UNIVERSITY OF LATVIA

Faculty of Biology

Monta Brīvība

DOCTORAL THESIS

MOLECULAR SIGNATURES FOR THERAPEUTIC EFFECTIVENESS AND RISK OF COMPLICATIONS IN TYPE 2 DIABETES MELLITUS

Promotion to the degree of Doctor of Biology

Molecular Biology

Supervisor: Dr. Biol., Prof. Jānis Kloviņš

Riga, 2022

The doctoral thesis was carried out in Latvian Biomedical Research and Study Centre, at Human Genetics and Molecular Medicine department, and Faculty of Biology, Department of Molecular Biology, University of Latvia from 2017 to 2021.

The research was supported by ERDF grant 1.1.1.1/16/A/091.

Form of the thesis: a collection of research papers in biology, subfield – molecular biology.

Supervisor: Prof., Dr. biol. Jānis Kloviņš

Reviewers:

1) PhD. biol. Dace Pjanova, Latvian Biomedical Research and Study Centre
2) PhD. biol., Prof. Una Riekstiņa, University of Latvia
3) PhD., Assoc. Prof. Elin Org, University of Tartu

The thesis will be defended at the public session of the Doctoral Committee of Biology, University of Latvia at 13:00 pm (EET), on March 4th, 2022 at Latvian Biomedical Research and Study Centre, Ratsupites Str. k-1.

The thesis is available at the Library of the University of Latvia, Kalpaka blvd. 4.

This thesis is accepted for the commencement of the degree of Doctor of Biology on December 20th, 2021 by the Doctoral Committee of Biology, University of Latvia.

Chairman of the Doctoral Committee _____/ *vārds, uzvārds/*
(paraksts)

Secretary of the Doctoral Committee _____/ *vārds, uzvārds/*
(paraksts)

ABSTRACT

The increasing prevalence of type 2 diabetes mellitus, a chronic disease associated with risk of disabling complications and premature mortality, requires continuous management and therefore has long become a serious public health concern causing major health expenditures globally. Since long-term complications are directly related to poor glycemic control, they may be prevented or at least delayed to some extent by early intervention. Metformin, a biguanide agent, is the first-line treatment for type 2 diabetes mellitus with an inter-individually variable glucose-lowering effect. Currently, there are no reliable biomarkers for either prediction of metformin response or early risk stratification for diabetic complications in a clinical setting.

The main objective of this thesis was to investigate the underlying mechanisms and molecular signatures representing the variability in the therapeutic response of metformin and complications associated with type 2 diabetes mellitus. We chose the longitudinal RNA-Seq-based comparative transcriptomics approach in peripheral blood cells to evaluate the systemic effect of metformin in healthy individuals and drug-naïve type 2 diabetes patients *in vivo*. Genome-wide genotyping was applied to identify genetic determinants for type 2 diabetes complications in the population of Latvia.

We discovered immediate metformin-induced transcriptional alterations in blood cell gene expression profiles of healthy individuals with notable subject-specific differences, and a strong association between metformin and immune pathways emphasizing the secretory immunoglobulin A-related intestinal immune responses. Our study highlighted potential predictive biomarkers of metformin response and pinpointed the expression of genes coding for NADH: ubiquinone oxidoreductase core subunits as key contributors in metformin effectiveness. Finally, ten novel genetic associations with type 2 diabetes complications were discovered in the population of Latvia. Furthermore, for five of the allelic variants, the nearby genes (*GYPA*, *PDE4DIP*, *NAT8*, *F5*, *RPS6KA2*) have been functionally linked to the pathogenesis of the diseases before. The obtained data provide detailed insight into potential molecular mechanisms underlying the well-known beneficial effects of metformin and potential key determinants of type 2 diabetes complications, highlighting the considerable advantage of the application of omics-based molecular targeting in clinical decision-making.

KOPSAVILKUMS

Pieaugošā 2. tipa cukura diabēta izplatība, kas ir saistīta ar palielinātu hronisku komplikāciju attīstības un priekšlaicīgas mirstības risku, prasa nepārtrauktu uzraudzību, tāpēc ir kļuvusi par nopietnu sabiedrības veselības problēmu, radot ievērojamu finansiālu slogu veselības aprūpes sistēmām visā pasaulē. Ilgtermiņa komplikāciju attīstības risks ir tieši saistīts ar nepietiekamu glikēmijas kontroli, tāpēc agrīna iejaukšanās var to novērst vai aizkavēt. Metformīns, biguanīdu grupas līdzeklis, ir pirmās izvēles medikaments 2. tipa cukura diabēta ārstēšanā, kam ir novērots variabls glikozes līmeni pazeminošs efekts starp 2. tipa cukura diabēta pacientiem. Klīniskajā praksē joprojām trūkst biomarķieru metformīna atbildes reakcijas prognozēšanai un agrīnai diabēta pacientu stratifikācijai pēc komplikāciju attīstības riska.

Šī pētījuma galvenais mērķis bija izpētīt molekulāros mehānismus, kas atspoguļo metformīna terapeitiskā efekta variabilitāti un 2. tipa cukura diabēta komplikāciju attīstības risku. Lai novērtētu metformīna sistēmisko efektu veselos indivīdos un 2. tipa cukura diabēta pacientos, kas iepriekš nav saņēmuši antidiabētisko terapiju, tika izmantota uz RNS sekvencēšanu balstīta *in vivo* transkriptomikas pieeja perifērajās asins šūnās. Ar genoma mēroga genotipēšanu tika identificēti tie ģenētiskie faktori, kas nosaka 2. tipa cukura diabēta komplikāciju risku Latvijas populācijā.

Pētījumā iegūtie dati norādīja uz tūlītējām, metformīna ierosinātām izmaiņām veselu indivīdu asins šūnu gēnu ekspresijas profilos ar ievērojamām indivīd-specifiskām atšķirībām, un ciešu saistību ar imūnsistēmas signālceļiem, tajā skaitā ar sekretoro imunoglobulīnu A saistītām zarnu imūnās atbildes reakcijām. Pētījuma īstenošanas laikā tika identificēti potenciāli, uz gēnu ekspresijas izmaiņām balstīti biomarķieri metformīna efektivitātes prognozēšanai, turklāt iegūtie dati norādīja uz būtisku NADH: ubikvinona oksidoreduktāzes kodola subvienību kodējošo gēnu ekspresijas lomu metformīna efektivitātes modulēšanā. Visbeidzot, Latvijas populācijā tika atklāti desmit jauni, ar diabēta komplikāciju attīstības risku saistīti lokusi, turklāt pieci no alēliskajiem variantiem tuvumā esošajiem gēniem (*GYPA, PDE4DIP, NAT8, F5, RPS6KA2*) ir funkcionāli saistīti ar slimību patoģenēzi. Iegūtie rezultāti sniedz papildu zināšanas par molekulārajiem mehānismiem, kas pamato metformīna labvēlīgos efektus, un diabēta komplikāciju attīstības risku noteicošajiem faktoriem, vienlaikus izceļot priekšrocības, ko sniedz uz omikas datiem balstītas lielapjoma molekulārās analīzes klīnisko lēmumu pieņemšanā.

TABLE OF CONTENTS

ABBREVIATIONS

ABCC2	ATP Binding Cassette Subfamily C Member 2
ACE	Angiotensin I Converting Enzyme
ACVR1B	Activin A Receptor Type 1B
ADIPOQ	Adiponectin, C1q and Collagen Domain Containing
AGEs	Advanced Glycation End Products
AKR1C3	Aldo-Keto Reductase Family 1 Member C3
ALDOC	Aldolase, Fructose-Bisphosphate C
ALMS1	Alstrom Syndrome 1
ALR2	Aldose Reductase
ALT	Alanine Aminotransferase
AMP	Adenosine Monophosphate
AMPK	5' Adenosine Monophosphate-Activated Protein Kinase
ANKH	Progressive Ankylosis Protein Homolog
AP-1	Activator Protein 1
APOB	Apolipoprotein B
APOL1	Apolipoprotein L1
ARHGAP22	Rho GTPase Activating Protein 22
ATF3	Activating Transcription Factor 3
ATM	Ataxia Telangiectasia Mutated
ATP	Adenosine Triphosphate
ATP5G1	ATP Synthase Membrane Subunit C Locus 1
ATXN1	Ataxin 1
BMI	Body Mass Index
CAMK1D	Calcium/Calmodulin Dependent Protein Kinase ID
cAMP	Cyclic Adenosine Monophosphate
CARS	Cysteinyl-tRNA Synthetase
CAT	Catalase
CCDC146	Coiled-Coil Domain Containing 146
CCL4	C-C Motif Chemokine Ligand 4
CCR2	C-C Motif Chemokine Receptor 2
CDC42	Cell Division Cycle 42
CEACAM1	Cea Cell Adhesion Molecule 1
Ch25h	Cholesterol-25-Hydroxylase
Chip-Seq	Chromatin Immunoprecipitation-Based Sequencing
CI	Confidence Interval
CpG	Cytosine-Phosphate-Guanine
CPM	Counts per million
CTGF	Connective Tissue Growth Factor
CTLA4	Cluster Of Differentiation 152
CXCL8	C-X-C Motif Chemokine Ligand 8
CXCR4	C-X-C Chemokine Receptor Type 4
DAVID	Database for Annotation, Visualization and Integrated Discovery
KEGG	Kyoto Encyclopedia of Genes and Genomes
KLF4	Krüppel-Like Factor 4
LD	Linkage disequilibrium
LDLR	Low-Density Lipoprotein Receptor
LGDB	Genome Database of Latvian Population
LPS	Lipopolysaccharide
LRP1	Low Density Lipoprotein Receptor-Related Protein 1
LRP1B	Low-Density Lipoprotein Receptor-Related Protein 1B
M0	Samples collected before metformin treatment
M24h	Samples collected 24 hours after the first metformin dose
M3m	Samples collected 3 months after starting the therapy
M7d	Samples collected 7 days after starting the therapy
MAF	Minor allele frequency
MAP3K14	Mitogen-Activated Protein Kinase Kinase Kinase 14
MAPK14	Mitogen-Activated Protein Kinase 14
MATE1	Multidrug And Toxin Extrusion Protein 1
MATE2	Multidrug And Toxin Extrusion Protein 2
miRNA	Microribonucleic acid
MKRN1	Makorin Ring Finger Protein 1
MODY	Maturity-Onset Diabetes of the Young
mRNA	Messenger RNA
MT-ATP6	Mitochondrially Encoded ATP Synthase Membrane Subunit 6
MT-ND2	Mitochondrially Encoded NADH:Ubiquinone Oxidoreductase Core Subunit 2
MT-ND4	Mitochondrially Encoded NADH:Ubiquinone Oxidoreductase Core Subunit 4
MT-ND4L	Mitochondrially Encoded NADH:Ubiquinone Oxidoreductase Core Subunit 4l
MT-ND6	Mitochondrially Encoded NADH:Ubiquinone Oxidoreductase Core Subunit 6
MYSM1	Histone H2a Deubiquitinase
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
NAT8	N-Acetyltransferase 8
NEUROD1	Neuronal Differentiation 1
NOS3	Nitric Oxide Synthase 3
OCT1	Organic Cation Transporter 1
OCT2	Organic Cation Transporter 2
OCT3	Organic Cation Transporter 3
OCTN1	Organic Cation Transporter, Novel, Type 1
OR	Odds ratio
OR2L13	Olfactory Receptor Family 2 Subfamily L Member 13
PAX4	Paired Box 4
PBMCs	Peripheral blood mononuclear cells
PCSK9	Proprotein Convertase Subtilisin/Kexin Type 9
PDE4DIP	Phosphodiesterase 4d Interacting Protein
PDGF	Platelet-Derived Growth Factor

DEG	Differentially expressed gene
DENND2C	DENN Domain Containing 2C
DMPK	Dystrophia Myotonica Protein Kinase
DNA	Deoxyribonucleic acid
DSC1	Desmocollin 1
ENO2	Enolase 2
ENO3	Enolase 3
EPOR	Erythropoietin Receptor
eQTL	Expression, quantitative trait locus
ER	Endoplasmic reticulum
ERBB3	Erb-B2 Receptor Tyrosine Kinase 3
ET	Endothelin
F5	Coagulation Factor V
FBPase	Fructose-1,6-Bisphosphatase
FDR	False discovery rate
FGL2	Fibrinogen Like 2
FOXO3a	Forkhead Box Protein O3
FRMD3	Ferm Domain Containing 3
GABRR1	Gamma-Aminobutyric Acid Type A Receptor Subunit Rho1
GAD antibodies	Group of diabetes-associated antibodies
GCK	Glucokinase
GLUT2	Glucose Transporter 2
GO	Gene ontology
GPX1	Glutathione Peroxidase 1
GRAF2	GTPase Regulator Associated With The Focal Adhesion Kinase
GRB2	Growth Factor Receptor Bound Protein 2
GTEx	Genotype tissue expression database
GWAS	Genome-wides association studies
GYPA	Glycophorin A
H295R cells	An Angiotensin-II-Responsive Steroid-Producing Adrenocortical Cell Line
HbA1c	Glycohemoglobin, Hemoglobin a1c
HHEX	Hematopoietically Expressed Homeobox
HLA-DQA1	Major Histocompatibility Complex, Class Ii, Dq Alpha 1
HNF1A	Hepatocyte Nuclear Factor 1-Alpha
HNF1B	Hepatocyte Nuclear Factor-1 Beta
HNF4A	Hepatocyte Nuclear Factor 4 Alpha
HOMA-IR	Homeostatic model assessment for insulin resistance
HR	Hazard ratio
HS6ST3	Heparan Sulfate 6-O-Sulfotransferase 3
HSD17B14	Hydroxysteroid 17-Beta Dehydrogenase 14
ICD-10	International Classification Of Diseases Diagnosis 10th Revision
IFIH1	Interferon Induced With Helicase C Domain 1
IFN	Interferon
IFN-α	Interferon Alpha 1
IFN-γ	Interferon Gamma
IGF2BP2	Insulin Like Growth Factor 2 mRNA Binding Protein 2
IGHA1	Immunoglobulin Heavy Constant Alpha 1
INS	Insulin
INSR	Insulin receptor
IQR	Interquartile range
IRS1	Insulin Receptor Substrate 1
IRS-2	Insulin Receptor Substrate 2
PHOSPHO1	Phosphoethanolamine/Phosphocholine Phosphatase 1
KCNJ11	Potassium Inwardly-Rectifying Channel, Subfamily J, Member 11
PDX1	Pancreatic and Duodenal Homeobox 1
PF4	Platelet Factor 4
PI3K	Phosphoinositide 3-Kinase
PKC	Protein Kinase C
PLCB4	Phospholipase C Beta 4
PLS-DA	Partial least squares discriminant analysis
PLXDC2	Plexin Domain Containing 2
PMAT	Plasma Membrane Monoamine Transporter
PRKAG2	Protein Kinase Amp-Activated Non-Catalytic Subunit Gamma 2
Ptger4	Prostaglandin E Receptor 4
PTGFRN	Prostaglandin F2 Receptor Inhibitor
PTPN22	Protein Tyrosine Phosphatase Non-Receptor Type 22
RIN	RNA integrity number
RNA	Ribonucleic acid
ROS	Reactive oxygen species
RP1-90L14.1	Long Non-Coding RNA RP1-90L14.1
Rpl13a	Ribosomal Protein L13a
RPS6KA2	Ribosomal Protein S6 Kinase A2
RR	Risk ratio
SCC	Spearman's correlation coefficient
SCN2A	Sodium Voltage-Gated Channel Alpha Subunit 2
sIgA	Secretory Immunoglobulin A
siRNA	Small interfering RNA
SLC22A1	Solute Carrier Family 22 Member 1
SLC22A2	Solute Carrier Family 22 Member 2
SLC22A3	Solute Carrier Family 22 Member 3
SLC22A4	Solute Carrier Family 22 Member 4
SLC29A4	Solute Carrier Family 29 Member 4
SLC2A2	Solute Carrier Family 2 Member 2
SLC46A1	Solute Carrier Family 46 Member
SLC47A1	Solute Carrier Family 47 Member 1
SLC47A2	Solute Carrier Family 47 Member 2
SNORD115	Small Nucleolar RNA, C/D Box 115-1
SNORD116	Small Nucleolar RNA, C/D Box 116 Cluster
SNP	Single nucleotide polymorphism
SSB	Small RNA Binding Exonuclease Protection Factor La
SULT2A1	Sulfotransferase Family 2a Member 1
T1DM	Type 1 Diabetes Mellitus
T2DM	Type 2 Diabetes Mellitus
TCF7L2	Transcription Factor 7 Like 2
TGF-β	Transforming Growth Factor-B
TGF-β1	Transforming Growth Factor-B 1
TNFRSF21	TNF Receptor Superfamily Member 21
TNF-α	Tumor Necrosis Factor Alpha
TNNT2	Troponin T2, Cardiac Type
TP53INP1	Tumor Protein P53 Inducible Nuclear Protein 1
UBE2O	Ubiquitin Conjugating Enzyme E2 O
UMOD	Uromodulin
VEGF	Vascular Endothelial Growth Factor
VIP	Variable importance of projection
ZNF519	Zinc Finger Protein 519

INTRODUCTION

The emergence of type 2 diabetes mellitus (T2DM) as a pandemic is raising the risk of long-term complications and premature death. T2DM control and management of complications have long become a serious public health concern generating substantial costs, and emphasizing the need for reliable biomarkers for early identification of patients at high risk for diabetic complications and prediction of therapy response. Metformin, an antihyperglycemic agent of the biguanide class, is still used as the first-line treatment of T2DM, despite the variable efficacy and tolerance exhibited among diabetic patients, highlighting its applicability for predictive biomarker studies. The collection of different types of omics data in well-designed clinical studies has already revealed a number of complex interacting pathways that are underlying the pathophysiology of T2DM, and these comprehensive approaches may be further applied in the personalization of diabetes care.

Importance of this work: Deep investigation of the driving and contributory mechanisms underlying the pathophysiology of diabetic complications and its treatment is important in the development of personalized diabetes care. Integration of omics-based biomarkers in T2DM management is needed to promote the early prediction of inter-patient variation in treatment response and develop reliable risk stratification strategies for diabetes complications, thus preventing the progression of the disease and increasing the life expectancy of the patients.

Aim of the study: To investigate the underlying mechanisms and molecular signatures representing the variability in the therapeutic response of metformin and diabetic complications.

Tasks to reach the aim:

1. To evaluate metformin-induced alterations in peripheral blood cell gene expression profiles and signaling pathways following metformin intervention of various duration in healthy individuals and T2DM patients.

2. To identify transcriptional divergence in different metformin response groups before and after administration of metformin, and determine possible biomarkers for early prediction of metformin effectiveness.

3. To explore the underlying mechanisms and genetic predisposition to diabetic complications in the population of Latvia.

1. LITERATURE REVIEW

1.1 Etiology and genetics of diabetes mellitus

Diabetes mellitus is a group of metabolic diseases and long-term conditions. It is characterized by hyperglycemia due to defective insulin secretion or inefficient use of it (Borgnakke 2019; Craig *et al.* 2009). According to the ninth edition of the International Diabetes Federation Atlas, there were around 463 million people with diabetes worldwide compiling around 9.3% of the global prevalence in 2019. Moreover, it is projected that by 2030 the global prevalence of diabetes will continue to rise to 10.2% (578.4 million cases), while in 2045 it may reach 10.9% (700.2 million cases) (Borgnakke 2019). Despite the efforts engaged in diabetes prevention, the disease is still among the top 10 leading causes of death worldwide, accounting for 4.2 million deaths in 2019 (Borgnakke 2019; Li *et al.* 2019), meanwhile causing substantial financial burden on both the patients and healthcare systems resulting from the complex requirements of diabetes management (Moucheraud *et al.* 2019). According to International Diabetes Federation Atlas, the total diabetes-related health expenditure in 2019 in Europe was approximately 161.4 billion USD, which is accounting for 21.2% of the amount devoted to diabetes care globally (Borgnakke 2019). Statistics of diabetes patients in Latvia indicate 91 571 (prevalence 7.4%) diabetes cases and 574 diabetes-related deaths confirmed in Latvia in 2019 (Latvia 2020), which is similar to the diabetes prevalence reported in Europe (8.9%) (Borgnakke 2019).

There are several pathogenic processes involved in the development of diabetes considered when assigning one of the precise types of the disease, nevertheless, all of the underlying mechanisms of diabetes mellitus are provoking raised plasma glucose concentrations and disturbances of glucose metabolism (Palicka 2002). The etiologic classification of diabetes mellitus involves the following types of the disease (Craig *et al.* 2009; WHO 2019):

1. Type 1 diabetes (T1DM), characterized by beta-cell destruction, usually leading to absolute insulin deficiency (immune-mediated and idiopathic);
2. Type 2 diabetes (T2DM) (range from predominantly insulin resistance with relative insulin deficiency to a predominantly secretory defect with insulin resistance);
3. Hybrid forms of diabetes including slowly evolving, immune-mediated diabetes of adults (GAD autoantibodies present) and ketosis-prone T2DM (presents with ketosis and insulin deficiency);
4. Other specific types:

4.1. Monogenic forms of diabetes caused either by mutations affecting beta-cell function or insulin action (e.g. MODY or maturity-onset diabetes of the young, mitochondrial DNA);
4.2. A disease of the exocrine pancreas (e.g. trauma, inflammation, and tumor);
4.3. Endocrine disorders (e.g. pancreatitis, trauma/pancreatectomy, neoplasia, cystic fibrosis, hemochromatosis, fibrocalculous pancreatopathy);
4.4. Drug or chemical induced (e.g. Vacor, pentamidine, nicotinic acid, glucocorticoids, thyroid hormone);
4.5. Infection-related diabetes (e.g. Congenital rubella, Cytomegalovirus);
4.6. Uncommon specific forms of immune-mediated diabetes (e.g. "Stiff-man" syndrome, anti-insulin receptor antibodies);
4.7. Other genetic syndromes sometimes associated with diabetes (e.g. Down syndrome, Klinefelter syndrome, Turner syndrome, Wolfram syndrome);

5. Unclassified diabetes (used temporarily in case of unclear diagnostic category);
6. Hyperglycaemia first detected during pregnancy :
   6.1. Diabetes mellitus in pregnancy (T1DM and T2DM first diagnosed during pregnancy);
   6.2. Gestational diabetes (hyperglycemia below diagnostic thresholds for diabetes in pregnancy.

The diagnosis of diabetes mellitus is set according to the elevated plasma glucose and HbA1c levels (Table 1).

Table 1. Diagnostic criteria for diabetes mellitus (WHO 2019).

Type of the disease	**Fasting plasma glucose**	**2-hour post-load plasma glucose**	**1-hour post-load plasma glucose**	**HbA1c**
Gestational diabetes	5.1–6.9 mmol/L	8.5–11.0 mmol/L	≥ 10.0 mmol/L	-
Other forms of diabetes	≥ 7.0 mmol/L	≥ 11.1 mmol/L	-	≥ 48 mmol/mol

Due to the potential application of genetic data in precision medicine by promoting the prediction of the clinical outcome or adjustment of the most effective therapeutic strategy, multiple studies aiming to explore the genetic predisposition of diabetes mellitus and reclassify the different types of the disease based on molecular data have been performed involving candidate gene approach, genetic linkage studies, and genome-wide association studies

(GWAS) (Guan *et al.* 2008; Tabara *et al.* 2009; Xue *et al.* 2018). Although the two most common forms of diabetes mellitus are caused by a combination of genetic and environmental risk factors, there are also rare forms of the disease that are directly inherited, which at least partially explains the highly divergent etiology of the disease (Sayed and Nabi 2021).

1.1.1 Type 1 diabetes mellitus

Insulin-dependent diabetes or juvenile-onset diabetes, which is also known as type 1 diabetes mellitus (T1DM) is caused by cellular-mediated autoimmune destruction of the beta-cells of the pancreas resulting in insufficient production of insulin or even absolute insulin deficiency (Palicka 2002). Insulin is an anabolic hormone and a key component in cell growth and the metabolism of glucose, proteins, and minerals. It promotes the glucose uptake in muscles and adipose tissue, glycogen storage and fatty acid synthesis in the liver, stimulates the uptake of amino acids and potassium in cells, and inhibits the breakdown of fat in adipose tissue, which underlies the relevance of this hormone in multiple vital processes conducted in the human body (Wong and Sul 2010). T1DM is more prevalent in children, in 2019 there were 128 900 new cases detected worldwide among children and adults aged 0-19 years (Borgnakke 2019). Although the exact etiology of T1DM and most of the other forms of diabetes mellitus is still unknown, it is generally believed that both genetics and environmental factors are implicated (Lucier and Weinstock 2021). There is strong evidence of genetic predisposition and notable heritability for T1DM, justified by the higher concordance among monozygotic (23% probandwise and 13% pairwise) than dizygotic twins (5% probandwise and 3% pairwise) (Kaprio *et al.* 1992). So far there are more than 40 risk loci identified for T1DM (e.g. *INS*, *CTLA4*, *PTPN22*, *IFIH1*, *ERBB3* (Steck and Rewers 2011)), many of them are found in GWAS studies (Pociot 2017), nevertheless, the human leukocyte antigen (HLA) region in chromosome 6p21 is still considered as the main player in providing genetic predisposition to T1DM. There are two specific HLA class 2 haplotypes (*HLA-DR3* and *HLA-DR4-DQ8),* that are associated with approximately 50% of disease heritability (Noble 2015). Meanwhile, enteroviral infection, higher maternal age, and dysbiosis are considered as environmental risk factors for T1DM (Rewers and Ludvigsson 2016). The main syptoms of T1DM in children are weight loss, polyuria, polydipsia, and even ketoacidosis, while the symptoms in adults vary more, therefore 5-15% of T1DM patients are misdiagnosed with T2DM (DiMeglio *et al.* 2018; Tuomi 2005). The main treatment of T1DM is insulin in various forms together with additional medications to lower the blood pressure or cholesterol levels (e.g. angiotensin-converting enzyme inhibitors, angiotensin II receptor blockers) (de Boer *et al.* 2008; Pathak *et al.* 2019).

1.1.2 Type 2 diabetes mellitus

The most common form of diabetes mellitus accounting for approximately 90% of all cases is T2DM with 462 million patients worldwide, the prevalence of 8 529 per 100 000 cases in Europe in 2019, and a tendency to get more prevalent globally (Khan *et al.* 2020). The incidence of the disease varies between geographical regions due to different environmental and lifestyle factors (Zimmet *et al.* 2001). There were 82 140 T2DM patients (4 171.8 per 100 000 individuals) detected back in 2015 in Latvia, out of them 6 091 were new cases. Despite the contribution of national-level initiatives for diabetes prevention, the incidence of T2DM in Latvia has not changed significantly and even got worse with 6 449 new cases in 2019 (LCDC 2020). Although multiple clinical strategies have been developed to improve T2DM management and disease control, T2DM patients still have almost two times increased mortality rates compared to healthy individuals (HR=1.93; 95% CI=1.89-1.97) (Mulnier *et al.* 2006).

Insulin resistance, in which insulin is ineffective and in later stages of the disease may also be lacking, is the main characteristic feature of T2DM leading to hyperglycemia (Goyal and Jialal 2021). Due to insulin resistance, the target tissues of insulin are lacking it, meanwhile, the secretion of insulin reduces along with the gradual destruction of beta-cells (Druet *et al.* 2006). The main destructive effect of insulin resistance is the decrease in glucose uptake in the target tissues and elevated glucose production via glycogenolysis, and gluconeogenesis (Basu *et al.* 2005). The increased glucose output is further supplemented by increased lipolysis and a high level of lipid metabolites as well as enhanced free fatty acid release (Roden *et al.* 2000; Saltiel and Kahn 2001). In addition, there are increased levels of plasma glucagon detected, which is released by pancreatic alpha-cells and stimulates excessive hepatic glucose production (Cherrington *et al.* 1987). Since insulin is suppressing the food intake and glucose regulation in the central nervous system, hypothalamic insulin resistance is underlying the impairment of both of these processes in T2DM (Cersosimo *et al.* 2000; Ono 2019).

The classic symptoms of T2DM are polyuria, polydipsia, and sudden weight loss, though the T2DM diagnosis is often also linked to obesity, hypertension, nephropathy, dyslipidemia, non-alcoholic fatty liver disease, and systemic inflammation (Rosenbloom *et al.* 2009). In long term, T2DM patients are dealing with chronic complications affecting multiple organ systems (Papatheodorou *et al.* 2018).

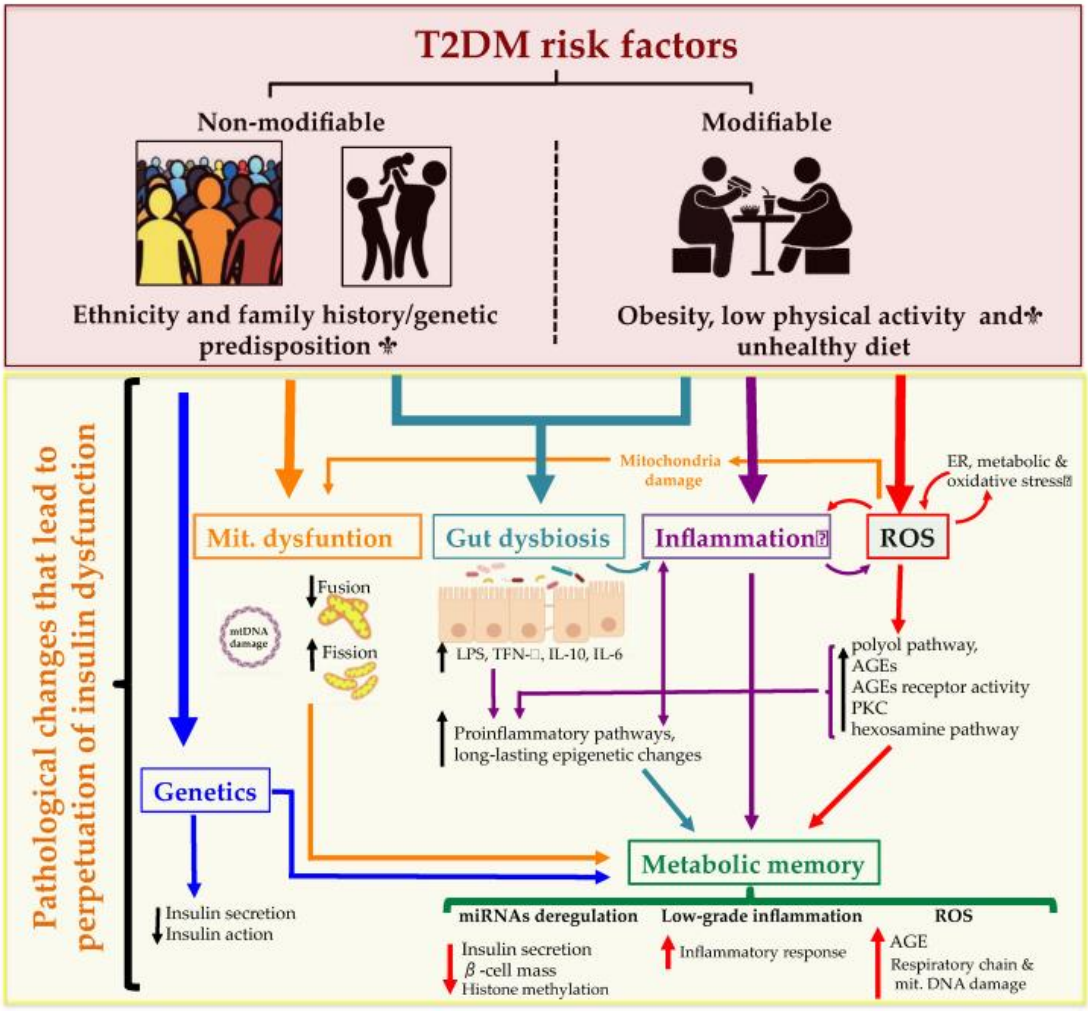

Figure 1. Risk factors of T2DM and characteristic pathological changes. ROS: reactive oxygen species; ER: endoplasmic reticulum; AGEs: advanced glycation end products; PKC: protein kinase C; LPS: lipopolysaccharide; miRNA: microRNA. Adapted from Galicia-Garcia *et al.*(Galicia-Garcia *et al.* 2020).

Since T2DM is a multifactorial disease, it is caused by a complex interaction between environmental and genetic factors (Ali 2013). Multiple risk factors of T2DM are identified so far, including obesity and sedentary lifestyle, smoking, poor diet, gut dysbiosis, and even certain medications affecting the sugar metabolism (e.g. statins, beta-blockers) (Figure 1) (Galicia-Garcia *et al.* 2020; IQWiG 2006). During the last decade, massive research has been performed to demonstrate the role of genetic factors in the development of T2DM using the linkage studies, candidate gene approach, but mostly the advantage of GWAS (Figure 2) (Ali 2013). The heritability of T2DM varies between 20%-80% depending on the population, age, and the particular study design (Almgren *et al.* 2011; Poulsen *et al.* 1999). A GWAS involving participants of European ancestry discovered 143 risk loci (e.g. *TCF7L2*, *PTGFRN*, *ANKH*, *CAMK1D*, *TP53INP1*, and *ATP5G1*) for T2DM, out of them 33 genes appeared to be differentially expressed providing functional evidence, and 3 genes were associated with regulatory function according to DNA methylation data (Xue *et al.* 2018). While the gene coding for transcription factor 7 like 2 (*TCF7L2*, rs7903146, OR=1.37; 95% CI=1.28-1.47) is

showing the strongest association with T2DM (Zeggini *et al.* 2008), there are many other risk loci identified so far, including hematopoietically expressed homeobox gene (*HHEX*, rs8050136, OR=1.13; 95% CI=1.08-1.17) (Diabetes Genetics Initiative of Broad Institute of *et al.* 2007), a gene coding for insulin-like growth factor 2 mRNA binding protein 2 (*IGF2BP2*, rs4402960, OR=1.17; 95% CI=1.10–1.25) (Zeggini *et al.* 2008), potassium inwardly rectifying channel subfamily j member 11 gene (*KCNJ11*, rs5219, OR=1.15; 95% CI=1.09-1.21) and others (Diabetes Genetics Initiative of Broad Institute of *et al.* 2007). Finally, combining data from 32 GWAS and around ~900,000 Europeans provided 243 risk loci for T2DM, including the low-frequency alleles with large effect sizes, such as *DENND2C* (rs184660829, MAF = 0.020%, OR = 8.1, *p*-value = 2.5×10^{-8}) (Mahajan *et al.* 2018). Although the previous studies have provided valuable knowledge of the genetic background and molecular mechanisms underlying the disease, they explain only a small proportion of heritability, suggesting the possible contribution of rare variants, gene-environment interactions, and epigenetics (Ali 2013).

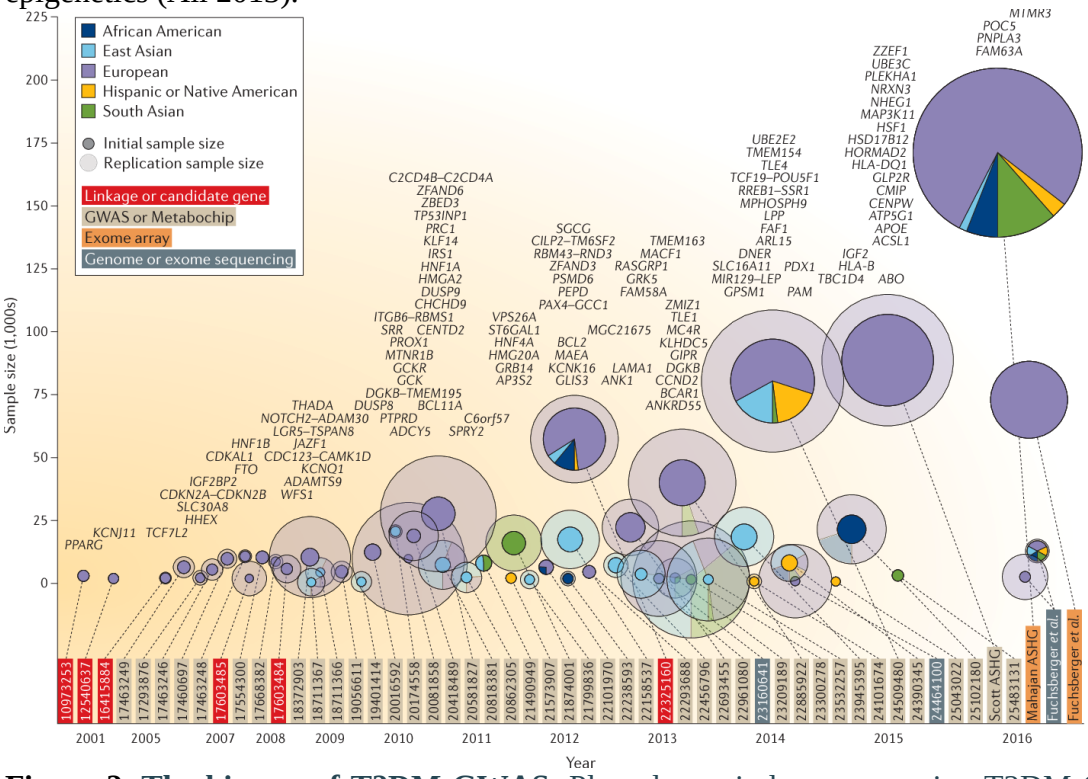

Figure 2. The history of T2DM GWAS. Plotted are circles representing T2DM GWAS as well as additional candidate gene or sequencing studies of note. The x-axis shows the year of publication, whereas the y-axis shows the discovery sample size. Darker circles are scaled in proportion to the discovery sample size, whereas the outer (lighter) circles are scaled in proportion to the total (discovery together with replication) sample size. Circles are colored according to the ethnic composition of the sample set: African American (dark blue), East Asian (light blue), European (purple), Hispanic or Native American (yellow), or South Asian (green). PubMed identifiers (or first author name) for each study are shown at the base of the figure and linked to the corresponding circle with a dotted line. Identifiers are colored according to the technology used in the study: linkage or candidate gene studies (red), GWAS or Metabochip (beige), exome array (orange), or sequencing (dark grey). The figure is adapted from Flannick and Florez, 2016 (Flannick and Florez 2016).

1.2 Genomics and transcriptomics of metformin treatment

1.2.1 Management of type 2 diabetes mellitus

Since T2DM is a complex disease characterized by high phenotypic variability and systemic manifestations, an integrated approach is applied for the treatment of this widespread condition. In T2DM therapy any strategy towards the improvement of blood glucose control involve patient engagement in self-management and lifestyle change, therefore education about the diet and physical activity in the context of blood glucose control as well as prophylactic screening for complications is ensured by a doctor regularly, nevertheless, most of the patients still need to combine the lifestyle changes with pharmacological therapy. The medications applied in T2DM treatment are targeting glucose levels and reducing HbA1c, and the global strategy is based on lifestyle changes, followed by administration of metformin, or using metformin with second antihyperglycemic agents such as inhibitors of dipeptidyl peptidase, gliflozins, glucagon-like peptide-1 receptor agonists, pioglitazone and sulphonylurea (Reusch and Manson 2017).

Sulfonylureas, including glyburide, glimepiride, and glipizide are among the top prescribed oral agents for T2DM, ensuring a reduction of HbA1c by around 1%. The mechanism underlying the glucose-lowering effect of sulfonylureas is based on increased insulin release from the pancreatic beta-cells, which also raises the risk of hypoglycemia. Moreover, due to characteristic beta-cell failure in T2DM, the effectiveness of these medications may diminish over time (Wright and Tylee 2016). Multiple studies suggest a significant association of sulfonylureas and cardiovascular disease and even cardiovascular death (RR=1.27; 95% CI=1.18–1.34), which is considered in therapeutic strategy decision making (Azoulay and Suissa 2017; Phung *et al.* 2013).

Another group of oral antidiabetic agents is dipeptidyl-peptidase 4 inhibitors, including sitagliptin, saxagliptin, linagliptin, and alogliptin. Agents corresponding to this specific group are targeting the incretin pathway, and therefore augmenting levels of intestinal hormones that are released during the meal. Dipeptidyl-peptidase 4 inhibitors limit the inactivation of the incretin hormone glucagon-like peptide 1 resulting in improved blood sugar control by promoting the release of insulin in response to food intake, nevertheless, the glucose-lowering effect of these agents is slightly lower than with other commonly prescribed diabetes medications (0.5% HbA1c reduction), therefore it is typically used as the second-line medication (Wright and Tylee 2016; Y. Zhang *et al.* 2014a).

Thiazolidinediones are targeting nuclear transcription factor peroxisome-proliferator-activated receptor gamma. Although this type of antidiabetic medication has no risk of hypoglycemia, the particular therapy has been associated with myocardial infarction (Wallach *et al.* 2020), edema (Mudaliar *et al.* 2003), bladder cancer (R. M. Turner *et al.* 2014) before, therefore it is considered as the third-line therapy for diabetes and is prescribed together with other anti-diabetic agents (Wright and Tylee 2016).

Sodium-glucose cotransporter 2 inhibitors are reducing the glucose reabsorption from urine which is mediated by sodium-glucose cotransporter 2. This inhibition results in reduced levels of fasting and postprandial blood glucose as well as weight loss and lower systolic blood pressure. The risk of hypoglycemia is usually increased only when used in combination with other anti-diabetic agents (Saisho 2020). As with the thiazolidinediones, the use of sodium-glucose cotransporter 2 inhibitors is linked to adverse effects, such as genital infections (Unnikrishnan *et al.* 2018), risk of diabetic ketoacidosis (Palmer *et al.* 2016), bone fracture (Azharuddin *et al.* 2018), and bladder cancer (Garcia *et al.* 2021), though the data proving particular associations ar conflicting. Nevertheless, in contrast to thiazolidinediones sodium-glucose cotransporter 2 inhibitors exert a protective effect against cardiovascular disease and death in T2DM patients (Arnott *et al.* 2020).

There are several types of injectable agents for the treatment of T2DM, such as glucagon-like peptide 1 receptor agonists, amylin, and different forms of insulin. The management of T2DM is complex and around 25% of the patient require insulin within 6 years of starting oral antidiabetic therapy (Ringborg *et al.* 2010). For patients with symptomatic hyperglycemia or difficulties lowering the HbA1c levels below 9.5% while using oral agents, insulin therapy is recommended including basal, prandial, or concentrated insulin depending on the efficacy of glycemic control and fluctuations. An additional class of injectable medications is glucagon-like peptide 1 receptor agonists (Wright and Tylee 2016). These agents promote supraphysiologic levels of glucagon-like peptide 1 activity, resulting in elevated insulin secretion and weight loss (Baggio and Drucker 2007).

1.2.2 Metformin

Metformin is a biguanide, which is globally used as the first-line oral treatment for T2DM due to its ability to effectively reduce both basal and postprandial blood glucose levels by decreasing intestinal absorption, promoting insulin sensitivity, and inhibiting the production of glucose in the liver, although data on the latter are still controversial (Corcoran and Jacobs 2021; Gormsen *et al.* 2019). The improved insulin sensitivity results in decreased luteinizing hormone and androgen levels, and stabilized menstrual cycle, therefore it has been widely used

in the treatment of polycystic ovary syndrome (Y. W. Wang *et al.* 2017). Long-term administration of metformin significantly reduces the risk of neurodegenerative diseases (adjusted HR=0.19; 95% CI= 0.12 - 0.31) (Shi *et al.* 2019), nevertheless, the reports discussing the neural therapeutic effects of the drug show conflicting data (Ping *et al.* 2020; Pusceddu *et al.* 2020). In addition, strong evidence supports the anticancer effect of metformin, which is demonstrated as inhibited cancer cell growth *in vitro*, delayed development of cancer in mice models, and even significantly reduced cancer risk in T2DM patients (HR=0.513; 95% CI=0.318–0.826) (H. J. Kim *et al.* 2018; Ko *et al.* 2016; Z. J. Zhang *et al.* 2014b). One of the main striking findings regarding the metformin pleiotropic effect is its ability to increase lifespan and attenuate the deleterious impact of aging in male mice, even more, metformin mimics calorie restriction phenotype (Martin-Montalvo *et al.* 2013). Finally, metformin inhibits the formation of atherosclerosis, normalizes blood pressure, lipid levels, and reduces body weight (Kyler *et al.* 2018; Thomopoulos *et al.* 2017).

Although it significantly reduces the body mass index (BMI) (weighted mass difference = −1.31; 95%CI = −2.07 to −0.54) the therapeutic efficacy of metformin against hyperglycemia is highly variable among different patients (Pu *et al.* 2020). A recent study revealed that only 59.5% of newly diagnosed T2DM patients developed their glycemic goal after the three months long therapy, while the rest of the patients were considered as non-responders (Rashid *et al.* 2019). In addition, around 25% of patients experience gastrointestinal side effects, such as diarrhea (62.1%) and retching (21.1%), and because of that around 5% of the patients discontinue the therapy (Florez *et al.* 2010; McCreight *et al.* 2016; Rashid *et al.* 2019). These data show an urgent need for the development of a precision medicine approach in T2DM care to maximize therapeutic benefit while limiting risks, therefore identification of biomarkers for early prediction of therapeutic response and significant adverse events is one of the main strategies in studies of T2DM (Fitipaldi *et al.* 2018). Genomics, transcriptomics, and other types of omics data are widely used in biomedical studies for both obtaining deeper knowledge about the pathogenesis mechanisms of diseases and biomarker discoveries (Z. Z. Hu *et al.* 2011).

1.2.2.1 Mechanism of action

The liver is the main site of action of metformin, where metformin inhibits gluconeogenic enzymes and stimulates glycolysis by altering the activity of multiple enzymes, resulting in around 75% reduction in hepatic glucose output (Pernicova and Korbonits 2014; Stumvoll *et al.* 1995). Metformin suppresses gluconeogenesis by upregulating the insulin receptor and insulin receptor substrate 2 (IRS-2) and suppressing the action of peptide hormone glucagon (Gunton *et al.* 2003). Metformin enhances insulin sensitivity and glucose uptake in the skeletal muscle, by improving the tyrosine kinase activity of insulin receptors and translocation of

glucose transporters on the plasma membrane (Matthaei *et al.* 1993). Moreover, metformin interacts with the incretin axis, by stimulating the expression and enhancing the action of glucagon-like peptide 1, which is further responsible for the secretion of insulin and reduction of glucagon levels in response to glucose (Vardarli *et al.* 2014).

Metformin acts mainly by reducing hepatic gluconeogenesis and increasing glucose uptake in skeletal muscle, though high interindividual variability is observed in metformin response and pharmacokinetics (Rashid *et al.* 2019). The cellular uptake of metformin is maintained by multiple membrane transporters (Table 2) (Pernicova and Korbonits 2014).

Table 2. Cell membrane transporters involved in metformin pharmacokinetics (Pernicova and Korbonits 2014).

