunsaturated and saturated fat overfeeding on the DNA-methylation pattern in human adipose tissue: a randomized controlled trial. *Am J Clin Nutr* 2017; **105**(4): 991-1000.
230. Davegårdh C, Broholm C, Perfilyev A, et al. Abnormal epigenetic changes during differentiation of human skeletal muscle stem cells from obese subjects. *BMC medicine* 2017; **15**(1): 39.
231. Volkov P, Bacos K, Ofori JK, et al. Whole-Genome Bisulfite Sequencing of Human Pancreatic Islets Reveals Novel Differentially Methylated Regions in Type 2 Diabetes Pathogenesis. *Diabetes* 2017; **66**(4): 1074-85.
232. Rönn T, Volkov P, Davegårdh C, et al. A six months exercise intervention influences the genome-wide DNA methylation pattern in human adipose tissue. *PLoS Genet* 2013; **9**(6): e1003572.
233. Hall E, Volkov P, Dayeh T, et al. Effects of palmitate on genome-wide mRNA expression and DNA methylation patterns in human pancreatic islets. *BMC medicine* 2014; **12**: 103.
234. Olsson AH, Volkov P, Bacos K, et al. Genome-wide associations between genetic and epigenetic variation influence mRNA expression and insulin secretion in human pancreatic islets. *PLoS Genet* 2014; **10**(11): e1004735.
235. Bacos K, Gillberg L, Volkov P, et al. Blood-based biomarkers of age-associated epigenetic changes in human islets associate with insulin secretion and diabetes. *Nat Commun* 2016; **7**: 11089.
236. Hall E, Dekker Nitert M, Volkov P, et al. The effects of high glucose exposure on global gene expression and DNA methylation in human pancreatic islets. *Mol Cell Endocrinol* 2018; **472**: 57-67.
237. Saxena A, Mathur N, Tiwari P, Mathur SK. Whole transcriptome RNA-seq reveals key regulatory factors involved in type 2 diabetes pathology in peripheral fat of Asian Indians. *Scientific Reports* 2021; **11**(1): 10632.

238. Chapter 17 - Diabetic Nephropathy. In: Clarkson MR, Magee CN, Brenner BM, eds. Pocket Companion to Brenner and Rector's The Kidney (Eighth Edition). Philadelphia: W.B. Saunders; 2011: 345-59.
239. Lote CJ. Principles of Renal Physiology; Fifth Edition.: Springer; 2012.
240. J. GORDON BETTS TJC, PETER DESAIX UONCACH, EDDIE JOHNSON COCC, et al. Anatomy & Physiology.
241. Oddsson Á, Patrakka J, Tryggvason K. Glomerular Filtration Barrier. Elsevier; 2014.
242. Breshears MA, Confer AW. Chapter 11 - The Urinary System1. In: Zachary JF, ed. Pathologic Basis of Veterinary Disease (Sixth Edition): Mosby; 2017: 617-81.e1.
243. Trimarchi H. Mechanisms of Podocyte Detachment, Podocyturia, and Risk of Progression of Glomerulopathies. *Kidney Diseases* 2020; **6**(5): 324-9.
244. NIDDK TNIoDaDaKD. Diabetic Kidney Disease. 2021.
245. Vallon V, Komers R. Pathophysiology of the diabetic kidney. *Compr Physiol* 2011; **1**(3): 1175-232.
246. Gheith O, Farouk N, Nampoory N, Halim MA, Al-Otaibi T. Diabetic kidney disease: world wide difference of prevalence and risk factors. *J Nephropharmacol* 2015; **5**(1): 49-56.
247. Comai G, Malvi D, Angeletti A, et al. Histological Evidence of Diabetic Kidney Disease Precede Clinical Diagnosis. *American journal of nephrology* 2019; **50**(1): 29-36.
248. Gu HF. Genetic and Epigenetic Studies in Diabetic Kidney Disease. *Frontiers in Genetics* 2019; **10**(507).
249. Böger CA, Gorski M, Li M, et al. Association of eGFR-Related Loci Identified by GWAS with Incident CKD and ESRD. *PLoS Genetics* 2011; **7**(9): e1002292.
250. Li M, Pezzolesi MG. Advances in understanding the genetic basis of diabetic kidney disease. *Acta Diabetologica* 2018; **55**(11): 1093-104.
251. Gu HF. Genetic and Epigenetic Studies in Diabetic Kidney Disease. *Frontiers in genetics* 2019; **10**: 507-.
252. Afkarian M. Diabetic kidney disease in children and adolescents. *Pediatric nephrology (Berlin, Germany)* 2015; **30**(1): 65-71.
253. Thomas MC, Brownlee M, Susztak K, et al. Diabetic kidney disease. *Nature Reviews Disease Primers* 2015; **1**(1): 15018.
254. van Zuydam NR, Ahlqvist E, Sandholm N, et al. A Genome-Wide Association Study of Diabetic Kidney Disease in Subjects With Type 2 Diabetes. *Diabetes* 2018; **67**(7): 1414-27.
255. Nakopoulou L, Lazaris AC, Boletis IN, et al. Evaluation of E-cadherin/catenin complex in primary and secondary glomerulonephritis. *American Journal of Kidney Diseases* 2002; **39**(3): 469-74.
256. Wahba IM, Mak RH. Obesity and Obesity-Initiated Metabolic Syndrome: Mechanistic Links to Chronic Kidney Disease. *Clinical Journal of the American Society of Nephrology* 2007; **2**(3): 550.