Transporter	**Gene**	**Function**
SLC22A1 or OCT1	*SLC22A1*	Main transporter for metformin uptake. Expressed in liver and kidney, considered as the main metformin transporter. Allelic variants within the gene are associated with impaired metformin uptake, increased elimination, and the following reduction of therapeutic response.
SLC22A2 or OCT2	*SLC22A2*	Mediates metformin secretion in the kidney, mediates around 80% of the total metformin clearance.
SLC22A3 or OCT3	*SLC22A3*	Expressed in liver, kidney, heart, skeletal muscle, brain, placenta and may be involved in the metformin uptake in muscle.
SLC22A4 or OCTN1	*SLC22A4*	Involved in the gastrointestinal absorption of metformin.
MATE1	*SLC47A1*	Mediates metformin secretion in the kidney and excretion into bile in the liver. Allelic variants in the gene alter the glucose-lowering effect of metformin in diabetic patients.
MATE2	*SLC47A2*	Mediates metformin secretion in the kidney.
PMAT	*SLC29A4*	Mediates renal and intestinal metformin uptake.

The cellular target of metformin is the mitochondrion, where the agent inhibits the mitochondrial respiratory chain complex 1, resulting in resulting in lower ATP production, increased cellular AMP:ATP and ADP:ATP ratios, the activation of 5' adenosine monophosphate-activated protein kinase (AMPK) and the following improvement in insulin sensitivity, together with reduced cyclic adenosine monophosphate (cAMP) levels and downregulation of gluconeogenetic enzymes via phosphorylation of the transcription factor cAMP response element binding protein (CREB). Activation of AMPK switches on the catabolic pathways that generate ATP and switches off anabolic pathways to restore the energy balance, leading to AMPK-driven phosphorylation of metabolic enzymes and transcription factors by altered gene expression, and finally inhibition of glucose, lipid, and protein synthesis and activation of fatty acid oxidation and glucose uptake in peripheral tissue (Vial *et al.* 2019; Y. Wang *et al.* 2019). The rise in AMP and ATP ratio inhibits fructose-1,6-bisphosphatase or

FBPase, inducing the acute inhibition of gluconeogenesis in an AMPK independent manner (Rena *et al.* 2017). Despite the large number of comprehensive studies aimed to reveal the complex mechanisms underlying the beneficial effects of the drug, controversy remains since there is strong evidence for other indirect mechanisms, such as the significant contribution of the gut microbiome underlying the glucose-lowering effect of the drug (Elbere *et al.* 2020).

1.2.2.2 Omics data in metformin studies

Genomics

So far multiple types of omics-based approaches have been applied in studies of metformin action and therapeutic efficacy (Udhane *et al.* 2017). One of the main approaches for studying metformin pharmacogenetics is GWAS. A comprehensive meta-analysis of 10 557 participants conducted by the MetGen consortium revealed a significant association of glucose transporter GLUT2 (*SLC2A2)* variant rs8192675 with metformin response (*p*-value=6.6×10^{-14}) which promoted the metformin-induced reduction in HbA1c levels by 0.17% (K. Zhou *et al.* 2016). Another genetic variant rs11212617 located in a locus of ataxia telangiectasia mutated gene (*ATM*) also reached the genome level significance in association with metformin response (OR=1.35; 95%CI=1.22-1.49, *p*-value=2.9×10^{-9}) and was identified as a result of collaboration between Genetics of Diabetes Audit and Research (GoDARTS) and the United Kingdom Prospective Diabetes Study (GoDarts *et al.* 2011). There are multiple other genetic factors such as organic cation transporter genes *OCT1* and *OCT2* associated with altered therapeutic efficacy of metformin (Shikata *et al.* 2007) and the overall heritability of the glycemic response of metformin expressed as the absolute reduction in HbA1c is 34% (95% CI=1–68; *p*-value=0.022) indicating on other non-genetic factors involved (K. Zhou *et al.* 2014a).

Metabolomics

A comprehensive metabolomics study on mouse embryonic fibroblast cells revealed both AMPK-dependent (upregulation of phospho-acetyl-CoA carboxylase) and independent (inhibition of phospho-mammalian target of rapamycin, phospho-S6 ribosomal protein, autophagy marker Lc3, adipose triglyceride lipase, and phospho- extracellular signal-related kinase) mechanisms of metformin action, resulting in metabolic reprogramming and reduction of energy state (Yan *et al.* 2019). Another study involving longitudinal metabolomics profiling of human plasma samples during the administration of metformin in a prospective trial revealed a reduction in levels of valine, tyrosine, and carnitine, that are involved in insulin resistance and mitochondrial dysfunction. Although the particular study did not reveal predictive biomarkers of metformin response, the hypothesis of an association between high levels of

leucine/isoleucine and/or carnitine C10:1 before anti-diabetic therapy and improved glycemic response to metformin was established (Safai *et al.* 2018).

Microbiome

The role of the gut microbiome in achieving the therapeutic effects of metformin has been demonstrated in various reports, therefore metagenomics is among the major data types studied in the context of metformin action. Metformin treatment has been previously associated with increased levels of *Escherichia, Akkermansia* spp., and decreased levels of *Intestinibacter* (Shin *et al.* 2014; Wu *et al.* 2017). A randomized, double-blind study of treatment-naïve T2DM patients followed by the humanization of germ-free mice by fecal microbiota provided evidence of metformin-altered microbiota to significantly improve glucose tolerance (Wu *et al.* 2017). Moreover, the shotgun metagenomic sequencing data of longitudinal stool samples of treatment-naïve T2DM patients have revealed the potential of the baseline composition of gut microbiota to influence the therapeutic efficacy of metformin meanwhile serving as a predictive biomarker for both efficacy and tolerance of the drug (Elbere *et al.* 2020).

Transcriptomics

Large-scale transcriptome profiling is proved to serve as a valuable source for molecular target discoveries. Among the various methods for determining the gene expression levels, RNA sequencing (RNA-Seq) is the state-of-the-art approach that is used to provide insight into the cellular transcriptome, it is applied in drug response profiling as well as efficacy biomarker discoveries (Geeleher *et al.* 2014). The RNA-Seq-based comparative transcriptomic approach has been already applied in metformin research by using animal models and cell cultures, nevertheless, the longitudinal *in vivo* studies reflecting the systemic effect of the drug in humans is lacking.

So far the gene expression-based studies have discovered multiple beneficial effects and molecular targets of metformin. Treating the primary human fibroblasts with metformin revealed significantly altered gene expression profiles in a concentration-dependent manner which is mediated by the transcription factor forkhead box protein O3 (*FOXO3a*), moreover, the treatment induced enrichment in the activator protein 1 (*AP-1*) transcription factor pathway, and cytokine-cytokine interaction pathway (Gillespie *et al.* 2019). In addition, the microarray-based gene expression analysis of metformin-treated human adrenal H295R cells (an angiotensin-II-responsive steroid-producing adrenocortical cell line) revealed the ability of metformin to modulate energy homeostasis via upregulation of genes coding for enolase 2 (*ENO2),* enolase 3 (*ENO3),* and aldolase, fructose-bisphosphate c (*ALDOC*), that are essential in glycolysis. This study showed that metformin also targets some of the regulators of sex steroid production, such as hydroxysteroid 17-beta dehydrogenase 14 (*HSD17B14),* aldo-keto

reductase family 1 member c3 (*AKR1C3)*, and sulfotransferase family 2a member 1 (*SULT2A1*), at least partially explaining the altered androgen production which is observed in such hyperandrogenic conditions as polycystic ovary syndrome after metformin therapy (Udhane *et al.* 2017). Together with various AMPK-dependent and independent molecular targets of metformin, the combination of RNA-Seq, chromatin immunoprecipitation-based sequencing (Chip-Seq), and functional validations by small interfering RNA (siRNA) knockdown demonstrated the potential role of activating transcription factor 3 (*ATF3*) in gluconeogenesis repression in primary human hepatocytes (Luizon *et al.* 2016). Meanwhile, metformin-treated breast cancer cells showed the implication of gene coding for cell division cycle 42 (*CDC42*) in metformin-induced suppression of cell proliferation and migration via AMPK-independent molecular mechanism (Athreya *et al.* 2017). Upregulation of krüppel-like factor 4 (*KLF4*) and cholesterol-25-hydroxylase (*CH25H*) coding genes as the opposite effect of high glucose levels were linked to metformin-induced suppression of endothelial-to-mesenchymal transition in human umbilical vein endothelial cells, which may serve as a beneficial effect in endothelial disfunction (Yu *et al.* 2020). Finally, the study on *DMPK* or dystrophia myotonica protein kinase gene-mutated mesodermal precursor cells suggested metformin as a modulator of alternative splicing of a subset of genes (e.g. *INSR* and *TNNT2*) (Laustriat *et al.* 2015).

In vivo animal models have provided more systemic insight into the molecular action of metformin. The study involving the rat model of obesity and insulin resistance highlighted a list of genes ensuring the cardiovascular benefits of metformin and clarified the heterogeneous effects of the drug across the arterial tree (Padilla *et al.* 2017). Moreover, a comprehensive study of mice models with different interventions including the administration of glipizide, rosiglitazone, and metformin, proved that metformin treatment for eight weeks mimics 75% of the gene expression effects of long-term calorie restriction and 92% of the 8 weeks long calorie restriction (Dhahbi *et al.* 2005). A recent study performed by Meng and colleagues obtained metformin-induced transcriptome profiles induced in 10 different tissue types (aorta, brown adipose, brain, eye, heart, liver, kidney, skeletal muscle, stomach, and testis) of healthy mice, revealing a clear tissue-specific effect of the drug on the gene expression profiles. The authors proved a significant correlation between the transcriptomic signature of metformin-treated normal mice tissue and calory restricted anti-aging intervention (Spearman's correlation coefficient (SCC)=0.258, false discovery rate (FDR)=2.72E-231), and a negative correlation with the model of dilated cardiomyopathy (prostaglandin E receptor 4, Ptger4-/-) (SCC=−0.054, FDR=4.19E-08). Moreover, the study provided evidence of hypertension as a potential side-effect of long-term metformin treatment in healthy subjects (Y. Meng *et al.* 2020). Controversially, a study reported by Zhu and colleagues proved that metformin fails to

increase the lifespan of aged female mice and the previously reported beneficial effect on cardiac metabolism is also lacking. The particular study involving RNA-Seq on heart tissue revealed the metformin-induced upregulation of extracellular matrix-related genes and downregulation of oxidative phosphorylation-related gene expression (Zhu *et al.* 2020). The microarray analysis of liver tissue obtained from obese diabetic db/db mice following a single metformin dose provided evidence of reduced expression of glucose-6-phosphatase coding gene and the following decreased activity of the enzyme, suggesting a potential mechanism at the mRNA level underlying the glucose-lowering effect of the drug (Heishi *et al.* 2006).

1.3 Genetic risk factors of diabetic complications

The main priority in T2DM management is to reduce the disease-related death rates (Borgnakke 2019). The relatively high mortality of T2DM patients is referred to the multiple chronic and acute complications of the disease, which also impairs the quality of life and imposes a significant financial burden on the healthcare systems globally (Z. Liu *et al.* 2010). The complication risk is associated with the diabetes duration and glycemic control, which is usually evaluated as HbA1c level representing glucose levels during the last 3 to 4 months (Lind *et al.* 2009). Hyperglycemia-induced ketoacidosis and diabetic coma are the main acute complications of diabetes (Fayfman *et al.* 2017), nevertheless, more than 20% of T2DM patients are also experiencing the chronic state of vascular issues classified as microvascular complications with small vessel damage and macrovascular complications caused by impaired function of arteries (Fowler 2008; Yadav *et al.* 2021). Microvascular complications include retinopathy, nephropathy, and neuropathy, while cardiovascular disease resulting in myocardial infarction and cerebrovascular dysfunction manifesting as stroke are considered macrovascular complications (Figure 3) (Forbes and Cooper 2013). The global crude prevalence of microvascular and macrovascular complications is 18.8% and 12.7%, respectively, according to the data collected during a 3-year, prospective, observational study program (Kosiborod *et al.* 2018).

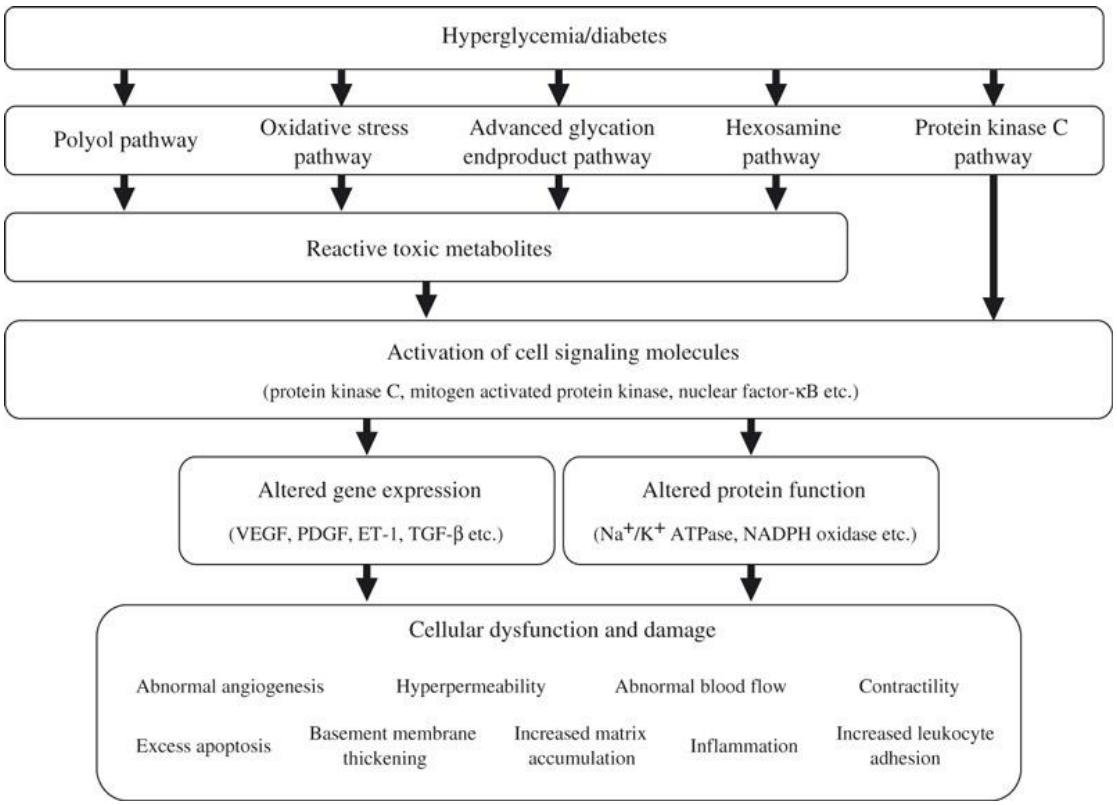

Figure 3. Mechanisms by which hyperglycemia induces diabetic vascular complications. ET-1: endothelin-1; NADPH: nicotinamide adenine dinucleotide phosphate; PDGF: platelet-derived growth factor: TGF-β, transforming growth factor-β; VEGF: vascular endothelial growth factor. Figure adapted from Khalil, 2016(Khalil 2017).

1.3.1 Diabetic nephropathy

Diabetic nephropathy is characterized as proteinuria, a decline in glomerular filtration rate, and eventually, it may cause end-stage renal failure (Mogensen *et al.* 1983). Although hyperglycemia is the main contributor to the development of nephropathy, only approximately 30-40% of T2DM patients experience this type of complication, suggesting the potential implication of other factors, including genetic predisposition since diabetic nephropathies tend to cluster in families (Brennan *et al.* 2013). Diabetic nephropathy usually begins as renal cellular hypertrophy and hyperfiltration, followed by progressive albuminuria and decreased glomerular filtration rate due to persistent hyperglycemia. Microalbuminuria which is defined as urinary albumin excretion rate 30-300 mg per day develops only 10-15 years after the onset of diabetes which is followed by macroalbuminuria (albumin excretion rate higher than 300 mg per day) (Chida *et al.* 2016). The progression of the disease is promoted by inflammation and hypertension. At the molecular level, several key modulators contributing to the development of renal fibrosis and diabetic nephropathy are identified so far including glucose itself, angiotensin II, reactive oxygen species, transforming growth factor beta 1 (TGF-β1), vascular

endothelial growth factor (VEGF), and connective tissue growth factor (CTGF) (Brennan *et al.* 2013). Although many genetic association studies have been conducted for diabetic nephropathy in both T1DM and T2DM patients, there are only several promising associations emerged so far, including rs2206136 (*PLCB4,* phospholipase c beta 4) and chronic kidney disease (OR=1.20; 95%CI=1.08-1.34), rs9942471 (*GABRR1,* gamma-aminobutyric acid type a receptor subunit rho1) and microalbuminuria (OR=1.24; 95%CI=1.15-1.34), rs11864909 (*UMOD*; uromodulin; β=2.42; 95%CI=1.28-3.56), rs1974990 (*SSB*; small RNA binding exonuclease protection factor la; β=4.07; 95%CI=2.61-5.52), rs10224002 (*PRKAG2*; protein kinase AMP-activated non-catalytic subunit gamma 2; β=1.75; 95%CI-0.85-2.66) and the estimated glomerular filtration rate (van Zuydam *et al.* 2018), rs10888287 (*OR2L13;* olfactory receptor family 2 subfamily l member 13) and nephropathy in T2DM (OR=0.04; 95%CI=0.80-0.99) (McDonough *et al.* 2011) and several others (Maeda *et al.* 2007; Taira *et al.* 2018). A GWAS by Pezzolesi *et al*. has revealed 13 different risk loci for diabetic nephropathy in T1DM patients some of them corresponding to the gene coding for FERM domain containing 3 (*FRMD3*; OR=1.45; 95%CI=1.25–1.67) and cysteinyl-tRNA synthetase (*CARS*; OR=1.36; 95%CI=1.19–1.56) that are expressed in the human kidney (Pezzolesi *et al.* 2009).

1.3.2 Diabetic neuropathy

Diabetic neuropathy is characterized by peripheral nerve dysfunction in diabetes patients when other clinical etiologies are excluded. Although there is a complex classification of diabetic neuropathy established including the categorization in a symmetric and asymmetric type of the disease, the typical and most common form of diabetic neuropathy is considered as chronic, symmetrical, distal sensorimotor polyneuropathy (Tesfaye *et al.* 2010). The distal symmetric polyneuropathy accounts for approximately 75% of all cases of diabetic neuropathy, and the sensory symptoms are present from toes to knees and later affect the palms, while asymmetrical neuropathies may involve cranial nerves, thoracic or limb nerves and are of acute onset (Bansal *et al.* 2006). Although the exact cause of diabetic neuropathy is still unknown, hyperglycemia is considered the key component in the pathogenesis of the disease. Due to hyperglycemia-induced rheological alterations, endothelial vascular resistance is increased and nerve blood flow is decreased, hyperglycemia induces oxidative stress, and affects the function of aldose reductase leading to the accumulation of metabolites (fructose and sorbitol) in the nerve and inducing non-enzymatic glycosylation of structural proteins. Altogether these changes induce hypoxia and impair the axonal transport causing the symptoms (Bansal *et al.* 2006). The duration and severity of hyperglycemia together with dyslipidemia, hypertension, and obesity are among the main risk factors for diabetic neuropathy, nevertheless few studies have proved the implication of genetic factors in the disease etiology (Tang *et al.* 2019; Van

Acker *et al.* 2009). Although not reaching the genome-wide significance threshold, an association between the locus next to the *GRAF2* (GTPase regulator associated with the focal adhesion kinase) gene (rs17428041; OR=0.67; 95%CI =0.57–0.78) and diabetic neuropathic pain has been identified (W. Meng *et al.* 2015). Similarly, a specific region in chromosome 2q24 is showing a strong protective effect (rs13417783; OR= 0.64; 95%CI 0.55–0.74) on peripheral diabetic neuropathy, moreover, the allele is also linked to increased expression of *SCN2A* gene coding for the sodium voltage-gated channel alpha subunit 2 in the tibial nerve, providing functional evidence (Tang *et al.* 2019).

1.3.3 Ophthalmic complications

Several ocular complications of diabetes have been identified so far such as diabetic retinopathy, diabetic papillopathy, glaucoma, cataract, and ocular surface diseases (Sayin *et al.* 2015). The most common type of microvascular complication of diabetes is diabetic retinopathy, affecting more than 90% of T1DM patients and around 60% of T2DM patients (Seema Garg 2009; Threatt *et al.* 2013). The main risk factor for diabetic retinopathy is chronic hyperglycemia, affecting multiple cell signaling pathways, including the activation of the polyol pathway, involving reduction of glucose into sorbitol and further osmotic damage followed by cell death (Gabbay 1975). Other molecular targets involve activation of protein kinase C pathway and renin-angiotensin system, upregulation of vascular, endothelial growth factor, inducing oxidative stress, and vascular endothelial dysfunction (Kowluru 2005; Sayin *et al.* 2015; Stitt 2003; Tarr *et al.* 2013). Some of these are serving as underlying mechanisms for glaucoma and cataract (Pollreisz and Schmidt-Erfurth 2010; Senthil *et al.* 2021). In addition to hyperglycemia, also hypertension and hyperlipidemia characterized by poor metabolic control are considered as the risk factors of diabetic retinopathy (Anonymous 1998). So far multiple GWAS reports have proven the significant role of genetic factors in the development of diabetic retinopathy, for instance, a study performed in the Taiwanese population revealed novel significant associations in five loci: histone H2A deubiquitinase: *MYSM1* (rs2811893, rs12092121, OR =1.50; 95%CI=1.03-2.20), plexin domain containing 2: *PLXDC2* (rs1571942, OR=1.67; 95%CI =1.06-2.65); Rho GTPase activating protein 22: *ARHGAP22* (rs4838605, rs11101355, rs11101357, OR =1.65; 95%CI=1.05-2.60), heparan sulfate 6-O-sulfotransferase 3: *HS6ST3* (rs2038823, OR=2.33; 95%CI=1.13-4.77), and gene with yet unknown function in the 5q were also associated with the disease (rs13163610, OR=3.63; 95%CI=1.38-9.58) (Huang *et al.* 2011). In the Japanese population the strongest signal was observed in *RP1-90L14.1* (rs9362054, OR=1.64; p-value$=1.4\times10^{-7}$) a long intergenic non-coding RNA (Awata *et al.* 2015), while in Australians the association with diabetic retinopathy was identified for

the variant rs9896052 near the growth factor receptor bound protein 2 coding gene (*GRB2)* in the T2DM cohort (OR=1.45; 95% CI=1.03-2.18) (Burdon *et al.* 2015).

1.3.4 Macrovascular complications

Macrovascular complications are considered as damage to large vessels, such as arteries and veins. These are atherosclerosis-related diseases, including coronary artery disease, peripheral vascular disease, and stroke (Kitada *et al.* 2010). Due to endothelial injury and inflammation, oxidized lipids accumulate in the endothelial wall of arteries, this oxidation is promoted by angiotensin II. Infiltration of monocytes stimulates activation of macrophages and T-lymphocytes, further, the proliferation of smooth muscles in the arterial walls together with collagen accumulation is induced, which leads to the formation of typical atherosclerotic lesions and in some cases even acute vascular infarction (Rader 2007). In comparison to microvascular complications, the major risk factor for macrovascular complications is insulin resistance, not the hyperglycemia leading to increased fatty acid flux in arterial endothelial cells from adipocytes, overproduction of reactive oxygen species, activation of protein kinase C pathway, and overall similar molecular etiology as in microvascular complications.

Cardiovascular disease is the leading cause of death in T2DM patients, accounting for approximately 70% of deaths in the particular patient group, moreover, T2DM patients have a four times higher risk of developing cardiovascular disease than individuals with no diabetes (Cade 2008). The main risk factors for cardiovascular disease are sex, age, smoking, the ratio of high-density lipoproteins, total cholesterol, hypertension, and HOMA-IR (homeostatic model assessment for insulin resistance) (Bonora *et al.* 2002). From all of the cardiovascular events, stroke is the third leading cause of death among T2DM patients. Although diabetes affects cerebrovascular circulation by increasing the risk of intracranial and extracranial atherosclerosis, diabetes mellitus and hyperglycemia are both proven to serve as independent risk factors for stroke (Davis *et al.* 1999). Finally, peripheral artery disease is also considered a macrovascular complication of diabetes, which is significantly associated with the severity and duration of diabetes, hyperglycemia, hypertension, obesity, and other factors (Wattanakit *et al.* 2005).

GWAS have provided evidence for multiple genetic loci modulating the risk for different phenotypes linked to the damage of large vessels. For instance, a GWAS performed in a population of African Americans with T2DM revealed 15 genetic variants showing significant association with coronary artery calcified atherosclerotic plaque which is a predictive factor for cardiovascular disease. The particular study suggested five novel risk loci: low-density lipoprotein receptor-related protein 1B, *LRP1B (*rs113533135; *p*-value=3.3×10^{-7}); ataxin 1,

ATXN1 (rs16879003; *p*-value=1.1×10^{-7}), *MAGI2* (rs113805659; *p*-value=1.4×10^{-7}); desmocollin 1, *DSC1* (rs4459623; *p*-value=5.3×10^{-7}) to be considered for further functional investigation (Divers *et al.* 2017). Similarly, a study conducted by Montesonti and colleagues revealed a protective allele in T2DM patients for ischemic cardiovascular disease and stroke located in adiponectin, C1Q, and collagen domain containing (*ADIPOQ*) gene (rs266729; OR=0.61, 95%CI=0.39-0.95) (Montesanto *et al.* 2018), and a variant on chromosome 1q25 and proved its association with coronary heart disease in T2DM patients (rs10911021; OR = 1.36; 95% CI =1.22-1.51) with additional evidence of its functional implication via differential gene expression (Qi *et al.* 2013). Since cardiovascular diseases are still the leading cause of death, there are multiple GWAS performed for coronary artery disease and myocardial infarction irrespective of diabetes (Erdmann *et al.* 2009; Mc Namara *et al.* 2019; Reilly *et al.* 2011).

2. MATERIALS AND METHODS

2.1 Study design described for each publication separately

The research described in all three publications was conducted in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki amended in Fortaleza, Brazil, October 2013) and The Convention for the protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine. Management of patient recruitment, collection of samples, and associated clinical data was ensured by Latvian Biomedical Research and Study Centre's core facility Genome Centre and the Genome Database of the Latvian Population (LGDB) following their standard procedures (Rovite *et al.* 2018).

2.1.1 Metformin strongly affects transcriptome of peripheral blood cells in healthy individuals

Twenty-five European descent volunteers with no history of chronic disease were involved in the study according to the exclusion/inclusion criteria set within the framework of the clinical trial 'Pharmacodynamics of antidiabetic drug metformin' protocol number MIKROMET16001, registration number of EU Clinical Trials Register: 2016-001092-74 (www.clinicaltrialsregister.eu) (Table S1). Fasting blood tests (measures of alanine aminotransferase, plasma glucose, creatinine levels) were performed in a certified clinical laboratory 1–3 days before metformin administration to evaluate general hematological and biochemical parameters and eligibility of volunteers. The study participants received a twice-daily oral 850 mg dose of metformin hydrochloride (*Metforal* 850mg, Berlin-Chemie AG) for 7 days.

RNA for transcriptome analysis was isolated from the peripheral blood samples collected at three time-points: (1) before administration of metformin (M0, morning, fasting state), 10 hours after the first metformin intake, though before the second dose (M10h, evening) and after 7 days long metformin course (M7d, morning, fasting state). The third blood sample was not collected from one out of 25 study subjects, due to the discontinuation of metformin treatment, nevertheless, the rest of the blood samples collected from the particular subject were included in the data analysis. The longitudinal study design of the open-label trial was chosen as the most suitable method for global gene expression analysis with high inter-individual variability expected. The study protocol was approved by the Central Medical Ethics committee of Latvia (1/16-05-12) and the State Agency of Medicines of the Republic of Latvia (17–1723). Written informed consent was obtained from every participant after a full explanation of the purpose and nature of all procedures used before their inclusion in the study.

2.1.2 Whole-blood transcriptome profiling reveals signatures of metformin and its therapeutic response

In total 17 T2DM patients of European descent fulfilling the inclusion/exclusion criteria (Table S2) were enrolled. The study was conducted within the framework of the ongoing observational, prospective and longitudinal study OPTIMED, which has been implemented since 2010 in collaboration with endocrinologists and general practitioners from the leading health care centers in Latvia, ensuring recruitment of newly diagnosed drug-naïve patients with ICD-10 (International classification of diseases diagnosis 10th revision) code E11 and follow-up data collection.

Metformin monotherapy (medication with metformin hydrochloride as the only active ingredient) was prescribed for each study participant by an endocrinologist for at least three months regardless of the research objectives. The drug manufacturer and dosage of metformin (varied from 850 mg to 2000 mg per day) were chosen by endocrinologists based on clinical experience, patient's health status, and manifestations of the disease. Measures of alanine aminotransferase, creatinine levels, HbA1c, triglycerides were performed in a certified clinical laboratory before the administration of metformin and after metformin therapy for three months to evaluate general hematological and biochemical parameters and eligibility of the subjects. Blood samples for RNA-Seq were collected at the same time points, hereinafter referred to as M0 (before administration of metformin) and M3m (after three months long metformin course). The study protocol was approved by the Central Medical Ethics Committee of Latvia (No. 01–29.1/22) and the Committee of Ethics in Pauls Stradins Clinical University Hospital (No.3000610 - 18L). Written informed consent was obtained from every participant after a full explanation of the purpose and nature of all procedures used before their inclusion in the study.

2.1.3 Novel susceptibility loci identified in a genome-wide association study of type 2 diabetes complications in population of Latvia

The study cohort consisted of 601 T2DM patients of European ancestry with and without a medical history of diabetes complications, selected from the participants of LGDB (recruited from June 2007 to November 2016) according to previously set inclusion criteria (Table S3). Associated clinical data, including the diagnosis date of T2DM, date and type of diabetes complications, HbA1c measures, and medications used, were obtained from the records of Diabetes registry, Latvian hospital inpatient discharges, outpatient progress notes, and pharmacy recipe records provided by The Centre for Disease Prevention and Control of Latvia and National Health Service of Latvia (Approval No. 3, Decision No. 7.1–3/3). The data about diabetic complications present for T2DM patients involved in LGDB were applied for accurate

stratification of 601 T2DM patients in four phenotype-based groups according to the type of complications experienced: diabetic neuropathy, diabetic nephropathy, ophthalmic complications, and macrovascular complications. The definition of phenotypes and patient stratification in different complication groups was done as follows:

1. Diabetic neuropathies: clinical diagnosis codes (ICD-10) E11.4 and E11.5, records of amputation of the leg/toe, gangrene, shunting and angioplasty, and presence of intermittent claudication or fresh ulcers since the diagnosis of T2DM.
2. Diabetic nephropathies: clinical diagnosis code E11.2 or records of kidney failure, kidney transplantation, renal replacement therapy, microalbuminuria, hemodialysis, peritoneal dialysis performed after the diagnosis of T2DM.
3. Ophthalmic complications: clinical diagnosis code E11.3 or records of photocoagulation, maculopathy, retinopathy, operative therapy, blindness made since the diagnosis of T2DM.
4. Macrovascular complications: clinical diagnosis codes I95, I20, I21, I24, I25, I50, I60, I61, I63, I64, and records of coronary shunting and angioplasty after the diagnosis of T2DM.

ICD-10-based phenotype definitions corresponding to the Latvian guidelines of diabetes management are generally used in clinical practice in Latvia. Subjects with the above-mentioned diagnosis codes or medical events recorded were considered as cases in their corresponding complication groups, while T2DM patients with no evidence of complications of interest during their follow-up period were recognized as controls in the particular group. Subjects experiencing specific diabetes complications before the set of T2DM diagnoses were excluded from the analysis of a particular complication group, explaining the variable total number of individuals among all complication groups tested. The follow-up period which coincides with diabetes duration was considered as time from the set of T2DM diagnosis until the date of diabetes complication recorded for cases or the date of the last entry in the National registry for control subjects. Administration of medications was considered in a group-specific manner, accounting for angiotensin II receptor blockers and angiotensin-converting enzyme inhibitors in the analysis of all complication groups analyzed and additional lipid-modifying agents in the analysis of macrovascular complications.

The study protocol was approved by the Central Medical Ethics Committee of Latvia (No. 01-29.1/2223). Written broad consent was obtained from every subject during the recruitment in LGDB.

2.2 Laboratory methods

2.2.1 RNA extraction, complementary DNA library preparation, and next-generation sequencing

For RNA isolation, 3 ml of whole blood were collected in Tempus™ Blood RNA Tubes (Thermo Fisher Scientific, USA) and further processed using either PerfectPure RNA Blood Kit (5Prime GmbH, Germany) for 74 samples of the cohort of healthy individuals or Tempus™ Spin RNA Isolation Kit (Thermo Fisher Scientific, USA) for 34 samples in T2DM patient cohort according to manufacturer's instructions. The quantity and quality of extracted RNA and prepared libraries were determined by Qubit Fluorometer (Thermo Fisher Scientific, USA) and Agilent 2100 Bioanalyzer systems (Agilent, USA), respectively. The integrity of RNA was evaluated by RNA integrity number (RIN) within the Agilent 2100 Bioanalyzer system (Agilent, USA). For depletion of ribosomal RNA 500 ng of total RNA from each sample were processed using Low Input RiboMinus™ Eukaryote System v2 (Thermo Fisher Scientific, USA). Complementary DNA library preparation was performed with Ion Total RNA-Seq Kit v2 (Thermo Fisher Scientific, USA). Ion Proton™ System (Post-Light™ Ion Semiconductor Sequencing, Thermo Fisher Scientific, USA) and Ion PI™ Chip (Thermo Fisher Scientific, USA) were used for 200-base-read single-end sequencing, following the manufacturer's instructions. Since the shot-gun RNA-Seq is considered to be the most accurate and desirable method for the quantification of the individual transcript and gene expression, additional methods for technical validation were not applied in this study (M. Zhang *et al.* 2019).

2.2.2 Detection of secretory IgA by ELISA in stool samples

Within the framework of the clinical trial two aliquots of stool samples were collected from each healthy individual at three time points: before administration of metformin (M0), 24 hours after the first dose (M24h), and 7 days after the first intake of metformin (M7d), except for two participants for whom the third stool sample was not available (n = 73 in total). From the T2DM patients, the stool samples were also collected: before administration of metformin (M0), 7 days after the first intake of metformin (M7d), and after metformin therapy for three months (M3m) (data not published). The samples were transferred at −80˚C within 24 hours since the collection. The concentration of secretory immunoglobulin A (sIgA) in 100 mg of each stool sample was determined by Immu- Chrom ELISA Kit (ImmuChrom GmbH, Germany), according to the manufacturer's instructions, the absorbance was read at 450nm and 620nm as the reference wavelength (U- Quant MQX200 Microplate reader, Bio-Tek, USA).

2.2.3 DNA isolation and genotyping

DNA was isolated from peripheral blood leukocytes using a phenol-chloroform extraction method according to LGDB standard procedures (Rovite *et al.* 2018). For genotyping, 601 DNA samples were collected according to study inclusion/exclusion criteria and the DNA quantity, which was determined by Qubit Fluorometer (Thermo Fisher Scientific, USA). DNA samples were genotyped with the Infinium Global Screening Array (Illumina, USA) on the iScan System microarray scanner (Illumina, USA) at the University of Tartu.

2.3 Bioinformatics and statistical analysis

2.3.1 Transcriptome data analysis

The trimming of sequencing reads was done using Trimmomatic 0.36 by applying window size 5 and a quality threshold of 10. After trimming reads had to have a minimum length of 30 bp and an average quality of 10 to be included in subsequent analyses. Sequencing reads were mapped against human reference genome GRCh38 release 90 and per-gene read counts were calculated with STAR (v.2.5.3a.). The reads were quantified if they match only one gene. The obtained read counts were normalized using trimmed mean normalization implemented in Bioconductor package edgeR in R (v.3.5.3). FilterByExpr function was applied for gene filtering in edgeR, taking into account the sample library sizes (Law *et al.* 2016).

2.3.2 Longitudinal comparison of transcriptome profiles in cohorts of healthy individuals and T2DM patients

To evaluate longitudinal metformin-induced alterations in the transcriptome profile differentially expressed genes (DEGs) were estimated using the Likelihood ratio test with added observation weights to reduce the influence of outliers, and sva (Surrogate Variable Analysis) package in R was used for removing batch effects (X. Zhou *et al.* 2014b). Each patient was set as a factor to account for the interrelationship between the samples.

In order to account for subject-specific expression in the cohort of healthy subjects, the quasi-likelihood F-test without any prior gene filtering was performed. Multiple testing correction was implemented using the Benjamini-Hochberg procedure, significant DEGs were determined using FDR < 0.05 cutoff regardless of the $\log_2$ fold change of expression for each gene in both of the tests applied (Benjamini 1995).

2.3.3 Comparison of transcriptome profiles in metformin responders against non-responders in T2DM cohort

Likelihood ratio test with added observation weights together with batch effect removal with sva was also used to compare transcriptome profiles between metformin responders and non-responders in each of the time points (M0 and M3m) separately (X. Zhou *et al.* 2014b). In the particular analysis sequencing run and baseline (M0) HbA1c levels were considered as covariates. Multiple testing correction was implemented using the Benjamini-Hochberg procedure (Benjamini 1995) and differential expression of the genes was determined using a FDR < 0.05 cutoff, regardless of the $\log_2$ fold change of expression for each gene. To identify key genes determining the metformin response, Partial least squares discriminant analysis (PLS-DA) was performed implemented in the mixOmics package of R (v.3.5.3). Counts per million (CPM) values (obtained in edgeR and adjusted for the impact of sequencing run and baseline HbA1c levels) were used in PLS-DA. Key genes contributing to a separation of patients in both metformin response groups were identified by using a cutoff of variable importance of projection (VIP) score >1 obtained from PLS-DA (Boulesteix and Strimmer 2007; Rohart *et al.* 2017).

Association between HbA1c levels and the log CPM expression values of each mitochondrial gene was performed with multiple linear regression using lm function in R. Sex and BMI were included in the model to account for their potential confounding.

2.3.4 Functional analysis of differentially expressed genes

Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were adopted as the functional terms. GO enrichment analysis was performed with R package Goseq, and KEGG pathway enrichment analysis was done using either the Goseq package or the online software Database for Annotation, Visualization and Integrated Discovery (DAVID) 6.8 (Huang da *et al.* 2009; Young *et al.* 2010).

Heat maps were constructed with Matplotlib and SciPy. Hierarchical clustering with the average linkage method implemented in SciPy was used for the clustering of genes according to their differences in CPM values (Hunter 2007; P 2007).

2.3.5 Analysis of the genotyping data

Illumina Genome Studio v2.0 was used to convert raw data into PLINK format, while for the quality control the workflow described in Marees *et al*. was used (Marees *et al.* 2018). SHAPEIT v2.r900 (Delaneau *et al.* 2013) and IMPUTE2 (Howie *et al.* 2012) were used for genotype phasing and genotype imputing. Imputed data were filtered using the following

parameters: marker correlation (INFO) > 0.8, hard call threshold 0.1, minor allele frequency > 1%, missingness < 2%.

Association analyses corresponding to four different complication groups (macrovascular complications, diabetic neuropathy, diabetic nephropathy, ophthalmic complications) were performed using PLINK v1.9 logistic regression with covariates: median HbA1c, sex, age at the diagnosis, diabetes duration, BMI, medications used. A genome-wide significance threshold of $P < 5 \times 10^{-8}$ was defined.

For the targeted analysis, candidate variants were selected from GWAS Central (Beck *et al.* 2020) (http://www.gwascentra l.org/) and GWAS Catalog (Buniello *et al.* 2019) (https://www.ebi.ac.uk/gwas/home), based on previously reported association with T2DM complications: diabetic neuropathy (4 allelic variants selected), macrovascular complications (43 allelic variants selected), ophthalmic complications (98 allelic variants selected) and diabetic nephropathy (49 allelic variants selected). A complete list of selected SNPs, their positions, and associated traits is provided in Table S4. FDR according to the Benjamini–Hochberg procedure was calculated to account for multiple testing and the threshold was set < 0.05.

For identification of the functional role of allelic variants expression, quantitative trait locus (eQTL) analysis was conducted by using the open-access Genotype Tissue Expression (GTEx) database (Consortium 2018). The tissue types for eQTL analysis were carefully selected considering the etiology of the diseases (artery, nerve, heart, skin, blood) (Forbes and Cooper 2013). The *p*-value threshold of 0.05 was used to discriminate significant associations. Variant Effect Predictor and Linkage Disequilibrium (LD) data from 1000 Genome project (Utah Residents (CEPH) with Northern and Western European Ancestry) were employed to explore the functional consequences of each variant and other variants in LD (Yates *et al.* 2020). To evaluate the potential single nucleotide polymorphism effects on quantitative phenotypes, analyses of variant association with HbA1C and BMI were performed using PLINK v1.9—assoc function for quantitative phenotypes, where the genome-wide significance threshold of $P < 5 \times 10^{-8}$ was used to identify the significant hits.

Manhattan plots and Q-Q plots were generated in R using the qqman package, while the Venn diagram was developed in the online visualization tool Venny 2.1.0.

2.3.6 Comparison of anthropometric measures and sIgA concentrations

Statistical analysis of anthropometric measures and biochemical data was performed in R by applying the Wilcoxon rank-sum test and Pearson's chi-squared test with a p-value threshold < 0.05. The four-parameter algorithm in GraphPad Prism 8 was used for the calculation of sIgA

concentrations in fecal samples and Wilcoxon signed-rank test in R was applied for the evaluation of fluctuations in fecal sIgA levels among different time points. For the correlation analysis between sIgA levels and CPM values reflecting the gene expression, Spearman's rank correlation test was performed in R, where the significance level was set at 0.05.

3. RESULTS

3.1 Metformin strongly affects transcriptome of peripheral blood cells in healthy individuals

Highlights

1. Significant transcriptomic changes in peripheral blood cells were observed already in 10 hours since the first metformin dose with a higher proportion of downregulated genes (364 of 479 DEGs in total).

2. Out of 561 genes showing significantly altered expression levels among all of the analyzed contrasts, 44 genes including *UBE2O*, *PHOSPHO1*, and *MKRN1* exhibited consistent metformin-evoked alterations in expression levels for 7 days.

2. Universal metformin-induced alterations of global gene expression profiles are associated with immune responses as evidenced by the enrichment of such signaling pathways as the Intestinal immune network for IgA production and Cytokine-cytokine receptor interaction.

3. An inter-individual variation of gene expression profiles was observed indicating subject-specific effects of the drug, which tends to be more permanent and related to the regulation of glucose and cholesterol levels due to differential expression of MODY-related genes (*INS*, *PDX1*, *PAX4*, *HNF4A*, *HNF1A*, *HNF1B*, *NEUROD1*, *GCK*), and hypercholesterolemia-related genes (*LDLR*, *PCSK9*, *APOB*).

4. The observed correlation between fecal sIgA concentration and immunity-related gene (*CXCR4*, *HLA-DQA1*, *MAP3K14*, *TNFRSF21*, *CCL4*, *ACVR1B*, *PF4*, *EPOR*, *CXCL8*) expression levels points at transcriptional shift as a potential constitutive component of the intestinal immunity-related effects of metformin.

PLOS ONE

RESEARCH ARTICLE

Metformin strongly affects transcriptome of peripheral blood cells in healthy individuals

Monta Ustinova[1], Ivars Silamikelis[1], Ineta Kalnina[1], Laura Ansone[1], Vita Rovite[1], Ilze Elbere[1], Ilze Radovica-Spalvina[1], Davids Fridmanis[1], Jekaterina Aladyeva[1], Ilze Konrade[2], Valdis Pirags[1], Janis Klovins[1]*

1 Latvian Biomedical Research and Study Centre, Riga, Latvia, **2** Riga Stradins University, Riga, Latvia

* Klovins@biomed.lu.lv

Citation: Ustinova M, Silamikelis I, Kalnina I, Ansone L, Rovite V, Elbere I, et al. (2019) Metformin strongly affects transcriptome of peripheral blood cells in healthy individuals. PLoS ONE 14(11): e0224835. https://doi.org/10.1371/journal.pone.0224835

Editor: Claudia Miele, Consiglio Nazionale delle Ricerche, ITALY

Received: May 13, 2019

Accepted: October 22, 2019

Published: November 8, 2019

Data Availability Statement: RNA-Seq raw data have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE137317 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE137317). All other relevant data underlying this study are within the manuscript and its Supporting Information files.

Funding: The study was supported by the European Regional Development Fund under the project "Investigation of interplay between multiple

Abstract

Metformin is a commonly used antihyperglycaemic agent for the treatment of type 2 diabetes mellitus. Nevertheless, the exact mechanisms of action, underlying the various therapeutic effects of metformin, remain elusive. The goal of this study was to evaluate the alterations in longitudinal whole-blood transcriptome profiles of healthy individuals after a one-week metformin intervention in order to identify the novel molecular targets and further prompt the discovery of predictive biomarkers of metformin response. Next generation sequencing-based transcriptome analysis revealed metformin-induced differential expression of genes involved in intestinal immune network for IgA production and cytokine-cytokine receptor interaction pathways. Significantly elevated faecal sIgA levels during administration of metformin, and its correlation with the expression of genes associated with immune response (*CXCR4*, *HLA-DQA1*, *MAP3K14*, *TNFRSF21*, *CCL4*, *ACVR1B*, *PF4*, *EPOR*, *CXCL8*) supports a novel hypothesis of strong association between metformin and intestinal immune system, and for the first time provide evidence for altered RNA expression as a contributing mechanism of metformin's action. In addition to universal effects, 4 clusters of functionally related genes with a subject-specific differential expression were distinguished, including genes relevant to insulin production (*HNF1B*, *HNF1A*, *HNF4A*, *GCK*, *INS*, *NEUROD1*, *PAX4*, *PDX1*, *ABCC8*, *KCNJ11*) and cholesterol homeostasis (*APOB*, *LDLR*, *PCSK9*). This inter-individual variation of the metformin effect on the transcriptional regulation goes in line with well-known variability of the therapeutic response to the drug.

Introduction

Metformin is the first-line antidiabetic agent used in pharmacotherapy of type 2 diabetes to improve glucose homeostasis[1]. Various additional therapeutic benefits beyond its antihyperglycaemic action have been highlighted lately, justifying the pleiotropic effect of the drug. In patients with type 2 diabetes metformin therapy is associated with reduced cardiovascular morbidity[2]. In addition, metformin exposure has a protective role against tumorigenesis in various types of cancers and it is proved to be beneficial in the preventive oncology regardless of the diabetic state[3, 4]. Furthermore, metformin therapy is often recommended for women with polycystic ovary syndrome to improve insulin sensitivity, facilitate menstrual regularity,

determinants influencing response to metformin: search for reliable predictors for efficacy of type 2 diabetes therapy" (Project No.: 1.1.1.1/16/A/091, https://ec.europa.eu/regional_policy/en/funding/erdf/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

induce ovulation, reduce circulating androgen levels and body weight[5, 6]. Likewise, metformin application in the treatment of such neurodegenerative disorders as Alzheimer's and Parkinson's diseases is currently under consideration[7–9].

Several possible molecular mechanisms of metformin action have been proposed, including inhibition of mitochondrial respiratory-chain complex 1, the reduction of cyclic adenosine monophosphate (AMP) levels, activation of AMP-activated protein kinase (AMPK) and recently described interaction with gut microbiota[10–13]. Although they partially explain major beneficial effects of the drug, exact mechanisms of metformin action remain unclear even after 60 years since its first clinical use.