257. Dai C, Stolz DB, Kiss LP, Monga SP, Holzman LB, Liu Y. Wnt/beta-catenin signaling promotes podocyte dysfunction and albuminuria. *Journal of the American Society of Nephrology : JASN* 2009; **20**(9): 1997-2008.
258. Fagerholm E, Ahlqvist E, Forsblom C, et al. SNP in the genome-wide association study hotspot on chromosome 9p21 confers susceptibility to diabetic nephropathy in type 1 diabetes. *Diabetologia* 2012; **55**(9): 2386-93.
259. Tang W, Fu YP, Figueroa JD, et al. Mapping of the UGT1A locus identifies an uncommon coding variant that affects mRNA expression and protects from bladder cancer. *Hum Mol Genet* 2012; **21**(8): 1918-30.
260. Alkayyali S, Lajer M, Deshmukh H, et al. Common variant in the HMGA2 gene increases susceptibility to nephropathy in patients with type 2 diabetes. *Diabetologia* 2013; **56**(2): 323-9.
261. Ahlqvist E, van Zuydam NR, Groop LC, McCarthy MI. The genetics of diabetic complications. *Nat Rev Nephrol* 2015; **11**(5): 277-87.
262. Reznichenko A, Böger CA, Snieder H, et al. UMOD as a susceptibility gene for end-stage renal disease. *BMC Medical Genetics* 2012; **13**(1): 78.
263. Trudu M, Janas S, Lanzani C, et al. Common noncoding UMOD gene variants induce salt-sensitive hypertension and kidney damage by increasing uromodulin expression. *Nature Medicine* 2013; **19**(12): 1655-60.
264. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nature Protocols* 2010; **5**(9): 1564-73.
265. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015; **4**: 7.
266. Shi S, Yuan N, Yang M, et al. Comprehensive Assessment of Genotype Imputation Performance. *Human Heredity* 2018; **83**(3): 107-16.
267. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* 2012; **44**(8): 955-9.
268. Fuchsberger C, Abecasis GR, Hinds DA. minimac2: faster genotype imputation. *Bioinformatics* 2015; **31**(5): 782-4.
269. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *American journal of human genetics* 2011; **88**(1): 76-82.
270. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* 2007; **81**(3): 559-75.
271. Li X, Li H, Jin M, J DG. Likelihood ratio and score tests to test the non-inferiority (or equivalence) of the odds ratio in a crossover study with binary outcomes. *Statistics in medicine* 2016; **35**(20): 3471-81.
272. Kraft P, Zeggini E, Ioannidis JPA. Replication in genome-wide association studies. *Stat Sci* 2009; **24**(4): 561-73.
273. Zeggini E, Ioannidis JPA. Meta-analysis in genome-wide association studies. *Pharmacogenomics* 2009; **10**(2): 191-201.