Despite the widespread application of metformin, approximately 30% of type 2 diabetes patients using the drug are failing to achieve the adequate glycemic control[14]. Moreover, 20%-30% of type 2 diabetes patients suffer metformin-associated gastrointestinal adverse events and about 5% discontinue the therapy because of severe intolerance[15, 16]. Heritability of the glycaemic response to metformin has been suggested to depend on many allelic variants with small to moderate effects[17]. Contribution of inheritance to variation in metformin response has gained great interest in the past years, encouraging numerous targeted studies investigating genes coding for organic cation transporters OCT1, OCT2, OCT3 and multidrug and toxin extrusion proteins MATE1 and MATE2-K[18–22]. Moreover, Genome-Wide Association Studies have revealed multiple genetic variations within *ATM*, *PRPF31*, *CPA6*, and *STAT* genes associated with metformin response[23–25]. However, genetic alterations explain only a small proportion of the heterogeneous response to metformin therapy, therefore omics-based investigation of the pleiotropic mechanism of the drug is needed to promote the development of biomarkers for therapeutic efficacy[26].

Previous studies have demonstrated metformin-mediated changes at the transcriptome level in various animal tissues, nevertheless studies of metformin-related transcriptome profiles in humans are lacking. For instance, recent study by Guo *et al.* discovered metformin-induced alterations of the coding transcriptome profile and non-coding RNAs in the liver of high-fat diet induced mouse model of non-alcoholic fatty liver disease[27]. Likewise, microarray analysis of mice liver and muscle tissues revealed the ability of metformin to mimic the calorie restriction-like transcriptome[28]. Furthermore, a distinct gene expression profile related to cardiovascular benefits of metformin, was observed in a rat model of obesity and insulin resistance[29]. Meanwhile, cell culture studies of adrenal H295R cell and MCF7 breast cancer cell transcriptome have revealed the association of metformin with complex cellular processes related with energy metabolism, steroidogenesis and the immune system as well as glycolysis and cancer-related pathways[30, 31].

To identify the genes targeted by metformin, we performed the whole-transcriptome analysis with total RNA sequencing on whole-blood samples, obtained from the healthy individuals undergoing a seven-day course of metformin. To the best of our knowledge, this is the first study providing information about the immediate effects of metformin administration on global gene expression in healthy individuals.

Materials and methods

Study design

The study enrolled 25 healthy European descent volunteers with no history of chronic disease, meeting exclusion/inclusion criteria (S4 Table) set within the framework of the ongoing clinical trial 'Pharmacodynamics of antidiabetic drug metformin' (50 individuals to be included in total), protocol number MIKROMET16001, registration number of EU Clinical Trials Register: 2016-001092-74 (www.clinicaltrialsregister.eu) (Fig 1). Participants received twice-daily

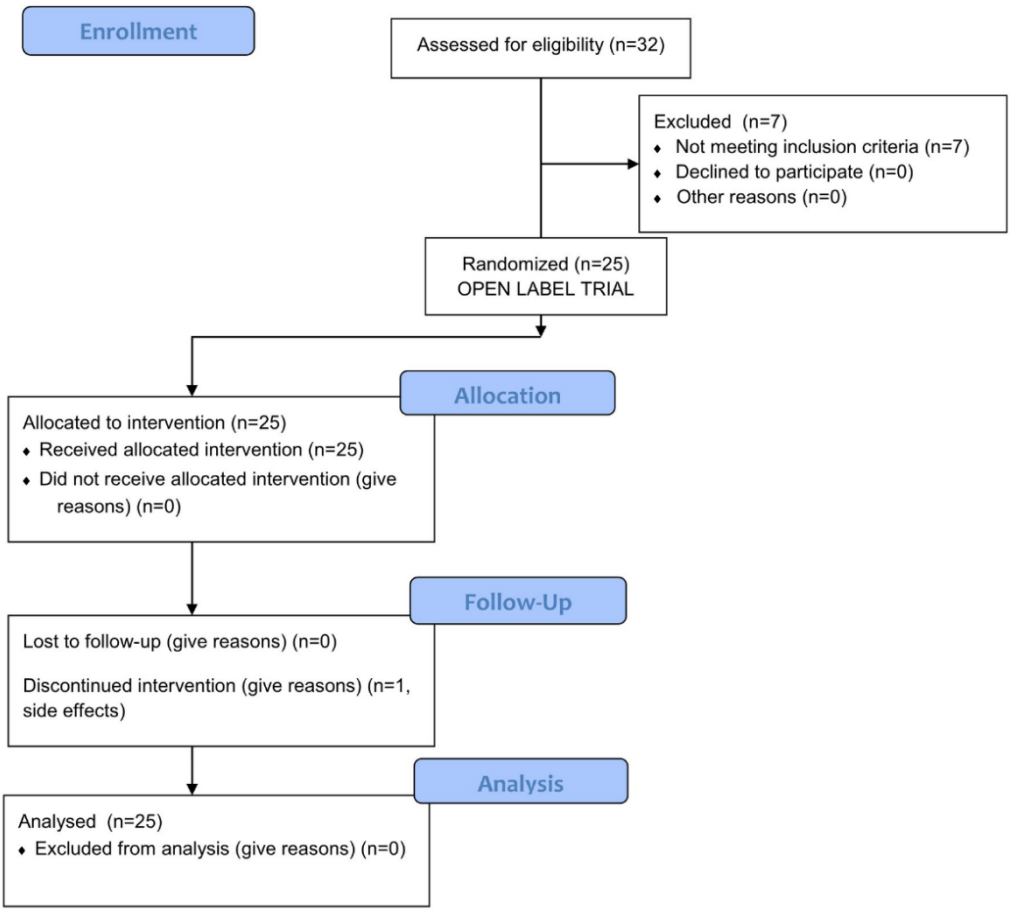

Fig 1. CONSORT flowchart of the open-label trial.

https://doi.org/10.1371/journal.pone.0224835.g001

oral 850mg dose of metformin hydrochloride (*Metforal* 850mg, Berlin-Chemie AG) for 7 days. Medication adherence was reported by each participant at the end of the active period of the study. Fasting blood tests (e.g. measures of ALT, plasma glucose, creatinine levels) were performed in certified clinical laboratory 1–3 days before metformin administration in order to evaluate general hematological and biochemical parameters, and eligibility of volunteers (Table 1). RNA for transcriptome analysis was isolated from the whole blood samples collected

in Latvian Biomedical Research and Study Centre at three time-points: (1) before administration of metformin (M0, morning, fasting state), 10 hours after the first metformin intake, but before the second dose (M10h, evening) and after 7 days long metformin course (M7d, morning, fasting state). The third blood sample was not collected from one out of 25 study subjects, due to the discontinuation of metformin treatment, nevertheless the rest of the blood samples collected from the particular subject were included in the data analysis. Longitudinal study design of the open-label trial was chosen as the most suitable method for global gene expression analysis with high inter-individual variability expected.

All individuals were concurrently involved in an ongoing analysis of gut microbiome and DNA methylation profiles according to the study protocol (S1 Text). The primary endpoint of the study was significantly shifted composition of the gut microbiome after administration of metformin. The secondary endpoint of the study was alterations in DNA methylation profiles and mRNA levels following metformin use.

Written informed consent was obtained from every participant and the study protocol was approved by the Central Medical Ethics committee of Latvia (1/16-05-12) and the State Agency of Medicines of the Republic of Latvia (17–1723). The research was conducted in accordance with the The Code of Ethics of the World Medical Association (Declaration of Helsinki) and International Conference on Harmonisation E6 (R2) Guidelines for Good Clinical Practice. All participants were included in the Genome Database of the Latvian Population[32].

RNA sample preparation and next generation sequencing

Venous blood samples (n = 74) were collected in Tempus Blood RNA Tubes, followed by total RNA isolation with PerfectPure RNA Blood Kit (5Prime GmbH, Germany), according to the manufacturer's instructions. The integrity of the extracted RNA was evaluated by RNA integrity number (RIN) within Agilent 2100 Bioanalyzer system (Agilent, USA). Ribosomal RNA depletion was done with Low Input RiboMinus Eukaryote System v2 (Thermo Fisher Scientific, USA) by processing 500ng of total RNA from each sample. For cDNA library preparation we used Ion Total RNA-Seq Kit v2 (Thermo Fisher Scientific, USA), sequencing was done on Ion Proton System and Ion PI Chip (Thermo Fisher Scientific, USA), following the manufacturer's instructions. Since shot-gun RNA sequencing is considered to be the most accurate and desirable method for quantification of expression of individual transcripts and genes, additional methods for technical validation were not applied in this study[33].

Stool sample collection and detection of secretory IgA by ELISA

Within the framework of the clinical trial two aliquots of stool samples were collected from each study participant at three time points: before administration of metformin (M0), 24 hours after the first dose (M24h) and 7 days after the first intake of metformin (M7d), except for two participants for whom the third stool sample was not available (n = 73 in total). The samples were transferred at −80˚C within 24 hours since the collection. The concentration of secretory immunoglobulin A (sIgA) in 100mg of each stool sample was determined by ImmuChrom ELISA Kit (ImmuChrom GmbH, Germany), according to the manufacturer's instructions, the absorbance was read at 450nm and 620nm as the reference wavelength.

Bioinformatic analysis

The sequencing reads were trimmed with Trimmomatic 0.36 using window size 5 and quality threshold 10. After trimming reads had to have a minimum length of 30bp and average quality of 10 to be included in subsequent analyzes. Sequencing reads were mapped against human reference genome GRCh38 release 90 and read counts were calculated with STAR 2.5.3a. The

obtained read counts were normalized using trimmed mean normalization implemented in Bioconductor package edgeR in R. Differentially expressed genes (DEGs) were estimated with two different methods. At first, likelihood ratio test with added observation weights was used to reduce the influence of outliers and to obtain a list of DEGs, where the counts per million (CPM) value had to be 1 or more in at least 24 samples for the gene to be included in the analysis (edgeR-robust)[34]. In order to account for subject-specific expression the quasi-likelihood F-test without any prior gene filtering was performed (edgeR-sensitive). Multiple testing correction was implemented using Benjamini-Hochberg procedure, significant DEGs were determined using false discovery rate (FDR) < 0.05 cutoff[35].

Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were adopted as the functional terms. GO and KEGG pathway enrichment analysis were performed with R package Goseq (1.30.0)[36]. KEGG pathways and GO terms with FDR < 0.05 were considered statistically significant.

Heat map was constructed with Matplotlib2 and SciPy3. Hierarchical clustering with average linkage method implemented in SciPy was used to cluster DEGs for contrasts M10h vs M0 and M7d vs M0 by their differences in read counts per million[37, 38].

Four-parameter algorithm in GraphPad Prism 8 was used for the calculation of sIgA concentrations in faecal samples and Wilcoxon signed rank test in R was applied for the evaluation of fluctuations in faecal sIgA levels among different time points. For the correlation analysis between sIgA levels and CPM values obtained in edgeR-robust analysis method, Spearman's rank correlation test was performed in R, where significance level was set at 0.05.

Results

Differential global gene expression induced by administration of metformin

In order to reveal the target genes and pathways affected by metformin, we performed a transcriptome analysis in whole-blood samples of 25 healthy volunteers receiving metformin for one week (Table 1). Venous blood samples were obtained at three consecutive time-points, hereinafter referred to as M0 (before administration of metformin), M10h (10 hours after the first metformin intake/before the second dose) and M7d (after 7 days long metformin course) in order to observe both, acute and sustained effects of metformin. RNA-Sequencing produced an average of 24.6 ± 8.9 million reads per sample, 83.5% of the reads were mapped to the reference genome.

We focused on DEGs at three contrasts comparing the samples collected at previously defined time-points: M10h vs M0, M7d vs M0 and M7d vs M10h. In total 561 unique genes showed significantly different expression levels among the analyzed contrasts (Fig 2A, 2B, 2C

Table 1. Characteristics of the study group.

Characteristic	Value
Female/ male, n (%)	18 (72.0%) / 7 (28.0%)
Mean age (years) ± SD	34.4 ± 10.8
Mean BMI ± SD	25.5 ± 3.1
Mean ALAT ± SD, U/l	24.8 ± 11.9
Mean creatinine ± SD, μmol/l	66.2 ± 9.0
Fasting plasma glucose (mmol/l), mean ± SD	5.2 ± 0.4

BMI—body mass index; SD—standard deviation; ALAT—alanine aminotransferase

https://doi.org/10.1371/journal.pone.0224835.t001

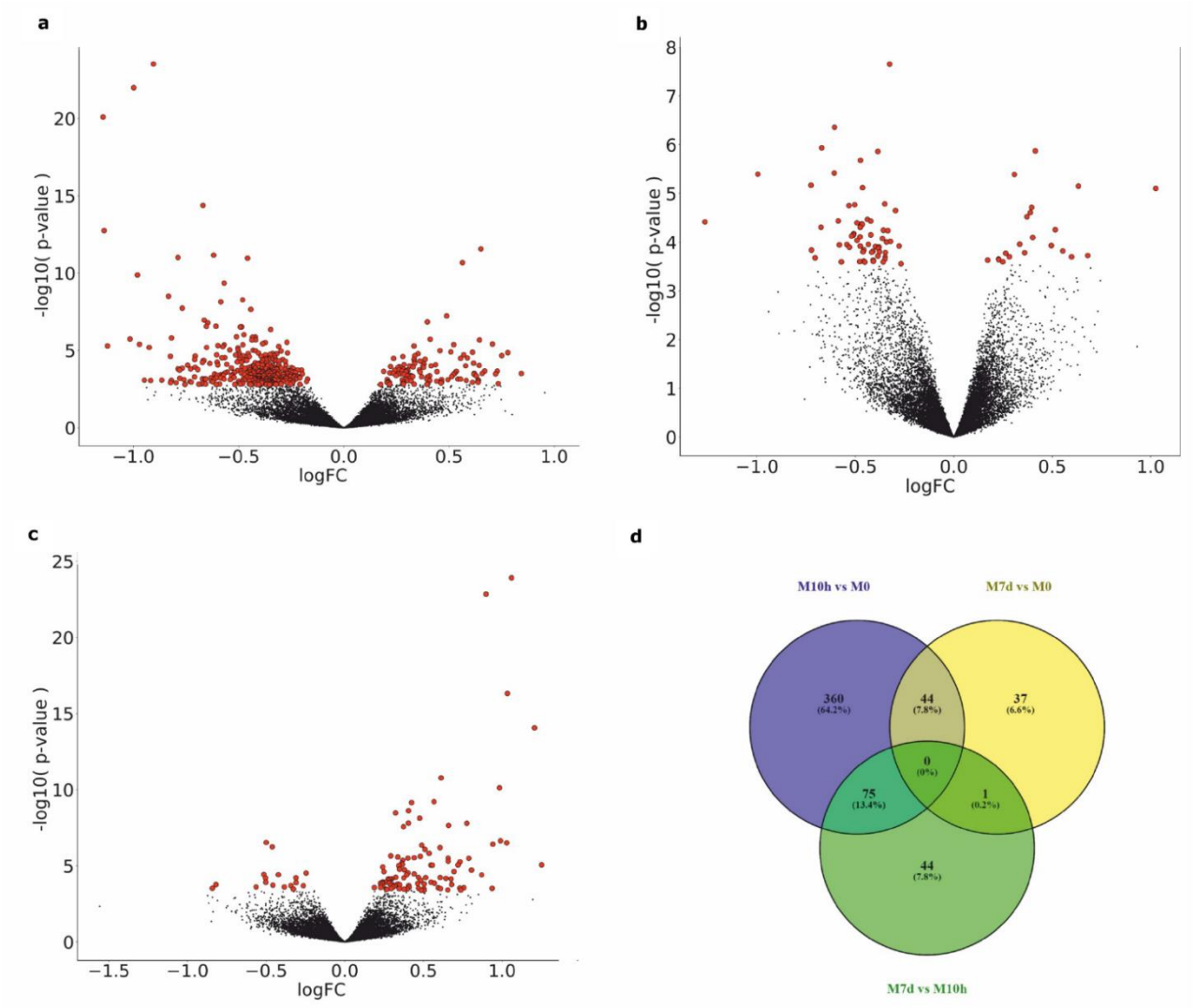

Fig 2. Metformin-induced alterations in gene expression profiles. Volcano plots showing the distribution of gene expression in the analyzed contrasts: (A)—M10h vs M0, (B)—M7d vs M0 and (C)—M7d vs M10h. Significance versus $\log_2$ fold change is plotted on the y and x axes, respectively. Red dots represent the significant DEGs (FDR < 0.05), black dots—nonsignificant genes. (D)—Venn diagram representing the number of total and overlapping significant DEGs in the analyzed contrasts obtained by edgeR-robust method.

https://doi.org/10.1371/journal.pone.0224835.g002

and 2D). The majority, 479 of DEGs appeared in the contrast M10h vs M0 (364 downregulated and 115 upregulated). Comparison of M7d vs M0 resulted in 82 DEGs (61 downregulated, 21 upregulated) and 120 DEGs were identified in the contrast M7d vs M10h with higher proportion of upregulated genes (17 downregulated, 103 upregulated) (Fig 3). The overlap of the two main contrasts (M10h vs M0 and M7d vs M0) consisted of 44 DEGs, including *UBE2O*, *PHOSPHO1*, *MKRN1*, possessing consistent metformin-evoked alterations in expression levels for 7 days. The complete lists of obtained DEGs are provided in S1 Table.

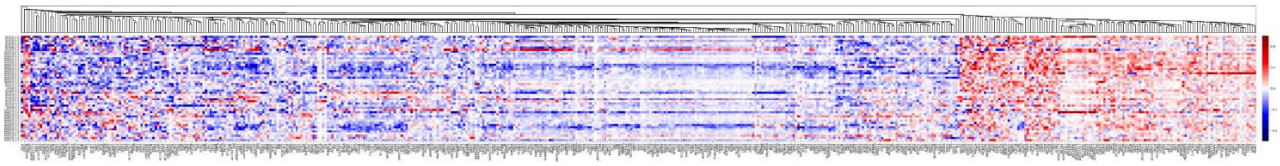

Fig 3. Heat map and hierarchical clustering of 517 DEGs in the contrasts M10h vs M0 and M7d vs M0. Each row corresponds to one subject in the respective contrast and each column represents a DEG. Normalized sequence read counts were rescaled to lie in range [0,1] and further used to estimate the difference between the gene expression levels in two time-points depending on the particular contrast. DEGs with analogous expression values were clustered at the column level, the list of DEGs were obtained by edgeR-robust method.

https://doi.org/10.1371/journal.pone.0224835.g003

Functional annotation of the identified DEGs

In order to gain insights into the molecular mechanisms underlying the short-term effects of metformin the KEGG pathway enrichment analysis was performed. Three lists of DEGs corresponding to each contrast were submitted in Goseq package of R. DEGs obtained from the contrasts M10h vs M0 and M7d vs M10h were assigned to pathways related with immune responses, while list of DEGs obtained by comparing M7d samples with M0 samples showed no pathways enriched (Table 2). In order to gain a greater understanding of the biological implications of the obtained DEGs Gene Ontology (GO) enrichment analysis was performed, see S2 Table and S3 Table for the complete list of GO terms and KEGG pathways corresponding to DEGs obtained in all of the contrasts analyzed.

Elevated faecal sIgA levels during the administration of metformin

Considering the enrichment of pathways related to intestinal immune responses, sIgA concentration was determined in stool samples collected from the study participants at three consecutive time points, analogous to the course of blood sample collection: before administration of metformin (M0; median sIgA concentration = 7969.93μg/ml; IQR = 15587.55), within 24

Table 2. Top KEGG pathways enriched by short-term metformin administration, ranked by statistical significance.

Contrast	Pathways	Count	Genes	Adjusted-p
M10h vs M0	Malaria	11	*GYPC*$^{\downarrow}$, *SELP*$^{\downarrow}$, *CD40LG*$^{\uparrow}$, *TLR2*$^{\downarrow}$, *CXCL8*$^{\downarrow}$, *HBA2*$^{\downarrow}$, *HBA1*$^{\downarrow}$, *CD40*$^{\uparrow}$, *THBS1*$^{\downarrow}$, *TGFB1*$^{\downarrow}$, ***SDC2***$^{\downarrow}$	7.88E-05
	Intestinal immune network for IgA production	10	*CXCR4*$^{\downarrow}$, *CD40LG*$^{\uparrow}$, *TNFSF13*$^{\downarrow}$, *CD40*$^{\uparrow}$, *HLA-DOA*$^{\uparrow}$, *MAP3K14*$^{\uparrow}$, ***HLA-DOB***$^{\uparrow}$, *TGFB1*$^{\downarrow}$, *HLA-DQA1*$^{\uparrow}$, *CD28*$^{\uparrow}$	7.88E-05
	Cytokine-cytokine receptor interaction	21	***TNFRSF21***$^{\downarrow}$, *IL1R2*$^{\downarrow}$, *FLT3*$^{\downarrow}$, *LEPR*$^{\downarrow}$, *CXCL8*$^{\downarrow}$, *TNFSF13*$^{\downarrow}$, *PF4*$^{\downarrow}$, *CD40*$^{\uparrow}$, *PF4V1*$^{\downarrow}$, *CCL4*$^{\downarrow}$, *TGFB1*$^{\downarrow}$, *ACVR1B*$^{\downarrow}$, *PPBP*$^{\downarrow}$, *CXCR4*$^{\downarrow}$, *CD40LG*$^{\uparrow}$, *EPOR*$^{\downarrow}$, *IL5RA*$^{\downarrow}$, *EGF*$^{\downarrow}$, *IL13RA1*$^{\downarrow}$, *IL3RA*$^{\downarrow}$, *CXCL5*$^{\downarrow}$.	7.88E-05
	Cell adhesion molecules (CAMs)	13	*ESAM*$^{\downarrow}$, ***CLDN5***$^{\downarrow}$, *HLA-DOB*$^{\uparrow}$, *SELP*$^{\downarrow}$, *HLA-DOA*$^{\uparrow}$, *CD28*$^{\uparrow}$, *PTPRF*$^{\downarrow}$, *CD8B*$^{\uparrow}$, *HLA-DQA1*$^{\uparrow}$, *CD40*$^{\uparrow}$, *VCAN*$^{\downarrow}$, *CD40LG*$^{\uparrow}$, *SDC2*$^{\downarrow}$.	6.49E-03
	Hematopoietic cell lineage	11	*MS4A1*$^{\uparrow}$, *IL5RA*$^{\downarrow}$, *IL3RA*$^{\uparrow}$, *CD8B*$^{\uparrow}$, ***FLT3***$^{\downarrow}$, *FCER2*$^{\uparrow}$, *CD19*$^{\uparrow}$, *GP9*$^{\downarrow}$, *IL1R2*$^{\downarrow}$, *ITGA2B*$^{\downarrow}$, *EPOR*$^{\downarrow}$.	6.49E-03
	Autoimmune thyroid disease	7	***HLA-DOB***$^{\uparrow}$, *HLA-DOA*$^{\uparrow}$, *CD28*$^{\uparrow}$, *GZMB*$^{\uparrow}$, *HLA-DQA1*$^{\uparrow}$, *CD40*$^{\uparrow}$, *CD40LG*$^{\uparrow}$.	6.49E-03
	Allograft rejection	7	***HLA-DOB***$^{\uparrow}$, *HLA-DOA*$^{\uparrow}$, *CD28*$^{\uparrow}$, *GZMB*$^{\uparrow}$, *HLA-DQA1*$^{\uparrow}$, *CD40*$^{\uparrow}$, *CD40LG*$^{\uparrow}$.	6.49E-03
	Rheumatoid arthritis	10	***HLA-DOB***$^{\uparrow}$, *HLA-DOA*$^{\uparrow}$, *CD28*$^{\uparrow}$, *TLR2*$^{\downarrow}$, *HLA-DQA1*$^{\uparrow}$, *CXCL5*$^{\downarrow}$, *TNFSF13*$^{\downarrow}$, *TGFB1*$^{\downarrow}$, *FOS*$^{\downarrow}$, *CXCL8*$^{\downarrow}$.	1.29E-02
	Graft-versus-host disease	7	*HLA-DOB*$^{\uparrow}$, *HLA-DOA*$^{\uparrow}$, *KLRD1*$^{\downarrow}$, *CD28*$^{\uparrow}$, *GZMB*$^{\uparrow}$, *HLA-DQA1*$^{\uparrow}$, ***KIR2DL1***$^{\downarrow}$.	1.49E-02
	Asthma	5	*HLA-DOB*$^{\uparrow}$, *HLA-DOA*$^{\uparrow}$, *HLA-DQA1*$^{\uparrow}$, *CD40*$^{\uparrow}$, *CD40LG*$^{\uparrow}$.	2.84E-02

Within each pathway one gene showing the highest expression variability, based on $\log_2$ fold change, is indicated in bold.

$^{\uparrow}$ Upregulated genes

$^{\downarrow}$ Downregulated genes

https://doi.org/10.1371/journal.pone.0224835.t002

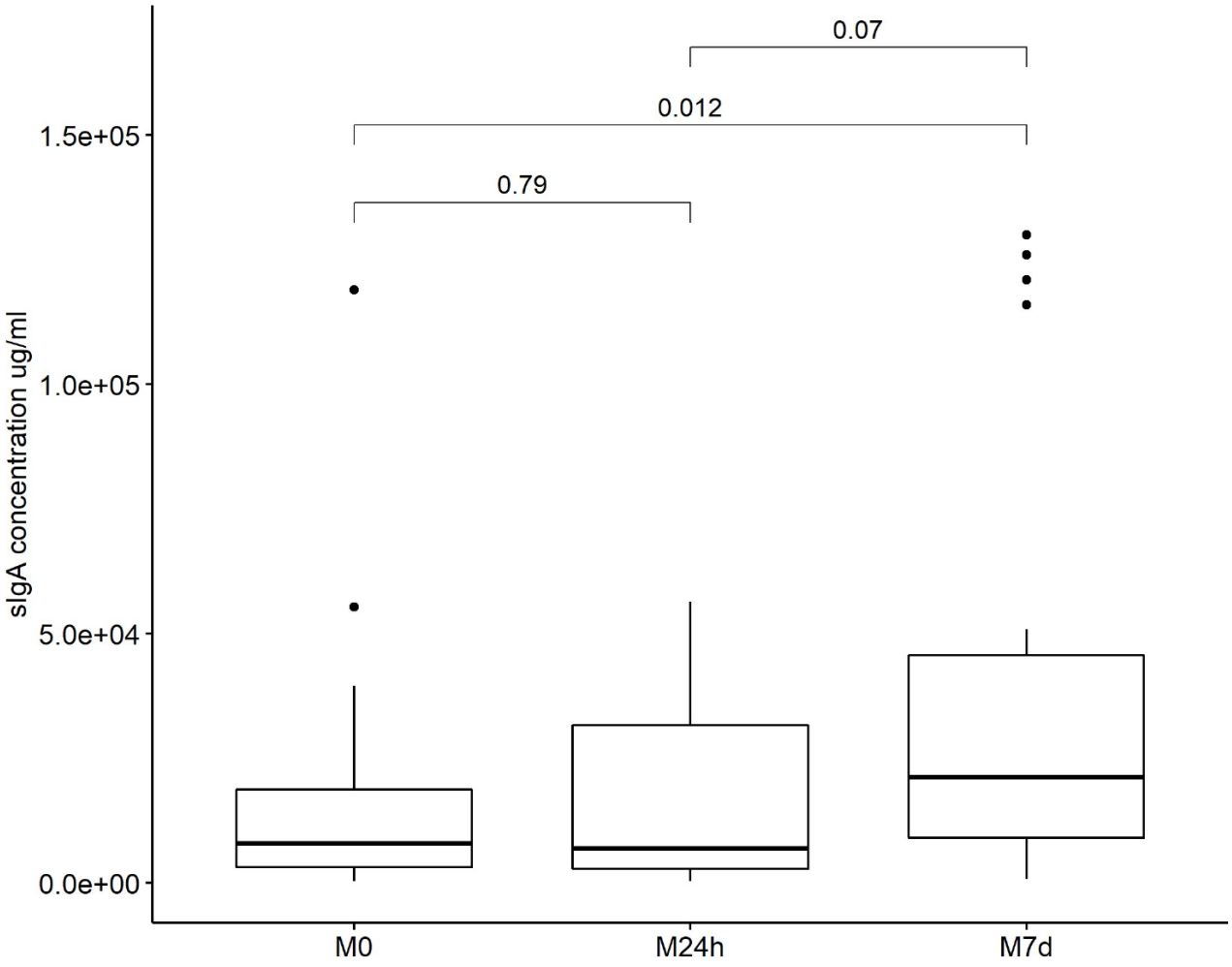

Fig 4. sIgA levels in stool samples during administration of metformin. Boxplot showing the difference in faecal sIgA levels at three time points: before administration of metformin (M0), 24 hours after the first dose (M24h) and 7 days after the first intake of metformin (M7d), measured by ELISA.

https://doi.org/10.1371/journal.pone.0224835.g004

hours after the first dose (M24h; median sIgA concentration = 6935.29μg/ml; IQR = 28953.46) and 7 days after the first intake of metformin (M7d; median sIgA concentration = 21207.64μg/ml; IQR = 36642.19), revealing significantly increased faecal sIgA levels comparing M7d samples vs M0 samples and M7d samples vs M24h samples (Fig 4).

The possible implication of specific immunity-related genes in intestinal IgA production in response to metformin administration was evaluated by Spearman's rank correlation test between faecal sIgA concentration and expression level of DEGs involved in Intestinal immune network for IgA production and Cytokine-cytokine receptor interaction pathways, revealing 9 significant correlations (Table 3).

Subject-specific effects of metformin

In order to clarify subject-specific metformin effects on the transcriptome profile, we performed an additional data analysis (quasi-likelihood F-test without any prior gene filtering

Table 3. Spearman's correlation between faecal sIgA levels and expression of immunity-related genes.

Pathway	Gene	Spearman's correlation coefficient	*p*-value
Intestinal immune network for IgA production	***CXCR4****	0.40	5.15E-04
	HLA-DQA1	0.30	9.11E-03
	MAP3K14	0.25	3.01E-02
	HLA-DOA	0.19	1.17E-01
	HLA-DOB	0.16	1.64E-01
	CD28	-0.15	2.01E-01
	*TNFSF13**	0.15	2.18E-01
	*CD40LG**	0.07	5.30E-01
	*CD40**	0.01	9.53E-01
	*TGFB1**	0.14	2.42E-01
Cytokine-cytokine receptor interaction	***TNFRSF21***	0.32	6.54E-03
	CCL4	0.30	8.90E-03
	ACVR1B	0.26	2.44E-02
	PF4	0.25	3.29E-02
	EPOR	0.24	3.73E-02
	CXCL8	-0.24	3.92E-02
	PF4V1	0.21	7.07E-02
	PPBP	0.20	9.80E-02
	FLT3	0.12	2.96E-01
	LEPR	-0.09	4.49E-01
	IL3RA	0.08	4.77E-01
	CXCL5	-0.07	5.47E-01
	EGF	-0.02	8.59E-01
	IL13RA1	-0.01	9.57E-01
	IL1R2	0.01	9.63E-01
	IL5RA	0.00	9.75E-01

*Genes involved in both Intestinal immune network for IgA production and Cytokine-cytokine receptor interaction pathways.
Genes showing significant correlation with faecal sIgA levels are indicated in bold.

https://doi.org/10.1371/journal.pone.0224835.t003

hereinafter referred to as edgeR-sensitive), which allowed us to identify 437 unique DEGs among all of the contrasts (S1 Table; S1A, S1B, S1C and S1D Fig). After a careful inspection of the generated heat map for the contrasts M10h vs M0 and M7d vs M0, we observed striking inter-individual variation in expression levels of most significant DEGs (S2 Fig). It was evident that the overall differences in expression levels for a number of gene clusters were influenced by extreme changes in expression observed only in few individuals from the study. Four such functional clusters of genes with similar expression profiles were distinguished after consideration of subject-specific effects. In each of the recognized clusters, genes sharing a common function were predominant, including upregulated genes, coding for small nuclear RNAs (snRNAs): *SNORA20*, *SCARNA5*, *SNORA80E*, *SNORA3B*, *SCARNA22*, *SCARNA6*, *SNORD8*, *SNORA7B*, *SNORD9*, *SNORD83A*, *SNORA23*, *SNORD67*, *SNORD46*, *SNORA71B*, *SNORD71*, *SNORD11B*, *SNORD17*, *SNORD116-15*, *SNORD67*; ribosomal genes and pseudogenes together with long non-coding RNAs (lncRNAs): *Y_RNA*, *RNY4P10*, *RNA5-8SN1*, *RNA5-8SN5*, *FP671120.4*, *FP236383.2*, *FP671120.3*, *FP236383.3*, *RPL3P4*, *AC008038.1*, *AL122020.1*, *RPL13AP5*, *RPS2P46*, *RPL6P27*, *RPL13AP25*, *AL590867*, *RPS7P1*, *AP001324.1*, *AC034236.1*, *RPL37P2*, *AC004453.1*, *RPL21P16*, *AC007969.1*, *RPS23P8*, *RPL15P3*, *RPL10AP6*; genes relevant to insulin production: *HNF1B*, *HNF1A*, *HNF4A*, *GCK*, *INS*, *NEUROD1*, *PAX4*, *PDX1*,

ABCC8, *KCNJ11* and downregulated genes involved in cholesterol homeostasis: *APOB*, *LDLR*, *PCSK9*.

Discussion

This longitudinal study of metformin administration in healthy subjects demonstrated that metformin strongly affects the gene expression profiles in blood cells as estimated by RNA sequencing. To our knowledge this is the first study accessing the metformin-mediated changes in RNA expression *in vivo* in humans. Moreover, the significant transcriptomic changes were observed even after 10 hours as a result of single metformin dose, indicating the pronounced and immediate influence on the cell functions. The most striking finding was the strong evidence for metformin-induced enrichment of immunity-related pathways resulting in elevated faecal IgA levels.

Multiple DEGs identified in this study represent the main functional groups associated with previously described therapeutic targets of metformin. For instance, downregulation of gene coding for ubiquitin conjugating enzyme E2 O (*UBE2O*) impairs the tumorigenesis, moreover combined treatment of *UBE2O* inhibitors and AMPK agonists, such as metformin, has been suggested as promising treatment strategy for cancer already before[39]. Similarly, suppression of *MKRN1* (makorin ring finger protein 1) activates AMPK, resulting in increased glucose consumption and reduced lipid accumulation, therefore *MKRN1*-mediated regulation of AMPK activity has been already considered as an attractive therapeutic approach for the treatment of metabolic disorders[40]. Finally, DNA methylation at the *PHOSPHO1* (phosphoethanolamine/phosphocholine phosphatase 1) locus in blood cells has been previously linked to decreased type 2 diabetes mellitus risk, which goes in line with the well-known anti-diabetic activity of metformin and metformin-evoked downregulation of *PHOSPHO1* observed in this study[41].

A group of discovered DEGs correspond to the pathways involved in immune response or regulation of inflammation. The pathway enrichment analysis showed comprehensively decreased expression levels of genes related to immune responses, which confirms the anti-inflammatory effect as a universal property of metformin. Here we convincingly support the previously described ability of metformin to suppress inflammatory cytokines and their receptors irrespective of the diabetes status, specifying the occurrence of this process at the level of mRNA[42]. The downregulation of *CXCL8* and *CXCR4*, coding for interleukin-8 and chemokine receptor type 4 respectively, has been previously attributed to the anticancer action of metformin in neoplastic cells, therefore the results of the present study describe the particular therapeutic effects as universal and characteristic also to the normal blood cells[43, 44].

Furthermore, RNA-Seq data revealed significant enrichment in the pathway of intestinal immune network for IgA production, which was further confirmed by elevated faecal sIgA in response to the metformin intervention. The observed correlation between faecal sIgA concentration and immunity-related gene expression levels points at transcriptional shift as a constitutive component of the intestinal immunity-related effects of metformin. IgA is responsible for bacteria-host interaction and is massively produced by mucosa in case of bacterial colonization; moreover, selective IgA deficiency is associated with mild intestinal dysbiosis and shifts in the microbial composition[45, 46]. Since the influence of metformin on gut microbiome is currently extensively studied[47], the potential implication of the intestinal immune network for IgA pathway in the metformin-microbiome interactions regarding the intestinal side effects should be considered.

In addition, we noticed subject-specific differences at gene expression levels, as number of individuals involved in study displayed overexpression or repression of specific functionally

related gene sets. Thus for example insulin coding gene (*INS*) showed an apparent gain of the mRNA expression levels (from 0.52 CPM before the treatment to 282.97 CPM after one week of metformin administration) in a single individual. Based on the hierarchical clustering and functionality of these genes we categorized 4 gene clusters: ribosomal genes and their pseudogenes, small nuclear RNAs, genes relevant to insulin production and cholesterol homeostasis. Interestingly, the changes in RNA expression of those genes are highly subject-specific with strongly altered expression in only one or two participants of the study.

Metformin-induced overexpression of Maturity onset diabetes of the young (MODY)—related genes (*INS*, *PDX1*, *PAX4*, *HNF4A*, *HNF1A*, *HNF1B*, *NEUROD1*, *GCK*) coding for transcription factors and regulators of β-cell function was observed in both main comparisons (M10h vs M0, M7d vs M0), but not in the contrast M7d vs M10h, suggesting that these alterations are likely to be associated with metformin intervention rather than discrepancies between fasting and feasting states at the time points of blood collection. Metformin-induced differential expression of MODY genes has been previously reported in the liver of spontaneously hypertensive rats, overlapping several homologues of human genes[48]. Moreover, subject-specific activation of the MODY genes may be the reason why the metformin ability to induce insulin secretion has not been observed before in human trials. One may speculate that metformin exerts insulin secretagogue ability only in subgroup of metabolically compromised individuals, however, to prove this additional research in patients with metabolic syndrome and diabetes is needed.

Very similar to our detection of the MODY cluster our study revealed a considerable downregulation of the genes coding for apolipoprotein-B (APOB), low-density lipoprotein receptor (LDLR) and proprotein convertase subtilisin/kexin type 9 (PSCK9) in one person from the study group. All of these genes are previously associated with cholesterol homeostasis and phenotype of familial hypercholesterinaemia[49–51]. To date, several reports have described a dose-dependent cholesterol lowering effect of metformin[52, 53]. Taken together, it would be reasonable to argue that the downregulation of *LDLR*, *PCSK9*, *APOB* might serve as potential mechanism of action, underlying beneficial cardiovascular properties of metformin, yet in a case-specific manner.

The following are some study limitations. These limitations include small sample size and only one week of the intervention time due to the fact that study was performed on healthy individuals. Another limitation is that this is an exploratory study without a placebo control arm that would be needed to draw definitive conclusions on the causality of observed transcriptional changes. Also a larger study group of diabetic subjects and longer observation time would provide information on the factors that may explain subject-specific differences in expression levels and relation of these differences to the treatment response including glucose control.

These limitations are compensated by the longitudinal design of the study in which the first sample from an individual was the control for further samples. We believe that this design and short-term observation should have minimized false associations and conclusions arising from unaccounted factors playing important role in human studies, meanwhile making reasonable the interpretation of observed inter-individual variability of gene expression profiles.

In conclusion, we were able to provide, for the first time, direct evidence of the effects of metformin on the immediate and strong transcriptome changes in whole-blood samples. Our results have pinpointed some important targets that need further investigation. First, the ability of metformin to induce extensive immune responses may be executed at the level of transcription and serve as the basis of common therapeutic effect of metformin. Second, the induction of IgA pathway may explain the widely discussed interaction of metformin with the gut microbiome. Third, the subject-specific response may explain the large percent of

unresponsiveness to the metformin therapy. Altogether these results may serve the ground for development of expression based biomarker sets to predict and/or monitor the treatment outcomes.

Supporting information

S1 Table. List of differentially expressed genes.
(XLSX)

S2 Table. Gene ontology analysis.
(XLSX)

S3 Table. KEGG pathway enrichment.
(XLSX)

S4 Table. Inclusion and exclusion criteria of the clinical trial 'Pharmacodynamics of anti-diabetic drug metformin'.
(PDF)

S5 Table. CONSORT Checklist.
(PDF)

S1 Text. Protocol of the clinical trial 'Pharmacodynamics of antidiabetic drug metformin'.
(PDF)

S1 Fig. Metformin-induced alterations in gene expression profiles, considering subject-specific effects. Volcano plots showing the distribution of gene expression in the analyzed contrasts: (A)—M10h vs M0, (B)—M7d vs M0 and (C)—M7d vs M10h. Significance versus log2 fold change is plotted on the y and x axes, respectively. Red dots represent the significant DEGs (FDR < 0.05), black dots—nonsignificant genes. (D)—Venn diagram representing the number of total and overlapping significant DEGs in the analyzed contrasts, DEGs are obtained in the edgeR-sensitive analysis.
(TIFF)

S2 Fig. Heat map and hierarchical clustering of 404 significant DEGs in the contrasts M10h vs M0 and M7d vs M0, accounting for subject-specific effects. Each row corresponds to one subject in the respective contrast and each column represents a DEG. Normalized sequence read counts were rescaled to lie in range [0,1] and further used to estimate the difference between the gene expression levels in two time-points depending on the particular contrast. DEGs with analogous expression values were clustered at the column level. Line plots show the expression levels (read counts per million) of the most representative genes of each subject-specific gene cluster in three blood sample collection time-points of one representative subject. DEGs are obtained in the edgeR-sensitive analysis.
(TIFF)

Acknowledgments

The authors would like to thank all the volunteers for their participation and acknowledge the Genome Database of the Latvian Population for providing biological material and data.

Author Contributions

Conceptualization: Janis Klovins.

Data curation: Ivars Silamikelis.

Formal analysis: Ivars Silamikelis.

Funding acquisition: Ineta Kalnina.

Investigation: Monta Ustinova, Laura Ansone, Ilze Radovica-Spalvina.

Project administration: Janis Klovins.

Resources: Vita Rovite, Ilze Elbere, Davids Fridmanis, Ilze Konrade, Valdis Pirags.

Supervision: Valdis Pirags, Janis Klovins.

Visualization: Jekaterina Aladyeva.

Writing – original draft: Monta Ustinova.

References

1. Standards of Medical Care in Diabetes-2017: Summary of Revisions. Diabetes Care. 2017; 40(Suppl 1):S4–S5. https://doi.org/10.2337/dc17-S003 PMID: 27979887
2. Johnson JA, Simpson SH, Toth EL, Majumdar SR. Reduced cardiovascular morbidity and mortality associated with metformin use in subjects with Type 2 diabetes. Diabet Med. 2005; 22(4):497–502. https://doi.org/10.1111/j.1464-5491.2005.01448.x PMID: 15787679
3. Quinn BJ, Kitagawa H, Memmott RM, Gills JJ, Dennis PA. Repositioning metformin for cancer prevention and treatment. Trends Endocrinol Metab. 2013; 24(9):469–80. https://doi.org/10.1016/j.tem.2013.05.004 PMID: 23773243
4. Currie CJ, Poole CD, Jenkins-Jones S, Gale EA, Johnson JA, Morgan CL. Mortality after incident cancer in people with and without type 2 diabetes: impact of metformin on survival. Diabetes Care. 2012; 35 (2):299–304. https://doi.org/10.2337/dc11-1313 PMID: 22266734
5. Velazquez EM, Mendoza S, Hamer T, Sosa F, Glueck CJ. Metformin therapy in polycystic ovary syndrome reduces hyperinsulinemia, insulin resistance, hyperandrogenemia, and systolic blood pressure, while facilitating normal menses and pregnancy. Metabolism. 1994; 43(5):647–54. https://doi.org/10.1016/0026-0495(94)90209-7 PMID: 8177055
6. Tang T, Lord JM, Norman RJ, Yasmin E, Balen AH. Insulin-sensitising drugs (metformin, rosiglitazone, pioglitazone, D-chiro-inositol) for women with polycystic ovary syndrome, oligo amenorrhoea and subfertility. Cochrane Database Syst Rev. 2012(5):CD003053. https://doi.org/10.1002/14651858.CD003053.pub5 PMID: 22592687
7. Gupta A, Bisht B, Dey CS. Peripheral insulin-sensitizer drug metformin ameliorates neuronal insulin resistance and Alzheimer's-like changes. Neuropharmacology. 2011; 60(6):910–20. https://doi.org/10.1016/j.neuropharm.2011.01.033 PMID: 21277873
8. Asadbegi M, Yaghmaei P, Salehi I, Ebrahim-Habibi A, Komaki A. Neuroprotective effects of metformin against Abeta-mediated inhibition of long-term potentiation in rats fed a high-fat diet. Brain Res Bull. 2016; 121:178–85. https://doi.org/10.1016/j.brainresbull.2016.02.005 PMID: 26861514
9. Mendelsohn AR, Larrick JW. Rapamycin as an antiaging therapeutic?: targeting mammalian target of rapamycin to treat Hutchinson-Gilford progeria and neurodegenerative diseases. Rejuvenation Res. 2011; 14(4):437–41. https://doi.org/10.1089/rej.2011.1238 PMID: 21851176
10. Zhou G, Myers R, Li Y, Chen Y, Shen X, Fenyk-Melody J, et al. Role of AMP-activated protein kinase in mechanism of metformin action. J Clin Invest. 2001; 108(8):1167–74. https://doi.org/10.1172/JCI13505 PMID: 11602624
11. Miller RA, Chu Q, Xie J, Foretz M, Viollet B, Birnbaum MJ. Biguanides suppress hepatic glucagon signalling by decreasing production of cyclic AMP. Nature. 2013; 494(7436):256–60. https://doi.org/10.1038/nature11808 PMID: 23292513
12. El-Mir MY, Nogueira V, Fontaine E, Averet N, Rigoulet M, Leverve X. Dimethylbiguanide inhibits cell respiration via an indirect effect targeted on the respiratory chain complex I. J Biol Chem. 2000; 275 (1):223–8. https://doi.org/10.1074/jbc.275.1.223 PMID: 10617608
13. Wu H, Esteve E, Tremaroli V, Khan MT, Caesar R, Manneras-Holm L, et al. Metformin alters the gut microbiome of individuals with treatment-naive type 2 diabetes, contributing to the therapeutic effects of the drug. Nat Med. 2017; 23(7):850–8. https://doi.org/10.1038/nm.4345 PMID: 28530702
14. Cook MN, Girman CJ, Stein PP, Alexander CM. Initial monotherapy with either metformin or sulphonylureas often fails to achieve or maintain current glycaemic goals in patients with Type 2 diabetes in UK

primary care. Diabet Med. 2007; 24(4):350–8. https://doi.org/10.1111/j.1464-5491.2007.02078.x PMID: 17335466

15. Kahn SE, Haffner SM, Heise MA, Herman WH, Holman RR, Jones NP, et al. Glycemic durability of rosiglitazone, metformin, or glyburide monotherapy. N Engl J Med. 2006; 355(23):2427–43. https://doi.org/10.1056/NEJMoa066224 PMID: 17145742

16. Garber AJ, Duncan TG, Goodman AM, Mills DJ, Rohlf JL. Efficacy of metformin in type II diabetes: results of a double-blind, placebo-controlled, dose-response trial. Am J Med. 1997; 103(6):491–7. https://doi.org/10.1016/s0002-9343(97)00254-4 PMID: 9428832

17. Zhou K, Donnelly L, Yang J, Li M, Deshmukh H, Van Zuydam N, et al. Heritability of variation in glycaemic response to metformin: a genome-wide complex trait analysis. Lancet Diabetes Endocrinol. 2014; 2 (6):481–7. https://doi.org/10.1016/S2213-8587(14)70050-6 PMID: 24731673

18. Tkac I, Klimcakova L, Javorsky M, Fabianova M, Schroner Z, Hermanova H, et al. Pharmacogenomic association between a variant in SLC47A1 gene and therapeutic response to metformin in type 2 diabetes. Diabetes Obes Metab. 2013; 15(2):189–91. https://doi.org/10.1111/j.1463-1326.2012.01691.x PMID: 22882994

19. Shikata E, Yamamoto R, Takane H, Shigemasa C, Ikeda T, Otsubo K, et al. Human organic cation transporter (OCT1 and OCT2) gene polymorphisms and therapeutic effects of metformin. J Hum Genet. 2007; 52(2):117–22. https://doi.org/10.1007/s10038-006-0087-0 PMID: 17111267

20. Choi JH, Yee SW, Ramirez AH, Morrissey KM, Jang GH, Joski PJ, et al. A common 5'-UTR variant in MATE2-K is associated with poor response to metformin. Clin Pharmacol Ther. 2011; 90(5):674–84. https://doi.org/10.1038/clpt.2011.165 PMID: 21956618

21. Chen L, Pawlikowski B, Schlessinger A, More SS, Stryke D, Johns SJ, et al. Role of organic cation transporter 3 (SLC22A3) and its missense variants in the pharmacologic action of metformin. Pharmacogenet Genomics. 2010; 20(11):687–99. https://doi.org/10.1097/FPC.0b013e32833fe789 PMID: 20859243

22. Jablonski KA, McAteer JB, de Bakker PI, Franks PW, Pollin TI, Hanson RL, et al. Common variants in 40 genes assessed for diabetes incidence and response to metformin and lifestyle intervention in the diabetes prevention program. Diabetes. 2010; 59(10):2672–81. https://doi.org/10.2337/db10-0543 PMID: 20682687

23. Rotroff DM, Yee SW, Zhou K, Marvel SW, Shah HS, Jack JR, et al. Genetic Variants in CPA6 and PRPF31 Are Associated With Variation in Response to Metformin in Individuals With Type 2 Diabetes. Diabetes. 2018; 67(7):1428–40. https://doi.org/10.2337/db17-1164 PMID: 29650774

24. GoDarts, Group UDPS, Wellcome Trust Case Control C, Zhou K, Bellenguez C, Spencer CC, et al. Common variants near ATM are associated with glycemic response to metformin in type 2 diabetes. Nat Genet. 2011; 43(2):117–20. https://doi.org/10.1038/ng.735 PMID: 21186350

25. Niu N, Liu T, Cairns J, Ly RC, Tan X, Deng M, et al. Metformin pharmacogenomics: a genome-wide association study to identify genetic and epigenetic biomarkers involved in metformin anticancer response using human lymphoblastoid cell lines. Hum Mol Genet. 2016; 25(21):4819–34. https://doi.org/10.1093/hmg/ddw301 PMID: 28173075

26. Pawlyk AC, Giacomini KM, McKeon C, Shuldiner AR, Florez JC. Metformin pharmacogenomics: current status and future directions. Diabetes. 2014; 63(8):2590–9. https://doi.org/10.2337/db13-1367 PMID: 25060887

27. Guo J, Zhou Y, Cheng Y, Fang W, Hu G, Wei J, et al. Metformin-Induced Changes of the Coding Transcriptome and Non-Coding RNAs in the Livers of Non-Alcoholic Fatty Liver Disease Mice. Cell Physiol Biochem. 2018; 45(4):1487–505. https://doi.org/10.1159/000487575 PMID: 29466788

28. Martin-Montalvo A, Mercken EM, Mitchell SJ, Palacios HH, Mote PL, Scheibye-Knudsen M, et al. Metformin improves healthspan and lifespan in mice. Nat Commun. 2013; 4:2192. https://doi.org/10.1038/ncomms3192 PMID: 23900241

29. Padilla J, Thorne PK, Martin JS, Rector RS, Akter S, Davis JW, et al. Transcriptomic effects of metformin in skeletal muscle arteries of obese insulin-resistant rats. Exp Biol Med (Maywood). 2017; 242 (6):617–24.

30. Udhane SS, Legeza B, Marti N, Hertig D, Diserens G, Nuoffer JM, et al. Combined transcriptome and metabolome analyses of metformin effects reveal novel links between metabolic networks in steroidogenic systems. Sci Rep. 2017; 7(1):8652. https://doi.org/10.1038/s41598-017-09189-y PMID: 28819133

31. Sacco F, Silvestri A, Posca D, Pirro S, Gherardini PF, Castagnoli L, et al. Deep Proteomics of Breast Cancer Cells Reveals that Metformin Rewires Signaling Networks Away from a Pro-growth State. Cell Syst. 2016; 2(3):159–71. https://doi.org/10.1016/j.cels.2016.02.005 PMID: 27135362

32. Rovite V, Wolff-Sagi Y, Zaharenko L, Nikitina-Zake L, Grens E, Klovins J. Genome Database of the Latvian Population (LGDB): Design, Goals, and Primary Results. J Epidemiol. 2018.

33. Zhang M, Liu YH, Chang CS, Zhi H, Wang S, Xu W, et al. Quantification of gene expression while taking into account RNA alternative splicing. Genomics. 2018.
34. Zhou X, Lindsay H, Robinson MD. Robustly detecting differential expression in RNA sequencing data using observation weights. Nucleic Acids Res. 2014; 42(11):e91. https://doi.org/10.1093/nar/gku310 PMID: 24753412
35. Benjamini YaH, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society, Series B (Methodological). 1995; 57(1).
36. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 2010; 11(2):R14. https://doi.org/10.1186/gb-2010-11-2-r14 PMID: 20132535
37. P JEOTP. SciPy: Open Source Scientific Tools for Python. Computing in Science & Engineering. 2007; 9:10–20.
38. Hunter JD. Matplotlib: A 2D graphics environment. Computing in Science & Engineering. 2007; 9:99–104.
39. Vila IK, Song SJ, Song MS. A new duet in cancer biology: AMPK the typical and UBE2O the atypical. Mol Cell Oncol. 2017; 4(3):e1304846. https://doi.org/10.1080/23723556.2017.1304846 PMID: 28616582
40. Lee MS, Han HJ, Han SY, Kim IY, Chae S, Lee CS, et al. Loss of the E3 ubiquitin ligase MKRN1 represses diet-induced metabolic syndrome through AMPK activation. Nat Commun. 2018; 9(1):3404. https://doi.org/10.1038/s41467-018-05721-4 PMID: 30143610
41. Dayeh T, Tuomi T, Almgren P, Perfilyev A, Jansson PA, de Mello VD, et al. DNA methylation of loci within ABCG1 and PHOSPHO1 in blood DNA is associated with future type 2 diabetes risk. Epigenetics. 2016; 11(7):482–8. https://doi.org/10.1080/15592294.2016.1178418 PMID: 27148772
42. Cameron AR, Morrison VL, Levin D, Mohan M, Forteath C, Beall C, et al. Anti-Inflammatory Effects of Metformin Irrespective of Diabetes Status. Circ Res. 2016; 119(5):652–65. https://doi.org/10.1161/CIRCRESAHA.116.308445 PMID: 27418629
43. Bruno S, Ledda B, Tenca C, Ravera S, Orengo AM, Mazzarello AN, et al. Metformin inhibits cell cycle progression of B-cell chronic lymphocytic leukemia cells. Oncotarget. 2015; 6(26):22624–40. https://doi.org/10.18632/oncotarget.4168 PMID: 26265439
44. Xiao Z, Wu W, Poltoratsky V. Metformin Suppressed CXCL8 Expression and Cell Migration in HEK293/TLR4 Cell Line. Mediators Inflamm. 2017; 2017:6589423. https://doi.org/10.1155/2017/6589423 PMID: 29147073
45. Gutzeit C, Magri G, Cerutti A. Intestinal IgA production and its role in host-microbe interaction. Immunol Rev. 2014; 260(1):76–85. https://doi.org/10.1111/imr.12189 PMID: 24942683
46. Fadlallah J, El Kafsi H, Sterlin D, Juste C, Parizot C, Dorgham K, et al. Microbial ecology perturbation in human IgA deficiency. Sci Transl Med. 2018; 10(439).
47. Elbere I, Kalnina I, Silamikelis I, Konrade I, Zaharenko L, Sekace K, et al. Association of metformin administration with gut microbiome dysbiosis in healthy volunteers. PLoS One. 2018; 13(9):e0204317. https://doi.org/10.1371/journal.pone.0204317 PMID: 30261008
48. Malinska H, Oliyarnyk O, Skop V, Silhavy J, Landa V, Zidek V, et al. Effects of Metformin on Tissue Oxidative and Dicarbonyl Stress in Transgenic Spontaneously Hypertensive Rats Expressing Human C-Reactive Protein. PLoS One. 2016; 11(3):e0150924. https://doi.org/10.1371/journal.pone.0150924 PMID: 26963617
49. Radovica-Spalvina I, Latkovskis G, Silamikelis I, Fridmanis D, Elbere I, Ventins K, et al. Next-generation-sequencing-based identification of familial hypercholesterolemia-related mutations in subjects with increased LDL-C levels in a latvian population. BMC Med Genet. 2015; 16:86. https://doi.org/10.1186/s12881-015-0230-x PMID: 26415676
50. Dixon JL, Ginsberg HN. Regulation of hepatic secretion of apolipoprotein B-containing lipoproteins: information obtained from cultured liver cells. J Lipid Res. 1993; 34(2):167–79. PMID: 8381452
51. Trapani L, Segatto M, Pallottini V. Regulation and deregulation of cholesterol homeostasis: The liver as a metabolic "power station". World J Hepatol. 2012; 4(6):184–90. https://doi.org/10.4254/wjh.v4.i6.184 PMID: 22761969
52. Wulffele MG, Kooy A, de Zeeuw D, Stehouwer CD, Gansevoort RT. The effect of metformin on blood pressure, plasma cholesterol and triglycerides in type 2 diabetes mellitus: a systematic review. J Intern Med. 2004; 256(1):1–14. https://doi.org/10.1111/j.1365-2796.2004.01328.x PMID: 15189360
53. Pentikainen PJ, Voutilainen E, Aro A, Uusitupa M, Penttila I, Vapaatalo H. Cholesterol lowering effect of metformin in combined hyperlipidemia: placebo controlled double blind trial. Ann Med. 1990; 22(5):307–12. https://doi.org/10.3109/07853899009147912 PMID: 2291838

3.2 Whole-blood transcriptome profiling reveals signatures of metformin and its therapeutic response

Highlights

1. In T2DM patients only 28 genes showed significantly altered expression levels after administration of metformin for three months, and enrichment of cell signaling pathways related to energy metabolism, immune responses, and lipid metabolism was observed.
2. Clear, transcriptome-based discrimination of study subjects into metformin responders and non-responders was noticed before the use of antidiabetic therapy, which was defined by the expression levels of 56 genes, including *CEACAM1*, *IRS1*, *ABCC2*, and *IGHA1*, explaining 13.9% of the variance.
3. The comparison of blood cell transcriptome profiles from responders with transcripts from non-responders after the metformin therapy for three months revealed significant downregulation of genes coding for NADH: ubiquinone oxidoreductase core subunits, suggesting the altered mitochondrial complex I activity as one of the mechanisms linked to the variability of metformin response.

PLOS ONE

RESEARCH ARTICLE

Whole-blood transcriptome profiling reveals signatures of metformin and its therapeutic response

Monta Ustinova[1], Laura Ansone[1], Ivars Silamikelis[1], Vita Rovite[1], Ilze Elbere[1], Laila Silamikele[1], Ineta Kalnina[1], Davids Fridmanis[1], Jelizaveta Sokolovska[2], Ilze Konrade[1,3], Valdis Pirags[1,2], Janis Klovins[1]*

1 Latvian Biomedical Research and Study Centre, Riga, Latvia, **2** Faculty of Medicine, University of Latvia, Riga, Latvia, **3** Faculty of Medicine, Riga Stradins University, Riga, Latvia

* Klovins@biomed.lu.lv

Citation: Ustinova M, Ansone L, Silamikelis I, Rovite V, Elbere I, Silamikele L, et al. (2020) Whole-blood transcriptome profiling reveals signatures of metformin and its therapeutic response. PLoS ONE 15(8): e0237400. https://doi.org/10.1371/journal.pone.0237400

Editor: Tao Huang, Chinese Academy of Sciences, CHINA

Received: March 25, 2020

Accepted: July 25, 2020

Published: August 11, 2020

Data Availability Statement: RNA-Seq raw data underlying the results presented in the study are available from NCBI's Gene Expression Omnibus, accession number (GSE153792). All other relevant data underlying this study are within the manuscript and its Supporting Information files.

Funding: The study was supported by the European Regional Development Fund under the project "Investigation of interplay between multiple determinants influencing response to metformin: search for reliable predictors for efficacy of type 2

Abstract

Metformin, a biguanide agent, is the first-line treatment for type 2 diabetes mellitus due to its glucose-lowering effect. Despite its wide application in the treatment of multiple health conditions, the glycemic response to metformin is highly variable, emphasizing the need for reliable biomarkers. We chose the RNA-Seq-based comparative transcriptomics approach to evaluate the systemic effect of metformin and highlight potential predictive biomarkers of metformin response in drug-naïve volunteers with type 2 diabetes *in vivo*. The longitudinal blood-derived transcriptome analysis revealed metformin-induced differential expression of novel and previously described genes involved in cholesterol homeostasis (*SLC46A1* and *LRP1*), cancer development *(CYP1B1*, *STAB1*, *CCR2*, *TMEM176B)*, and immune responses (*CD14*, *CD163*) after administration of metformin for three months. We demonstrate for the first time a transcriptome-based molecular discrimination between metformin responders (delta HbA1c $\geq$ 1% or 12.6 mmol/mol) and non-responders (delta HbA1c < 1% or 12.6 mmol/mol), that is determined by expression levels of 56 genes, explaining 13.9% of the variance in the therapeutic efficacy of the drug. Moreover, we found a significant upregulation of *IRS2* gene (log_2FC 0.89) in responders compared to non-responders before the use of metformin. Finally, we provide evidence for the mitochondrial respiratory complex I as one of the factors related to the high variability of the therapeutic response to metformin in patients with type 2 diabetes mellitus.

Introduction

Diabetes mellitus is a chronic disease affecting approximately 463 million people worldwide, which is nearly 9.3% of the global population [1]. Type 2 diabetes mellitus (T2DM) is the most common type of diabetes accounting for approximately 90% of all cases. The persistent hyperglycemia and insulin resistance, a characteristic of T2DM patients, is associated with an increased risk of serious microvascular and macrovascular complications, including

diabete therapy" (Project No.: 1.1.1.1/16/A/091, https://ec.europa.eu/regional_policy/en/funding/erdf/) to JK. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

nephropathy, retinopathy, neuropathy, myocardial infarction, and stroke, which may be reduced by early initiation of antidiabetic therapy [2–4]. Metformin is the first-line medication for treating hyperglycemia in T2DM with beneficial effects in the treatment of multiple non-diabetes related conditions, such as polycystic ovary syndrome, cancer, and neurodegenerative disorders [5–7]. Despite the pleiotropic effects of the drug, the variable efficacy and gastrointestinal side-effects observed cause a significant non-compliance and discontinuation of the therapy, justifying a need for studies exploring molecular mechanisms of metformin action, and biomarkers predicting both treatment response and tolerance of the drug [8].

The mechanism of metformin action is generally considered to involve modulation of the activity of mitochondrial complex I, activation of 5′ AMP-activated protein kinase (AMPK)-dependent mechanisms, and increased AMP concentrations, though some controversy remains since multiple studies are providing evidence for other indirect mechanisms, such as the significant contribution of the gut microbiome underlying the glucose-lowering effect of the drug [9–11].

RNA sequencing (RNA-Seq) is the state-of-the-art approach that may be used to profile drug response and efficacy biomarkers [12, 13]. So far, transcriptome datasets obtained from cell cultures and tissue samples of animal models are extensively used in studies describing molecular mechanisms of metformin concerning various conditions, nevertheless, longitudinal data of *in vivo* studies in humans are still lacking. RNA-Seq has revealed various novel effects and therapeutic targets of metformin, such as enrichment of the transcriptional regulator forkhead box O3a (*FOXO3a*) in primary human fibroblasts [14], upregulation of activating transcription factor 3 (*ATF3*) in primary human hepatocytes [15], downregulation of cell division control protein 42 homolog (*CDC42*) in breast cancer cells [16], upregulation of krüppel-like factor 4 (*KLF4*) resulting in suppressed endothelial dysfunction [17], and even modulated alternative splicing in embryonic stem cells [18]. Moreover, multiple animal-based studies have reported metformin-specific signatures in gene expression profiles of rat arteries and mice epididymal fat, liver and muscle tissue, nevertheless, they still do not explain many beneficial effects of the drug [19–21].

Although the relatively high proportion (>30%) of patients failing to achieve glycemic control during metformin therapy has been partially explained by the contribution of genetic inheritance (allelic variants of organic cation transporters *OCT1*, *OCT2*, etc.) [22–24], recent studies report that heterogeneity of metformin response may be both patient and cell type-specific, suggesting the presence of yet unknown, non-genetic and selective manifestations of the drug [25, 26]. The main objective of the study was to assess the systemic effect of metformin in T2DM patients and reveal potential biomarkers for accurate prediction of its therapeutic efficacy. We have previously reported direct evidence of the effects of metformin on the immediate and strong transcriptome changes in whole-blood samples of healthy subjects, though we considered the diabetic state as a significant confounding factor, therefore the study was continued in a well-characterized, prospective T2DM patient cohort with prolonged observational time, providing much wider applicability of the study results [27]. We believe that our strategy will promote the development of biomarker-based approaches for monitoring treatment outcomes and early identification of metformin responders, moving towards precision medicine.

Materials and methods

Study design

The study was conducted within the framework of the ongoing observational, prospective and longitudinal study OPTIMED, which has been implemented since 2010 in collaboration with endocrinologists and general practitioners from the leading health care centers in Latvia, ensuring recruitment of newly diagnosed drug-naïve patients with ICD-10 diagnosis code E11

and follow-up data collection. Written informed consent was obtained from every participant after full explanation of the purpose and nature of all procedures used before their inclusion in the study, and the study protocol was approved by Central Medical Ethics Committee of Latvia (No. 01–29.1/22) and Committee of Ethics in Pauls Stradins Clinical University Hospital (No.3000610 - 18L). The research was conducted in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki amended in Fortaleza, Brazil, October 2013) and The Convention for the protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine. Management of patient recruitment, collection of samples and associated clinical data was ensured by Latvian Biomedical Research and Study Centre's core facility Genome Centre and the Genome Database of the Latvian Population following their standard procedures [28].

In total, 17 patients of European descent fulfilling the following inclusion criteria were enrolled: (1a) newly diagnosed type 2 diabetes mellitus (ICD-10 code E11) requiring oral antidiabetic therapy, or (1b) previously diagnosed type 2 diabetes mellitus but no oral antidiabetic therapy or insulin has been used for the last three years, or (1c) newly diagnosed type 2 diabetes mellitus and intensive insulin therapy initiated in a hospital for acute glycemic normalization; (2) the patient is not currently involved and is not planning to enroll in clinical trials during the OPTIMED study; (3) the patient has attained 18 years of age; (4) the patient is not pregnant at the time of application; (5) the patient meets the criteria for the diagnosis of type 2 diabetes mellitus: (a) fasting blood glucose level $\geq$ 7 mmol/l, (b) a blood glucose level $\geq$ 11.1 mmol/l for a two-hour glucose tolerance test with 75 g intake. The exclusion criteria of the study were as follows: (1) the patient is receiving oral antidiabetic therapy on a regular basis or has received the therapy during the last three years; (2) the patient is receiving insulin therapy at the time of application; (3) the patient is pregnant. According to the observational study design, the randomization procedure was not performed and metformin monotherapy (medication with metformin hydrochloride as the only active ingredient) was prescribed for each study participant by an endocrinologist for at least three months regardless of the research objectives. The drug manufacturer and dosage of metformin (varied from 850 mg to 2000 mg per day) were chosen by endocrinologists based on clinical experience, patient's health status, and manifestations of the disease. Enrollment in the study did not affect the choice of treatment strategy made by endocrinologists. Blood tests (e.g. measures of ALT, creatinine levels, HbA1c) were performed in a certified clinical laboratory before the administration of metformin and after metformin therapy for three months to evaluate general hematological and biochemical parameters and eligibility of the subjects (Table 1). Blood samples for RNA-Seq were collected at the same time points, hereinafter referred to as M0 (before administration of metformin) and M3m (after three months long metformin course). Considering the high inter-

Table 1. Characteristics of the study group.

Characteristic	Value
Female/ male, n (%)	11 (64.7%) / 6 (35.3%)
Mean age (years) ± SD	61.12± 9.57
Mean BMI ± SD	34.94 ± 4.70
Mean ALAT ± SD, μkat/L	0.74 ± 0.51
Mean creatinine ± SD, μmol/l	63.25 ± 12.60
Mean triglycerides ± SD, mmol/l	2.51 ± 1.86
HbA1c level before the therapy ± SD, mmol/mol	60 ± 14
HbA1c level after 3 months of metformin therapy ± SD, mmol/mol	46 ± 6

BMI, body mass index; SD, standard deviation; ALAT, alanine aminotransferase; HbA1c, glycated hemoglobin.

https://doi.org/10.1371/journal.pone.0237400.t001

individual variability expected in RNA-Seq data, longitudinally repeated measurements were chosen as the most suitable approach for global gene expression analysis.

HbA1c measurements were made for all participants at two consecutive time points analogous to the blood collection for RNA extraction (M0 and M3m). Study participants were further stratified in two subgroups based on metformin response, which was defined according to metformin-induced alterations in glycated hemoglobin (HbA1c) levels comparing measurements made in the time points M0 and M3m: responders (delta HbA1c $\geq$ 1% or 12.6 mmol/mol), non-responders (delta HbA1c $<$ 1% or 12.6 mmol/mol) [29].

Sample processing and RNA sequencing

For RNA isolation, 3 ml of whole blood were collected in Tempus™ Blood RNA Tubes (Thermo Fisher Scientific, USA) and further processed using Tempus™ Spin RNA Isolation Kit (Thermo Fisher Scientific, USA) according to manufacturer's instructions. The quantity and quality of extracted RNA and prepared libraries were determined by Qubit Fluorometer (Thermo Fisher Scientific, USA) and Agilent 2100 Bioanalyzer systems (Agilent, USA), respectively. The integrity of RNA was evaluated by RNA integrity number (RIN) within the Agilent 2100 Bioanalyzer system (Agilent, USA). For depletion of ribosomal RNA 500 ng of total RNA from each sample were processed using Low Input RiboMinus™ Eukaryote System v2 (Thermo Fisher Scientific, USA). Complementary DNA library preparation was performed with Ion Total RNA-Seq Kit v2 (Thermo Fisher Scientific, USA). Ion Proton™ System (Post-Light™ Ion Semiconductor Sequencing, Thermo Fisher Scientific, USA) and Ion PI™ Chip (Thermo Fisher Scientific, USA) was used for 200-base-read single-end sequencing, following the manufacturer's instructions. Since the shot-gun RNA-Seq is considered to be the most accurate and desirable method for the quantification of the individual transcript and gene expression, additional methods for technical validation were not applied in this study [30].

Data analysis

Trimmomatic 0.36 was used for read trimming applying window size 5 and quality threshold of 10. After trimming reads had to have a minimum length of 30 bp and an average quality of 10 to be included in subsequent analyses. Sequencing reads were mapped against human reference genome GRCh38 release 90 and per-gene read counts were calculated with STAR (v.2.5.3a.). STAR (v.2.5.3a.) outputs read quantification per gene while performing read mapping. The reads were quantified if they match only one gene. The obtained read counts were then normalized using trimmed mean normalization implemented in Bioconductor package edgeR in R (v.3.5.3). Differentially expressed genes (DEGs) were estimated using the Likelihood ratio test with added observation weights to reduce the influence of outliers, and sva (Surrogate Variable Analysis) package in R (v.3.5.3) was used for removing batch effects [31]. In order to evaluate metformin-induced alterations in the transcriptome profile (comparison of samples M3m vs M0), each sample was set as a factor considering longitudinal study design. When comparing responders against non-responders in each time point separately, sequencing run and baseline (M0) HbA1c levels were considered as covariates. FilterByExpr function was applied for gene filtering in edgeR, taking into account the sample library sizes [32]. Multiple testing correction was implemented using the Benjamini-Hochberg procedure and differential expression of the genes was determined using a false discovery rate (FDR) $<$ 0.05 cutoff, regardless of the $\log_2$ fold change of expression for each gene [33]. In order to identify key genes determining the metformin response, Partial least squares discriminant analysis (PLS-DA) was performed implemented in the mixOmics package of R (v.3.5.3). CPM values (obtained in edgeR and adjusted for the impact of sequencing run and baseline HbA1c levels)

were used in PLS-DA. Key genes contributing to a separation of patients in both metformin response groups were identified by using a cutoff of variable importance of projection (VIP) score >1 obtained from PLS-DA [34]. Association between HbA1c levels and the log CPM expression values of each mitochondrial gene was performed with multiple linear regression using lm function in R (v.3.5.3). Sex and body mass index were included in the model to account for their potential confounding.

Genes showing the p-value <0.05 for differential expression were further used in the functional analysis. Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were adopted as the functional terms. GO enrichment analysis was performed with R package Goseq (v.1.38.0), and KEGG pathway enrichment analysis was done using an online software Database for Annotation, Visualization and Integrated Discovery (DAVID) 6.8., the threshold value of enrichment was selected by a P-value <0.05 [35, 36].

Heat maps were constructed with Matplotlib and SciPy. Hierarchical clustering with average linkage method implemented in SciPy was used for the clustering of genes according to their differences in CPM values [37, 38]. Statistical analysis of anthropometric measures and biochemical data was performed in R (v3.5.3.) by applying the Wilcoxon rank-sum test and Pearson's chi-squared test with a p-value threshold < 0.05.

Results

Identification of metformin-induced differential expression of genes in blood cells

We characterized the transcriptome profiles of whole-blood samples obtained from 17 drug-naïve T2D patients (characteristics of patients provided in Table 1) before any antidiabetic therapy (M0) and after three months of metformin monotherapy (M3m) by RNA-Seq technique to detect sustained transcriptional alterations in blood cells at the diabetic state due to the administration of the drug. The median of read counts per sample produced by RNA sequencing was 21.1 (IQR = 6.3) and 87.9% of the reads (median = 18.2; IQR = 5.1) were further mapped to the human reference genome.

Differential expression analysis performed by comparing transcriptome profiles of blood samples collected after metformin administration for three months against samples collected before the use of any antidiabetic therapy (M3m vs M0) revealed 28 DEGs (FDR < 0.05) from the pool of 9992 transcripts identified in total. Out of them, 20 genes were significantly down-regulated and 8 genes showed significant up-regulation after administration of metformin (Fig 1, Table 2, S1 Fig).

Functional enrichment characteristic to metformin therapy

In order to describe the implementation of metformin-modulated transcriptome profiles in cell signaling pathways and core biologic functions, the KEGG pathway and GO enrichment analysis was performed. Over-representation of biological pathways (e.g. amino sugar and nucleotide sugar metabolism, antigen processing and presentation) and GO Terms (e.g. helper T cell chemotaxis and lipoprotein particle receptor activity) related to energy metabolism, immune responses and lipid metabolism were observed, see S1 and S2 Tables for the complete list of GO terms and enriched KEGG pathways.

Determining genes involved in differential metformin responsiveness

Considering the potential contribution of distinct molecular mechanisms mediating variable metformin response, all participants were stratified in two efficacy groups, according to

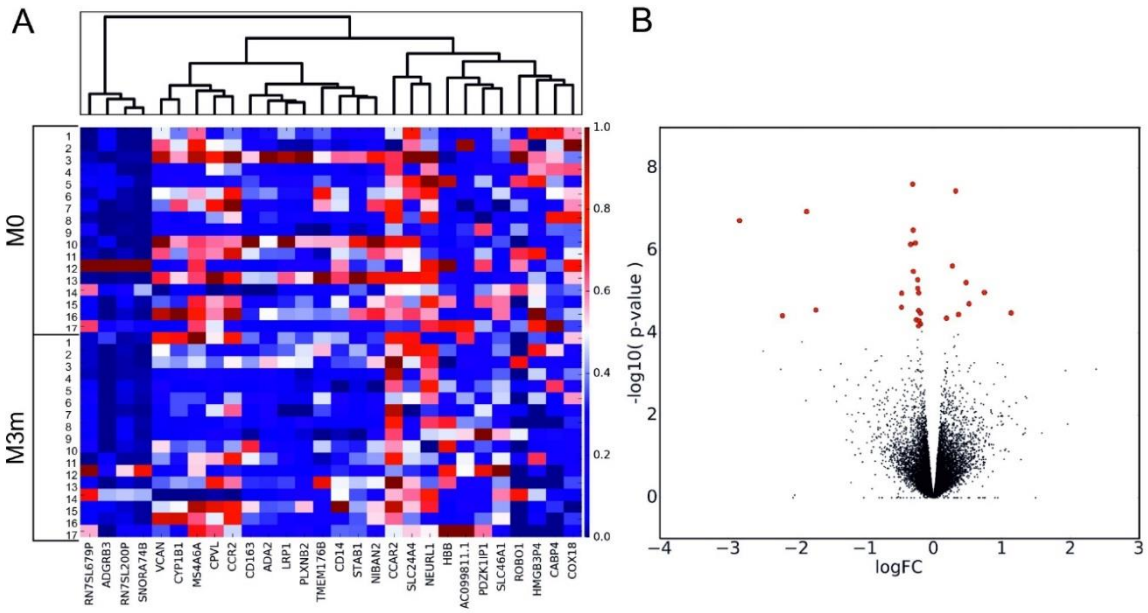

Fig 1. Metformin-induced alterations of gene expression profiles. (A)—Heat map and hierarchical clustering of 28 DEGs identified. Each row corresponds to one subject in the respective time-point and each column represents a DEG. Normalized sequence read counts were rescaled to lie in the range [0,1] and further used to estimate the difference between the gene expression levels in two time-points. DEGs with analogous expression values were clustered at the column level. (B)—Volcano plot showing the distribution of gene expression in the analyzed contrast. Significance versus $\log_2$ fold change is plotted on the y and x axes, respectively, calculated using likelihood ratio test and edgeR. Red dots represent the significant DEGs (FDR < 0.05), black dots–non-significant genes. M0 –time point of blood collection before administration of metformin; M3m –time point of blood collection after metformin therapy for three months.

https://doi.org/10.1371/journal.pone.0237400.g001

metformin-induced changes in their HbA1c levels (responders: delta HbA1c ≥1% or 12.6 mmol/mol; non-responders: delta HbA1c <1% or 12.6 mmol/mol) (Table 3).

Genes associated with variable metformin responsiveness were determined by differential gene expression analysis comparing the number of RNA-Seq reads from responders against non-responders in each time point of blood collection (M0 and M3m) separately. In total, 27 significant DEGs were identified contrasting responders against non-responders before the administration of metformin (M0). We observed a notable portion of downregulated genes coding for small nucleolar RNAs and upregulated insulin receptor substrate 2 (*IRS2*) gene in responders, revealing distinctive features between response groups already before the use of any antidiabetic therapy (Fig 2C), (Table 4).

PLS-DA was applied to understand whether patients could be clustered based on their gene expression profiles before the administration of metformin. PLS-DA revealed pronounced discrimination of samples, based on the obtained transcriptome profiles, explaining 13.9% of the variance in total (11% of variance explained by latent variable 1 and 2.9% of variance explained by latent variable 2) (Fig 3B). In total, 56 discriminatory genes showing the strongest separation of different metformin response groups were identified, based on variable importance for projection (VIP>1) value in two latent variables (components), generated by PLS-DA (Fig 3A). See S5 Table for the full list and VIP values of discriminatory genes.

Comparison of gene expression profiles between metformin responders and non-responders after the use of metformin for three months (M3m) showed differential expression of 15 genes (12 up-regulated, 3 down-regulated), including 5 mitochondrial genes, which may be associated with mechanisms underlying variable efficacy of the drug (Table 5) (Fig 2D).

Table 2. List of differentially expressed genes due to administration of metformin, ranked by log_2 fold change.

Gene symbol	Full name	log_2FC	FDR
ROBO1	Roundabout guidance receptor 1	-2.83	7.54E-04
ADGRB3	Adhesion G protein-coupled receptor B3	-2.2	2.67E-02
CABP4	Calcium binding protein 4	-1.85	6.08E-04
HMGB3P4	High mobility group box 3 pseudogene 4	-1.72	2.45E-02
NEURL1	Neuralized E3 Ubiquitin Protein Ligase 1	-0.46	2.26E-02
COX18	Cytochrome C Oxidase Assembly Factor	-0.46	1.17E-02
NIBAN2	Niban apoptosis regulator 2	-0.33	1.64E-03
STAB1	Stabilin 1	-0.3	2.89E-04
CPVL	Carboxypeptidase vitellogenic like	-0.29	1.03E-03
TMEM176B	Transmembrane protein 176B	-0.29	5.72E-03
SLC24A4	Solute Carrier Family 24 Member 4	-0.26	1.64E-03
CYP1B1	Cytochrome P450 family 1 subfamily B member 1	-0.25	3.10E-02
LRP1	LDL receptor related protein 1	-0.23	8.23E-03
CCR2	C-C motif chemokine receptor 2	-0.23	1.11E-02
CD14	CD14 molecule	-0.22	3.88E-02
VCAN	Versican	-0.21	2.45E-02
MS4A6A	Membrane Spanning 4-Domains A6A	-0.21	1.17E-02
CD163	CD163 molecule	-0.21	3.15E-02
PLXNB2	Plexin B2	-0.19	2.57E-02
ADA2	Adenosine deaminase 2	-0.19	3.62E-02
CCAR2	Cell cycle and apoptosis regulator 2	0.19	2.95E-02
HBB†	Hemoglobin subunit beta	0.28	4.77E-03
PDZK1IP1†	PDZK1-interacting protein 1	0.33	2.89E-04
SNORA74B	Small nucleolar RNA, H/ACA box 74B	0.37	2.61E-02
RN7SL679P†	RNA, 7SL, cytoplasmic 679, pseudogene	0.48	8.77E-03
RN7SL200P	RNA, 7SL, cytoplasmic 200, pseudogene	0.52	1.96E-02
AC099811.1	AC099811.2 (novel transcript, sense intronic to STAT5B)	0.74	1.17E-02
SLC46A1	Solute carrier family 46 member 1	1.14	2.57E-02

Log_2FC, log_2 fold change; FDR, false discovery rate.

†Genes showing significant differential expression due to the administration of metformin also in healthy individuals [27].

https://doi.org/10.1371/journal.pone.0237400.t002

Table 3. Characteristics of responders and non-responders.

Characteristic	Responders (n = 10)	Non-responders (n = 7)	P-value
Female/ male, n (%)	6(60.00%)/4(40.00%)	5(71.43%)/2(28.57%)	1.00E+00
Mean age (years) ± SD†	61.14±11.35	61.10±8.77	8.84E-01
Mean BMI ± SD†	33.70±4.50	35.81±5.12	2.61E-01
Mean ALAT ± SD, μkat/L†	0.91±0.70	0.61±0.27	4.91E-01
Mean creatinine ± SD, μmol/l†	59.57±13.35	66.11±12.75	2.23E-01
Mean triglycerides ± SD, mmol/l†	3.65±2.57	1.75±0.67	8.78E-02
HbA1c level before the therapy ± SD, mmol/mol†	76±15	48±4	4.48E-03
HbA1c level after 3 months of metformin therapy ± SD, mmol/mol	48±7	45±4	5.57E-01

BMI—body mass index; SD—standard deviation; ALAT—alanine aminotransferase; HbA1c - glycated hemoglobin.

†Measured before the administration of metformin (M0).

https://doi.org/10.1371/journal.pone.0237400.t003

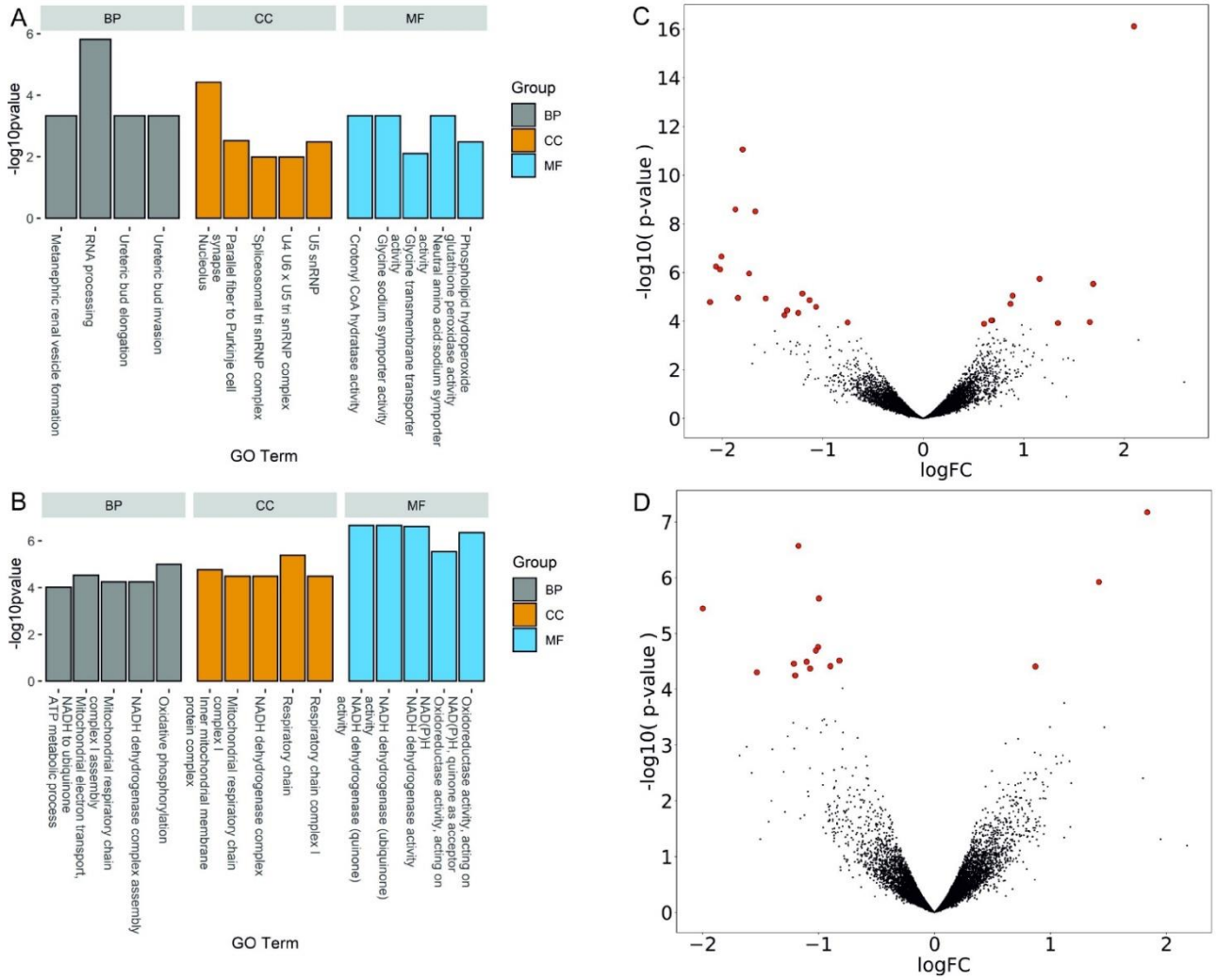

Fig 2. Differentially expressed genes and their representation in Gene Ontologies. Bar plots showing the top 5 enriched Gene Ontology terms in the comparison of responders against non-responders before administration of metformin (A) and three months after metformin therapy (B). Volcano plot represents the distribution of gene expression comparing responders against non-responders before administration of metformin (C) and three months after metformin therapy (D). Significance versus $\log_2$ fold change is plotted on the y and x axes, respectively, calculated using likelihood ratio test and edgeR. Red dots represent the significant DEGs (FDR < 0.05), black dots–non-significant genes. BP–biological process; CC–cellular component; MF–molecular function.

https://doi.org/10.1371/journal.pone.0237400.g002

Moreover, a multiple linear regression model revealed a significant association between the expression of two out of five mitochondrial genes and HbA1c levels after metformin therapy for three months (*MT-ND4*, p-value = 0.047; *MT-ND4L*, p-value = 0.032) (S7 Table).

Biological functions contributing to metformin responsiveness

GO analysis was performed for each list of DEGs (responders against non-responders at time points M0 and M3m separately). RNA processing and negative regulation of fatty acid transport was revealed among the enriched biological processes using DEG list obtained by comparison of responders against non-responders before administration of metformin (Fig 2A).

Table 4. List of differentially expressed genes comparing responders against non-responders before administration of metformin.

Gene symbol	Full name	log_2FC	FDR
RNU5A-1	RNA, U5A small nuclear 1	-2.12	1.07E-02
SNORA20	Small nucleolar RNA, H/ACA box 20	-2.06	9.85E-04
SNORD82	Small nucleolar RNA, C/D box 82	-2.02	1.11E-03
SNORA5C	Small nucleolar RNA, H/ACA box 5C	-2.01	4.63E-04
SNORA28	Small nucleolar RNA, H/ACA box 28	-1.87	8.00E-06
RNU5B-1	RNA, U5B small nuclear 1	-1.84	8.63E-03
SNORD20	Small nucleolar RNA, C/D box 20	-1.8	4.58E-08
RNY4	RNA, Ro60-associated Y4	-1.73	1.44E-03
SNORD90	Small nucleolar RNA, C/D box 90	-1.67	8.00E-06
SNORA75	Small nucleolar RNA, H/ACA box 75	-1.57	8.63E-03
SNORA74B	Small nucleolar RNA, H/ACA box 74B	-1.38	2.80E-02
SNORD91A	Small nucleolar RNA, C/D box 91A	-1.35	1.98E-02
SNORA37	Small nucleolar RNA, H/ACA box 37	-1.24	2.39E-02
SNORA66	Small nucleolar RNA, H/ACA box 66	-1.2	7.00E-03
SNORA14B	Small nucleolar RNA, H/ACA box 14B	-1.13	9.45E-03
CLC	Charcot-Leyden crystal galectin	-1.07	1.49E-02
RALGPS2†	Ral GEF with PH domain and SH3 binding motif 2	-0.75	4.74E-02
GPX4	Glutathione peroxidase 4	0.61	4.95E-02
PCMTD2	Protein-L-isoaspartate (D-aspartate) O-methyltransferase domain containing 2	0.68	4.23E-02
MARCH2	Membrane associated ring-CH-type finger 2	0.69	4.23E-02
CDYL	Chromodomain Y like	0.87	1.19E-02
IRS2	Insulin receptor substrate 2	0.89	7.81E-03
SIRPG	Signal regulatory protein gamma	1.16	2.13E-03
SLC6A9	Solute carrier family 6 member 9	1.34	4.83E-02
FMN1†	Formin 1	1.66	4.74E-02
HEBP1†	Heme binding protein 1	1.69	3.06E-03
CHI3L1	Chitinase 3 like 1	2.1	8.03E-13

Log2FC, log2 fold change; FDR, false discovery rate.

†Genes showing significant differential expression comparing responders against non-responders also after metformin therapy for 3 months.

https://doi.org/10.1371/journal.pone.0237400.t004

Contrasting responders against non-responders after three months of metformin use identified enrichment of GO terms strongly related to the mitochondrial activity (e.g. mitochondrial respiratory chain complex I, ATP metabolic process) (Fig 2B). See S3 and S4 Tables for the full list of enriched GO terms identified. Although KEGG pathway analysis of DEG list, obtained from the baseline analysis (M0), did not reveal any significantly enriched cell signaling pathway, comparison of responders against non-responders after three months of metformin use showed differential expression of 5 mitochondrial genes (*MT-ATP6*, *MT-ND2*, *MT-ND4*, *MT-ND4L*, *MT-ND6*). All of these genes provided a significant enrichment of the following pathways: hsa00190: Oxidative phosphorylation (p-value = 4.47E-06), hsa05012: Parkinson's disease (p-value = 5.81E-06) and hsa01100: Metabolic pathways (p-value = 2.18E-02).

Discussion

The results of our study showed metformin-specific signatures in blood cell transcriptome profiles associated with some of its well-known properties: the ability to improve energy metabolism, influence immune responses, and inhibit cancer progression. This study has demonstrated a gene expression-based molecular discrimination between metformin responders

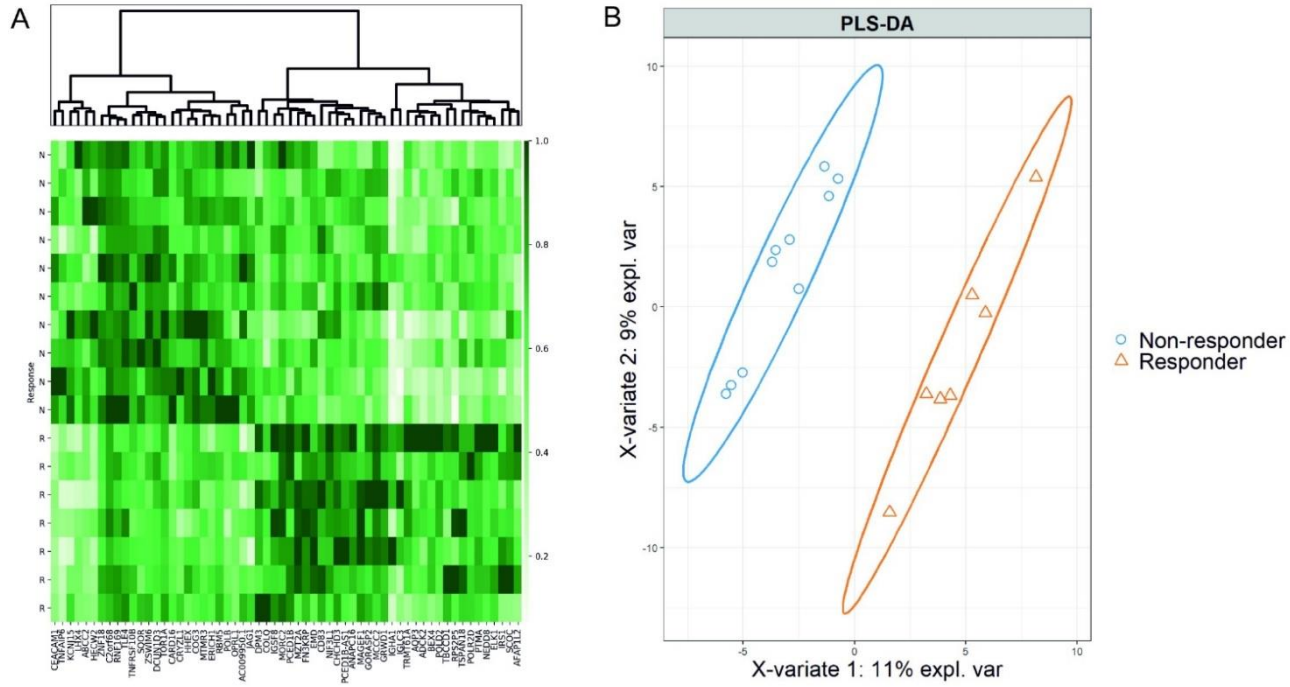

Fig 3. The partial least-square discriminant analysis of the RNA-Seq data obtained before the administration of metformin. (A) Heat map and hierarchical clustering of 56 key genes contributing the most to the patient separation in metformin response groups, selected by VIP score threshold >1 from PLS-DA. Each row corresponds to one subject (N: non-responder; R: responder) and each column represents a gene. Normalized sequence read counts were rescaled to lie in the range [0,1], genes with analogous expression values were clustered at the column level. (B) PLS-DA plot of RNA-Seq data showing clear transcriptome-based discrimination of patients with different metformin responses. Each point represents the transcriptome signature of one patient, the confidence level is set to 95% for ellipses. The separation of samples in the PLS-DA model is based on latent variables (X-variate 1 on and X-variate 2 on x and y axes, respectively). T2DM patients with different metformin responses are projected into distinct clusters indicating the difference in their transcriptome profiles.

https://doi.org/10.1371/journal.pone.0237400.g003

and non-responders, and suggested that mitochondrial respiratory complex I may be associated with metformin efficacy. To the best of our knowledge, this is the first longitudinal study focusing on metformin-induced transcriptional alterations of drug-naïve T2DM patients *in vivo*.