274. Mägi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* 2010; **11**(1): 288.
275. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; **26**(17): 2190-1.
276. Willer CJ, Sanna S, Jackson AU, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet* 2008; **40**(2): 161-9.
277. Sanna S, Jackson AU, Nagaraja R, et al. Common variants in the GDF5-UQCC region are associated with variation in human height. *Nat Genet* 2008; **40**(2): 198-203.
278. Karssen LC, van Duijn CM, Aulchenko YS. The GenABEL Project for statistical genomics. *F1000Research* 2016; **5**: 914-.
279. Boughton AP, Welch RP, Flickinger M, et al. LocusZoom.js: Interactive and embeddable visualization of genetic association study results. *Bioinformatics (Oxford, England)* 2021; **37**(18): 3017-8.
280. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. *Genome Biology* 2016; **17**(1): 122.
281. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nature Biotechnology* 2011; **29**(1): 24-6.
282. Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP. Variant Review with the Integrative Genomics Viewer. *Cancer Research* 2017; **77**(21): e31-e4.
283. Gamazon ER, Segrè AV, van de Bunt M, et al. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nature Genetics* 2018; **50**(7): 956-67.
284. reviewed]. F1000Research , 5:2607 (poster) (https://doi.org/10.7490/f1000research.1113352.1) CMaAMPTDTtdkpapftdohgivnp. 2016.
285. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015; **12**(3): e1001779.
286. Duncan LE, Shen H, Ballon JS, Hardy KV, Noordsy DL, Levinson DF. Genetic Correlation Profile of Schizophrenia Mirrors Epidemiological Results and Suggests Link Between Polygenic and Rare Variant (22q11.2) Cases of Schizophrenia. *Schizophr Bull* 2018; **44**(6): 1350-61.
287. Zheng J, Erzurumluoglu AM, Elsworth BL, et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics (Oxford, England)* 2017; **33**(2): 272-9.
288. Xu C, Jackson SA. Machine learning and complex biological data. *Genome Biology* 2019; **20**(1): 76.
289. Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. *Nature Reviews Molecular Cell Biology* 2021.

290. Liu C, Che D, Liu X, Song Y. Applications of machine learning in genomics and systems biology. *Computational and mathematical methods in medicine* 2013; **2013**: 587492.
291. Feltes BC, Grisci BI, Poloni JF, Dorn M. Perspectives and applications of machine learning for evolutionary developmental biology. *Molecular omics* 2018; **14**(5): 289-306.
292. Liudmila Prokhorenkova GG, Aleksandr Vorobev, Anna Veronika Dorogush, Andrey Gulin. CatBoost: unbiased boosting with categorical features. (arXiv:1706.09516v5 [cs.LG]).
293. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; **25**(1): 25-9.
294. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Research* 2013; **42**(D1): D199-D205.
295. Szklarczyk D, Gable AL, Nastou KC, et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* 2021; **49**(D1): D605-d12.
296. Croft D, O'Kelly G, Wu G, et al. Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res* 2011; **39**(Database issue): D691-7.
297. Li B, Wang Z, Chen Q, et al. GPCards: An integrated database of genotype–phenotype correlations in human genetic diseases. *Computational and Structural Biotechnology Journal* 2021; **19**: 1603-11.
298. Raynal C, Corvol H. Variant classifications, databases and genotype-phenotype correlations. *Archives de pediatrie : organe officiel de la Societe francaise de pediatrie* 2020; **27 Suppl 1**: eS13-eS8.
299. BioCarta. *Biotech Software & Internet Report* 2001; **2**(3): 117-20.
300. Lamparter D, Marbach D, Rueedi R, Kutalik Z, Bergmann S. Fast and Rigorous Computation of Gene and Pathway Scores from SNP-Based Summary Statistics. *PLOS Computational Biology* 2016; **12**(1): e1004714.
301. Pers TH, Karjalainen JM, Chan Y, et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nature Communications* 2015; **6**(1): 5890.
302. Jassal B, Matthews L, Viteri G, et al. The reactome pathway knowledgebase. *Nucleic Acids Res* 2020; **48**(D1): D498-d503.
303. Textor J, van der Zander B, Gilthorpe MS, Liśkiewicz M, Ellison GT. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *International journal of epidemiology* 2017; **45**(6): 1887-94.
304. Fisher RA. Statistical methods in genetics. 1951. *International journal of epidemiology* 2010; **39**(2): 329-35.
305. Udler MS, McCarthy MI, Florez JC, Mahajan A. Genetic Risk Scores for Diabetes Diagnosis and Precision Medicine. *Endocrine reviews* 2019; **40**(6): 1500-20.
306. Noble JA, Valdes AM. Genetics of the HLA region in the prediction of type 1 diabetes. *Current diabetes reports* 2011; **11**(6): 533-42.

307. Nguyen C, Varney MD, Harrison LC, Morahan G. Definition of High-Risk Type 1 Diabetes HLA-DR and HLA-DQ Types Using Only Three Single Nucleotide Polymorphisms. *Diabetes* 2013; **62**(6): 2135-40.
308. Ebenibo S, Edeoga C, Owei I, Dagogo-Jack S. Basal and Dynamic Leptin Secretion: Association with Cardiometabolic Risk and Body Weight Trajectories in African-Americans and European-Americans. *Front Endocrinol (Lausanne)* 2018; **9**: 12.
309. Buzga M, Zavadilova V, Holeczy P, et al. Dietary intake and ghrelin and leptin changes after sleeve gastrectomy. *Wideochir Inne Tech Maloinwazyjne* 2014; **9**(4): 554-61.
310. Breij LM, Mulder MT, van Vark-van der Zee LC, Hokken-Koelega ACS. Appetite-regulating hormones in early life and relationships with type of feeding and body composition in healthy term infants. *Eur J Nutr* 2017; **56**(4): 1725-32.
311. Becer E, Mehmetcik G, Bareke H, Serakinci N. Association of leptin receptor gene Q223R polymorphism on lipid profiles in comparison study between obese and non-obese subjects. *Gene* 2013; **529**(1): 16-20.
312. Mack R, Skurnick B, Sterling-Jean Y, Pedra-Nobre M, Bigg D. Fasting insulin levels as a measure of insulin resistance in American blacks. *J Med* 2003; **34**(1-6): 31-8.
313. Then C, Then H, Lechner A, et al. Serum uromodulin is inversely associated with the metabolic syndrome in the KORA F4 study. *Endocr Connect* 2019; **8**(10): 1363-71.
314. Olivier-Van Stichelen S, Guinez C, Mir AM, et al. The hexosamine biosynthetic pathway and O-GlcNAcylation drive the expression of beta-catenin and cell proliferation. *Am J Physiol Endocrinol Metab* 2012; **302**(4): E417-24.
315. Ha T-S. Roles of adaptor proteins in podocyte biology. *World J Nephrol* 2013; **2**(1): 1-10.
316. Zhou F, Huo J, Liu Y, et al. Elevated glucose levels impair the WNT/beta-catenin pathway via the activation of the hexosamine biosynthesis pathway in endometrial cancer. *J Steroid Biochem Mol Biol* 2016; **159**: 19-25.
317. Li Y, Chen Q, Ran D, et al. Changes in the levels of 12/15-lipoxygenase, apoptosis-related proteins and inflammatory factors in the cortex of diabetic rats and the neuroprotection of baicalein. *Free Radic Biol Med* 2019; **134**: 239-47.
318. Lager S, Ramirez VI, Acosta O, et al. Docosahexaenoic Acid Supplementation in Pregnancy Modulates Placental Cellular Signaling and Nutrient Transport Capacity in Obese Women. *J Clin Endocrinol Metab* 2017; **102**(12): 4557-67.
319. Opstad TB, Arnesen H, Pettersen AA, Seljeflot I. Combined Elevated Levels of the Proinflammatory Cytokines IL-18 and IL-12 Are Associated with Clinical Events in Patients with Coronary Artery Disease: An Observational Study. *Metab Syndr Relat Disord* 2016; **14**(5): 242-8.
320. Valverde AM, Lorenzo M, Teruel T, Benito M. Alterations in the insulin signaling pathway induced by immortalization and H-ras transformation of brown adipocytes. *Endocrinology* 1997; **138**(8): 3195-206.
321. Goossens GH. The Renin-Angiotensin System in the Pathophysiology of Type 2 Diabetes. *Obesity Facts* 2012; **5**(4): 611-24.