Our study revealed metformin-induced differential expression of genes involved in cholesterol homeostasis, such as solute carrier family 46 member 1 *(SLC46A1)*, which is involved in the intestinal folate absorption affecting plasma high-density lipoprotein levels [39], and lipoprotein receptor-related protein 1 (*LRP1*), a crucial protein for cholesterol uptake. Metformin-induced reduction of hepatic *LRP1* expression level has been reported before [40], though here we report a similar effect in blood cells for the first time. Metformin-induced differential expression of both genes may serve as a contributing factor for the cholesterol-lowering effect of the drug. In our data, the same mechanism was supported by the enrichment of lipoprotein particle receptor activity among identified GO terms.

In addition, we found significant downregulation of multiple cancer-related genes coding for cytochrome P450 1B1 (*CYP1B1*), C-C chemokine receptor type 2 (*CCR2*), stabilin-1 (*STAB1*) and transmembrane protein 176B (*TMEM176B*). Some of the observed alterations have been previously identified in different tissue types, such as the downregulation of *CCR2*, which has been explained by the ability of metformin to block M2-like polarization of tumor-associated macrophages providing the anti-metastatic effect of the drug [41], or downregulation of *CYP1B1* in breast cancer cells where due to its crucial role in estrogen metabolism,

Table 5. List of differentially expressed genes comparing responders against non-responders after metformin therapy for three months.

Gene symbol	Full name	log_2FC	FDR
S100P	S100 calcium binding protein P	-2.00	7.54E-03
FP671120.7	Novel transcript, similar to YY1 associated myogenesis RNA 1 YAM1	-1.53	3.77E-02
PAX5	Paired box 5	-1.21	3.45E-02
TNFRSF13C	TNF receptor superfamily member 13C	-1.20	4.01E-02
IGHM	Immunoglobulin heavy constant mu	-1.17	1.43E-03
MT-ND6	Mitochondrially encoded NADH: ubiquinone oxidoreductase core subunit 6	-1.10	3.45E-02
MT-ND4L	Mitochondrially encoded NADH: ubiquinone oxidoreductase core subunit 4L	-1.07	3.46E-02
MT-ATP6	Mitochondrially encoded ATP synthase membrane subunit 6	-1.02	3.07E-02
MT-ND4	Mitochondrially encoded NADH: ubiquinone oxidoreductase core subunit 4	-1.00	3.07E-02
MT-ND2	Mitochondrially encoded NADH: ubiquinone oxidoreductase core subunit 2	-1.00	6.27E-03
CD79A	CD79a molecule	-0.90	3.45E-02
RALGPS2†	Ral GEF with PH domain and SH3 binding motif 2	-0.82	3.45E-02
WARS	Tryptophanyl-trna synthetase	0.87	3.45E-02
HEBP1†	Heme binding protein 1	1.42	4.23E-03
FMN1†	Formin 1	1.83	7.12E-04

Log2FC, log2 fold change; FDR, false discovery rate.

†Genes showing significant differential expression comparing responders against non-responders also before administration of metformin.

https://doi.org/10.1371/journal.pone.0237400.t005

metformin has been suggested as a potential chemopreventive agent against carcinogenesis [42]. Multiple studies have reported that aberrant expression of *STAB1* is related to the tumor progression in various types of cancer, serving as a potential molecular target for cancer therapy [43, 44]. Similarly, elevated protein levels of TMEM176B are detected in multiple malignancies, moreover, inhibition of TMEM176B has already been proved to promote $CD8^+$ T cell-mediated tumor growth control, enhancing the therapeutic efficacy of cancers [45, 46]. Therefore, the observed downregulation of both, *STAB1* and *TMEM176B* may highlight novel players in the anti-cancer activity of metformin.

Finally, the enrichment of antigen processing and presentation pathway and significantly reduced expression of genes coding for the cluster of differentiation 14 (*CD14*), which was already proved to be differentially expressed in metformin-treated monocyte cells, and the cluster of differentiation 163 (CD163) [47], a scavenger receptor which has been previously associated with insulin resistance in patients with T2DM, altogether may explain well-known participation of metformin in the inflammatory and immune responses [48].

According to our previous study, there are only three genes (*HBB*, *PDZK1IP1*, and *RN7SL679P*) showing metformin-induced differential expression in blood cells obtained from both, T2DM patients and healthy volunteers. Although both studies were longitudinal, the duration of the therapy (7 days for healthy volunteers and 3 months for T2DM patients) and the dose of metformin (850 mg twice a day for healthy volunteers and variable dose in the T2DM patient group), differed between both studies, which may explain the observed variability of metformin effects on the blood cell transcriptome profiles among both study groups. Moreover, the general health status including the presence of T2DM may serve as the major confounding factor to modulate the cell metabolism and effects of metformin in the patient group compared to healthy individuals [49].

One of the main findings of our study was the clear, transcriptome-based discrimination of study subjects into metformin responders and non-responders before the administration of any anti-diabetic therapy. Differential expression analysis before the use of metformin revealed significant upregulation of *IRS2* gene coding for insulin receptor-2 (log_2FC 0.89) in responders

compared to non-responders. So far, multiple studies have demonstrated the ability of metformin to activate the hepatic insulin receptor and the IRS2/PI3K/Akt pathway resulting in reduced insulin resistance and increased glucose uptake [50, 51]. The results of our study suggest that the activity of *IRS2* in blood cells prior to the administration of metformin may be related to the efficacy of the therapy. Nevertheless, this hypothesis must be tested in a larger longitudinal cohort.

PLS-DA analysis also provided strong evidence for transcriptome-based patient stratification which was mainly explained by a single latent variable (component) and defined by expression of 56 key genes (VIP>1). We found carcinoembryonic antigen-related cell adhesion molecule 1 (*CEACAM1*), Insulin receptor substrate 1 (*IRS1*), *ABCC2* gene coding for multidrug resistance protein 2 and *IGHA1* gene coding for the constant segment of immunoglobulin A heavy chain among the marker genes determining the distribution of patients in metformin response groups. So far, there are no studies reporting the role of multidrug resistance protein 2 in metformin efficacy, nevertheless, its homolog mitochondrial multidrug resistance protein 1 has been already associated with resistance to metformin in human malignant mesothelioma cells [52]. Interestingly, *CEACAM1* is a mediator of insulin clearance in the liver [53] and *IRS1* is playing a key role in glucose homeostasis [54], though no reports were linking their activity with metformin efficacy so far. In our previous studies, we have already reported the contribution of IgA, the most abundant intestinal antibody shaping the gut microbiome composition, in metformin action [27], and here we suggest a potential implication of IgA in microbiome-mediated metformin response [55]. Together these genes showed notable patient segregation in metformin response groups also when visualized in CPM-based heat map, therefore they may be considered as candidates for further studies of biomarkers for metformin response.

The comparison of blood cell transcriptome profiles from responders with transcripts from non-responders after the metformin therapy for three months revealed significant downregulation of mitochondrial genes (*MT-ATP6*, *MT-ND2*, *MT-ND4*, *MT-ND4L*, *MT-ND6*). Furthermore, *MT-ND4* and *MT-ND4L* also showed a positive association with HbA1c levels according to the multiple regression model. Metformin directly acts on mitochondria and limits respiration by inhibiting mitochondrial respiratory-chain complex 1 (NADH: ubiquinone oxidoreductase), which is among the top targets of the drug, and it also catalyzes oxidative phosphorylation in mammalian mitochondria [56, 57]. Thus, the observed downregulation of genes coding for NADH: ubiquinone oxidoreductase core subunits in responders and enrichment of oxidative phosphorylation seems rational. These data suggest altered mitochondrial complex I activity as one of the mechanisms linked to the variability of metformin response.

There are a few limitations in this study that could be addressed by future research. First, a small sample size due to the essential inclusion criterion of drug-naïve T2DM patients. Second, prolonged observation time (three months) that provides an advantage to evaluate long-term metformin effects, while causing an issue of possible accumulation of uncontrollable factors affecting gene expression and masking the true effects of the drug. Third, the lack of control arm allowing the elimination of these confounding factors. Nevertheless, we believe that RNA-Seq, a highly sensitive and accurate method for gene expression, together with the repeated measures within the longitudinal study design and strictly defined inclusion criteria ensures the absence of potential influence of other anti-diabetic therapies and improves the validity of our results.

In conclusion, the current study applying RNA-Seq for the discovery of transcriptional effects of metformin in drug-naïve T2DM patients provided detailed insight into potential molecular mechanisms underlying well-known beneficial effects of metformin. However, since there are no data confirming the accumulation of metformin in other blood cells than

erythrocytes [58], it seems reasonable to assume that the observed effects of metformin on peripheral blood cells most likely are indirect and reflects the systemic consequence of the therapy. Nevertheless, we suggest that blood-derived transcriptome profiles may be used for evaluation of therapeutic efficacy and specific genes may be further applied in the development of biomarkers for metformin response.

Supporting information

S1 Table. Gene ontology analysis for the list of differentially expressed genes identified by comparing M3m samples with M0 samples.
(XLSX)

S2 Table. KEGG pathway enrichment analysis for the list of differentially expressed genes identified by comparing M3m samples with M0 samples.
(XLSX)

S3 Table. Gene ontology analysis for the list of differentially expressed genes identified by comparing responders against non-responders before administration of metformin (M0).
(XLSX)

S4 Table. Gene ontology analysis for the list of differentially expressed genes identified by comparing responders against non-responders after administration of metformin for 3 months (M3m).
(XLSX)

S5 Table. The variable importance of projection score for key genes contributing the most to the patient separation in metformin response groups before the use of any antidiabetic therapy.
(XLSX)

S6 Table. Summary of sequencing statistics.
(XLSX)

S7 Table. Association between HbA1c and gene expression levels obtained by multiple linear regression model after metformin therapy for three months.
(XLSX)

S1 Fig. Violin plots showing the $\log_2$ fold change of 28 differentially expressed genes after metformin use. Each dot represents the log2 fold change for a particular gene of one study subject. The shape of the plot represents the distribution of the data obtained using kernel density estimate and Scott's rule for bandwidth selection.
(TIFF)

S2 Fig. CONSORT flowchart of the observational prospective study.
(TIFF)

Acknowledgments

The authors would like to thank all the patients for their participation and acknowledge the Genome Database of the Latvian Population for providing biological material and data.

Author Contributions

Conceptualization: Valdis Pirags, Janis Klovins.

Formal analysis: Ivars Silamikelis.

Funding acquisition: Ineta Kalnina, Janis Klovins.

Investigation: Monta Ustinova, Laura Ansone.

Resources: Monta Ustinova, Vita Rovite, Ilze Elbere, Laila Silamikele, Davids Fridmanis, Jelizaveta Sokolovska, Ilze Konrade.

Supervision: Janis Klovins.

Visualization: Monta Ustinova, Ivars Silamikelis.

Writing – original draft: Monta Ustinova.

Writing – review & editing: Laura Ansone, Ivars Silamikelis, Vita Rovite, Ilze Elbere, Laila Silamikele, Ineta Kalnina, Davids Fridmanis, Jelizaveta Sokolovska, Ilze Konrade, Valdis Pirags, Janis Klovins.

References

1. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9(th) edition. Diabetes research and clinical practice. 2019; 157:107843. https://doi.org/10.1016/j.diabres.2019.107843 PMID: 31518657
2. Chawla A, Chawla R, Jaggi S. Microvasular and macrovascular complications in diabetes mellitus: Distinct or continuum? Indian journal of endocrinology and metabolism. 2016; 20(4):546–51. https://doi.org/10.4103/2230-8210.183480 PMID: 27366724
3. Holman N, Young B, Gadsby R. Current prevalence of Type 1 and Type 2 diabetes in adults and children in the UK. Diabetic medicine: a journal of the British Diabetic Association. 2015; 32(9):1119–20.
4. Holman RR, Paul SK, Bethel MA, Matthews DR, Neil HA. 10-year follow-up of intensive glucose control in type 2 diabetes. The New England journal of medicine. 2008; 359(15):1577–89. https://doi.org/10.1056/NEJMoa0806470 PMID: 18784090
5. Zi F, Zi H, Li Y, He J, Shi Q, Cai Z. Metformin and cancer: An existing drug for cancer prevention and therapy. Oncology letters. 2018; 15(1):683–90. https://doi.org/10.3892/ol.2017.7412 PMID: 29422962
6. Lashen H. Role of metformin in the management of polycystic ovary syndrome. Therapeutic advances in endocrinology and metabolism. 2010; 1(3):117–28. https://doi.org/10.1177/2042018810380215 PMID: 23148156
7. Rotermund C, Machetanz G, Fitzgerald JC. The Therapeutic Potential of Metformin in Neurodegenerative Diseases. Frontiers in endocrinology. 2018; 9:400. https://doi.org/10.3389/fendo.2018.00400 PMID: 30072954
8. Garber AJ, Duncan TG, Goodman AM, Mills DJ, Rohlf JL. Efficacy of metformin in type II diabetes: results of a double-blind, placebo-controlled, dose-response trial. The American journal of medicine. 1997; 103(6):491–7. https://doi.org/10.1016/s0002-9343(97)00254-4 PMID: 9428832
9. Vallianou NG, Stratigou T, Tsagarakis S. Metformin and gut microbiota: their interactions and their impact on diabetes. Hormones. 2019; 18(2):141–4. https://doi.org/10.1007/s42000-019-00093-w PMID: 30719628
10. Elbere I, Kalnina I, Silamikelis I, Konrade I, Zaharenko L, Sekace K, et al. Association of metformin administration with gut microbiome dysbiosis in healthy volunteers. PloS one. 2018; 13(9):e0204317. https://doi.org/10.1371/journal.pone.0204317 PMID: 30261008
11. Foretz M, Hebrard S, Leclerc J, Zarrinpashneh E, Soty M, Mithieux G, et al. Metformin inhibits hepatic gluconeogenesis in mice independently of the LKB1/AMPK pathway via a decrease in hepatic energy state. The Journal of clinical investigation. 2010; 120(7):2355–69. https://doi.org/10.1172/JCI40671 PMID: 20577053
12. Geeleher P, Cox NJ, Huang RS. Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. Genome biology. 2014; 15(3):R47. https://doi.org/10.1186/gb-2014-15-3-r47 PMID: 24580837
13. Verbist B, Klambauer G, Vervoort L, Talloen W, Consortium Q, Shkedy Z, et al. Using transcriptomics to guide lead optimization in drug discovery projects: Lessons learned from the QSTAR project. Drug discovery today. 2015; 20(5):505–13. https://doi.org/10.1016/j.drudis.2014.12.014 PMID: 25582842

14. Gillespie ZE, Wang C, Vadan F, Yu TY, Ausio J, Kusalik A, et al. Metformin induces the AP-1 transcription factor network in normal dermal fibroblasts. Scientific reports. 2019; 9(1):5369. https://doi.org/10.1038/s41598-019-41839-1 PMID: 30926854

15. Luizon MR, Eckalbar WL, Wang Y, Jones SL, Smith RP, Laurance M, et al. Genomic Characterization of Metformin Hepatic Response. PLoS genetics. 2016; 12(11):e1006449. https://doi.org/10.1371/journal.pgen.1006449 PMID: 27902686

16. Athreya AP, Kalari KR, Cairns J, Gaglio AJ, Wills QF, Niu N, et al. Model-based unsupervised learning informs metformin-induced cell-migration inhibition through an AMPK-independent mechanism in breast cancer. Oncotarget. 2017; 8(16):27199–215. https://doi.org/10.18632/oncotarget.16109 PMID: 28423712

17. Yu B, Wu Y, Li Z. KLF4/Ch25h axis activated by metformin suppresses EndoMT in human umbilical vein endothelial cells. Biochemical and biophysical research communications. 2020; 522(4):838–44. https://doi.org/10.1016/j.bbrc.2019.11.181 PMID: 31801667

18. Laustriat D, Gide J, Barrault L, Chautard E, Benoit C, Auboeuf D, et al. In Vitro and In Vivo Modulation of Alternative Splicing by the Biguanide Metformin. Molecular therapy Nucleic acids. 2015; 4:e262. https://doi.org/10.1038/mtna.2015.35 PMID: 26528939

19. Padilla J, Thorne PK, Martin JS, Rector RS, Akter S, Davis JW, et al. Transcriptomic effects of metformin in skeletal muscle arteries of obese insulin-resistant rats. Experimental biology and medicine. 2017; 242(6):617–24. https://doi.org/10.1177/1535370216689825 PMID: 28114814

20. Jung J, Kim GW, Lee W, Mok C, Chung SH, Jang W. Meta- and cross-species analyses of insulin resistance based on gene expression datasets in human white adipose tissues. Scientific reports. 2018; 8(1):3747. https://doi.org/10.1038/s41598-017-18082-7 PMID: 29487289

21. Lee SH, Min KJ. Caloric restriction and its mimetics. BMB reports. 2013; 46(4):181–7. https://doi.org/10.5483/bmbrep.2013.46.4.033 PMID: 23615258

22. Shikata E, Yamamoto R, Takane H, Shigemasa C, Ikeda T, Otsubo K, et al. Human organic cation transporter (OCT1 and OCT2) gene polymorphisms and therapeutic effects of metformin. Journal of human genetics. 2007; 52(2):117–22. https://doi.org/10.1007/s10038-006-0087-0 PMID: 17111267

23. Cook MN, Girman CJ, Stein PP, Alexander CM. Initial monotherapy with either metformin or sulphonylureas often fails to achieve or maintain current glycaemic goals in patients with Type 2 diabetes in UK primary care. Diabetic medicine: a journal of the British Diabetic Association. 2007; 24(4):350–8.

24. Rashid M, Shahzad M, Mahmood S, Khan K. Variability in the therapeutic response of Metformin treatment in patients with type 2 diabetes mellitus. Pakistan journal of medical sciences. 2019; 35(1):71–6. https://doi.org/10.12669/pjms.35.1.100 PMID: 30881399

25. Yang X, Xu Z, Zhang C, Cai Z, Zhang J. Metformin, beyond an insulin sensitizer, targeting heart and pancreatic beta cells. Biochimica et biophysica acta Molecular basis of disease. 2017; 1863(8):1984–90. https://doi.org/10.1016/j.bbadis.2016.09.019 PMID: 27702625

26. Asiedu MK, Barron M, Aubry MC, Wigle DA. Patient- and Cell Type-Specific Heterogeneity of Metformin Response. Basic & clinical pharmacology & toxicology. 2018; 122(2):214–22.

27. Ustinova M, Silamikelis I, Kalnina I, Ansone L, Rovite V, Elbere I, et al. Metformin strongly affects transcriptome of peripheral blood cells in healthy individuals. PloS one. 2019; 14(11):e0224835. https://doi.org/10.1371/journal.pone.0224835 PMID: 31703101

28. Rovite V, Wolff-Sagi Y, Zaharenko L, Nikitina-Zake L, Grens E, Klovins J. Genome Database of the Latvian Population (LGDB): Design, Goals, and Primary Results. Journal of epidemiology. 2018; 28(8):353–60. https://doi.org/10.2188/jea.JE20170079 PMID: 29576601

29. Kashi Z, Mahrooz A, Kianmehr A, Alizadeh A. The Role of Metformin Response in Lipid Metabolism in Patients with Recent-Onset Type 2 Diabetes: HbA1c Level as a Criterion for Designating Patients as Responders or Nonresponders to Metformin. PloS one. 2016; 11(3):e0151543. https://doi.org/10.1371/journal.pone.0151543 PMID: 26978661

30. Zhang M, Liu YH, Chang CS, Zhi H, Wang S, Xu W, et al. Quantification of gene expression while taking into account RNA alternative splicing. Genomics. 2018.

31. Zhou X, Lindsay H, Robinson MD. Robustly detecting differential expression in RNA sequencing data using observation weights. Nucleic acids research. 2014; 42(11):e91. https://doi.org/10.1093/nar/gku310 PMID: 24753412

32. Law CW, Alhamdoosh M, Su S, Dong X, Tian L, Smyth GK, et al. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. F1000Research. 2016; 5.

33. Benjamini YaH, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society, Series B (Methodological). 1995; 57(1).

34. Rohart F, Gautier B, Singh A, Le Cao KA. mixOmics: An R package for 'omics feature selection and multiple data integration. PLoS computational biology. 2017; 13(11):e1005752. https://doi.org/10.1371/journal.pcbi.1005752 PMID: 29099853

35. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols. 2009; 4(1):44–57. https://doi.org/10.1038/nprot.2008.211 PMID: 19131956

36. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic acids research. 2009; 37(1):1–13. https://doi.org/10.1093/nar/gkn923 PMID: 19033363

37. JEOTP P. SciPy: Open Source Scientific Tools for Python. Computing in Science & Engineering. 2007; 9:10–20.

38. Hunter JD. Matplotlib: A 2D graphics environment. Computing in Science & Engineering. 2007; 9:99–104.

39. Clifford AJ, Rincon G, Owens JE, Medrano JF, Moshfegh AJ, Baer DJ, et al. Single nucleotide polymorphisms in CETP, SLC46A1, SLC19A1, CD36, BCMO1, APOA5, and ABCA1 are significant predictors of plasma HDL in healthy adults. Lipids in health and disease. 2013; 12:66. https://doi.org/10.1186/1476-511X-12-66 PMID: 23656756

40. Geerling JJ, Boon MR, van der Zon GC, van den Berg SA, van den Hoek AM, Lombes M, et al. Metformin lowers plasma triglycerides by promoting VLDL-triglyceride clearance by brown adipose tissue in mice. Diabetes. 2014; 63(3):880–91. https://doi.org/10.2337/db13-0194 PMID: 24270984

41. Ding L, Liang G, Yao Z, Zhang J, Liu R, Chen H, et al. Metformin prevents cancer metastasis by inhibiting M2-like polarization of tumor associated macrophages. Oncotarget. 2015; 6(34):36441–55. https://doi.org/10.18632/oncotarget.5541 PMID: 26497364

42. Do MT, Kim HG, Tran TT, Khanal T, Choi JH, Chung YC, et al. Metformin suppresses CYP1A1 and CYP1B1 expression in breast cancer cells by down-regulating aryl hydrocarbon receptor expression. Toxicology and applied pharmacology. 2014; 280(1):138–48. https://doi.org/10.1016/j.taap.2014.07.021 PMID: 25110054

43. Tu W, Luo M, Wang Z, Yan W, Xia Y, Deng H, et al. Upregulation of SATB1 promotes tumor growth and metastasis in liver cancer. Liver international: official journal of the International Association for the Study of the Liver. 2012; 32(7):1064–78.

44. Mir R, Pradhan SJ, Galande S. Chromatin organizer SATB1 as a novel molecular target for cancer therapy. Current drug targets. 2012; 13(13):1603–15. https://doi.org/10.2174/138945012803530008 PMID: 22998183

45. Cuajungco MP, Podevin W, Valluri VK, Bui Q, Nguyen VH, Taylor K. Abnormal accumulation of human transmembrane (TMEM)-176A and 176B proteins is associated with cancer pathology. Acta histochemica. 2012; 114(7):705–12. https://doi.org/10.1016/j.acthis.2011.12.006 PMID: 22244448

46. Segovia M, Russo S, Jeldres M, Mahmoud YD, Perez V, Duhalde M, et al. Targeting TMEM176B Enhances Antitumor Immunity and Augments the Efficacy of Immune Checkpoint Blockers by Unleashing Inflammasome Activation. Cancer cell. 2019; 35(5):767–81 e6. https://doi.org/10.1016/j.ccell.2019.04.003 PMID: 31085177

47. Tsuei AC, Martinus RD. Metformin induced expression of Hsp60 in human THP-1 monocyte cells. Cell stress & chaperones. 2012; 17(1):23–8.

48. Kawarabayashi R, Motoyama K, Nakamura M, Yamazaki Y, Morioka T, Mori K, et al. The Association between Monocyte Surface CD163 and Insulin Resistance in Patients with Type 2 Diabetes. Journal of diabetes research. 2017; 2017:6549242. https://doi.org/10.1155/2017/6549242 PMID: 29445750

49. Sambol NC, Chiang J, O'Conner M, Liu CY, Lin ET, Goodman AM, et al. Pharmacokinetics and pharmacodynamics of metformin in healthy subjects and patients with noninsulin-dependent diabetes mellitus. Journal of clinical pharmacology. 1996; 36(11):1012–21. https://doi.org/10.1177/009127009603601105 PMID: 8973990

50. Xu H, Zhou Y, Liu Y, Ping J, Shou Q, Chen F, et al. Metformin improves hepatic IRS2/PI3K/Akt signaling in insulin-resistant rats of NASH and cirrhosis. The Journal of endocrinology. 2016; 229(2):133–44. https://doi.org/10.1530/JOE-15-0409 PMID: 26941037

51. Gunton JE, Delhanty PJ, Takahashi S, Baxter RC. Metformin rapidly increases insulin receptor activation in human liver and signals preferentially through insulin-receptor substrate-2. The Journal of clinical endocrinology and metabolism. 2003; 88(3):1323–32. https://doi.org/10.1210/jc.2002-021394 PMID: 12629126

52. Hwang SH, Kim MC, Ji S, Yang Y, Jeong Y, Kim Y. Glucose starvation induces resistance to metformin through the elevation of mitochondrial multidrug resistance protein 1. Cancer science. 2019; 110(4):1256–67. https://doi.org/10.1111/cas.13952 PMID: 30689265

53. DeAngelis AM, Heinrich G, Dai T, Bowman TA, Patel PR, Lee SJ, et al. Carcinoembryonic antigen-related cell adhesion molecule 1: a link between insulin and lipid metabolism. Diabetes. 2008; 57 (9):2296–303. https://doi.org/10.2337/db08-0379 PMID: 18544705

54. Dong X, Park S, Lin X, Copps K, Yi X, White MF. Irs1 and Irs2 signaling is essential for hepatic glucose homeostasis and systemic growth. The Journal of clinical investigation. 2006; 116(1):101–14. https://doi.org/10.1172/JCI25735 PMID: 16374520

55. Whang A, Nagpal R, Yadav H. Bi-directional drug-microbiome interactions of anti-diabetics. EBioMedicine. 2019; 39:591–602. https://doi.org/10.1016/j.ebiom.2018.11.046 PMID: 30553752

56. Andrzejewski S, Gravel SP, Pollak M, St-Pierre J. Metformin directly acts on mitochondria to alter cellular bioenergetics. Cancer & metabolism. 2014; 2:12.

57. Bridges HR, Jones AJ, Pollak MN, Hirst J. Effects of metformin and other biguanides on oxidative phosphorylation in mitochondria. The Biochemical journal. 2014; 462(3):475–87. https://doi.org/10.1042/BJ20140620 PMID: 25017630

58. Robert F, Fendri S, Hary L, Lacroix C, Andrejak M, Lalau JD. Kinetics of plasma and erythrocyte metformin after acute administration in healthy subjects. Diabetes & metabolism. 2003; 29(3):279–83.

3.3 Novel susceptibility loci identified in a genome-wide association study of type 2 diabetes complications in population of Latvia

Highlights

1. GWAS of 601 T2DM patients revealed ten novel susceptibility loci for T2DM complications reaching genome-wide significance level. Out of them rs1132787 (*GYPA*) and rs522521 (*LOC105371557*) showed an association with diabetic neuropathy, rs2477088 (*PDE4DIP*), rs522521 (*LOC105371557*), rs4852954 (*NAT8*), rs6032 (*F5*), rs6935464 (*RPS6KA2*), rs7236163 (*ZNF519*), rs3095447 (*CCDC146*) were significantly associated with macrovascular complications, and only variant rs3095447 (*CCDC146*) was related to a greater risk of ophthalmic complications.

2. Three risk loci (rs4852954, rs7236163, rs3095447) for macrovascular complications proved to be significantly associated with the expression levels of multiple nearby genes (e.g. *FGL2*, *ALMS1*) with previously described functional importance in cardiovascular disease-related traits.

3. The targeted approach confirmed a strong genetic association for diabetic neuropathy (*MAPK14:* rs3761980, rs80028505), and diabetic nephropathy (*APOL1:* rs136161), proving the contribution of these risk loci in the pathogenesis of diabetic complications in the population of Latvia.

Ustinova *et al. BMC Med Genomics* (2021) 14:18
https://doi.org/10.1186/s12920-020-00860-4

BMC Medical Genomics

RESEARCH ARTICLE Open Access

Novel susceptibility loci identified in a genome-wide association study of type 2 diabetes complications in population of Latvia

Monta Ustinova[1], Raitis Peculis[1], Raimonds Rescenko[1], Vita Rovite[1], Linda Zaharenko[1], Ilze Elbere[1], Laila Silamikele[1], Ilze Konrade[1,3], Jelizaveta Sokolovska[2], Valdis Pirags[1,2] and Janis Klovins[1*]

Abstract

Background: Type 2 diabetes complications cause a serious emotional and economical burden to patients and healthcare systems globally. Management of both acute and chronic complications of diabetes, which dramatically impair the quality of patients' life, is still an unsolved issue in diabetes care, suggesting a need for early identification of individuals with high risk for developing diabetes complications.

Methods: We performed a genome-wide association study in 601 type 2 diabetes patients after stratifying them according to the presence or absence of four types of diabetes complications: diabetic neuropathy, diabetic nephropathy, macrovascular complications, and ophthalmic complications.

Results: The analysis revealed ten novel associations showing genome-wide significance, including rs1132787 (*GYPA*, OR = 2.71; 95% CI = 2.02–3.64) and diabetic neuropathy, rs2477088 (*PDE4DIP*, OR = 2.50; 95% CI = 1.87–3.34), rs4852954 (*NAT8*, OR = 2.27; 95% CI = 2.71–3.01), rs6032 (*F5*, OR = 2.12; 95% CI = 1.63–2.77), rs6935464 (*RPS6KA2*, OR = 2.25; 95% CI = 6.69–3.01) and macrovascular complications, rs3095447 (*CCDC146*, OR = 2.18; 95% CI = 1.66–2.87) and ophthalmic complications. By applying the targeted approach of previously reported susceptibility loci we managed to replicate three associations: *MAPK14* (rs3761980, rs80028505) and diabetic neuropathy, *APOL1* (rs136161) and diabetic nephropathy.

Conclusions: Together these results provide further evidence for the implication of genetic factors in the development of type 2 diabetes complications and highlight several potential key loci, able to modify the risk of developing these conditions. Moreover, the candidate variant approach proves a strong and consistent effect for multiple variants across different populations.

Keywords: Type 2 diabetes mellitus, Genome-wide genotyping, Diabetic complications

Background

The past few decades have shown a marked increase in the number of patients with diabetes rising from 151 million (4.6% of the global population) in 2000 to 463 million (9.3%) in 2019 [1]. The risk of type 2 diabetes (T2DM), the most common type of diabetes, is modified by a strong interaction between environmental and genetic factors [2, 3]. T2DM is a multifactorial disease with a population-specific heritability (26% in the European population) [4]. A number of common variants implicated in the pathogenesis and genetic architecture of T2DM have been identified so far, some of them also capable of modifying the pharmacologic response to antidiabetic drugs [5, 6].

*Correspondence: klovins@biomed.lu.lv
[1] Latvian Biomedical Research and Study Centre, Ratsupites iela 1, Riga 1067, Latvia
Full list of author information is available at the end of the article

Uncontrolled T2DM can lead to long-term illnesses or chronic health conditions divided into microvascular complications, such as diabetic retinopathy, nephropathy, neuropathy, and macrovascular complications, including stroke, heart disease, and peripheral vascular disease [7], accounting for 53% of direct health costs of diabetes with cardiovascular and renal complications contributing to the greatest financial burden [8, 9]. Moreover, diabetic patients show increased all-cause mortality rates, especially cardiovascular deaths (HR 2.6, 95% CI 1.4–4.7) [10].

Studies show evidence of considerable genetic component predisposing to diabetic complications, explaining even around 50% of the risk of proliferative retinopathy [11]. In the last few decades, genetic research including genome-wide association studies (GWAS), linkage analysis, and candidate gene approach has revealed several susceptibility loci for diabetic retinopathy and nephropathy (*VEGF*, *CAT*, *FTO*, *UCP1,* and *INSR*), and also macrovascular complications (*ADIPOQ*). Nevertheless, they explain only a small proportion of the phenotypic variation observed in T2DM patients [12–17], justifying a need for identification of novel genetic risk factors for T2DM complications and improvement of knowledge about molecular mechanisms underlying these comorbid conditions.

Since the high impact of population specificity for the discrimination of genetic variants and their contribution to the phenotype of interest is evidenced by a number of SNPs that failed to replicate in different populations, both discovery and replication studies in populations of different ancestries are needed [18, 19]. An example of population-specific allele frequency is rs61736969, the risk variant of T2DM, which is located in the *TBC1D4* (TBC1 Domain Family Member 4) gene. It is highly frequent (minor allele frequency of 17%) in the Greenlandic population, nevertheless, it has not been present so far in European individuals, most probably due to different linkage disequilibrium patterns [20]. Although poor glycemic control is considered to be the driving factor of T2DM complications, early genotype-based identification of individuals at high risk of diabetic complications may promote the prevention or, at least, delay of the disease [21]. In this study, we aimed to discriminate novel susceptibility loci for T2DM complications and replicate the findings of other GWAS in the study cohort from the Genome Database of the Latvian Population (LGDB) [22].

Methods

Study group and phenotype definitions

In total, the study cohort consisted of 601 T2DM patients of European ancestry with and without a medical history of diabetes complications, selected from the participants of LGDB (recruited from June 2007 to November 2016) according to the following inclusion criteria: (1) clinically confirmed diagnosis of T2DM (E11 diagnosis code, ICD-10), (2) information on age at diagnosis, sex, weight, and height available, (3) records of national diabetes registry and Latvian hospital inpatient discharges available. Written broad consent was obtained from every subject during the recruitment in LGDB.

The collection of blood samples and relevant anthropometric data was ensured by LGDB according to their standard procedures [22]. Associated clinical data, including the diagnosis date of T2DM, date and type of diabetes complications, HbA1c measures, and medications used, were obtained from the records of Diabetes registry, Latvian hospital inpatient discharges, outpatient progress notes, and pharmacy recipe records, provided by The Centre for Disease Prevention and Control of Latvia and National Health Service of Latvia (Approval No. 3, Decision No. 7.1–3/3). The data about diabetic complications present for T2DM patients involved in LGDB were applied for accurate stratification of 601 T2DM patients in four phenotype-based groups according to the type of complications experienced: diabetic neuropathy, diabetic nephropathy, ophthalmic complications, and macrovascular complications. The definition of phenotypes and patient stratification in different complication groups was done as follows:

1. Diabetic neuropathies: clinical diagnosis codes (ICD-10) E11.4 and E11.5, records of amputation of the leg/toe, gangrene, shunting and angioplasty, and presence of intermittent claudication or fresh ulcers since the diagnosis of T2DM.
2. Diabetic nephropathies: clinical diagnosis code E11.2 or records of kidney failure, kidney transplantation, renal replacement therapy, microalbuminuria, hemodialysis, peritoneal dialysis performed after the diagnosis of T2DM.
3. Ophthalmic complications: clinical diagnosis code E11.3 or records of photocoagulation, maculopathy, retinopathy, operative therapy, blindness made since the diagnosis of T2DM.
4. Macrovascular complications: clinical diagnosis codes I95, I20, I21, I24, I25, I50, I60, I61, I63, I64, and records of coronary shunting and angioplasty after the diagnosis of T2DM.

ICD-10-based phenotype definitions corresponding to the Latvian guidelines of diabetes management are generally used in clinical practice in Latvia. Subjects with the above-mentioned diagnosis codes or medical events recorded were considered as cases in their corresponding

complication groups, while T2DM patients with no evidence of complications of interest during their follow-up period were recognized as controls in the particular group. Subjects experiencing specific diabetes complications before the set of T2DM diagnosis were excluded from the analysis of a particular complication group, explaining the variable total number of individuals among all complication groups tested. The follow-up period which coincides with diabetes duration was considered as time since the set of T2DM diagnosis until the date of diabetes complication recorded for cases or the date of the last entry in the National registry for control subjects. Administration of medications was considered in a group-specific manner, accounting for angiotensin II receptor blockers and angiotensin-converting enzyme inhibitors in the analysis of all complication groups analyzed and additional lipid-modifying agents in the analysis of macrovascular complications. In order to adjust for the inter-individual variability of glycaemic control, a key factor in the development of T2DM complications, the median HbA1c level during the observation period was fitted as a covariate, irrespective of the antidiabetic therapy used. Sex, age at the diagnosis of T2DM, body mass index (BMI), diabetes duration, and use of particular medications were also included among the covariates.

DNA extraction and genotyping

Within the framework of this study DNA samples from 601 T2DM patients were used. DNA was isolated from peripheral blood leukocytes using a phenol–chloroform extraction method according to LGDB standard procedures [22]. DNA samples were genotyped with the Infinium Global Screening Array (Illumina, USA) on the iScan System microarray scanner (Illumina, USA). Illumina Genome Studio v2.0 was used to convert raw data into PLINK format and workflow described in Marees et al. [23]. used for data quality control. SHAPEIT v2.r900 [24] and IMPUTE2 [25] were used for genotype phasing and genotype imputing. Imputed data were filtered using the following parameters: marker correlation (INFO) > 0.8, hard call threshold 0.1, minor allele frequency > 1%, Missingness < 2%.

Statistical analysis

Association analyses corresponding to four different complication groups (macrovascular complications, diabetic neuropathy, diabetic nephropathy, ophthalmic complications) were performed using PLINK v1.9 logistic regression with covariates: median HbA1c, sex, age at the diagnosis, diabetes duration, BMI, medications used. A genome-wide significance threshold of $P < 5 \times 10^{-8}$ was defined.

For the targeted analysis candidate variants were selected from GWAS Central [26] (http://www.gwascentral.org/) and GWAS Catalog [27] (https://www.ebi.ac.uk/gwas/home), based on previously reported association with T2DM complications: diabetic neuropathy (4 allelic variants selected), macrovascular complications (43 allelic variants selected), ophthalmic complications (98 allelic variants selected) and diabetic nephropathy (49 allelic variants selected). A complete list of selected SNPs, their positions, and associated traits is provided in Additional file 1. False discovery rate (FDR) according to the Benjamini–Hochberg procedure was calculated to account for multiple testing and the threshold was set at 0.05.

Manhattan plots and Q-Q plots were generated in R v3.5.3 using the qqman package, while the Venn diagram was developed in the online visualization tool Venny 2.1.0. Statistical analysis of anthropometric measures and biochemical data was performed in R v3.5.3. by applying the Wilcoxon rank-sum test and Pearson's chi-squared test with a p-value threshold < 0.05. For identification of the functional role of allelic variants expression quantitative trait locus (eQTL) analysis was conducted by using the open-access Genotype Tissue Expression (GTEx) database [28]. The tissue types for eQTL analysis were carefully selected considering the etiology of the diseases (artery, nerve, heart, skin, blood) [29]. The p-value threshold of 0.05 was used to discriminate significant associations. Variant Effect Predictor and Linkage Disequilibrium (LD) data from 1000 Genome project (Utah Residents (CEPH) with Northern and Western European Ancestry) were employed to explore the functional consequences of each variant and other variants in LD [30]. To evaluate the potential SNP effects on quantitative phenotypes, analyses of variant association with HbA1C and BMI were performed using PLINK v1.9—assoc function for quantitative phenotypes, where the genome-wide significance threshold of $P < 5 \times 10^{-8}$ was used to identify the significant hits.

Results

Genome-wide association analysis

We studied a cohort of 601 T2DM patients of which 241 were men and 360 women aged 22 to 82 years (average age 56.86 ± 10.24 years) stratified as controls or cases in four complication groups tested (diabetic nephropathy n = 601, diabetic neuropathy n = 600, ophthalmic complications n = 601 and macrovascular complications n = 559), based on corresponding diagnosis or medical events experienced after the onset of T2DM. A phenotype-based distribution of subjects among different complication groups is shown in Fig. 1. After inspection of patient anthropometric data, we observed that T2DM

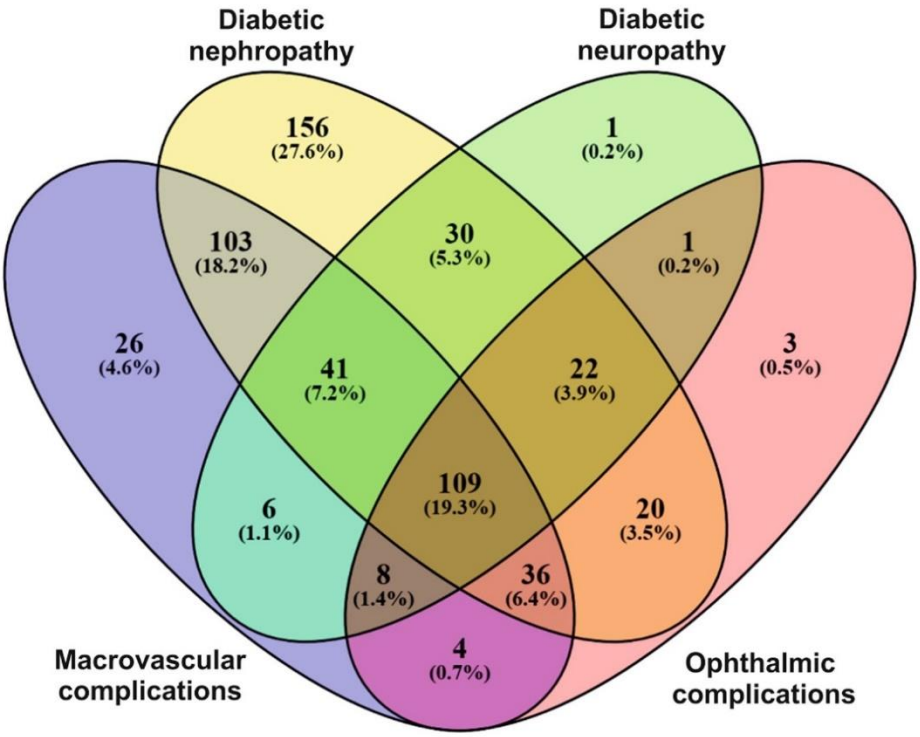

Fig. 1 Venn diagram showing the distribution and overlap of cases among the analyzed T2DM complication groups

patients experiencing ophthalmic complications and diabetic neuropathy were significantly younger compared to the group of patients without particular complications while those with diabetic nephropathy, macrovascular and ophthalmic complications were characterized by longer duration of diabetes. As expected, median glycated hemoglobin levels were higher in cases (patients with specific complications) compared to controls (patients without the same diabetes complication) except for the diabetic nephropathy group (Table 1).

After quality control and filtering 5 378 539 SNPs were used for further testing in each T2DM complication group. The total genotyping rate was >0.99 in all of the tested T2DM complication groups. The genomic inflation factor was negligible in all data sets based on median chi-squared statistics: 1.02 for neuropathies, 1.02 for macrovascular complications, 1.01 for ophthalmic complications, and 1.00 for nephropathies. After adjustments for age, sex, BMI, diabetes duration, median HbA1c, and medications used ten susceptibility loci were identified for different T2DM complications (Fig. 2), among them rs1132787 (*GYPA*) and rs522521 (*LOC105371557)* showed an association with diabetic neuropathy, rs2477088 (*PDE4DIP),* rs522521 (*LOC105371557),* rs4852954 (*NAT8)*, rs6032 (*F5*), rs6935464 (*RPS6KA2*), rs7236163 (*ZNF519),* rs3095447 (*CCDC146*) were significantly associated with macrovascular complications, and only variant rs3095447 (*CCDC146*) was related to a greater risk of ophthalmic complications, while no significant hits were found for diabetic nephropathy (Table 2, Additional file 2).

The functional consequences of the identified variants were evaluated by the eQTL analysis, which was conducted in GTEx database by focusing only on those tissue types that may be relevant or even damaged according to the etiology of the disease [29]. Three out of eight novel allelic variants identified in this study appeared to be significantly associated with the expression of multiple nearby genes: rs4852954 (*ALMS1*, *DUSP11*, *NAT8, ALMS1P1, TPRKB, ALMS1-IT1, RP11-434P11.2),* rs7236163 (*ZNF519, RP11-411B10.2*), and rs3095447 (*RP11-467H10.1, FGL2, PMS2P9, GSAP, SPDYE18, FAM185BP, CCDC146, UPK3BP1*). See Additional file 3 for the full list of the identified associations.

Targeted analysis

In order to investigate if previously reported associations in other populations are also true in our study

Table 1 Characteristics of the study participants

	Neuropathies			Nephropathies			Ophthalmic complications			Macrovascular complications		
	Cases (n = 218)	Controls (n = 382)	*P* value	Cases (n = 517)	Controls (n = 84)	*P* value	Cases (n = 203)	Controls (n = 398)	*P* value	Cases (n = 333)	Controls (n = 226)	*P*-value
Mean age, years ± SD	55.28 ± 10.24	57.72 ± 10.13	4.61E−03	56.59 ± 10.25	58.57 ± 10.02	1.02E−01	54.38 ± 10.65	58.10 ± 9.80	6.59E−05	56.36 ± 9.88	56.52 ± 10.63	7.58E−01
Female/male, n (%)	120/98 (55.05/44.95)	239/143 (62.57/37.43)	7.85E−02	312/205 (60.35/39.65)	47/37 (55.92/44.05)	5.08E−01	114/89 (56.16/43.84)	245/153 (61.56/38.44)	2.21E−01	190/142 (57.23/42.77)	149/77(65.93/34.07)	4.80E−02
Mean BMI, kg/m^2 ± SD	33.27 ± 6.29	33.09 ± 5.96	7.39E−01	33.37 ± 6.00	32.21 ± 6.50	5.18E−02	33.23 ± 6.22	33.2 ± 6.02	8.11E−01	33.46 ± 6.13	32.79 ± 5.97	2.41E−01
Median HbA1c, % (IQR)	7.20 (1.70)	6.90 (1.39)	3.93E−05	7.10 (1.54)	7.00 (1.83)	4.15E−01	7.30 (1.95)	6.91 (1.35)	1.55E−05	7.10(1.85)	6.90(1.15)	3.08E−05
Diabetes duration, years ± SD	6.72 ± 6.09	6.30 ± 4.50	4.07E−01	7.05 ± 6.32	4.79 ± 5.31	3.29E−05	5.95 ± 5.53	6.29 ± 4.45	1.21E−02	5.40 ± 6.09	5.74 ± 4.97	5.88E-03

SD standard deviation, *BMI* body mass index, *HbA1c* hemoglobin A1C, *IQR* interquartile range

(See figure on next page.)

Fig. 2 Manhattan plots for GWAS of T2DM complications. **a** Diabetic neuropathies, **b** macrovascular complications, **c** ophthalmic complications. X-axis shows chromosomal positions. Y-axis shows –log10 *P*-values. The red lines indicate a genome-wide significant threshold of $P < 5 \times 10^{-8}$, and the blue lines denote a suggestive significance threshold of $P < 5 \times 10^{-5}$. Association signals that reached genome-wide significance are denoted by reference SNP ID number

cohorts, we performed a targeted analysis in the same subjects and stratification in complication groups (diabetic neuropathy, macrovascular complications, ophthalmic complications, and diabetic nephropathy) (Table 1) by performing an association analysis for 194 candidate variants in total. Information on previously known allele-trait associations reported in both GWAS Catalog and GWAS Central was used for the selection of risk variants associated with at least one group of the tested T2DM complications (see Additional file 1 for a full list of selected candidate SNPs). The genotyping rate in all of the complication groups analyzed was > 0.99. By applying the targeted approach we managed to replicate two significant associations for diabetic neuropathy rs3761980 and rs80028505, both mapping to *MAPK14* loci and one significant hit (rs136161, *APOL1*) for diabetic nephropathy (Table 3) in our study cohort.

Discussion

Here we present the results of the genome-wide association study for T2DM complications performed in a population of Latvia for the first time, revealing 10 susceptibility loci for T2DM complications, including diabetic neuropathy, macrovascular and ophthalmic complications. As in other reports aimed to identify the risk factors of T2DM complications [15, 32], the control group of our study consisted of T2DM patients with no evidence of the complication type of interest instead of conventional healthy subjects, since the implementation of healthy controls would rather reveal genetic associations with the diagnosis of T2DM itself, not the T2DM complications.

We found two novel variants (rs1132787 and rs522521) associated with diabetic neuropathy and none of them have been linked to any disease or specific phenotype before. Variant rs1132787 is located within the 3′ UTR of a gene coding for glycophorin A coding (*GYPA*). Glycophorin A is the major erythrocyte membrane sialoglycoprotein. Although it has not been directly associated with susceptibility to any T2DM-related condition before, studies report a significant upregulation of the *GYPA* gene in the dorsal root ganglia of a mouse model of T2DM and the metabolic syndrome, and even type 1 diabetes with diabetic polyneuropathy [33, 34]. Moreover, copy number variation overlapping *GYPA* has been already linked to body mass index, obesity, and obesity-related traits, such as weight, hip circumference, and waist circumference, providing more evidence for the potential contribution of *GYPA* in the development of diabetic complications [35]. Because of the multiple evidence of the implication of *GYPA* gene in neuropathies and metabolic traits, and the potential functional consequence of the top variant rs1132787 located within the 3′UTR of the gene, we consider the *GYPA* gene as the first candidate for future functional validation studies. Nevertheless, credible evidence for this association should be established first by performing a replication study with larger sample size. The other identified risk variant rs522521 is located near the poorly characterized gene *LOC105371557* with yet unknown function.

In total, seven variants reached genome-wide significance for the association with macrovascular complications of T2DM. The strongest association was exhibited by an intron variant of the Phosphodiesterase 4D Interacting Protein coding gene (*PDE4DIP*). Although the rs2477088 variant has not been previously linked to any T2DM manifestations, *PDE4DIP*, also known as myomegalin or cardiomyopathy-associated protein 2, is a well-known contributor of the microtubule control process [36] and some previous evidence exist indicating on the potential role of the gene in macrovascular diseases and T2DM. The exome sequencing has revealed a rare variant of *PDE4DIP*, which significantly increases the risk of ischemic stroke [37], moreover, CpG island methylation in leukocytes annotated to *PDE4DIP* contributes to the epigenetic fingerprint of myocardial infarction [38], and finally, the gene is also significantly downregulated in liver of T2DM patients [39]. Another risk allele (rs4852954) identified in the analysis of macrovascular complications is located near the N-Acetyltransferase 8 coding gene (*NAT8*) and has been previously associated with systolic blood pressure and renal function in the Estonian population [40] which is genetically close to the Latvian population [41]. Although *NAT8* gene has not been linked with T2DM macrovascular complications before, it is considered to be a susceptibility locus for diabetic kidney disease [42]. In addition, we found the coagulation factor 5 coding gene (*F5*) among the risk loci of macrovascular complications. The variant rs6032 is located only around 7 kb from the Factor V Leiden (rs6025), which has been strongly associated with ischemic stroke and

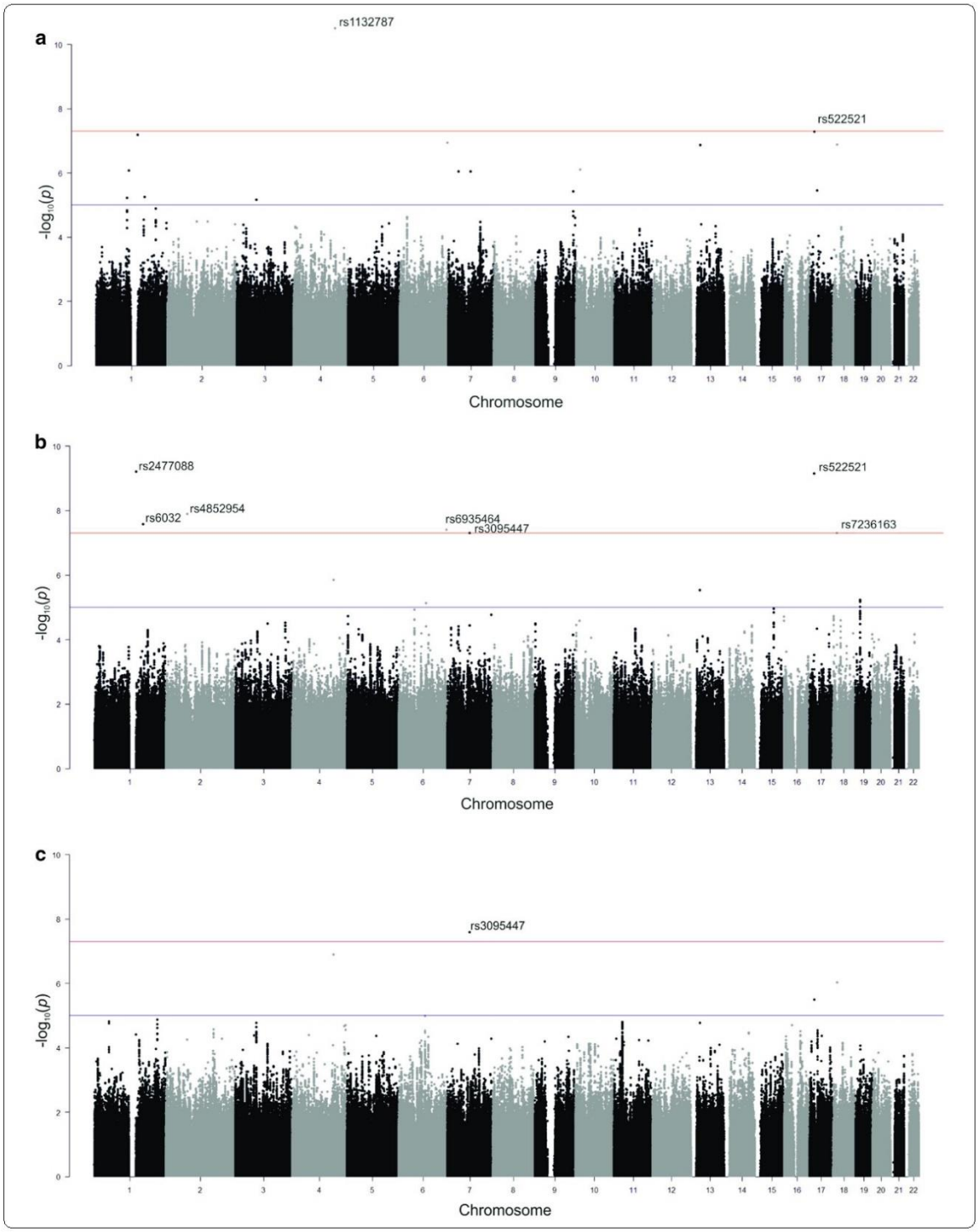
a
rs1132787
rs522521
-log10(p)
Chromosome
b
rs2477088
rs522521
rs6032
rs4852954
rs6935464
rs3095447
rs7236163
-log10(p)
Chromosome
c
rs3095447
-log10(p)
Chromosome

Table 2 Susceptibility loci associated with type 2 diabetes mellitus complications

CHR	Closest gene	SNP	A1/A2	OR (95% CI)	*P*-value	MAF cases/controls	MAF European
Diabetic neuropathy							
4:145030546	*GYPA*	rs1132787	T/C	2.71 (2.02–3.64)	3.23E−11	0.37/0.16	0.31
17:15733545	*LOC105371557*	rs522521	A/C	0.49 (0.38–0.64)	5.07E−08	0.37/0.56	0.70
Macrovascular complications							
1:144936353	*PDE4DIP*	rs2477088	T/C	2.50 (1.87–3.34)	6.11E−10	0.51/0.32	0.58
17:15733545	*LOC105371557*	rs522521	A/C	0.42 (17.32–0.56)	6.95E−10	0.40/0.60	0.70
2:73870010	*NAT8*	rs4852954	T/C	2.27 (2.71–3.01)	1.26E−08	0.50/0.31	0.61
1:169511555	*F5*	rs6032	T/C	2.12 (1.63–2.77)	2.62E−08	0.56/0.37	0.73
6:167114208	*RPS6KA2*	rs6935464	A/G	2.25 (6.69–3.01)	3.89E−08	0.44/0.26	0.57
18:14150724	*ZNF519*	rs7236163	T/C	2.14 (18.63–2.82)	4.97E−08	0.46/0.27	0.58
7:76764970	*CCDC146*	rs3095447	A/C	2.16 (7.64–2.84)	4.98E−08	0.55/0.37	0.79
Ophthalmic complications							
7:76764970	*CCDC146*	rs3095447	A/C	2.18 (1.66–2.87)	2.55E−08	0.61/0.40	0.79

CHR chromosome and base pair position in Human Genome build hg19, *A1* minor allele, *A2* major allele, *OR* odds ratio for the minor allele, *CI* confidence interval 95%, *MAF* minor allele frequency, *MAF European* minor allele frequency observed in the European population [31]

Table 3 Candidate variants showing a significant association with type 2 diabetes complications

CHR	Closest gene	SNP	A1/A2	MAF cases/controls	MAF European	OR (95% CI)	*P*-value	FDR	References
Neuropathies									
6:35993906	*MAPK14, SLC26A8*	rs3761980	G/A	0.13/0.09	0.10	1.58 (1.08–2.33)	1.94E-02	3.88E-02	Meng et al. [49]
6:35998388	*MAPK14*	rs80028505	T/C	0.13/0.09	0.11	1.58 (1.08–2.33)	1.94E-02	3.88E-02	Meng et al. [49]
Nephropathies									
22:36657432	*APOL1*	rs136161	G/C	0.47/ 0.29	0.40	2.00 (1.40–2.86)	1.41E-04	6.93E-03	Iyengar et al. [50]

CHR chromosome and base pair position in Human Genome build hg19, *A1* minor allele, *A2* major allele, *OR* odds ratio for the minor allele, *CI* confidence interval 95%, *FDR* false discovery rate by Benjamini & Hochberg method, *MAF* minor allele frequency, *MAF* European: minor allele frequency observed in the European population [31]

incident venous thrombosis before [43, 44], therefore we may speculate that in rs6032 carriers manifestation of the trait are amplified by the presence of T2DM. Although rs6032 is the only missense variant among the top hits, it is categorized as benign according to SIFT and PolyPhen. Nevertheless, the variant is in the LD with rs4524, the risk variant for venous thromboembolism (OR = 1.14; CI = 1.11–1.16) [45].

Finally, 4 more loci showed genome-wide significance for the association with macrovascular complications, among them rs6935464, located within the *RPS6KA2* gene coding for Ribosomal Protein S6 Kinase A2, which is involved in cardiac myocyte stress responses and even considered as a therapeutic target for the prevention of heart failure [46], and rs3095447, an intron variant of Coiled-Coil Domain Containing 146 gene (*CCDC146*) which is also the only significant hit for ophthalmic complications in our data. The last two variants (rs522521 and rs7236163) identified in the analysis of macrovascular complications are located in the intergenic regions near genes *LOC105371557* and *ZNF519* respectively, and both have not been linked to any disease before.

We noticed two of the variants (rs522521, rs3095447) appearing among the top hits of multiple complication groups, which seems rational since a number of patients had experienced more than one type of complication, resulting in a notable overlap of patients among four tested phenotype groups (Fig. 1). This finding may be also explained by similar etiologic characteristics between microvascular diabetic complications involving small vessels (neuropathy, nephropathy, ophthalmic complications) and macrovascular complications related to large vessel damage, with chronic hyperglycemia as the main cause of all these comorbidities. Moreover, microvascular and macrovascular complications tend to be strongly interconnected, and the damage of small vessels may contribute to the manifestations of heart disease in diabetes [7], which coincides with our data showing only a small number of T2DM patients corresponding to one complication group only, while 18% (n = 109) of T2DM patients

included in this study had experienced multiple manifestations during the observation period, corresponding to all four analyzed complication types.

According to Ensembl Variant Effect Predictor, most of the identified variants are intronic, though rs1132787, which is located in the enhancer site, and rs4852954 laying within the promoter flanking region may disrupt the functions of regulatory elements and therefore modulate the gene expression patterns. Some of the genes affected by the identified eQTLs (Additional file 3) are previously associated with the performance of the vascular system, for instance, expression of Fibrinogen-like protein 2 coding gene (*FGL2*) in endothelial cells has been previously linked to microthrombosis and cardiac impairment in rats with T2DM [47], and Alstrom Syndrome Protein 1 coding gene (*ALMS1*) is associated with Alström syndrome and characteristic dilated cardiomyopathy [48]. According to GTEx data, rs3095447 negatively correlates with *FGL2* expression in heart and artery, while rs4852954 is linked to lower *ALMS1* levels in whole blood and skin tissue, suggesting the possible functional implication of these variants in the pathogenesis of the disease. In order to explore additional effects of the top hits, quantitative phenotype analyses were performed, revealing significant associations between rs6935464, rs6032, rs3095447, rs4852954, rs7236163, rs522521, and median HbA1c levels. These data suggest that the identified genetic loci may serve as markers for both, development of T2DM complications and alterations in HbA1c levels, though the particular analysis should be repeated in a different cohort where the variable impact of different anti-diabetic medications should also be considered.

By performing targeted analysis of candidate variants we managed to replicate three associations with T2DM complications in the population of Latvia. Both variants (rs3761980 and rs80028505) showing an association with T2DM neuropathy are located near the Mitogen-activated protein kinase 14 coding gene (*MAPK14*) and have been previously linked to increased risk of diabetic foot ulcers in the report of The Genetics of Diabetes Audit and Research in Tayside Scotland (GoDARTS) project [49]. In addition, we found a significant association of rs136161 located in the Apolipoprotein L1 coding gene (*APOL1*) with T2DM nephropathy, which has been already linked to an advanced diabetic kidney disease across multiple ethnic groups [50]. Although we have observed smaller effect sizes for rs3761980 and rs80028505 comparing to other GWAS and much larger effect for rs136161, the previously reported odds ratio values for all three variants fit within the 95% confidence intervals calculated in our study [49, 50].

This study has several limitations, though the small sample size is the primary limiting factor in the risk variant discovery, which may result in an insufficient statistical power for the detection of rare variants with small effect sizes. This may also explain why variants with large effect sizes (OR up to 2.71) are markedly represented among the significant results of our study. Although the retrospective distribution of cases and controls in the analysis of T2DM neuropathy and ophthalmic complications was relatively balanced, the lack of significant hits in the analysis of diabetic nephropathy may be explained by the high incidence of the specific type of complication among the study participants (86%), leaving the size of the control group too small for the identification of true associations. Additionally, the suboptimal case-to-control ratio in the analysis of macrovascular complications may affect the statistical power of the study and explain the relatively high number of associations identified in the specific complication group. Due to the limited number of study subjects, the follow-up period was not fixed or set as inclusion criteria, though it was fitted as a covariate in the association analysis. Since the duration of diabetes is one of the strongest risk factors for the development of vascular complications [51], the use of an equal observation period in all cases would reduce the residual variability and improve the quality of this study.

Successful integration of genotyping data with longitudinal phenotypic information produced from several national health registries has provided strong support for 10 loci showing a genome-wide significance for the association with T2DM complications, some of them with already known importance to the comorbid conditions analyzed. We believe that these findings provide deeper insight into the pathogenesis of T2DM complications and suggest novel candidate genes for further functional studies, while our targeted approach highlights several susceptibility loci showing a directionally consistent impact on phenotype in multiple populations.

Conclusions

Using the genome-wide genotyping approach this study identified ten novel associations with T2DM complications, including *GYPA* (rs1132787) in diabetic neuropathy, *PDE4DIP* (rs2477088), *NAT8* (rs4852954), *F5* (rs6032), *RPS6KA2* (rs6935464) in macrovascular complications, and *CCDC146* (rs3095447) in ophthalmic complications. Meanwhile, the candidate gene analysis demonstrated a strong association for diabetic neuropathy (*MAPK14:* rs3761980, rs80028505), and diabetic nephropathy (*APOL1:* rs136161), proving the contribution of these risk loci in the pathogenesis of diabetic complications across various populations.

Supplementary information

The online version contains supplementary material available at https://doi.org/10.1186/s12920-020-00860-4.

Additional file 1: List of all candidate SNPs selected for targeted association analysis.

Additional file 2: Quantile-quantile plots for GWAS of T2DM complications.

Additional file 3: List of significant associations obtained in eQTL analysis.

Abbreviations

BMI: Body mass index; CI: Confidence interval; eQTL: Expression quantitative trait locus; GWAS: Genome-wide association study; HbA1c: Hemoglobin A1C; HR: Hazard ratio; ICD-10: 10Th revision of the International Statistical Classification of Diseases and Related Health Problems.; LGDB: Genome Database of the Latvian Population; MAF: Minor allele frequency; OR: Odds ratio; SD: Standard deviation; SNP: Single nucleotide polymorphism; T2DM: Type 2 diabetes mellitus; UTR: Untranslated region.

Acknowledgements

We thank the participants for their contribution and acknowledge the Genome Database of the Latvian Population for providing biological material and data.

Authors' contributions

MU, RP, VR, and JK made substantial contributions to the design of the work; RP, RR, LZ, and MU performed data analysis and interpretation; VR, IK, IE, LS, and LZ were involved in the DNA processing and acquisition of genotyping data; VP, JS, IE, LS, IK were involved in the acquisition of registry data, analysed clinical outcomes and established definition of phenotypes for analysis; MU and RP drafted the manuscript; VR, JK provided critical revision of the manuscript. All authors read and approved the final manuscript, and have agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature. All authors read and approved the final manuscript.

Funding

The study was supported by European Regional Development Fund (ERDF) On Implementation of Activity 1.1.1.2 "Post-doctoral Research Aid" of the Specific Aid Objective 1.1.1 "To increase the research and innovative capacity of scientific institutions of Latvia and the ability to attract external financing, investing in human resources and infrastructure" of the Operational Programme "Growth and Employment". Project No 1.1.1.2/VIAA/2/18/287 "Identification of clinical subgroups of Type 2 diabetes mellitus and application of pharmacogenetics in the development of personalized antidiabetic therapy". The funding body was not involved in the design of the study, collection, analysis or interpretation of data, or writing the manuscript.

Availability of data and materials

The majority of the data generated and analysed during this study are included in this published article and its supplementary information files. The raw genotyping data are under restricted access from LGDB, nevertheless, they are available for research purposes from the corresponding author on reasonable request.

Ethics approval and consent to participate

The study protocol is approved by the Central Medical Ethics Committee of Latvia (No. 01-29.1/2223). The study was conducted in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) and The Convention for the protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine. All study participants gave broad written informed consent for the recruitment in LGDB involving the use of their samples for research purposes. The acquisition of clinical data from the records of the national diabetes registry and Latvian hospital inpatient discharges was approved by the Latvian Centre for Disease Prevention and Control (No. 3/10.07.2017.) according to the regulations of Latvia.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1] Latvian Biomedical Research and Study Centre, Ratsupites iela 1, Riga 1067, Latvia. [2] Faculty of Medicine, University of Latvia, Jelgavas iela 3, Riga 1004, Latvia. [3] Faculty of Medicine, Riga Stradins University, Dzirciema iela 16, Riga 1007, Latvia.

Received: 25 November 2020 Accepted: 20 December 2020
Published online: 11 January 2021

References

1. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. Diabetes Res Clin Pract. 2019;157:107843.
2. Murea M, Ma L, Freedman BI. Genetic and environmental factors associated with type 2 diabetes and diabetic vascular complications. Rev Diabet Stud. 2012;9(1):6–22.
3. Ali O. Genetics of type 2 diabetes. World J Diabetes. 2013;4(4):114–23.
4. Poulsen P, Kyvik KO, Vaag A, Beck-Nielsen H. Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance–a population-based twin study. Diabetologia. 1999;42(2):139–45.
5. Zhao W, Rasheed A, Tikkanen E, Lee JJ, Butterworth AS, Howson JMM, et al. Identification of new susceptibility loci for type 2 diabetes and shared etiological pathways with coronary heart disease. Nat Genet. 2017;49(10):1450–7.
6. Prudente S, Di Paola R, Pezzilli S, Garofolo M, Lamacchia O, Filardi T, et al. Pharmacogenetics of oral antidiabetes drugs: evidence for diverse signals at the IRS1 locus. Pharmacogenomics J. 2018;18(3):431–5.
7. Chawla A, Chawla R, Jaggi S. Microvasular and macrovascular complications in diabetes mellitus: Distinct or continuum? Indian J Endocrinol Metab. 2016;20(4):546–51.
8. von Ferber L, Koster I, Hauner H. Medical costs of diabetic complications total costs and excess costs by age and type of treatment results of the German CoDiM Study. Exp Clin Endocrinol Diabetes. 2007;115(2):97–104.
9. Clarke P, Gray A, Legood R, Briggs A, Holman R. The impact of diabetes-related complications on healthcare costs: results from the United Kingdom Prospective Diabetes Study (UKPDS Study No. 65). Diabet Med. 2003;20(6):442–50.
10. Barr EL, Zimmet PZ, Welborn TA, Jolley D, Magliano DJ, Dunstan DW, et al. Risk of cardiovascular and all-cause mortality in individuals with diabetes mellitus, impaired fasting glucose, and impaired glucose tolerance: the Australian Diabetes, Obesity, and Lifestyle Study (AusDiab). Circulation. 2007;116(2):151–7.
11. Hietala K, Forsblom C, Summanen P, Groop PH, FinnDiane Study G. Heritability of proliferative diabetic retinopathy. Diabetes. 2008;57(8):2176–80.
12. Vionnet N, Tregouet D, Kazeem G, Gut I, Groop PH, Tarnow L, et al. Analysis of 14 candidate genes for diabetic nephropathy on chromosome 3q in European populations: strongest evidence for association with a variant in the promoter region of the adiponectin gene. Diabetes. 2006;55(11):3166–74.
13. Montesanto A, Bonfigli AR, Crocco P, Garagnani P, De Luca M, Boemi M, et al. Genes associated with Type 2 Diabetes and vascular complications. Aging (Albany NY). 2018;10(2):178–96.
14. Looker HC, Nelson RG, Chew E, Klein R, Klein BE, Knowler WC, et al. Genome-wide linkage analyses to identify Loci for diabetic retinopathy. Diabetes. 2007;56(4):1160–6.
15. Liu C, Chen G, Bentley AR, Doumatey A, Zhou J, Adeyemo A, et al. Genome-wide association study for proliferative diabetic retinopathy in Africans. NPJ Genom Med. 2019;4:20.

16. Jardeleza MS, Miller JW. Review of anti-VEGF therapy in proliferative diabetic retinopathy. Semin Ophthalmol. 2009;24(2):87–92.
17. Cheung CY, Hui EY, Lee CH, Kwok KH, Gangwani RA, Li KK, et al. Impact of genetic loci identified in genome-wide association studies on diabetic retinopathy in Chinese patients with type 2 diabetes. Invest Ophthalmol Vis Sci. 2016;57(13):5518–24.
18. Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, et al. Genetic analyses of diverse populations improves discovery for complex traits. Nature. 2019;570(7762):514–8.
19. Kochi Y, Suzuki A, Yamada R, Yamamoto K. Genetics of rheumatoid arthritis: underlying evidence of ethnic differences. J Autoimmun. 2009;32(3–4):158–62.
20. Moltke I, Grarup N, Jorgensen ME, Bjerregaard P, Treebak JT, Fumagalli M, et al. A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. Nature. 2014;512(7513):190–3.
21. Maghbooli Z, Pasalar P, Keshtkar A, Farzadfar F, Larijani B. Predictive factors of diabetic complications: a possible link between family history of diabetes and diabetic retinopathy. J Diabetes Metab Disord. 2014;13:55.
22. Rovite V, Wolff-Sagi Y, Zaharenko L, Nikitina-Zake L, Grens E, Klovins J. Genome database of the latvian population (LGDB): design, goals, and primary results. J Epidemiol. 2018;28(8):353–60.
23. Marees AT, de Kluiver H, Stringer S, Vorspan F, Curis E, Marie-Claire C, et al. A tutorial on conducting genome-wide association studies: quality control and statistical analysis. Int J Methods Psychiatr Res. 2018;27(2):e1608.
24. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods. 2013;10(1):5–6.
25. 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.
26. Beck T, Shorter T, Brookes AJ. GWAS Central: a comprehensive resource for the discovery and comparison of genotype and phenotype data from genome-wide association studies. Nucleic Acids Res. 2020;48(D1):D933–40.
27. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 2019;47(D1):D1005–12.
28. Consortium GT. Erratum: Genetic effects on gene expression across human tissues. Nature. 2018;553(7689):530.
29. Forbes JM, Cooper ME. Mechanisms of diabetic complications. Physiol Rev. 2013;93(1):137–88.
30. Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, et al. Ensembl 2020. Nucleic Acids Res. 2020;48(D1):D682–8.
31. Phan L,Jin Y,Zhang H, Qiang W, Shekhtman E, Shao D, Revoe D, Villamarin R, Ivanchenko E, Kimura M, Wang ZY, Hao L, Sharopova N, Bihan M, Sturcke A, Lee M, Popova N, Wu W, Bastiani C, Ward M, Holmes JB, Lyoshin V, Kaur K, Moyer E, Feolo M, Kattman BL. ALFA: Allele Frequency Aggregator: National Center for Biotechnology Information, U.S. National Library of Medicine; 2020 [10 Mar]. Available from: www.ncbi.nlm.nih.gov/snp/docs/gsr/alfa/.
32. Meng W, Shah KP, Pollack S, Toppila I, Hebert HL, McCarthy MI, et al. A genome-wide association study suggests new evidence for an association of the NADPH Oxidase 4 (NOX4) gene with severe diabetic retinopathy in type 2 diabetes. Acta Ophthalmol. 2018;96(7):e811–9.
33. Hur J, Dauch JR, Hinder LM, Hayes JM, Backus C, Pennathur S, et al. The metabolic syndrome and microvascular complications in a murine model of type 2 diabetes. Diabetes. 2015;64(9):3294–304.
34. Cheng C, Kobayashi M, Martinez JA, Ng H, Moser JJ, Wang X, et al. Evidence for epigenetic regulation of gene expression and function in chronic experimental diabetic neuropathy. J Neuropathol Exp Neurol. 2015;74(8):804–17.
35. Zhao W, Wineinger NE, Tiwari HK, Mosley TH, Broeckel U, Arnett DK, et al. Copy number variations associated with obesity-related traits in African Americans: a joint analysis between GENOA and HyperGEN. Obesity (Silver Spring). 2012;20(12):2431–7.
36. Roubin R, Acquaviva C, Chevrier V, Sedjai F, Zyss D, Birnbaum D, et al. Myomegalin is necessary for the formation of centrosomal and Golgi-derived microtubules. Biol Open. 2013;2(2):238–50.
37. Auer PL, Nalls M, Meschia JF, Worrall BB, Longstreth WT Jr, Seshadri S, et al. Rare and coding region genetic variants associated with risk of ischemic stroke: the NHLBI exome sequence project. JAMA Neurol. 2015;72(7):781–8.
38. Ward-Caviness CK, Agha G, Chen BH, Pfeiffer L, Wilson R, Wolf P, et al. Analysis of repeated leukocyte DNA methylation assessments reveals persistent epigenetic alterations after an incident myocardial infarction. Clin Epigenet. 2018;10(1):161.
39. Li L, Pan Z, Yang X. Key genes and co-expression network analysis in the livers of type 2 diabetes patients. J Diabetes Investig. 2019;10(4):951–62.
40. Juhanson P, Kepp K, Org E, Veldre G, Kelgo P, Rosenberg M, et al. N-acetyltransferase 8, a positional candidate for blood pressure and renal regulation: resequencing, association and in silico study. BMC Med Genet. 2008;9:25.
41. Pliss L, Timsa L, Rootsi S, Tambets K, Pelnena I, Zole E, et al. Y-chromosomal lineages of latvians in the context of the genetic variation of the Eastern-Baltic region. Ann Hum Genet. 2015;79(6):418–30.
42. Kottgen A, Pattaro C, Boger CA, Fuchsberger C, Olden M, Glazer NL, et al. New loci associated with kidney function and chronic kidney disease. Nat Genet. 2010;42(5):376–84.
43. Simone B, De Stefano V, Leoncini E, Zacho J, Martinelli I, Emmerich J, et al. Risk of venous thromboembolism associated with single and combined effects of Factor V Leiden, Prothrombin 20210A and Methylenetethraydrofolate reductase C677T: a meta-analysis involving over 11,000 cases and 21,000 controls. Eur J Epidemiol. 2013;28(8):621–47.
44. Casas JP, Hingorani AD, Bautista LE, Sharma P. Meta-analysis of genetic studies in ischemic stroke: thirty-two genes involving approximately 18,000 cases and 58,000 controls. Arch Neurol. 2004;61(11):1652–61.
45. Lindstrom S, Wang L, Smith EN, Gordon W, van Hylckama VA, de Andrade M, et al. Genomic and transcriptomic association studies identify 16 novel susceptibility loci for venous thromboembolism. Blood. 2019;134(19):1645–57.
46. Martinez EC, Passariello CL, Li J, Matheson CJ, Dodge-Kafka K, Reigan P, et al. RSK3: a regulator of pathological cardiac remodeling. IUBMB Life. 2015;67(5):331–7.
47. Ding Y, Liu K, Wang Y, Su G, Deng H, Zeng Q, et al. Expression and significance of fgl2 prothrombinase in cardiac microvascular endothelial cells of rats with type 2 diabetes. J Huazhong Univ Sci Technolog Med Sci. 2010;30(5):575–81.
48. Hearn T. ALMS1 and Alstrom syndrome: a recessive form of metabolic, neurosensory and cardiac deficits. J Mol Med (Berl). 2019;97(1):1–17.
49. Meng W, Veluchamy A, Hebert HL, Campbell A, Colhoun HM, Palmer CNA. A genome-wide association study suggests that MAPK14 is associated with diabetic foot ulcers. Br J Dermatol. 2017;177(6):1664–70.
50. Iyengar SK, Sedor JR, Freedman BI, Kao WH, Kretzler M, Keller BJ, et al. Genome-wide association and trans-ethnic meta-analysis for advanced diabetic kidney disease: family investigation of nephropathy and diabetes (FIND). PLoS Genet. 2015;11(8):e1005352.
51. Zoungas S, Woodward M, Li Q, Cooper ME, Hamet P, Harrap S, et al. Impact of age, age at diagnosis and duration of diabetes on the risk of macrovascular and microvascular complications and death in type 2 diabetes. Diabetologia. 2014;57(12):2465–74.

Publisher's Note

4. DISCUSSION

The increasing prevalence of T2DM, a chronic disease associated with risk of long-term complications and premature mortality, requires continuous management and therefore has long become a serious public health concern causing major health expenditures globally (van Dieren *et al.* 2010). Since the long-term complications, especially microvascular disease, have been directly related to poor glycemic control, diabetic complications may be prevented or at least delayed to some extent by early intervention (Stolar 2010). Despite the notable variety of available glucose-lowering agents, around 37% of patients still do not maintain a glycaemic control, which may be improved by implying individualized therapeutic strategies (de Pablos-Velasco *et al.* 2014), therefore many efforts have been made to discover reliable clinical and molecular signatures for the early prediction of anti-diabetic therapeutic efficacy and side-effects (Sattar 2012). The collection of different types of omics data in well-designed clinical studies has already highlighted the complex interacting pathways that are underlying the pathophysiology of the disease, and are still providing the necessary information to understand the factors determining the therapeutic response meanwhile promoting the personalization of diabetes care (Pearson 2016). Due to the promising evidence of beneficial properties of metformin in other non-metabolic disorders (Saraei *et al.* 2019) and high variability in the therapeutic response (Rashid *et al.* 2019; Sportelli *et al.* 2020), intensive research has been done to gain a clear understanding of its effects resulting in 10 753 research papers published in the PubMed database (National Library of Medicine, https://pubmed.ncbi.nlm.nih.gov/) during the last 5 years.

4.1 Immune-modulatory effects of metformin

Although multiple ongoing phase I clinical trials aim to estimate the bioavailability, bioequivalence of different formulations of metformin hydrochloride and the pharmacokinetics of various combination therapies (https://www.clinicaltrialsregister.eu/), longitudinal omics-based studies on healthy individuals with no background of hyperglycemia describing systemic effects of the drug are still lacking. We were the first ones revealing signatures of metformin in whole-blood transcriptomics and gut metagenomics data from 25 healthy subjects within the clinical trial MIKROMET16001. Exploratory RNA-Seq analysis in healthy subjects showed significant enrichment of immunity-related pathways, including the Cytokine-cytokine receptor interaction pathway already 10 hours after the first metformin dose, supporting the previously described ability of metformin to suppress inflammatory cytokines at the mRNA level (Cameron *et al.* 2016). Significant metformin-induced enrichment of the same pathway was recently also found in *P. gingivalis* lipopolysaccharide-treated cells, serving as a simulation of

periodontitis (Tan *et al.* 2020). The ability of metformin to modify immune responses at mRNA level was also recently confirmed in the *in vivo* study of 11 healthy individuals by Lachmandas *et al*., reporting the anti-inflammatory properties of metformin by downregulation of IFN-α, IFN-γ, and TNF-α particularly following the stimulation of human peripheral blood mononuclear cells (PBMCs) with *M. tuberculosis.* Together with *in vitro* data authors proved that metformin inhibits the *M. tuberculosis*–induced type 1 IFN response and inflammation in PBMCs (Lachmandas *et al.* 2019). These recent findings of immune-modulatory properties of metformin in different disease models and also healthy subjects with no infections reported at the time of sample collection pinpoint the possible non-specificity of these transcriptional responses, meanwhile complementing the knowledge of pleiotropic effects of metformin irrespective of diabetes status.

The longitudinal blood transcriptome profiling of the T2DM patient cohort showed transcriptional regulation as a mechanism exerting the well-known properties of metformin, including glucose homeostasis, impact on immune responses, and tumor prevention. Our data shows metformin-induced downregulation of another immunity-related gene coding for C-C Motif Chemokine Receptor 2 (*CCR2*) among the top hits. Targeting *CCR2* expression may explain the pleiotropic effect of metformin since CCR2 serves as a receptor of monocyte chemoattractant proteins and a key driver of monocyte/macrophage trafficking with a significant role in tumor pathogenesis (Hao *et al.* 2020) and inflammation of adipose tissue (J. Kim *et al.* 2016). Interestingly, the study of Kim *et al*. showed no effect of Ccr2 deficiency on several metabolic traits including insulin sensitivity, while the mouse model of diet-induced obesity Ccr2+/+ showed higher insulin resistance with significant improvement after treatment with CCR2 antagonist (Weisberg *et al.* 2006). The evidence of the ability of CCR2 deficiency to attenuate the obesity-induced changes in adipose tissue gene expression (e.g. Tnfa) suggests that the beneficial effect of energy metabolism observed in metformin therapy may be at least partly achieved by downregulation of *CCR2*.

4.2 The implication of secretory immunoglobulin A - mediated immunity

Multiple studies show the contrasting impact of metformin on the complex network on the immune system from protective anti-inflammatory effect in autoimmune (S. Y. Lee *et al.* 2017) and infectious diseases including COVID-19 (X. Chen *et al.* 2020) to immune activation in malignancies (Saito *et al.* 2020). So far we are the first ones reporting metformin involvement in immune responses via the ability to enhance sIgA function in humans evidenced by the enrichment of intestinal immune network for IgA production pathway in gene expression data and significantly increased fecal sIgA levels in healthy subjects. The relationship between

metformin administration and sIgA production was later also observed in the cohort of T2DM patients (data not published) measured before administration of metformin (median=18270.03μg/ml, IQR=47353.55μg/ml), after metformin therapy for 7 days (median=20661.78μg/ml, IQR=31431.56μg/ml) and after metformin therapy for 3 months (median=34676.19μg/ml, IQR=41876.33μg/ml), although it did not reach the statistical significance, which may be explained by insufficient sample size for such a long observation period and increased probability of the presence of uncontrolled time-varying confounders. A strong association between metabolic diseases and IgA functions has been recognized already before. Oikawa *et al*. showed that diabetes mellitus reduces sIgA secretion rates in saliva which may explain the susceptibility to infections in diabetes patients (Oikawa *et al.* 2015). Moreover, Brown and colleagues discovered altered relative abundance in the sIgA coated fraction of gut and oral microbiome between normoglycemic, prediabetic, and diabetic samples distinct from the overall microbiome samples and an association of *Escherichia_Shigella* and *Pseudomonas* (*Proteobacteria*) with IgA+ and IgA- fractions respectively was demonstrated, suggesting a significant and disease-specific role of sIgA in shaping the composition of the microbiome (Brown *et al.* 2020). IgA plays a critical role in modulating the composition of gut microbiota, and IgA dysfunction is linked to an overrepresentation of *Proteobacteria* and metabolic syndrome including obesity (Guo *et al.* 2021). Moreover, the comprehensive animal study published by Luck *et al*. showed that IgA-coated immune cells modulate the glucose regulation in a mouse model of diet-induced obesity, and IgA deficiency not only induces dysbiosis but also worsens insulin resistance via inflammation of visceral adipose tissue and contribution of gut microbiota. The particular study also proved the ability of metformin to prevent the high fat diet-induced reduction of IgA producing cells and increase the levels of fecal sIgA (Luck *et al.* 2019).

To investigate the true metformin effects and underlying mechanisms under controlled conditions, we implemented a well-designed animal experiment involving the development of a high-fat diet-induced type 2 diabetes mouse model followed by ten weeks-long metformin treatment in specific pathogen-free conditions, which was focused on metformin-induced alterations in mice gut microbiome composition (Silamikele *et al.* 2021) and transcriptome profiles in the target tissue (data not published). Similar to Luck and colleagues we observed significantly reduced baseline levels of fecal sIgA in the high-fat diet-fed animals compared to control diet-fed animals (*p*-value=0.036), followed by a notable though not statistically significant increment of fecal sIgA levels after metformin intervention in both control diet-fed animals (increase from median=60.45μg/ml, IQR=29.04μg/ml to median=64.03μg/ml, IQR=32.75μg/ml) and high-fat diet-fed animals (from median=38.11μg/ml, IQR=69.21 μg/ml

to median=48.92μg/ml, IQR=6.20 μg/ml). Nevertheless, we also observed a sIgA gain in control animals not receiving any antidiabetic treatment or diet shift pointing to other overlooked sIgA-modulating factors such as aging itself (Diebel *et al.* 2011; Senda *et al.* 1988). Although Luck and colleagues did not observe any difference in IgA+ cell counts in the small bowel of high-fat diet-fed mice after metformin treatment in drinking water, our RNA-Seq data analysis of the distal part of the small intestine exhibited significant differential expression of 46 genes, including the gene coding for immunoglobulin heavy variable V1-7 (Ighv1-7, logFC = 5.76, FDR=0.04) when comparing samples obtained from high-fat diet-fed animals receiving metformin with the ones from high-fat diet-fed animals not receiving any antidiabetic therapy. Moreover, we noticed the Intestinal immune network for IgA production among the top enriched pathways in the same comparison, though it did not reach statistical significance (*p*-value = 0.084).

4.3 Subject specificity in MODY and cholesterol homeostasis-related transcriptional effects of metformin

The transcriptome profiles of blood cells obtained from the healthy cohort indicated the subject-specific effect of metformin, with metformin-induced overexpression of MODY-related genes in two individuals among the most surprising findings. Although multiple genetic studies have identified common risk alleles for the development of MODY, T2DM, and other glycemic traits (Holmkvist *et al.* 2008; Winckler *et al.* 2007), only a few studies have addressed the modifying effect of MODY-gene variants in anti-diabetic therapy. For instance, a randomized clinical trial revealed two variants in MODY genes (rs3212185, *HNF4A*, and rs6719578, *NEUROD1*) significantly influencing metformin response in individuals with high T2DM risk (Billings *et al.* 2017). Metformin-induced differential expression of MODY genes has been previously described in hypertensive rats (Malinska *et al.* 2016), nevertheless, we are the first ones reporting a selectively altered expression of these genes following metformin intervention in humans.

Among the subject-specific effects of metformin observed in the cohort of healthy individuals, we also noticed a strong downregulation of three genes (*APOB*, *LDLR*, *PCSK9*) involved in cholesterol homeostasis after 10h and 7 days of metformin intervention in one individual only. The implication of *PCSK9* was recently explored by Hu and colleagues, who proved that metformin exerts its cholesterol-lowering effect by transcriptional downregulation of *PCSK9* in a dose and time-dependent as well as AMPK-independent manner. Moreover, these data showed that *PCSK9* downregulation, which was also noticed in hepatic mRNA levels, is related to the metformin-induced reduction of intracellular glucose levels and

specifically targeting the ChREBP glucose sensor. In the study of Hu *et al.,* the reduction of circulating *PCSK9* due to metformin therapy was proved in non-diabetic dyslipidemia patients (D. Hu *et al.* 2021). Although in our study there was no history of dyslipidemia and use of lipid-lowering medications reported before, and the low-density lipoprotein cholesterol levels were near-optimal (2.36 mmol/l) at the time of application in the clinical trial for the subject showing metformin-induced downregulation of cholesterol modulatory genes, the obtained RNA-Seq data support the hypothesis of metformin-induced *PCSK9* downregulation at least in a case-specific manner. One may speculate that metformin exerts insulin secretagogue and cholesterol-lowering properties only in a subgroup of metabolically compromised individuals, though substantiation of such a hypothesis should be done by additional studies.

The whole-blood RNA-Seq analysis in our T2DM patient cohort did not highlight the subject-specific effects of metformin, though there was a clear universal impact on the mRNA levels of other cholesterol-related genes, such as *LRP1* and *SLC46A1*. Reduced surface LRP1 protein levels have been reported in N2a neuroblastoma cells treated with metformin and downregulation of *LRP1* gene has also been shown in liver tissue of mouse model for human-like lipoprotein metabolism (Y. Chen *et al.* 2009; Geerling *et al.* 2014). Nevertheless, Lu and colleagues reported a negligible influence of metformin on *LRP1* expression in brain tissue of transgenic mouse model of Alzheimer's disease, suggesting a distinct effect of the drug that may be tissue-specific (Y. Meng *et al.* 2020), depending on the duration of the therapy or even backgrounding disease (X. Y. Lu *et al.* 2020).

4.4 Large-scale gene expression profiling reveals biological markers for early stratification in different metformin response groups

Although the first attempts to predict the metformin response dates back to the 90ties (Hermann *et al.* 1994), there are still no promising biomarkers currently used in a clinical setting. Multiple studies have applied metabolomics (Park *et al.* 2018), genomics (K. Zhou *et al.* 2016), and metagenomics (Elbere *et al.* 2020) data to predict metformin response. In our study, two different approaches, the likelihood ratio test and PLS-DA were used for the identification of transcriptional differences in metformin responders and non-responders and finding potential predictive biomarkers. Whole-blood transcriptional profiling of T2DM patients discriminated in two metformin response-based groups showed differential expression of genes in both time points analyzed by the likelihood ratio test (27 DEGs before metformin administration, 15 DEGs after therapy for 3 months), facilitating the idea of early whole-blood transcriptome-based stratification of T2DM patients in different metformin efficacy groups and revealing transcriptomic factors affecting the achievement of glycaemic goals. The well-known

insulin receptor *IRS-2* showed significant upregulation between responders vs non-responders before any antidiabetic therapy. Although the main site of metformin action, involving the PI3K (phosphoinositide 3-kinase) pathway, promoting the reduction of insulin resistance, is the liver (Yuan *et al.* 2003), our data suggest the prior activity of insulin signaling transduction genes in blood cells that may be associated to metformin response.

Twelve genes representing a subgroup of small non-coding RNAs named small nucleolar RNAs prevailed in the list of differentially expressed genes acquired in the comparison of responders against non-responders before antidiabetic therapy. These molecules guide chemical modifications of structural RNAs, and although the functional significance of small nucleolar RNAs in the pathogenesis of human diseases is only beginning to emerge, several similar markers have already been paid more attention in metabolic diseases, for instance, SNORD115 (small nucleolar RNA, C/D box 115-1) and SNORD116 (small nucleolar RNA, C/D box 116 cluster) in respect to obesity and metabolic syndrome-related Prader-Willi syndrome (Galiveti *et al.* 2014). Similarly, a study conducted by Rimer *et al*. shows that inflammation promotes the vesicle-mediated release of specific Rpl13a (ribosomal protein L13a) snoRNAs (U32a, U33, U34, and U35a) that are ensuring protein expression and cellular growth by targeting nascent ribosomal RNAs for 2′-*O*-methylation. The authors managed to prove that these molecules can travel through the circulation, may be uptaken by other cells, and function in distant tissues (J. Lee *et al.* 2016), making small nucleolar RNAs both promising predictive biomarkers and future therapeutic agents.