322. Zhou MS, Schulman IH, Zeng Q. Link between the renin-angiotensin system and insulin resistance: implications for cardiovascular disease. *Vascular medicine (London, England)* 2012; **17**(5): 330-41.
323. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 2020; **369**(6509): 1318-30.
324. Isidor MS, Dong W, Servin-Uribe RI, et al. Insulin resistance rewires the metabolic gene program and glucose utilization in human white adipocytes. *International Journal of Obesity* 2021.
325. Obradovic M, Sudar-Milovanovic E, Soskic S, et al. Leptin and Obesity: Role and Clinical Implication. *Frontiers in Endocrinology* 2021; **12**(563).
326. Adamska-Patruno E, Ostrowska L, Goscik J, Pietraszewska B, Kretowski A, Gorska M. The relationship between the leptin/ghrelin ratio and meals with various macronutrient contents in men with different nutritional status: a randomized crossover study. *Nutrition Journal* 2018; **17**(1): 118.
327. Klok MD, Jakobsdottir S, Drent ML. The role of leptin and ghrelin in the regulation of food intake and body weight in humans: a review. *Obes Rev* 2007; **8**(1): 21-34.
328. Huynh FK, Levi J, Denroche HC, et al. Disruption of hepatic leptin signaling protects mice from age- and diet-related glucose intolerance. *Diabetes* 2010; **59**(12): 3032-40.
329. Makarova EN, Chepeleva EV, Panchenko PE, Bazhan NM. Influence of abnormally high leptin levels during pregnancy on metabolic phenotypes in progeny mice. *Am J Physiol Regul Integr Comp Physiol* 2013; **305**(11): R1268-80.
330. Enwere E, Shingo T, Gregg C, Fujikawa H, Ohta S, Weiss S. Aging results in reduced epidermal growth factor receptor signaling, diminished olfactory neurogenesis, and deficits in fine olfactory discrimination. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 2004; **24**(38): 8354-65.
331. Bastian C, Zaleski J, Stahon K, et al. NOS3 Inhibition Confers Post-Ischemic Protection to Young and Aging White Matter Integrity by Conserving Mitochondrial Dynamics and Miro-2 Levels. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 2018; **38**(28): 6247-66.

Glossary

1. **DNA:** Deoxyribonucleic acid consists of two complementary polynucleotide chains that coil to form a double helix.
2. **RNA:** Ribonucleic acid is a single polynucleotide chain and regulates gene expression.
3. **Nucleosome:** DNA wrapped around histone proteins.
4. **Chromatin:** compact dense package of long DNA wrapped around proteins in the nucleus of eukaryotic cells.
5. **Chromosome:** Long DNA strands that carry genetic information. Each diploid cell has 22 pairs of autosomal chromosomes and 1 pair of sex chromosomes. Haploid cells have 23 chromosomes; 22 autosomal and 1 sex chromosome.
6. **Nucleotide:** Constitutes the main unit of DNA and RNA. It is made up of sugar, phosphate, and a nitrogenous base. The sugar is deoxyribose and ribose in DNA AND RNA respectively. The nitrogenous bases in DNA are adenine and its complementary base thymine, guanine, and its complementary base cytosine. RNA has uracil instead of thymine.
7. **Cell cycle:** The four stages of cell division; interphase (G1), synthesis phase (S), condensation phase (G2), and mitosis phase (M).
8. **Mitosis:** Each diploid cell divides into two genetically identical diploid daughter cells. Prophase, anaphase, metaphase, and telophase are the four phases of mitosis.
9. **Meiosis:** Takes place in the ovaries and testis to produce gametes, where each diploid cell divides into four haploid cells (gametes). It has eight steps; the first four are for the crossover and replication of chromosomes and the second four are the same as mitosis.
10. **Gametes:** Ova and sperms for the female and males, respectively. They are genetically different due to the crossover and recombination in meiosis.
11. **Crossing over:** The exchange of genetic material between homologous chromosomes in prophase I of meiosis.

12. **Recombination:** The new combination of chromosomes after crossing over and is important for the genetic diversity of the offspring.
13. **Mutations:** Variation in the DNA**.**
14. **Somatic mutations:** Mutations that take place in the normal cells and not in the gametes.
15. **Germline mutations:** Mutations that take place in the gametes and are passed to the offspring.
16. **Aneuploidy:** Change in chromosome number.
17. **Copy number variations / structural variation:** Number of repeats of a certain part of the genome that varies between individuals.
18. **Point mutations:** Changes in the DNA that occur at one position (nucleotide) on the chromosome and are also called single nucleotide polymorphism (SNP). This change can be within the non-coding regions of the gene; introns or within the coding region; exons.
19. **Exon mutations:** Mutations in the genetic code on the transcribed mRNA that could stop the formation of the encoded protein; stop-codon or the formation of an entirely different protein
20. **Missense mutations:** Changes in the genetic code on the transcribed mRNA could lead to the formation of a different protein.
21. **Insertions/Deletions/Duplications:** Changes where nucleotides are added or subtracted from the DNA sequence.
22. **Translocations:** Abnormal chromosome breaks and rearrangements between non-homologous chromosomes.
23. **Karyotyping:** The technique for ordering and pairing the organism's chromosomes.
24. **Shotgun sequencing:** Sequencing of random DNA strands of an organism's genome by breaking the DNA into small fragments that are sequenced separately.
25. **Menedelian inheritance:** Laws set by Gregor Mendel in 1865.

    Genetic characters are unitary (discrete) and have alternate forms (alleles), each allele is inherited from one parent. The phenotype is described by the dominant allele and assumes independent assortment; genes are not linked.
26. **Menedelian traits:** Inherited monogenic phenotypes caused by one copy of the dominant allele or two copies of the recessive allele.

27. **Genomics:** The study of the genome including all genes, gene-gene interaction, gene-environment interactions.
28. **Epigenomics:** The study of epigenetic modifications (markers that tag the DNA and DNA-associated proteins).
29. **Transcriptomics:** The study of the mRNA and the tissue-specific gene expression.
30. **Gene expression:** Formation of functional protein from the gene.
31. **Transcription:** Formation of messenger RNA complementary to the DNA in the gene.
32. **Transcription factors:** Proteins that bind to the DNA at the promoter region and regulate gene expression.
33. **Proteomics:** The study of structural and functional proteins in the cell, tissue, or organism.
34. **Metabolomics:** The study of the metabolic pathways and the metabolites involved in cell metabolism.
35. **Functional genomics:** The integration of genomics, transcriptomics, and proteomics to understand cell physiology.
36. **System biology:** The use of omics to understand the biological systems.

Printed by Media-Tryck, Lund 2022 NORDIC SWAN ECOLABEL 3041 0903

FACULTY OF MEDICINE

Department of Clinical Sciences Malmö

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2022:5
ISBN 978-91-8021-166-6
ISSN 1652-8220