One of the main discoveries was the differential expression of mitochondrial genes (*MT-ATP6, MT-ND2, MT-ND4, MT-ND4L, MT-ND6*) comparing metformin responders against non-responders after metformin therapy for 3 months, which was also reflected as enrichment of oxidative phosphorylation pathway. In addition, we managed to prove a clear positive association between *MT-ND4* and *MT-ND4L* expression and HbA1c levels. Although the inhibition of mitochondrial respiratory chain complex 1 is a well-known target of the drug, only a few studies are addressing this mechanism at the mRNA level. Recently Lord and colleagues described the ability of metformin to target the oxidative phosphorylation pathway in the whole-transcriptome RNA-Seq data of breast cancer tissue (Lord *et al.* 2020). In addition, Yang and colleagues described the ability of metformin to change the transcriptomic profile of mitochondrial genes targeting the mitochondrial calcium processes, mitochondrial protein quality control, and downregulating mitochondrial DNA replication and translation in the intestinal cultures (Yang *et al.* 2021). To the best of our knowledge, we are the first ones suggesting the downregulation of mitochondrial genes including the ones coding for the subunits of NADH: ubiquinone oxidoreductase core, which results in the enrichment of

oxidative phosphorylation pathway, as not only one of the key mechanisms of metformin action, but also the driving force determining better glycemic response to metformin.

The whole-blood RNA-Seq data obtained from a relatively small group of well-characterized T2DM patients incorporated in the PLS-DA in our study showed clear transcriptome-based discrimination of metformin responders and nonresponders prior administration of any antidiabetic therapy, mediated by the differential expression of 56 genes. Although the model needs additional validation in a larger cohort, we managed to explain 13.9% of the variability in metformin response. Recently in collaboration with colleagues from Lund University, we implemented a blood-based epigenetic marker analysis revealing 11 differentially methylated sites predicting metformin efficacy and 4 sites ensuring discrimination for the future tolerance of the drug (Park *et al.* 2018). Together, after careful validation, these markers explained even 19% to 73% of the variation in glycemic response depending on the cohort. Although in the report of Park *et al.* we managed to confirm the differential methylation of several blood-based markers also in adipocytes and transcriptional functionality of several CpG sites on the metformin transporter genes in cultured hepatocytes, the differentially expressed genes originating from the RNA-Seq data comparison of responders and non-responders described in the particular thesis were not overlapping with the top CpG site-associated genes described previously by Park *et al.*, probably due to tissue specificity. In addition to the recently verified integration of epigenetic modifications, strong evidence implicates also the gut microbiota as a site of action of metformin. Thus, we have recently applied the shot-gun metagenomic sequencing-based microbial abundance patterns to predict the glycemic response to metformin in T2DM patients. With an analogous approach for highlighting discriminative features as described in the RNA-Seq dataset, we managed to explain 5.4% of the variance by 43 taxonomic groups (Elbere *et al.* 2020). Although both, the RNA-Seq and metagenomic sequencing data-based prediction models must be validated in a larger cohort involving the omics data integration techniques, so far both of the applications provide a notable contribution towards the development of personalized antidiabetic therapy.

4.5 Genetic loci shaping the risk of T2DM complications

The T2DM care and dealing with long-term complications has attracted even more attention during the COVID-19 pandemic since T2DM patients show poorer clinical outcomes and higher mortality rates after infection of SARS-CoV-2 (You *et al.* 2020), therefore early action is needed to address the clinical consequences of the long-term manifestation of T2DM especially when the risk of infection is elevated globally. GWAS studies have provided novel insights on the etiology of diabetic complications, nevertheless a universal set of biomarkers

allowing early patient stratification is still not developed. We managed to highlight 10 genetic loci showing a genome-wide significance for the association with one of four T2DM complication types studied (neuropathy, nephropathy, ophthalmopathy, macrovascular complications). Although among the reported genetic susceptibility studies of diabetic complications, GWAS of diabetic retinopathy has appeared to be the most productive in terms of significant hits (Buniello *et al.* 2019), the majority of the loci identified in our cohort were predicting the risk of macrovascular complications. So far, the traditional risk factors for macrovascular complications are elevated low-density lipoprotein cholesterol levels or blood pressure, bearing a significant genetic determination themselves (Reilly *et al.* 2011; R. C. Turner *et al.* 1998). We were the first ones reporting the association of allelic variants of the gene coding for Phosphodiesterase 4D Interacting Protein (*PDE4DIP*) also known as cardiomyopathy-associated protein 2 with macrovascular complications. Interestingly, deletions in this locus have been previously linked to atrial fibrillation phenotype and more severe and complex manifestations of Alström syndrome which is strongly related to cardiovascular complications including stroke (Abou Ziki *et al.* 2021; Lombardo *et al.* 2020). Similarly, we were the first ones proposing the role of coagulation factor V coding gene *F5* in the risk of macrovascular complications, while the rest of the loci identified in the particular analysis reside in gene regions that were not previously suspected in the pathogenesis of macrovascular events.

Diabetic neuropathy is the most common persistent manifestation among T2DM patients. Nevertheless, the exact cause of the disease, as well as the risk factors, are poorly understood (X. Liu *et al.* 2019b). Multiple genetic loci in a close association with the disease have been reported in the proximity of the genes involved in key molecular pathways such as oxidative stress (catalase, *CAT;* glutathione peroxidase 1, *GPX1;* nitric oxide synthase 3, *NOS3;* aldose reductase, *ALR2)* and neurovascular impairments *(*vascular endothelial growth factor, *VEGF*; angiotensin I converting enzyme, *ACE*) (Jankovic *et al.* 2021). While the implication of gene coding for glycophorin A (*GYPA*), an intrinsic membrane protein of erythrocytes, in the context of neurodegenerative complications is so far described only at the level of mRNA (Cheng *et al.* 2015; Hur *et al.* 2015; Kobayashi and Zochodne 2018) or via copy number variation (Zhao *et al.* 2012), our findings suggest that allelic variations of the gene may also modulate the risk of neuropathies in T2DM patients. Our data showed one genetic association with ophthalmic complications (*CCDC146*: rs3095447), which overlaps with the results of the macrovascular complication group possibly due to similar etiologic characteristics between both diseases. Although there were no significant hits at the genome-wide significance level in the analysis of diabetic nephropathy, we managed to replicate strong associations for diabetic neuropathy

(*MAPK14*: rs3761980, rs80028505), and diabetic nephropathy (*APOL1*: rs136161), which proved the contribution of these risk loci in the pathogenesis of diabetic complications across various populations.

By using The Genotype-Tissue Expression data we found genes that are affected by the expression quantitative trait loci, suggesting the possible functional implication of the variants in the pathogenesis of the disease also at the mRNA level. We found a negative association between rs3095447 and *FGL2* expression in the heart and artery, while rs4852954 was linked to lower *ALMS1* levels in whole blood and skin tissue. Both of these genes have been functionally related to cardiovascular phenotypes before (Ding *et al.* 2010; Hearn 2019).

4.6 Limitations of the study

The main limitation of our study is the relatively small sample size of the cohorts studied, which is mainly caused by limited attainability of individuals corresponding the strictly defined inclusion criteria involving metformin monotherapy for at least 3 months in the T2DM cohort or giving consent for a week-long metformin intervention in the cohort of healthy individuals. The limitation of sample size is especially pronounced in transcriptomic datasets, where it is overcome by applying the longitudinal study design. The introduction of repeated measures of samples obtained from the same individual provides higher statistical power, minimizes the potential interference of individual-level confounding variables such as age and sex, meanwhile assuring the possibility to detect subject-specific effects (N. Lu *et al.* 2013).

Another aspect one may consider as a limitation is the lack of conventional controls. In the metformin study, incorporation of the blood RNA-Seq data of healthy controls and T2DM patients with no metformin intervention would reflect the innate transcriptomic fluctuations, which could be further used to reduce the risk of type I errors and therefore prove the causality of the observed effects. We believe that the longitudinal approach allows at least partially to control for the intra-individual variation in transcriptomic data. Moreover, to focus on true metformin effects, the lists of significant DEGs were carefully inspected for the possible presence of genes driving the manifestations of circadian rhythms. In the GWAS study, the incorporation of healthy subjects was not considered. Since the main objective of this study was to identify genetic risk loci for T2DM complications, the control subjects were carefully selected as T2DM patients with no evidence of the complication of interest during the follow-up period. In this case, the control group does not reflect the general healthy population, since the implementation of such controls would rather reveal genetic associations with the diagnosis of T2DM itself, not the T2DM complications. A similar approach has been widely used in GWAS before (C. Liu *et al.* 2019a; W. Meng *et al.* 2018).

It is well-known that metformin exerts different effects for various diseases, which may be the reason for the difference in metformin-induced blood cell transcriptome profiles between healthy subjects and T2DM patients. Nevertheless, the difference in the duration of the therapy and the variable dose of metformin between both cohorts studied may, to a certain degree, also explain the inconsistency in the observed metformin effects between both cohorts studied. We believe that rigorous synchronization of the study designs in terms of the duration of metformin intervention and the regimen of metformin therapy would reveal more overlapping transcripts showing metformin-induced differential expression. However, since the clinically unreasonable use of any drug poses a risk to human health, we decided to focus on short-term effects of metformin in the cohort of healthy subjects, and investigate long-term response in the T2DM patient cohort, meanwhile leaving the decision on the antidiabetic treatment strategy to the endocrinologists regardless of the research objectives.

4.7 Future perspectives

There are a number of gaps in our knowledge about the underlying mechanisms of both, diabetes complications and metformin therapy, that would benefit from further research. Comprehensive data obtained during this study together with the data obtained from other ongoing T2DM-related scientific projects raises a number of opportunities for future research. Validation of the gene expression-based classification and prediction model in a larger cohort of T2DM patients may be the first rational step towards the identification of relevant biomarkers for the therapeutic efficacy of metformin. In addition, integration of the available multi-omics data (e.g. gut microbiome, genomics, epigenomics data) may be beneficial, allowing more holistic investigation of the underlying mechanisms. In-depth *in vivo* analysis of metformin-induced transcriptional alterations of the target tissue of metformin such as muscles and liver may reveal the general information about the extent to which the whole-blood transcriptomics data are reflecting the gene expression profiles of the main target tissue. Here, the single-cell RNA sequencing and even spatial transcriptomics approach may be applied to simultaneously measure the expression level of either the entire genome or specific marker genes, ensuring the careful characterization of transcriptome variability in subpopulations of cells. Moreover, to examine the functionality of specific genes in modifying the metformin response additional animal experiments may be implemented involving the RNA interference technique.

Meanwhile, for the GWAS of T2DM complications validation of the candidate variants in a larger, independent cohort is essential. Since the development of a phenotypically similar cohort involving T2DM patients with different types of complications is challenging, it may be reasonable to use some of the existing major massive databases such as the UK Biobank

resource. Moreover, the ongoing '1+ Million Genomes' Initiative has promoted the establishment of a national genomic reference cohort by using whole-genome sequencing data, which will address the issue of population specificity by improving imputation quality and therefore mitigating the bias of linkage disequilibrium differences between different populations. After careful validation of the significant loci, the genotyping data may be used to calculate polygenic risk scores and estimate the lifetime risk of T2DM complications at an individual level.

5. CONCLUSIONS

1. In healthy individuals, metformin-induced alterations of global blood cell gene expression profiles are associated with immune responses, while subject-specific effects are involved in metabolic pathways.

2. There is a strong association between metformin and the intestinal immune system, which may be guided by induction of immune network for IgA production pathway resulting in elevated secretory IgA levels in stool samples.

3. Metformin-specific signatures in blood cell transcriptome profiles of T2DM patients point out some of its well-known properties: the ability to improve energy metabolism, influence immune responses, and inhibit cancer progression.

4. Early blood cell gene expression-based T2DM patient stratification into different metformin response groups may facilitate future biomarker discoveries and clinical applications.

5. Altered expression of genes coding for NADH: ubiquinone oxidoreductase core subunits and enrichment of oxidative phosphorylation may be associated with metformin efficacy in T2DM patients.

6. We have identified ten novel allelic variants associated with T2DM complications in the population of Latvia, including five variants located nearby genes (*GYPA*, *PDE4DIP*, *NAT8*, *F5*, *RPS6KA2*) that have been functionally linked to the pathogenesis of the diseases before.

7. Causal genetic effect of allelic variants in *MAPK14* (rs3761980, rs80028505) in diabetic neuropathy and *APOL1* (rs136161) in diabetic nephropathy is consistent across different populations.

6. THESIS

1. Metformin has a strong and immediate effect on the transcriptome of peripheral blood cells showing interindividual gene expression variability.

2. Metformin universally induces alterations in the expression of genes regulating immune responses resulting in elevated fecal sIgA levels.

3. Estimation of the transcriptome-based biomarkers at the time of T2DM diagnosis enables an early stratification of T2DM patients in different metformin response groups.

4. Ten novel allelic variants are associated with the development of T2DM complications in the population of Latvia, suggesting new candidate genes for further functional studies.

7. PUBLICATIONS

1. **Ustinova M**, Silamikelis I, Kalnina I, Ansone L, Rovite V, Elbere I, Radovica-Spalvina I, Fridmanis D, Aladyeva J, Konrade I, Pirags V, Klovins J. Metformin strongly affects transcriptome of peripheral blood cells in healthy individuals. PLoS One. 2019 Nov 8;14(11):e0224835. doi: 10.1371/journal.pone.0224835.

2. **Ustinova M**, Ansone L, Silamikelis I, Rovite V, Elbere I, Silamikele L, Kalnina I, Fridmanis D, Sokolovska J, Konrade I, Pirags V, Klovins J. Whole-blood transcriptome profiling reveals signatures of metformin and its therapeutic response. PLoS One. 2020 Aug 11;15(8):e0237400. doi: 10.1371/journal.pone.0237400.

3. **Ustinova M**, Peculis R, Rescenko R, Rovite V, Zaharenko L, Elbere I, Silamikele L, Konrade I, Sokolovska J, Pirags V, Klovins J. Novel susceptibility loci identified in a genome-wide association study of type 2 diabetes complications in population of Latvia. BMC Med Genomics. 2021 Jan 11;14(1):18. doi: 10.1186/s12920-020-00860-4.

8. APPROBATION OF RESEARCH

1. M. Ustinova, I. Silamiķelis, I. Kalniņa, L. Ansone, V. Rovīte, I. Elbere, I. Radoviča-Spalviņa, D. Fridmanis, J. Aladyeva, I. Konrāde, V. Pīrāgs, J. Kloviņš. *Metformin-induced alterations of human blood cell transcriptome*, FEBS3+, Riga, Latvia, 2019. Oral presentation.

2. I. Konrade, M. Ustinova, R. Peculis, V. Rovite, L. Zaharenko, J. Klovins. *GWAS for type 2 diabetes complications identifies loci associated with neuropathy.* International Diabetes Federation Congress 2019, Busan, South Korea, 2019. Poster presentation, Abstract book.

3. M. Ustinova, I. Silamiķelis, I. Kalniņa, L. Ansone, V. Rovīte, I. Elbere, I. Radoviča-Spalviņa, D. Fridmanis, J. Aladyeva, I. Konrāde, V. Pīrāgs, J. Kloviņš. *Metformīna ierosinātas izmaiņas gēnu ekspresijas profilos veselās personās.* 76th international scientific conference of the University of Latvia, (Scientific session of the scholarship holders supported by the patrons of the University of Latvia Foundation), Riga, Latvia, 2019. Oral presentation.

4. M .Ustinova , I. Silamiķelis , I. Elbere , I. Kalniņa, L. Zaharenko , R. Pečulis , I. Konrāde, V. Pīrāgs, J. Kloviņš. *Metformin-induced alterations of transcriptome profile in healthy individuals*. European Association for the Study of Diabetes Annual Meeting 2018, Berlin, Germany, 2018. Poster presentation, Diabetologia: 61, S1, S341-S341.

5. M .Ustinova, I. Silamiķelis, I. Elbere, I. Kalniņa, L. Zaharenko, R. Pečulis, I. Konrāde, V. Pīrāgs, J. Kloviņš. *Metformin induces alterations in transcriptome profile of helathy individuals*. European Asssociation of Human Genetics Conference 2018, Milan, Italy, 2018. Poster presnetation, Eur J Hum Genet 27, 1–688 (2019).

6. M .Ustinova, I. Silamiķelis, I. Elbere, I. Kalniņa, L. Zaharenko, R. Pečulis, I. Konrāde, V. Pīrāgs, J.Kloviņš. *Metformīna ierosinātas izmaiņas gēnu ekspresijas profilos veselās personās*, 76th international scientific conference of the University of Latvia, Section of Molecular Biology, Riga, Latvia, 2018. Oral presentation.

9. ACKNOWLEDGEMENTS

This research was supported by the European Regional Development Fund under the project “Investigation of interplay between multiple determinants influencing response to metformin: search for reliable predictors for efficacy of type 2 diabetes therapy” (Project No. 1.1.1.1/16/A/091).

I would like to thank all of the study participants for their contribution and acknowledge the Genome Database of the Latvian Population for organizing patient recruitment, providing biological material and associated data.

I would like to express my sincere gratitude to my supervisor Prof. Jānis Kloviņš for his invaluable advice, continuous support, and experience shared during the implementation of this work. I am also grateful to my colleagues in the Latvian Biomedical Research and Study Centre for their helpful contributions and practical suggestions, especially Ilze Elbere for close and productive collaboration.

Finally, I would like to thank my husband, family, and friends for their tremendous understanding, encouragement, and patients all through my studies.

REFERENCES

Abou Ziki, M. D., Bhat, N., Neogi, A., Driscoll, T. P., Ugwu, N., Liu, Y., Smith, E., Abboud, J. M., Chouairi, S., Schwartz, M. A., Akar, J. G., and Mani, A. (2021), 'Epistatic interaction of PDE4DIP and DES mutations in familial atrial fibrillation with slow conduction', *Hum Mutat,* 42 (10), 1279-93.

Ali, O. (2013), 'Genetics of type 2 diabetes', *World J Diabetes,* 4 (4), 114-23.

Almgren, P., Lehtovirta, M., Isomaa, B., Sarelin, L., Taskinen, M. R., Lyssenko, V., Tuomi, T., Groop, L., and Botnia Study, Group (2011), 'Heritability and familiality of type 2 diabetes and related quantitative traits in the Botnia Study', *Diabetologia,* 54 (11), 2811-9.

Anonymous (1998), 'Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). UK Prospective Diabetes Study (UKPDS) Group', *Lancet,* 352 (9131), 837-53.

Arnott, C., Li, Q., Kang, A., Neuen, B. L., Bompoint, S., Lam, C. S. P., Rodgers, A., Mahaffey, K. W., Cannon, C. P., Perkovic, V., Jardine, M. J., and Neal, B. (2020), 'Sodium-Glucose Cotransporter 2 Inhibition for the Prevention of Cardiovascular Events in Patients With Type 2 Diabetes Mellitus: A Systematic Review and Meta-Analysis', *J Am Heart Assoc,* 9 (3), e014908.

Athreya, A. P., Kalari, K. R., Cairns, J., Gaglio, A. J., Wills, Q. F., Niu, N., Weinshilboum, R., Iyer, R. K., and Wang, L. (2017), 'Model-based unsupervised learning informs metformin-induced cell-migration inhibition through an AMPK-independent mechanism in breast cancer', *Oncotarget,* 8 (16), 27199-215.

Awata, T., Yamashita, H., Kurihara, S., Morita-Ohkubo, T., Miyashita, Y., Katayama, S., Mori, K., Yoneya, S., Kohda, M., Okazaki, Y., Maruyama, T., Shimada, A., Yasuda, K., Nishida, N., Tokunaga, K., and Koike, A. (2015), 'Correction: A Genome-Wide Association Study for Diabetic Retinopathy in a Japanese Population: Potential Association with a Long Intergenic Non-Coding RNA', *PLoS One,* 10 (4), e0126789.

Azharuddin, M., Adil, M., Ghosh, P., and Sharma, M. (2018), 'Sodium-glucose cotransporter 2 inhibitors and fracture risk in patients with type 2 diabetes mellitus: A systematic literature review and Bayesian network meta-analysis of randomized controlled trials', *Diabetes Res Clin Pract,* 146, 180-90.

Azoulay, L. and Suissa, S. (2017), 'Sulfonylureas and the Risks of Cardiovascular Events and Death: A Methodological Meta-Regression Analysis of the Observational Studies', *Diabetes Care,* 40 (5), 706-14.

Baggio, L. L. and Drucker, D. J. (2007), 'Biology of incretins: GLP-1 and GIP', *Gastroenterology,* 132 (6), 2131-57.

Bansal, V., Kalita, J., and Misra, U. K. (2006), 'Diabetic neuropathy', *Postgrad Med J,* 82 (964), 95-100.

Basu, R., Chandramouli, V., Dicke, B., Landau, B., and Rizza, R. (2005), 'Obesity and type 2 diabetes impair insulin-induced suppression of glycogenolysis as well as gluconeogenesis', *Diabetes,* 54 (7), 1942-8.

Beck, T., Shorter, T., and Brookes, A. J. (2020), 'GWAS Central: a comprehensive resource for the discovery and comparison of genotype and phenotype data from genome-wide association studies', *Nucleic Acids Res,* 48 (D1), D933-D40.

Benjamini, Y. and Hochberg, Y. (1995), 'Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing', *Journal of the Royal Statistical Society, Series B (Methodological),* 57 (1).

Billings, L. K., Jablonski, K. A., Warner, A. S., Cheng, Y. C., McAteer, J. B., Tipton, L., Shuldiner, A. R., Ehrmann, D. A., Manning, A. K., Dabelea, D., Franks, P. W., Kahn, S. E., Pollin, T. I., Knowler, W. C., Altshuler, D., Florez, J. C., and Diabetes Prevention Program Research, Group (2017), 'Variation in Maturity-Onset Diabetes of the Young Genes Influence Response to Interventions for Diabetes Prevention', *J Clin Endocrinol Metab,* 102 (8), 2678-89.

Bonora, E., Formentini, G., Calcaterra, F., Lombardi, S., Marini, F., Zenari, L., Saggiani, F., Poli, M., Perbellini, S., Raffaelli, A., Cacciatori, V., Santi, L., Targher, G., Bonadonna, R., and Muggeo, M. (2002), 'HOMA-estimated insulin resistance is an independent predictor of cardiovascular disease in type 2 diabetic subjects: prospective data from the Verona Diabetes Complications Study', *Diabetes Care,* 25 (7), 1135-41.

Borgnakke, W. S. (2019), 'IDF Diabetes Atlas: Diabetes and oral health - A two-way relationship of clinical importance', *Diabetes Res Clin Pract,* 157, 107839.

Boulesteix, A. L. and Strimmer, K. (2007), 'Partial least squares: a versatile tool for the analysis of high-dimensional genomic data', *Brief Bioinform,* 8 (1), 32-44.

Brennan, E., McEvoy, C., Sadlier, D., Godson, C., and Martin, F. (2013), 'The genetics of diabetic nephropathy', *Genes (Basel),* 4 (4), 596-619.

Brown, E. L., Essigmann, H. T., Hoffman, K. L., Palm, N. W., Gunter, S. M., Sederstrom, J. M., Petrosino, J. F., Jun, G., Aguilar, D., Perkison, W. B., Hanis, C. L., and DuPont, H. L. (2020), 'Impact of Diabetes on the Gut and Salivary IgA Microbiomes', *Infect Immun,* 88 (12).

Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., Suveges, D., Vrousgou, O., Whetzel, P. L., Amode, R., Guillen, J. A., Riat, H. S., Trevanion, S. J., Hall, P., Junkins, H., Flicek, P., Burdett, T., Hindorff, L. A., Cunningham, F., and Parkinson, H. (2019), 'The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019', *Nucleic Acids Res,* 47 (D1), D1005-D12.

Burdon, K. P., Fogarty, R. D., Shen, W., Abhary, S., Kaidonis, G., Appukuttan, B., Hewitt, A. W., Sharma, S., Daniell, M., Essex, R. W., Chang, J. H., Klebe, S., Lake, S. R., Pal, B., Jenkins, A., Govindarjan, G., Sundaresan, P., Lamoureux, E. L., Ramasamy, K., Pefkianaki, M., Hykin, P. G., Petrovsky, N., Brown, M. A., Gillies, M. C., and Craig, J. E. (2015), 'Genome-wide association study for sight-threatening diabetic retinopathy reveals association with genetic variation near the GRB2 gene', *Diabetologia*, 58 (10), 2288-97.

Cade, W. T. (2008), 'Diabetes-related microvascular and macrovascular diseases in the physical therapy setting', *Phys Ther*, 88 (11), 1322-35.

Cameron, A. R., Morrison, V. L., Levin, D., Mohan, M., Forteath, C., Beall, C., McNeilly, A. D., Balfour, D. J., Savinko, T., Wong, A. K., Viollet, B., Sakamoto, K., Fagerholm, S. C., Foretz, M., Lang, C. C., and Rena, G. (2016), 'Anti-Inflammatory Effects of Metformin Irrespective of Diabetes Status', *Circ Res*, 119 (5), 652-65.

Cersosimo, E., Triplitt, C., Solis-Herrera, C., Mandarino, L. J., and DeFronzo, R. A. (2000), 'Pathogenesis of Type 2 Diabetes Mellitus', in K. R. Feingold, et al. (eds.), *Endotext* (South Dartmouth (MA)).

Chen, Y., Zhou, K., Wang, R., Liu, Y., Kwak, Y. D., Ma, T., Thompson, R. C., Zhao, Y., Smith, L., Gasparini, L., Luo, Z., Xu, H., and Liao, F. F. (2009), 'Antidiabetic drug metformin (GlucophageR) increases biogenesis of Alzheimer's amyloid peptides via up-regulating BACE1 transcription', *Proc Natl Acad Sci U S A*, 106 (10), 3907-12.

Chen, X., Guo, H., Qiu, L., Zhang, C., Deng, Q., and Leng, Q. (2020), 'Immunomodulatory and Antiviral Activity of Metformin and Its Potential Implications in Treating Coronavirus Disease 2019 and Lung Injury', *Front Immunol*, 11, 2056.

Cheng, C., Kobayashi, M., Martinez, J. A., Ng, H., Moser, J. J., Wang, X., Singh, V., Fritzler, M. J., and Zochodne, D. W. (2015), 'Evidence for Epigenetic Regulation of Gene Expression and Function in Chronic Experimental Diabetic Neuropathy', *J Neuropathol Exp Neurol*, 74 (8), 804-17.

Cherrington, A. D., Stevenson, R. W., Steiner, K. E., Davis, M. A., Myers, S. R., Adkins, B. A., Abumrad, N. N., and Williams, P. E. (1987), 'Insulin, glucagon, and glucose as regulators of hepatic glucose uptake and production in vivo', *Diabetes Metab Rev*, 3 (1), 307-32.

Chida, S., Fujita, Y., Ogawa, A., Hayashi, A., Ichikawa, R., Kamata, Y., Takeuchi, A., Takano, K., and Shichiri, M. (2016), 'Levels of albuminuria and risk of developing macroalbuminuria in type 2 diabetes: historical cohort study', *Sci Rep*, 6, 26380.

Consortium, G. TEx (2018), 'Erratum: Genetic effects on gene expression across human tissues', *Nature*, 553 (7689), 530.

Corcoran, C. and Jacobs, T. F. (2021), 'Metformin', *StatPearls* (Treasure Island (FL)).

Craig, M. E., Hattersley, A., and Donaghue, K. C. (2009), 'Definition, epidemiology and classification of diabetes in children and adolescents', *Pediatr Diabetes*, 10 Suppl 12, 3-12.

Davis, T. M., Millns, H., Stratton, I. M., Holman, R. R., and Turner, R. C. (1999), 'Risk factors for stroke in type 2 diabetes mellitus: United Kingdom Prospective Diabetes Study (UKPDS) 29', *Arch Intern Med*, 159 (10), 1097-103.

de Boer, I. H., Kestenbaum, B., Rue, T. C., Steffes, M. W., Cleary, P. A., Molitch, M. E., Lachin, J. M., Weiss, N. S., Brunzell, J. D., Diabetes, Control, Complications Trial /Epidemiology of Diabetes, Interventions, and Complications Study Research, Group (2008), 'Insulin therapy, hyperglycemia, and hypertension in type 1 diabetes mellitus', *Arch Intern Med*, 168 (17), 1867-73.

de Pablos-Velasco, P., Parhofer, K. G., Bradley, C., Eschwege, E., Gonder-Frederick, L., Maheux, P., Wood, I., and Simon, D. (2014), 'Current level of glycaemic control and its associated factors in patients with type 2 diabetes across Europe: data from the PANORAMA study', *Clin Endocrinol (Oxf)*, 80 (1), 47-56.

Delaneau, O., Zagury, J. F., and Marchini, J. (2013), 'Improved whole-chromosome phasing for disease and population genetic studies', *Nat Methods*, 10 (1), 5-6.

Dhahbi, J. M., Mote, P. L., Fahy, G. M., and Spindler, S. R. (2005), 'Identification of potential caloric restriction mimetics by microarray profiling', *Physiol Genomics*, 23 (3), 343-50.

Diabetes Genetics Initiative of Broad Institute of, Harvard, Mit, Lund University, Novartis Institutes of BioMedical, Research, Saxena, R., Voight, B. F., Lyssenko, V., Burtt, N. P., de Bakker, P. I., Chen, H., Roix, J. J., Kathiresan, S., Hirschhorn, J. N., Daly, M. J., Hughes, T. E., Groop, L., Altshuler, D., Almgren, P., Florez, J. C., Meyer, J., Ardlie, K., Bengtsson Bostrom, K., Isomaa, B., Lettre, G., Lindblad, U., Lyon, H. N., Melander, O., Newton-Cheh, C., Nilsson, P., Orho-Melander, M., Rastam, L., Speliotes, E. K., Taskinen, M. R., Tuomi, T., Guiducci, C., Berglund, A., Carlson, J., Gianniny, L., Hackett, R., Hall, L., Holmkvist, J., Laurila, E., Sjogren, M., Sterner, M., Surti, A., Svensson, M., Svensson, M., Tewhey, R., Blumenstiel, B., Parkin, M., Defelice, M., Barry, R., Brodeur, W., Camarata, J., Chia, N., Fava, M., Gibbons, J., Handsaker, B., Healy, C., Nguyen, K., Gates, C., Sougnez, C., Gage, D., Nizzari, M., Gabriel, S. B., Chirn, G. W., Ma, Q., Parikh, H., Richardson, D., Ricke, D., and Purcell, S. (2007), 'Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels', *Science*, 316 (5829), 1331-6.

Diebel, M. E., Diebel, L. N., and Liberati, D. M. (2011), 'Gender dimorphism in the gut: mucosal protection by estrogen stimulation of IgA transcytosis', *J Trauma*, 71 (2), 474-9.

DiMeglio, L. A., Evans-Molina, C., and Oram, R. A. (2018), 'Type 1 diabetes', *Lancet*, 391 (10138), 2449-62.

Ding, Y., Liu, K., Wang, Y., Su, G., Deng, H., Zeng, Q., Liao, Y., and Wang, Z. (2010), 'Expression and significance of fgl2 prothrombinase in cardiac microvascular endothelial cells of rats with type 2 diabetes', *J Huazhong Univ Sci Technolog Med Sci*, 30 (5), 575-81.

Divers, J., Palmer, N. D., Langefeld, C. D., Brown, W. M., Lu, L., Hicks, P. J., Smith, S. C., Xu, J., Terry, J. G., Register, T. C., Wagenknecht, L. E., Parks, J. S., Ma, L., Chan, G. C., Buxbaum, S. G., Correa, A., Musani, S., Wilson, J. G., Taylor, H. A., Bowden, D. W., Carr, J. J., and Freedman, B. I. (2017), 'Genome-wide association study of coronary artery calcified atherosclerotic plaque in African Americans with type 2 diabetes', *BMC Genet*, 18 (1), 105.

Druet, C., Tubiana-Rufi, N., Chevenne, D., Rigal, O., Polak, M., and Levy-Marchal, C. (2006), 'Characterization of insulin secretion and resistance in type 2 diabetes of adolescents', *J Clin Endocrinol Metab*, 91 (2), 401-4.

Elbere, I., Silamikelis, I., Dindune, II, Kalnina, I., Ustinova, M., Zaharenko, L., Silamikele, L., Rovite, V., Gudra, D., Konrade, I., Sokolovska, J., Pirags, V., and Klovins, J. (2020), 'Baseline gut microbiome composition predicts metformin therapy short-term efficacy in newly diagnosed type 2 diabetes patients', *PLoS One*, 15 (10), e0241338.

Erdmann, J., Grosshennig, A., Braund, P. S., Konig, I. R., Hengstenberg, C., Hall, A. S., Linsel-Nitschke, P., Kathiresan, S., Wright, B., Tregouet, D. A., Cambien, F., Bruse, P., Aherrahrou, Z., Wagner, A. K., Stark, K., Schwartz, S. M., Salomaa, V., Elosua, R., Melander, O., Voight, B. F., O'Donnell, C. J., Peltonen, L., Siscovick, D. S., Altshuler, D., Merlini, P. A., Peyvandi, F., Bernardinelli, L., Ardissino, D., Schillert, A., Blankenberg, S., Zeller, T., Wild, P., Schwarz, D. F., Tiret, L., Perret, C., Schreiber, S., El Mokhtari, N. E., Schafer, A., Marz, W., Renner, W., Bugert, P., Kluter, H., Schrezenmeir, J., Rubin, D., Ball, S. G., Balmforth, A. J., Wichmann, H. E., Meitinger, T., Fischer, M., Meisinger, C., Baumert, J., Peters, A., Ouwehand, W. H., Italian Atherosclerosis, Thrombosis, Vascular Biology Working, Group, Myocardial Infarction Genetics, Consortium, Wellcome Trust Case Control, Consortium, Cardiogenics, Consortium, Deloukas, P., Thompson, J. R., Ziegler, A., Samani, N. J., and Schunkert, H. (2009), 'New susceptibility locus for coronary artery disease on chromosome 3q22.3', *Nat Genet*, 41 (3), 280-2.

Fayfman, M., Pasquel, F. J., and Umpierrez, G. E. (2017), 'Management of Hyperglycemic Crises: Diabetic Ketoacidosis and Hyperglycemic Hyperosmolar State', *Med Clin North Am*, 101 (3), 587-606.

Fitipaldi, H., McCarthy, M. I., Florez, J. C., and Franks, P. W. (2018), 'A Global Overview of Precision Medicine in Type 2 Diabetes', *Diabetes*, 67 (10), 1911-22.

Flannick, J. and Florez, J. C. (2016), 'Type 2 diabetes: genetic data sharing to advance complex disease research', *Nat Rev Genet*, 17 (9), 535-49.

Florez, H., Luo, J., Castillo-Florez, S., Mitsi, G., Hanna, J., Tamariz, L., Palacio, A., Nagendran, S., and Hagan, M. (2010), 'Impact of metformin-induced gastrointestinal symptoms on quality of life and adherence in patients with type 2 diabetes', *Postgrad Med*, 122 (2), 112-20.

Forbes, J. M. and Cooper, M. E. (2013), 'Mechanisms of diabetic complications', *Physiol Rev*, 93 (1), 137-88.

Fowler, Michael J. (2008), 'Microvascular and Macrovascular Complications of Diabetes', *Clinical Diabetes*, Apr; (26(2)), 77-82.

Gabbay, K. H. (1975), 'Hyperglycemia, polyol metabolism, and complications of diabetes mellitus', *Annu Rev Med*, 26, 521-36.

Galicia-Garcia, U., Benito-Vicente, A., Jebari, S., Larrea-Sebal, A., Siddiqi, H., Uribe, K. B., Ostolaza, H., and Martin, C. (2020), 'Pathophysiology of Type 2 Diabetes Mellitus', *Int J Mol Sci*, 21 (17).

Galiveti, C. R., Raabe, C. A., Konthur, Z., and Rozhdestvensky, T. S. (2014), 'Differential regulation of non-protein coding RNAs from Prader-Willi Syndrome locus', *Sci Rep*, 4, 6445.

Garcia, M., Arteche-Martinez, U., Lertxundi, U., and Aguirre, C. (2021), 'SGLT2 Inhibitors and Bladder Cancer: Analysis of Cases Reported in the European Pharmacovigilance Database', *J Clin Pharmacol*, 61 (2), 187-92.

Geeleher, P., Cox, N. J., and Huang, R. S. (2014), 'Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines', *Genome Biol*, 15 (3), R47.

Geerling, J. J., Boon, M. R., van der Zon, G. C., van den Berg, S. A., van den Hoek, A. M., Lombes, M., Princen, H. M., Havekes, L. M., Rensen, P. C., and Guigas, B. (2014), 'Metformin lowers plasma triglycerides by promoting VLDL-triglyceride clearance by brown adipose tissue in mice', *Diabetes*, 63 (3), 880-91.

Gillespie, Z. E., Wang, C., Vadan, F., Yu, T. Y., Ausio, J., Kusalik, A., and Eskiw, C. H. (2019), 'Metformin induces the AP-1 transcription factor network in normal dermal fibroblasts', *Sci Rep*, 9 (1), 5369.

GoDarts, Group, Ukpds Diabetes Pharmacogenetics Study, Wellcome Trust Case Control, Consortium, Zhou, K., Bellenguez, C., Spencer, C. C., Bennett, A. J., Coleman, R. L., Tavendale, R., Hawley, S. A., Donnelly, L. A., Schofield, C., Groves, C. J., Burch, L., Carr, F., Strange, A., Freeman, C., Blackwell, J. M., Bramon, E., Brown, M. A., Casas, J. P., Corvin, A., Craddock, N., Deloukas, P., Dronov, S., Duncanson, A., Edkins, S., Gray, E., Hunt, S., Jankowski, J., Langford, C., Markus, H. S., Mathew, C. G., Plomin, R., Rautanen, A., Sawcer, S. J., Samani, N. J., Trembath, R., Viswanathan, A. C., Wood, N. W., investigators, Magic, Harries, L. W., Hattersley, A. T., Doney, A. S., Colhoun, H., Morris, A. D., Sutherland, C., Hardie, D. G., Peltonen, L., McCarthy, M. I., Holman, R. R., Palmer, C. N., Donnelly, P.,

and Pearson, E. R. (2011), 'Common variants near ATM are associated with glycemic response to metformin in type 2 diabetes', *Nat Genet*, 43 (2), 117-20.

Goyal, R. and Jialal, I. (2021), 'Diabetes Mellitus Type 2', *StatPearls* (Treasure Island (FL)).

Gormsen, L. C., Sondergaard, E., Christensen, N. L., Brosen, K., Jessen, N., and Nielsen, S. (2019), 'Metformin increases endogenous glucose production in non-diabetic individuals and individuals with recent-onset type 2 diabetes', *Diabetologia*, 62 (7), 1251-56.

Guan, W., Pluzhnikov, A., Cox, N. J., Boehnke, M., and International Type 2 Diabetes Linkage Analysis, Consortium (2008), 'Meta-analysis of 23 type 2 diabetes linkage studies from the International Type 2 Diabetes Linkage Analysis Consortium', *Hum Hered*, 66 (1), 35-49.

Gunton, J. E., Delhanty, P. J., Takahashi, S., and Baxter, R. C. (2003), 'Metformin rapidly increases insulin receptor activation in human liver and signals preferentially through insulin-receptor substrate-2', *J Clin Endocrinol Metab*, 88 (3), 1323-32.

Guo, J., Han, X., Huang, W., You, Y., and Jicheng, Z. (2021), 'Interaction between IgA and gut microbiota and its role in controlling metabolic syndrome', *Obes Rev*, 22 (4), e13155.

Hao, Q., Vadgama, J. V., and Wang, P. (2020), 'CCL2/CCR2 signaling in cancer pathogenesis', *Cell Commun Signal*, 18 (1), 82.

Hearn, T. (2019), 'ALMS1 and Alstrom syndrome: a recessive form of metabolic, neurosensory and cardiac deficits', *J Mol Med (Berl)*, 97 (1), 1-17.

Heishi, M., Ichihara, J., Teramoto, R., Itakura, Y., Hayashi, K., Ishikawa, H., Gomi, H., Sakai, J., Kanaoka, M., Taiji, M., and Kimura, T. (2006), 'Global gene expression analysis in liver of obese diabetic db/db mice treated with metformin', *Diabetologia*, 49 (7), 1647-55.

Hermann, L. S., Schersten, B., and Melander, A. (1994), 'Antihyperglycaemic efficacy, response prediction and dose-response relations of treatment with metformin and sulphonylurea, alone and in primary combination', *Diabet Med*, 11 (10), 953-60.

Holmkvist, J., Almgren, P., Lyssenko, V., Lindgren, C. M., Eriksson, K. F., Isomaa, B., Tuomi, T., Nilsson, P., and Groop, L. (2008), 'Common variants in maturity-onset diabetes of the young genes and future risk of type 2 diabetes', *Diabetes*, 57 (6), 1738-44.

Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., and Abecasis, G. R. (2012), 'Fast and accurate genotype imputation in genome-wide association studies through pre-phasing', *Nat Genet*, 44 (8), 955-9.

Hu, D., Guo, Y., Wu, R., Shao, T., Long, J., Yu, B., Wang, H., Luo, Y., Lu, H., Zhang, J., Chen, Y. E., and Peng, D. (2021), 'New Insight Into Metformin-Induced Cholesterol-Lowering Effect Crosstalk Between Glucose and Cholesterol Homeostasis via ChREBP (Carbohydrate-Responsive Element-Binding Protein)-Mediated PCSK9 (Proprotein Convertase Subtilisin/Kexin Type 9) Regulation', *Arterioscler Thromb Vasc Biol*, 41 (4), e208-e23.

Hu, Z. Z., Huang, H., Wu, C. H., Jung, M., Dritschilo, A., Riegel, A. T., and Wellstein, A. (2011), 'Omics-based molecular target and biomarker identification', *Methods Mol Biol*, 719, 547-71.

Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009), 'Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources', *Nat Protoc*, 4 (1), 44-57.

Huang, Y. C., Lin, J. M., Lin, H. J., Chen, C. C., Chen, S. Y., Tsai, C. H., and Tsai, F. J. (2011), 'Genome-wide association study of diabetic retinopathy in a Taiwanese population', *Ophthalmology*, 118 (4), 642-8.

Hunter, J. D. (2007), 'Matplotlib: A 2D graphics environment', *Computing in Science & Engineering*, 9, 99-104.

Hur, J., Dauch, J. R., Hinder, L. M., Hayes, J. M., Backus, C., Pennathur, S., Kretzler, M., Brosius, F. C., 3rd, and Feldman, E. L. (2015), 'The Metabolic Syndrome and Microvascular Complications in a Murine Model of Type 2 Diabetes', *Diabetes*, 64 (9), 3294-304.

IQWiG (2006), 'Type 2 diabetes: Overview.'.

Yadav, R., Jain, N., Raizada, N., Jhamb, R., Rohatgi, J., and Madhu, S. V. (2021), 'Prevalence of diabetes related vascular complications in subjects with normal glucose tolerance, prediabetes, newly detected diabetes and known diabetes', *Diabetes Metab Syndr*, 15 (5), 102226.

Yan, M., Qi, H., Xia, T., Zhao, X., Wang, W., Wang, Z., Lu, C., Ning, Z., Chen, H., Li, T., Tekcham, D. S., Liu, X., Liu, J., Chen, D., Liu, X., Xu, G., and Piao, H. L. (2019), 'Metabolomics profiling of metformin-mediated metabolic reprogramming bypassing AMPKalpha', *Metabolism*, 91, 18-29.

Yang, M., Darwish, T., Larraufie, P., Rimmington, D., Cimino, I., Goldspink, D. A., Jenkins, B., Koulman, A., Brighton, C. A., Ma, M., Lam, B. Y. H., Coll, A. P., O'Rahilly, S., Reimann, F., and Gribble, F. M. (2021), 'Inhibition of mitochondrial function by metformin increases glucose uptake, glycolysis and GDF-15 release from intestinal cells', *Sci Rep*, 11 (1), 2529.

Yates, A. D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Azov, A. G., Bennett, R., Bhai, J., Billis, K., Boddu, S., Marugan, J. C., Cummins, C., Davidson, C., Dodiya, K., Fatima, R., Gall, A., Giron, C. G., Gil, L., Grego, T., Haggerty, L., Haskell, E., Hourlier, T., Izuogu, O. G., Janacek, S. H., Juettemann, T., Kay, M., Lavidas, I., Le, T., Lemos, D., Martinez, J. G., Maurel, T., McDowall, M., McMahon, A., Mohanan, S., Moore, B., Nuhn, M., Oheh, D. N., Parker, A., Parton, A., Patricio, M., Sakthivel, M. P., Abdul Salam, A. I., Schmitt, B. M., Schuilenburg, H., Sheppard, D., Sycheva, M., Szuba, M., Taylor, K., Thormann, A., Threadgold, G., Vullo, A., Walts, B., Winterbottom, A., Zadissa, A., Chakiachvili, M., Flint, B., Frankish, A., Hunt, S. E., G, I. Isley,

Kostadima, M., Langridge, N., Loveland, J. E., Martin, F. J., Morales, J., Mudge, J. M., Muffato, M., Perry, E., Ruffier, M., Trevanion, S. J., Cunningham, F., Howe, K. L., Zerbino, D. R., and Flicek, P. (2020), 'Ensembl 2020', *Nucleic Acids Res*, 48 (D1), D682-D88.

You, J. H., Lee, S. A., Chun, S. Y., Song, S. O., Lee, B. W., Kim, D. J., and Boyko, E. J. (2020), 'Clinical Outcomes of COVID-19 Patients with Type 2 Diabetes: A Population-Based Study in Korea', *Endocrinol Metab (Seoul)*, 35 (4), 901-08.

Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010), 'Gene ontology analysis for RNA-seq: accounting for selection bias', *Genome Biol*, 11 (2), R14.

Yu, B., Wu, Y., and Li, Z. (2020), 'KLF4/Ch25h axis activated by metformin suppresses EndoMT in human umbilical vein endothelial cells', *Biochem Biophys Res Commun*, 522 (4), 838-44.

Yuan, L., Ziegler, R., and Hamann, A. (2003), 'Metformin modulates insulin post-receptor signaling transduction in chronically insulin-treated Hep G2 cells', *Acta Pharmacol Sin*, 24 (1), 55-60.

Jankovic, M., Novakovic, I., Nikolic, D., Mitrovic Maksic, J., Brankovic, S., Petronic, I., Cirovic, D., Ducic, S., Grajic, M., and Bogicevic, D. (2021), 'Genetic and Epigenomic Modifiers of Diabetic Neuropathy', *Int J Mol Sci*, 22 (9).

Kaprio, J., Tuomilehto, J., Koskenvuo, M., Romanov, K., Reunanen, A., Eriksson, J., Stengard, J., and Kesaniemi, Y. A. (1992), 'Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus in a population-based cohort of twins in Finland', *Diabetologia*, 35 (11), 1060-7.

Khalil, H. (2017), 'Diabetes microvascular complications-A clinical update', *Diabetes Metab Syndr*, 11 Suppl 1, S133-S39.

Khan, M. A. B., Hashim, M. J., King, J. K., Govender, R. D., Mustafa, H., and Al Kaabi, J. (2020), 'Epidemiology of Type 2 Diabetes - Global Burden of Disease and Forecasted Trends', *J Epidemiol Glob Health*, 10 (1), 107-11.

Kim, H. J., Lee, S., Chun, K. H., Jeon, J. Y., Han, S. J., Kim, D. J., Kim, Y. S., Woo, J. T., Nam, M. S., Baik, S. H., Ahn, K. J., and Lee, K. W. (2018), 'Metformin reduces the risk of cancer in patients with type 2 diabetes: An analysis based on the Korean National Diabetes Program Cohort', *Medicine (Baltimore)*, 97 (8), e0036.

Kim, J., Chung, K., Choi, C., Beloor, J., Ullah, I., Kim, N., Lee, K. Y., Lee, S. K., and Kumar, P. (2016), 'Silencing CCR2 in Macrophages Alleviates Adipose Tissue Inflammation and the Associated Metabolic Syndrome in Dietary Obese Mice', *Mol Ther Nucleic Acids*, 5, e280.

Kitada, M., Zhang, Z., Mima, A., and King, G. L. (2010), 'Molecular mechanisms of diabetic vascular complications', *J Diabetes Investig*, 1 (3), 77-89.

Kyler, K. E., Kadakia, R. B., Palac, H. L., Kwon, S., Ariza, A. J., and Binns, H. J. (2018), 'Use of Metformin for Weight Management in Children and Adolescents With Obesity in the Clinical Setting', *Clin Pediatr (Phila)*, 57 (14), 1677-85.

Ko, Y., Choi, A., Lee, M., and Lee, J. A. (2016), 'Metformin displays in vitro and in vivo antitumor effect against osteosarcoma', *Korean J Pediatr*, 59 (9), 374-80.

Kobayashi, M. and Zochodne, D. W. (2018), 'Diabetic neuropathy and the sensory neuron: New aspects of pathogenesis and their treatment implications', *J Diabetes Investig*, 9 (6), 1239-54.

Kosiborod, M., Gomes, M. B., Nicolucci, A., Pocock, S., Rathmann, W., Shestakova, M. V., Watada, H., Shimomura, I., Chen, H., Cid-Ruzafa, J., Fenici, P., Hammar, N., Surmont, F., Tang, F., Khunti, K., and investigators, Discover (2018), 'Vascular complications in patients with type 2 diabetes: prevalence and associated factors in 38 countries (the DISCOVER study program)', *Cardiovasc Diabetol*, 17 (1), 150.

Kowluru, R. A. (2005), 'Diabetic retinopathy: mitochondrial dysfunction and retinal capillary cell death', *Antioxid Redox Signal*, 7 (11-12), 1581-87.

Lachmandas, E., Eckold, C., Bohme, J., Koeken, Vacm, Marzuki, M. B., Blok, B., Arts, R. J. W., Chen, J., Teng, K. W. W., Ratter, J., Smolders, E. J., Van den Heuvel, C., Stienstra, R., Dockrell, H. M., Newell, E., Netea, M. G., Singhal, A., Cliff, J. M., and Van Crevel, R. (2019), 'Metformin Alters Human Host Responses to Mycobacterium tuberculosis in Healthy Subjects', *J Infect Dis*, 220 (1), 139-50.

Latvia, The Centre for Disease Prevention and Control of (2020), 'Health Statistics Database'.

Laustriat, D., Gide, J., Barrault, L., Chautard, E., Benoit, C., Auboeuf, D., Boland, A., Battail, C., Artiguenave, F., Deleuze, J. F., Benit, P., Rustin, P., Franc, S., Charpentier, G., Furling, D., Bassez, G., Nissan, X., Martinat, C., Peschanski, M., and Baghdoyan, S. (2015), 'In Vitro and In Vivo Modulation of Alternative Splicing by the Biguanide Metformin', *Mol Ther Nucleic Acids*, 4, e262.

Law, C. W., Alhamdoosh, M., Su, S., Dong, X., Tian, L., Smyth, G. K., and Ritchie, M. E. (2016), 'RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR', *F1000Res*, 5.

LCDC, The Centre for Disease Prevention and Control of Latvia (2020), *Health Statistics Database*. The Centre for Disease Prevention and Control of Latvia.

Lee, J., Harris, A. N., Holley, C. L., Mahadevan, J., Pyles, K. D., Lavagnino, Z., Scherrer, D. E., Fujiwara, H., Sidhu, R., Zhang, J., Huang, S. C., Piston, D. W., Remedi, M. S., Urano, F., Ory, D. S., and Schaffer, J. E. (2016), 'Rpl13a small nucleolar RNAs regulate systemic glucose metabolism', *J Clin Invest*, 126 (12), 4616-25.

Lee, S. Y., Moon, S. J., Kim, E. K., Seo, H. B., Yang, E. J., Son, H. J., Kim, J. K., Min, J. K., Park, S. H., and Cho, M. L. (2017), 'Metformin Suppresses Systemic Autoimmunity in Roquin(san/san) Mice through Inhibiting B Cell Differentiation into Plasma Cells via Regulation of AMPK/mTOR/STAT3', *J Immunol*, 198 (7), 2661-70.

Li, S., Wang, J., Zhang, B., Li, X., and Liu, Y. (2019), 'Diabetes Mellitus and Cause-Specific Mortality: A Population-Based Study', *Diabetes Metab J*, 43 (3), 319-41.

Lind, M., Oden, A., Fahlen, M., and Eliasson, B. (2009), 'The true value of HbA1c as a predictor of diabetic complications: simulations of HbA1c variables', *PLoS One*, 4 (2), e4412.

Liu, C., Chen, G., Bentley, A. R., Doumatey, A., Zhou, J., Adeyemo, A., Yang, J., and Rotimi, C. (2019a), 'Genome-wide association study for proliferative diabetic retinopathy in Africans', *NPJ Genom Med*, 4, 20.

Liu, X., Xu, Y., An, M., and Zeng, Q. (2019b), 'The risk factors for diabetic peripheral neuropathy: A meta-analysis', *PLoS One*, 14 (2), e0212574.

Liu, Z., Fu, C., Wang, W., and Xu, B. (2010), 'Prevalence of chronic complications of type 2 diabetes mellitus in outpatients - a cross-sectional hospital based survey in urban China', *Health Qual Life Outcomes*, 8, 62.

Lombardo, B., D'Argenio, V., Monda, E., Vitale, A., Caiazza, M., Sacchetti, L., Pastore, L., Limongelli, G., Frisso, G., and Mazzaccara, C. (2020), 'Genetic analysis resolves differential diagnosis of a familial syndromic dilated cardiomyopathy: A new case of Alstrom syndrome', *Mol Genet Genomic Med*, 8 (7), e1260.

Lord, S. R., Collins, J. M., Cheng, W. C., Haider, S., Wigfield, S., Gaude, E., Fielding, B. A., Pinnick, K. E., Harjes, U., Segaran, A., Jha, P., Hoefler, G., Pollak, M. N., Thompson, A. M., Roy, P. G., English, R., Adams, R. F., Frezza, C., Buffa, F. M., Karpe, F., and Harris, A. L. (2020), 'Transcriptomic analysis of human primary breast cancer identifies fatty acid oxidation as a target for metformin', *Br J Cancer*, 122 (2), 258-65.

Lu, N., Han, Y., Chen, T., Gunzler, D. D., Xia, Y., Lin, J. Y., and Tu, X. M. (2013), 'Power analysis for cross-sectional and longitudinal study designs', *Shanghai Arch Psychiatry*, 25 (4), 259-62.

Lu, X. Y., Huang, S., Chen, Q. B., Zhang, D., Li, W., Ao, R., Leung, F. C., Zhang, Z., Huang, J., Tang, Y., and Zhang, S. J. (2020), 'Metformin Ameliorates Abeta Pathology by Insulin-Degrading Enzyme in a Transgenic Mouse Model of Alzheimer's Disease', *Oxid Med Cell Longev*, 2020, 2315106.

Lucier, J. and Weinstock, R. S. (2021), 'Diabetes Mellitus Type 1', *StatPearls* (Treasure Island (FL)).

Luck, H., Khan, S., Kim, J. H., Copeland, J. K., Revelo, X. S., Tsai, S., Chakraborty, M., Cheng, K., Tao Chan, Y., Nohr, M. K., Clemente-Casares, X., Perry, M. C., Ghazarian, M., Lei, H., Lin, Y. H., Coburn, B., Okrainec, A., Jackson, T., Poutanen, S., Gaisano, H., Allard, J. P., Guttman, D. S., Conner, M. E., Winer, S., and Winer, D. A. (2019), 'Gut-associated IgA(+) immune cells regulate obesity-related insulin resistance', *Nat Commun*, 10 (1), 3650.

Luizon, M. R., Eckalbar, W. L., Wang, Y., Jones, S. L., Smith, R. P., Laurance, M., Lin, L., Gallins, P. J., Etheridge, A. S., Wright, F., Zhou, Y., Molony, C., Innocenti, F., Yee, S. W., Giacomini, K. M., and Ahituv, N. (2016), 'Genomic Characterization of Metformin Hepatic Response', *PLoS Genet*, 12 (11), e1006449.

Maeda, S., Osawa, N., Hayashi, T., Tsukada, S., Kobayashi, M., and Kikkawa, R. (2007), 'Genetic variations associated with diabetic nephropathy and type II diabetes in a Japanese population', *Kidney Int Suppl*, (106), S43-8.

Mahajan, A. andTaliun, D. andThurner, M. andRobertson, N. R. andTorres, J. M. andRayner, N. W. andPayne, A. J. andSteinthorsdottir, V. andScott, R. A. andGrarup, N. andCook, J. P. andSchmidt, E. M. andWuttke, M. andSarnowski, C. andMagill, R. andNano, J. andGieger, C. andTrompet, S. andLecoeur, C. andPreuss, M. H. andPrins, B. P. andGuo, X. Q. andBielak, L. F. andBelow, J. E. andBowden, D. W. andChambers, J. C. andKim, Y. J. andNg, M. C. Y. andPetty, L. E. andSim, X. L. andZhang, W. H. andBennett, A. J. andBork-Jensen, J. andBrummett, C. M. andCanouil, M. andKardt, K. U. E. andFischer, K. andKardia, S. L. R. andKronenberg, F. andLall, K. andLiu, C. T. andLocke, A. E. andLuan, J. A. andNtalla, L. andNylander, V. andSchonherr, S. andSchurmann, C. andYengo, L. andBottinger, E. P. andBrandslund, I. andChristensen, C. andDedoussis, G. andFlorez, J. C. andFord, I. andFrance, O. H. andFrayling, T. M. andGiedraitis, V. andHackinger, S. andHattersley, A. T. andHerder, C. andIkram, M. A. andIngelsson, M. andJorgensen, M. E. andJorgensen, T. andKriebel, J. andKuusisto, J. andLigthart, S. andLindgren, C. M. andLinneberg, A. andLyssenko, V. andMamakou, V. andMeitinger, T. andMohlke, K. L. andMorris, A. D. andNadkarni, G. andPankow, J. S. andPeters, A. andSattar, N. andStancakova, A. andStrauch, K. andTaylor, K. D. andThorand, B. andThorleifsson, G. andThorsteinsdottir, U. andTuomilehto, J. andWitte, D. R. andDupuis, J. andPeyser, P. A. andZeggini, E. andLoos, R. J. F. andFroguel, P. andIngelsson, E. andLind, L. andGroop, L. andLaakso, M. andCollins, F. S. andJukema, J. W. andPalmer, C. N. A. andGrallert, H. andMetspalu, A. andDehghan, A. andKottgen, A. andAbecasis, G. R. andMeigs, J. B. andRotter, J. I. andMarchini, J. andPedersen, O. andHansen, T. andLangenberg, C. andWareham, N. J. andStefansson, K. andGloyn, A. L. andMorris, A. P. andBoehnke, M. and McCarthy, M. I. (2018), 'Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps', *Nature Genetics*, 50 (11), 1505-+.

Malinska, H., Oliyarnyk, O., Skop, V., Silhavy, J., Landa, V., Zidek, V., Mlejnek, P., Simakova, M., Strnad, H., Kazdova, L., and Pravenec, M. (2016), 'Effects of Metformin on Tissue Oxidative and Dicarbonyl Stress in Transgenic Spontaneously Hypertensive Rats Expressing Human C-Reactive Protein', *PLoS One*, 11 (3), e0150924.

Marees, A. T., de Kluiver, H., Stringer, S., Vorspan, F., Curis, E., Marie-Claire, C., and Derks, E. M. (2018), 'A tutorial on conducting genome-wide association studies: Quality control and statistical analysis', *Int J Methods Psychiatr Res*, 27 (2), e1608.

Martin-Montalvo, A., Mercken, E. M., Mitchell, S. J., Palacios, H. H., Mote, P. L., Scheibye-Knudsen, M., Gomes, A. P., Ward, T. M., Minor, R. K., Blouin, M. J., Schwab, M., Pollak, M., Zhang, Y., Yu, Y., Becker, K. G., Bohr, V. A., Ingram, D. K., Sinclair, D. A., Wolf, N. S., Spindler, S. R., Bernier, M., and de Cabo, R. (2013), 'Metformin improves healthspan and lifespan in mice', *Nat Commun*, 4, 2192.

Matthaei, S., Reibold, J. P., Hamann, A., Benecke, H., Haring, H. U., Greten, H., and Klein, H. H. (1993), 'In vivo metformin treatment ameliorates insulin resistance: evidence for potentiation of insulin-induced translocation and increased functional activity of glucose transporters in obese (fa/fa) Zucker rat adipocytes', *Endocrinology*, 133 (1), 304-11.

Mc Namara, K., Alzubaidi, H., and Jackson, J. K. (2019), 'Cardiovascular disease as a leading cause of death: how are pharmacists getting involved?', *Integr Pharm Res Pract*, 8, 1-11.

McCreight, L. J., Bailey, C. J., and Pearson, E. R. (2016), 'Metformin and the gastrointestinal tract', *Diabetologia*, 59 (3), 426-35.

McDonough, C. W., Palmer, N. D., Hicks, P. J., Roh, B. H., An, S. S., Cooke, J. N., Hester, J. M., Wing, M. R., Bostrom, M. A., Rudock, M. E., Lewis, J. P., Talbert, M. E., Blevins, R. A., Lu, L., Ng, M. C., Sale, M. M., Divers, J., Langefeld, C. D., Freedman, B. I., and Bowden, D. W. (2011), 'A genome-wide association study for diabetic nephropathy genes in African Americans', *Kidney Int*, 79 (5), 563-72.

Meng, Y., Xiang, R., Yan, H., Zhou, Y., Hu, Y., Yang, J., Zhou, Y., and Cui, Q. (2020), 'Transcriptomic landscape profiling of metformin-treated healthy mice: Implication for potential hypertension risk when prophylactically used', *J Cell Mol Med*, 24 (14), 8138-50.

Meng, W., Deshmukh, H. A., van Zuydam, N. R., Liu, Y., Donnelly, L. A., Zhou, K., Wellcome Trust Case Control, Consortium, Surrogate Markers for, Micro, Macro-Vascular Hard Endpoints for Innovative Diabetes Tools Study, Group, Morris, A. D., Colhoun, H. M., Palmer, C. N., and Smith, B. H. (2015), 'A genome-wide association study suggests an association of Chr8p21.3 (GFRA2) with diabetic neuropathic pain', *Eur J Pain*, 19 (3), 392-9.

Meng, W., Shah, K. P., Pollack, S., Toppila, I., Hebert, H. L., McCarthy, M. I., Groop, L., Ahlqvist, E., Lyssenko, V., Agardh, E., Daniell, M., Kaidonis, G., Craig, J. E., Mitchell, P., Liew, G., Kifley, A., Wang, J. J., Christiansen, M. W., Jensen, R. A., Penman, A., Hancock, H. A., Chen, C. J., Correa, A., Kuo, J. Z., Li, X., Chen, Y. I., Rotter, J. I., Klein, R., Klein, B., Wong, T. Y., Morris, A. D., Doney, A. S. F., Colhoun, H. M., Price, A. L., Burdon, K. P., Groop, P. H., Sandholm, N., Grassi, M. A., Sobrin, L., Palmer, C. N. A., Wellcome Trust Case Control Consortium, Surrogate markers for Micro, and Macro-vascular hard endpoints for Innovative diabetes Tools study, group (2018), 'A genome-wide association study suggests new evidence for an association of the NADPH Oxidase 4 (NOX4) gene with severe diabetic retinopathy in type 2 diabetes', *Acta Ophthalmol*, 96 (7), e811-e19.

Mogensen, C. E., Christensen, C. K., and Vittinghus, E. (1983), 'The stages in diabetic renal disease. With emphasis on the stage of incipient diabetic nephropathy', *Diabetes*, 32 Suppl 2, 64-78.

Montesanto, A., Bonfigli, A. R., Crocco, P., Garagnani, P., De Luca, M., Boemi, M., Marasco, E., Pirazzini, C., Giuliani, C., Franceschi, C., Passarino, G., Testa, R., Olivieri, F., and Rose, G. (2018), 'Genes associated with Type 2 Diabetes and vascular complications', *Aging (Albany NY)*, 10 (2), 178-96.

Moucheraud, C., Lenz, C., Latkovic, M., and Wirtz, V. J. (2019), 'The costs of diabetes treatment in low- and middle-income countries: a systematic review', *BMJ Glob Health*, 4 (1), e001258.

Mudaliar, S., Chang, A. R., and Henry, R. R. (2003), 'Thiazolidinediones, peripheral edema, and type 2 diabetes: incidence, pathophysiology, and clinical implications', *Endocr Pract*, 9 (5), 406-16.

Mulnier, H. E., Seaman, H. E., Raleigh, V. S., Soedamah-Muthu, S. S., Colhoun, H. M., and Lawrenson, R. A. (2006), 'Mortality in people with type 2 diabetes in the UK', *Diabet Med*, 23 (5), 516-21.

Noble, J. A. (2015), 'Immunogenetics of type 1 diabetes: A comprehensive review', *J Autoimmun*, 64, 101-12.

Oikawa, J., Ukawa, S., Ohira, H., Kawamura, T., Wakai, K., Ando, M., Hata, A., and Tamakoshi, A. (2015), 'Diabetes Mellitus is Associated With Low Secretion Rates of Immunoglobulin A in Saliva', *J Epidemiol*, 25 (7), 470-4.

Ono, H. (2019), 'Molecular Mechanisms of Hypothalamic Insulin Resistance', *Int J Mol Sci*, 20 (6).

P, Jones E.; Oliphant T.; Peterson (2007), 'SciPy: Open Source Scientific Tools for Python', *Computing in Science & Engineering*, 9, 10-20.

Padilla, J., Thorne, P. K., Martin, J. S., Rector, R. S., Akter, S., Davis, J. W., Laughlin, M. H., and Jenkins, N. T. (2017), 'Transcriptomic effects of metformin in skeletal muscle arteries of obese insulin-resistant rats', *Exp Biol Med (Maywood)*, 242 (6), 617-24.

Palicka, V. (2002), 'Pathophysiology of Diabetes Mellitus', *EJIFCC*, 13 (5), 140-44.

Palmer, B. F., Clegg, D. J., Taylor, S. I., and Weir, M. R. (2016), 'Diabetic ketoacidosis, sodium glucose transporter-2 inhibitors and the kidney', *J Diabetes Complications*, 30 (6), 1162-6.

Papatheodorou, K., Banach, M., Bekiari, E., Rizzo, M., and Edmonds, M. (2018), 'Complications of Diabetes 2017', *J Diabetes Res*, 2018, 3086167.

Park, J. E., Jeong, G. H., Lee, I. K., Yoon, Y. R., Liu, K. H., Gu, N., and Shin, K. H. (2018), 'A Pharmacometabolomic Approach to Predict Response to Metformin in Early-Phase Type 2 Diabetes Mellitus Patients', *Molecules*, 23 (7).

Pathak, V., Pathak, N. M., O'Neill, C. L., Guduric-Fuchs, J., and Medina, R. J. (2019), 'Therapies for Type 1 Diabetes: Current Scenario and Future Perspectives', *Clin Med Insights Endocrinol Diabetes*, 12, 1179551419844521.

Pearson, E. R. (2016), 'Personalized medicine in diabetes: the role of 'omics' and biomarkers', *Diabet Med*, 33 (6), 712-7.

Pernicova, I. and Korbonits, M. (2014), 'Metformin--mode of action and clinical implications for diabetes and cancer', *Nat Rev Endocrinol*, 10 (3), 143-56.

Pezzolesi, M. G., Poznik, G. D., Mychaleckyj, J. C., Paterson, A. D., Barati, M. T., Klein, J. B., Ng, D. P., Placha, G., Canani, L. H., Bochenski, J., Waggott, D., Merchant, M. L., Krolewski, B., Mirea, L., Wanic, K., Katavetin, P., Kure, M., Wolkow, P., Dunn, J. S., Smiles, A., Walker, W. H., Boright, A. P., Bull, S. B., Group, Dcct Edic Research, Doria, A., Rogus, J. J., Rich, S. S., Warram, J. H., and Krolewski, A. S. (2009), 'Genome-wide association scan for diabetic nephropathy susceptibility genes in type 1 diabetes', *Diabetes*, 58 (6), 1403-10.

Phung, O. J., Schwartzman, E., Allen, R. W., Engel, S. S., and Rajpathak, S. N. (2013), 'Sulphonylureas and risk of cardiovascular disease: systematic review and meta-analysis', *Diabet Med*, 30 (10), 1160-71.

Ping, F., Jiang, N., and Li, Y. (2020), 'Association between metformin and neurodegenerative diseases of observational studies: systematic review and meta-analysis', *BMJ Open Diabetes Res Care*, 8 (1).

Pociot, F. (2017), 'Type 1 diabetes genome-wide association studies: not to be lost in translation', *Clin Transl Immunology*, 6 (12), e162.

Pollreisz, A. and Schmidt-Erfurth, U. (2010), 'Diabetic cataract-pathogenesis, epidemiology and treatment', *J Ophthalmol*, 2010, 608751.

Poulsen, P., Kyvik, K. O., Vaag, A., and Beck-Nielsen, H. (1999), 'Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance--a population-based twin study', *Diabetologia*, 42 (2), 139-45.

Pu, R., Shi, D., Gan, T., Ren, X., Ba, Y., Huo, Y., Bai, Y., Zheng, T., and Cheng, N. (2020), 'Effects of metformin in obesity treatment in different populations: a meta-analysis', *Ther Adv Endocrinol Metab*, 11, 2042018820926000.

Pusceddu, S., Vernieri, C., Prinzi, N., Torchio, M., Coppa, J., Antista, M., Niger, M., Milione, M., Giacomelli, L., Corti, F., Prisciandaro, M., Monteleone, M., Colombo, E., Di Bartolomeo, M., and de Braud, F. (2020), 'The potential role of metformin in the treatment of patients with pancreatic neuroendocrine tumors: a review of preclinical to clinical evidence', *Therap Adv Gastroenterol*, 13, 1756284820927271.

Qi, L., Qi, Q., Prudente, S., Mendonca, C., Andreozzi, F., di Pietro, N., Sturma, M., Novelli, V., Mannino, G. C., Formoso, G., Gervino, E. V., Hauser, T. H., Muehlschlegel, J. D., Niewczas, M. A., Krolewski, A. S., Biolo, G., Pandolfi, A., Rimm, E., Sesti, G., Trischitta, V., Hu, F., and Doria, A. (2013), 'Association between a genetic variant related to glutamic acid metabolism and coronary heart disease in individuals with type 2 diabetes', *JAMA*, 310 (8), 821-8.

Rader, D. J. (2007), 'Effect of insulin resistance, dyslipidemia, and intra-abdominal adiposity on the development of cardiovascular disease and diabetes mellitus', *Am J Med*, 120 (3 Suppl 1), S12-8.

Rashid, M., Shahzad, M., Mahmood, S., and Khan, K. (2019), 'Variability in the therapeutic response of Metformin treatment in patients with type 2 diabetes mellitus', *Pak J Med Sci*, 35 (1), 71-76.

Reilly, M. P., Li, M., He, J., Ferguson, J. F., Stylianou, I. M., Mehta, N. N., Burnett, M. S., Devaney, J. M., Knouff, C. W., Thompson, J. R., Horne, B. D., Stewart, A. F., Assimes, T. L., Wild, P. S., Allayee, H., Nitschke, P. L., Patel, R. S., Myocardial Infarction Genetics, Consortium, Wellcome Trust Case Control, Consortium, Martinelli, N., Girelli, D., Quyyumi, A. A., Anderson, J. L., Erdmann, J., Hall, A. S., Schunkert, H., Quertermous, T., Blankenberg, S., Hazen, S. L., Roberts, R., Kathiresan, S., Samani, N. J., Epstein, S. E., and Rader, D. J. (2011), 'Identification of ADAMTS7 as a novel locus for coronary atherosclerosis and association of ABO with myocardial infarction in the presence of coronary atherosclerosis: two genome-wide association studies', *Lancet*, 377 (9763), 383-92.

Rena, G., Hardie, D. G., and Pearson, E. R. (2017), 'The mechanisms of action of metformin', *Diabetologia*, 60 (9), 1577-85.

Reusch, J. E. and Manson, J. E. (2017), 'Management of Type 2 Diabetes in 2017: Getting to Goal', *JAMA*, 317 (10), 1015-16.

Rewers, M. and Ludvigsson, J. (2016), 'Environmental risk factors for type 1 diabetes', *Lancet*, 387 (10035), 2340-48.

Ringborg, A., Lindgren, P., Yin, D. D., Martinell, M., and Stalhammar, J. (2010), 'Time to insulin treatment and factors associated with insulin prescription in Swedish patients with type 2 diabetes', *Diabetes Metab*, 36 (3), 198-203.

Roden, M., Stingl, H., Chandramouli, V., Schumann, W. C., Hofer, A., Landau, B. R., Nowotny, P., Waldhausl, W., and Shulman, G. I. (2000), 'Effects of free fatty acid elevation on postabsorptive endogenous glucose production and gluconeogenesis in humans', *Diabetes*, 49 (5), 701-7.

Rohart, F., Gautier, B., Singh, A., and Le Cao, K. A. (2017), 'mixOmics: An R package for 'omics feature selection and multiple data integration', *PLoS Comput Biol*, 13 (11), e1005752.

Rosenbloom, A. L., Silverstein, J. H., Amemiya, S., Zeitler, P., and Klingensmith, G. J. (2009), 'Type 2 diabetes in children and adolescents', *Pediatr Diabetes*, 10 Suppl 12, 17-32.

Rovite, V., Wolff-Sagi, Y., Zaharenko, L., Nikitina-Zake, L., Grens, E., and Klovins, J. (2018), 'Genome Database of the Latvian Population (LGDB): Design, Goals, and Primary Results', *J Epidemiol*, 28 (8), 353-60.

Safai, N., Suvitaival, T., Ali, A., Spegel, P., Al-Majdoub, M., Carstensen, B., Vestergaard, H., Ridderstrale, M., and Group, Cimt Trial (2018), 'Effect of metformin on plasma metabolite profile in the Copenhagen Insulin and Metformin Therapy (CIMT) trial', *Diabet Med*, 35 (7), 944-53.

Saisho, Y. (2020), 'SGLT2 Inhibitors: the Star in the Treatment of Type 2 Diabetes?', *Diseases*, 8 (2).

Saito, A., Kitayama, J., Horie, H., Koinuma, K., Ohzawa, H., Yamaguchi, H., Kawahira, H., Mimura, T., Lefor, A. K., and Sata, N. (2020), 'Metformin changes the immune microenvironment of colorectal cancer in patients with type 2 diabetes mellitus', *Cancer Sci*, 111 (11), 4012-20.

Sayed, S. and Nabi, Ahmn (2021), 'Diabetes and Genetics: A Relationship Between Genetic Risk Alleles, Clinical Phenotypes and Therapeutic Approaches', *Adv Exp Med Biol*, 1307, 457-98.

Sayin, N., Kara, N., and Pekel, G. (2015), 'Ocular complications of diabetes mellitus', *World J Diabetes*, 6 (1), 92-108.

Saltiel, A. R. and Kahn, C. R. (2001), 'Insulin signalling and the regulation of glucose and lipid metabolism', *Nature*, 414 (6865), 799-806.

Saraei, P., Asadi, I., Kakar, M. A., and Moradi-Kor, N. (2019), 'The beneficial effects of metformin on cancer prevention and therapy: a comprehensive review of recent advances', *Cancer Manag Res*, 11, 3295-313.

Sattar, N. (2012), 'Biomarkers for diabetes prediction, pathogenesis or pharmacotherapy guidance? Past, present and future possibilities', *Diabet Med*, 29 (1), 5-13.

Seema Garg, Richard M. Davis (2009), 'Diabetic Retinopathy Screening Update', *Clinical Diabetes*, Oct (27(4)), 140-45.

Senda, S., Cheng, E., and Kawanishi, H. (1988), 'Aging-associated changes in murine intestinal immunoglobulin A and M secretions', *Scand J Immunol*, 27 (2), 157-64.

Senthil, S., Dada, T., Das, T., Kaushik, S., Puthuran, G. V., Philip, R., Rani, P. K., Rao, H., Singla, S., and Vijaya, L. (2021), 'Neovascular glaucoma - A review', *Indian J Ophthalmol*, 69 (3), 525-34.

Shi, Q., Liu, S., Fonseca, V. A., Thethi, T. K., and Shi, L. (2019), 'Effect of metformin on neurodegenerative disease among elderly adult US veterans with type 2 diabetes mellitus', *BMJ Open*, 9 (7), e024954.

Shikata, E., Yamamoto, R., Takane, H., Shigemasa, C., Ikeda, T., Otsubo, K., and Ieiri, I. (2007), 'Human organic cation transporter (OCT1 and OCT2) gene polymorphisms and therapeutic effects of metformin', *J Hum Genet*, 52 (2), 117-22.

Shin, N. R., Lee, J. C., Lee, H. Y., Kim, M. S., Whon, T. W., Lee, M. S., and Bae, J. W. (2014), 'An increase in the Akkermansia spp. population induced by metformin treatment improves glucose homeostasis in diet-induced obese mice', *Gut*, 63 (5), 727-35.

Silamikele, L., Silamikelis, I., Ustinova, M., Kalnina, Z., Elbere, I., Petrovska, R., Kalnina, I., and Klovins, J. (2021), 'Metformin Strongly Affects Gut Microbiome Composition in High-Fat Diet-Induced Type 2 Diabetes Mouse Model of Both Sexes', *Front Endocrinol (Lausanne)*, 12, 626359.

Sportelli, C., Urso, D., Jenner, P., and Chaudhuri, K. R. (2020), 'Metformin as a Potential Neuroprotective Agent in Prodromal Parkinson's Disease-Viewpoint', *Front Neurol*, 11, 556.

Steck, A. K. and Rewers, M. J. (2011), 'Genetics of type 1 diabetes', *Clin Chem*, 57 (2), 176-85.

Stitt, A. W. (2003), 'The role of advanced glycation in the pathogenesis of diabetic retinopathy', *Exp Mol Pathol*, 75 (1), 95-108.

Stolar, M. (2010), 'Glycemic control and complications in type 2 diabetes mellitus', *Am J Med*, 123 (3 Suppl), S3-11.

Stumvoll, M., Nurjhan, N., Perriello, G., Dailey, G., and Gerich, J. E. (1995), 'Metabolic effects of metformin in non-insulin-dependent diabetes mellitus', *N Engl J Med*, 333 (9), 550-4.

Tabara, Y., Osawa, H., Kawamoto, R., Onuma, H., Shimizu, I., Miki, T., Kohara, K., and Makino, H. (2009), 'Replication study of candidate genes associated with type 2 diabetes based on genome-wide screening', *Diabetes*, 58 (2), 493-8.

Taira, M., Imamura, M., Takahashi, A., Kamatani, Y., Yamauchi, T., Araki, S. I., Tanaka, N., van Zuydam, N. R., Ahlqvist, E., Toyoda, M., Umezono, T., Kawai, K., Imanishi, M., Watada, H., Suzuki, D., Maegawa, H., Babazono, T., Kaku, K., Kawamori, R., Consortium, Summit, Groop, L. C., McCarthy, M. I., Kadowaki, T., and Maeda, S. (2018), 'A variant within the FTO confers susceptibility to diabetic nephropathy in Japanese patients with type 2 diabetes', *PLoS One*, 13 (12), e0208654.

Tan, Y., Chen, J., Jiang, Y., Chen, X., Li, J., Chen, B., and Gao, J. (2020), 'The anti-periodontitis action of metformin via targeting NLRP3 inflammasome', *Arch Oral Biol*, 114, 104692.

Tang, Y., Lenzini, P. A., Pop-Busui, R., Ray, P. R., Campbell, H., Perkins, B. A., Callaghan, B., Wagner, M. J., Motsinger-Reif, A. A., Buse, J. B., Price, T. J., Mychaleckyj, J. C., Cresci, S., Shah, H., and Doria, A. (2019), 'A Genetic Locus on Chromosome 2q24 Predicting Peripheral Neuropathy Risk in Type 2 Diabetes: Results From the ACCORD and BARI 2D Studies', *Diabetes*, 68 (8), 1649-62.

Tarr, J. M., Kaul, K., Chopra, M., Kohner, E. M., and Chibber, R. (2013), 'Pathophysiology of diabetic retinopathy', *ISRN Ophthalmol*, 2013, 343560.

Tesfaye, S., Boulton, A. J., Dyck, P. J., Freeman, R., Horowitz, M., Kempler, P., Lauria, G., Malik, R. A., Spallone, V., Vinik, A., Bernardi, L., Valensi, P., and Toronto Diabetic Neuropathy Expert, Group (2010), 'Diabetic neuropathies: update on definitions, diagnostic criteria, estimation of severity, and treatments', *Diabetes Care*, 33 (10), 2285-93.

Thomopoulos, C., Katsimagklis, G., and Makris, T. (2017), 'Metformin and blood pressure lowering: a questioned association', *J Hypertens*, 35 (1), 27-28.

Threatt, J., Williamson, J. F., Huynh, K., Davis, R. M., and Hermayer, K. (2013), 'Ocular disease, knowledge and technology applications in patients with diabetes', *Am J Med Sci*, 345 (4), 266-70.

Tuomi, T. (2005), 'Type 1 and type 2 diabetes: what do they have in common?', *Diabetes*, 54 Suppl 2, S40-5.

Turner, R. C., Millns, H., Neil, H. A., Stratton, I. M., Manley, S. E., Matthews, D. R., and Holman, R. R. (1998), 'Risk factors for coronary artery disease in non-insulin dependent diabetes mellitus: United Kingdom Prospective Diabetes Study (UKPDS: 23)', *BMJ*, 316 (7134), 823-8.

Turner, R. M., Kwok, C. S., Chen-Turner, C., Maduakor, C. A., Singh, S., and Loke, Y. K. (2014), 'Thiazolidinediones and associated risk of bladder cancer: a systematic review and meta-analysis', *Br J Clin Pharmacol*, 78 (2), 258-73.

Udhane, S. S., Legeza, B., Marti, N., Hertig, D., Diserens, G., Nuoffer, J. M., Vermathen, P., and Fluck, C. E. (2017), 'Combined transcriptome and metabolome analyses of metformin effects reveal novel links between metabolic networks in steroidogenic systems', *Sci Rep*, 7 (1), 8652.

Unnikrishnan, A. G., Kalra, S., Purandare, V., and Vasnawala, H. (2018), 'Genital Infections with Sodium Glucose Cotransporter-2 Inhibitors: Occurrence and Management in Patients with Type 2 Diabetes Mellitus', *Indian J Endocrinol Metab*, 22 (6), 837-42.

Van Acker, K., Bouhassira, D., De Bacquer, D., Weiss, S., Matthys, K., Raemen, H., Mathieu, C., and Colin, I. M. (2009), 'Prevalence and impact on quality of life of peripheral neuropathy with or without neuropathic pain in type 1 and type 2 diabetic patients attending hospital outpatients clinics', *Diabetes Metab*, 35 (3), 206-13.

van Dieren, S., Beulens, J. W., van der Schouw, Y. T., Grobbee, D. E., and Neal, B. (2010), 'The global burden of diabetes and its complications: an emerging pandemic', *Eur J Cardiovasc Prev Rehabil*, 17 Suppl 1, S3-8.

van Zuydam, N. R., Ahlqvist, E., Sandholm, N., Deshmukh, H., Rayner, N. W., Abdalla, M., Ladenvall, C., Ziemek, D., Fauman, E., Robertson, N. R., McKeigue, P. M., Valo, E., Forsblom, C., Harjutsalo, V., Finnish Diabetic Nephropathy, Study, Perna, A., Rurali, E., Marcovecchio, M. L., Igo, R. P., Jr., Salem, R. M., Perico, N., Lajer, M., Karajamaki, A., Imamura, M., Kubo, M., Takahashi, A., Sim, X., Liu, J., van Dam, R. M., Jiang, G., Tam, C. H. T., Luk, A. O. Y., Lee, H. M., Lim, C. K. P., Szeto, C. C., So, W. Y., Chan, J. C. N., Hong Kong Diabetes Registry Theme-based Research Scheme Project, Group, Ang, S. F., Dorajoo, R., Wang, L., Clara, T. S. H., McKnight, A. J., Duffy, S., Warren, Genetics of Kidneys in Diabetes Study, Group, Pezzolesi, M. G., Consortium, Genie, Marre, M., Gyorgy, B., Hadjadj, S., Hiraki, L. T., Diabetes, Control, Complications Trial /Epidemiology of Diabetes, Interventions, Complications Research, Group, Ahluwalia, T. S., Almgren, P., Schulz, C. A., Orho-Melander, M., Linneberg, A., Christensen, C., Witte, D. R., Grarup, N., Brandslund, I., Melander, O., Paterson, A. D., Tregouet, D., Maxwell, A. P., Lim, S. C., Ma, R. C. W., Tai, E. S., Maeda, S., Lyssenko, V., Tuomi, T., Krolewski, A. S., Rich, S. S., Hirschhorn, J. N., Florez, J. C., Dunger, D., Pedersen, O., Hansen, T., Rossing, P., Remuzzi, G., Micro, S. Urrogate markers for, Macrovascular hard endpoints for Innovative diabetes Tools, Consortium, Brosnan, M. J., Palmer, C. N. A., Groop, P. H., Colhoun, H. M., Groop, L. C., and McCarthy, M. I. (2018), 'A Genome-Wide Association Study of Diabetic Kidney Disease in Subjects With Type 2 Diabetes', *Diabetes*, 67 (7), 1414-27.

Vardarli, I., Arndt, E., Deacon, C. F., Holst, J. J., and Nauck, M. A. (2014), 'Effects of sitagliptin and metformin treatment on incretin hormone and insulin secretory responses to oral and "isoglycemic" intravenous glucose', *Diabetes*, 63 (2), 663-74.

Vial, G., Detaille, D., and Guigas, B. (2019), 'Role of Mitochondria in the Mechanism(s) of Action of Metformin', *Front Endocrinol (Lausanne)*, 10, 294.

Wallach, J. D., Wang, K., Zhang, A. D., Cheng, D., Grossetta Nardini, H. K., Lin, H., Bracken, M. B., Desai, M., Krumholz, H. M., and Ross, J. S. (2020), 'Updating insights into rosiglitazone and cardiovascular risk through shared data: individual patient and summary level meta-analyses', *BMJ*, 368, l7078.

Wang, Y., An, H., Liu, T., Qin, C., Sesaki, H., Guo, S., Radovick, S., Hussain, M., Maheshwari, A., Wondisford, F. E., O'Rourke, B., and He, L. (2019), 'Metformin Improves Mitochondrial Respiratory Activity through Activation of AMPK', *Cell Rep*, 29 (6), 1511-23 e5.

Wang, Y. W., He, S. J., Feng, X., Cheng, J., Luo, Y. T., Tian, L., and Huang, Q. (2017), 'Metformin: a review of its potential indications', *Drug Des Devel Ther*, 11, 2421-29.

Wattanakit, K., Folsom, A. R., Selvin, E., Weatherley, B. D., Pankow, J. S., Brancati, F. L., and Hirsch, A. T. (2005), 'Risk factors for peripheral arterial disease incidence in persons with diabetes: the Atherosclerosis Risk in Communities (ARIC) Study', *Atherosclerosis*, 180 (2), 389-97.

Weisberg, S. P., Hunter, D., Huber, R., Lemieux, J., Slaymaker, S., Vaddi, K., Charo, I., Leibel, R. L., and Ferrante, A. W., Jr. (2006), 'CCR2 modulates inflammatory and metabolic effects of high-fat feeding', *J Clin Invest*, 116 (1), 115-24.

WHO, Geneva: (2019), 'Classification of diabetes mellitus'.

Winckler, W., Weedon, M. N., Graham, R. R., McCarroll, S. A., Purcell, S., Almgren, P., Tuomi, T., Gaudet, D., Bostrom, K. B., Walker, M., Hitman, G., Hattersley, A. T., McCarthy, M. I., Ardlie, K. G., Hirschhorn, J. N., Daly, M. J., Frayling, T. M., Groop, L., and Altshuler, D. (2007), 'Evaluation of common variants in the six known maturity-onset diabetes of the young (MODY) genes for association with type 2 diabetes', *Diabetes*, 56 (3), 685-93.

Wong, R. H. and Sul, H. S. (2010), 'Insulin signaling in fatty acid and fat synthesis: a transcriptional perspective', *Curr Opin Pharmacol*, 10 (6), 684-91.

Wright, J. J. and Tylee, T. S. (2016), 'Pharmacologic Therapy of Type 2 Diabetes', *Med Clin North Am*, 100 (4), 647-63.

Wu, H., Esteve, E., Tremaroli, V., Khan, M. T., Caesar, R., Manneras-Holm, L., Stahlman, M., Olsson, L. M., Serino, M., Planas-Felix, M., Xifra, G., Mercader, J. M., Torrents, D., Burcelin, R., Ricart, W., Perkins, R., Fernandez-Real, J. M., and Backhed, F. (2017), 'Metformin alters the gut microbiome of individuals with treatment-naive type 2 diabetes, contributing to the therapeutic effects of the drug', *Nat Med*, 23 (7), 850-58.

Xue, A., Wu, Y., Zhu, Z., Zhang, F., Kemper, K. E., Zheng, Z., Yengo, L., Lloyd-Jones, L. R., Sidorenko, J., Wu, Y., e, QTLGen Consortium, McRae, A. F., Visscher, P. M., Zeng, J., and Yang, J. (2018), 'Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes', *Nat Commun*, 9 (1), 2941.

Zeggini, E., Scott, L. J., Saxena, R., Voight, B. F., Marchini, J. L., Hu, T., de Bakker, P. I., Abecasis, G. R., Almgren, P., Andersen, G., Ardlie, K., Bostrom, K. B., Bergman, R. N., Bonnycastle, L. L., Borch-Johnsen, K., Burtt, N. P., Chen, H., Chines, P. S., Daly, M. J., Deodhar, P., Ding, C. J., Doney, A. S., Duren, W. L., Elliott, K. S., Erdos, M. R., Frayling, T. M., Freathy, R. M., Gianniny, L., Grallert, H., Grarup, N., Groves, C. J., Guiducci, C., Hansen, T., Herder, C., Hitman, G. A., Hughes, T. E., Isomaa, B., Jackson, A. U., Jorgensen, T., Kong, A., Kubalanza, K., Kuruvilla, F. G., Kuusisto, J., Langenberg, C., Lango, H., Lauritzen, T., Li, Y., Lindgren, C. M., Lyssenko, V., Marvelle, A. F., Meisinger, C., Midthjell, K., Mohlke, K. L., Morken, M. A., Morris, A. D., Narisu, N., Nilsson, P., Owen, K. R., Palmer, C. N., Payne, F., Perry, J. R., Pettersen, E., Platou, C., Prokopenko, I., Qi, L., Qin, L., Rayner, N. W., Rees, M., Roix, J. J., Sandbaek, A., Shields, B., Sjogren, M., Steinthorsdottir, V., Stringham, H. M., Swift, A. J., Thorleifsson, G., Thorsteinsdottir, U., Timpson, N. J., Tuomi, T., Tuomilehto, J., Walker, M., Watanabe, R. M., Weedon, M. N., Willer, C. J., Wellcome Trust Case Control, Consortium, Illig, T., Hveem, K., Hu, F. B., Laakso, M., Stefansson, K., Pedersen, O., Wareham, N. J., Barroso, I., Hattersley, A. T., Collins, F. S., Groop, L., McCarthy, M. I., Boehnke, M., and Altshuler, D. (2008), 'Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes', *Nat Genet*, 40 (5), 638-45.

Zhang, Y., Hong, J., Chi, J., Gu, W., Ning, G., and Wang, W. (2014a), 'Head-to-head comparison of dipeptidyl peptidase-IV inhibitors and sulfonylureas - a meta-analysis from randomized clinical trials', *Diabetes Metab Res Rev*, 30 (3), 241-56.

Zhang, M., Liu, Y. H., Chang, C. S., Zhi, H., Wang, S., Xu, W., Smith, C. W., and Zhang, H. B. (2019), 'Quantification of gene expression while taking into account RNA alternative splicing', *Genomics*, 111 (6), 1517-28.

Zhang, Z. J., Bi, Y., Li, S., Zhang, Q., Zhao, G., Guo, Y., and Song, Q. (2014b), 'Reduced risk of lung cancer with metformin therapy in diabetic patients: a systematic review and meta-analysis', *Am J Epidemiol*, 180 (1), 11-4.

Zhao, W., Wineinger, N. E., Tiwari, H. K., Mosley, T. H., Broeckel, U., Arnett, D. K., Kardia, S. L., Kabagambe, E. K., and Sun, Y. V. (2012), 'Copy number variations associated with obesity-related traits in African Americans: a joint analysis between GENOA and HyperGEN', *Obesity (Silver Spring)*, 20 (12), 2431-7.

Zhou, K., Donnelly, L., Yang, J., Li, M., Deshmukh, H., Van Zuydam, N., Ahlqvist, E., Wellcome Trust Case Control, Consortium, Spencer, C. C., Groop, L., Morris, A. D., Colhoun, H. M., Sham, P. C., McCarthy, M. I., Palmer, C. N., and Pearson, E. R. (2014a), 'Heritability of variation in glycaemic response to metformin: a genome-wide complex trait analysis', *Lancet Diabetes Endocrinol*, 2 (6), 481-7.

Zhou, K., Yee, S. W., Seiser, E. L., van Leeuwen, N., Tavendale, R., Bennett, A. J., Groves, C. J., Coleman, R. L., van der Heijden, A. A., Beulens, J. W., de Keyser, C. E., Zaharenko, L., Rotroff, D. M., Out, M., Jablonski, K. A., Chen, L., Javorsky, M., Zidzik, J., Levin, A. M., Williams, L. K., Dujic, T., Semiz, S., Kubo, M., Chien, H. C., Maeda, S., Witte, J. S., Wu, L., Tkac, I., Kooy, A., van Schaik, R. H. N., Stehouwer, C. D. A., Logie, L., MetGen, Investigators, Investigators, D. P. P., Investigators, Accord, Sutherland, C., Klovins, J., Pirags, V., Hofman, A., Stricker, B. H., Motsinger-Reif, A. A., Wagner, M. J., Innocenti, F., t Hart, L. M., Holman, R. R., McCarthy, M. I., Hedderson, M. M., Palmer, C. N. A., Florez, J. C., Giacomini, K. M., and Pearson, E. R. (2016), 'Variation in the glucose transporter gene SLC2A2 is associated with glycemic response to metformin', *Nat Genet*, 48 (9), 1055-59.

Zhou, X., Lindsay, H., and Robinson, M. D. (2014b), 'Robustly detecting differential expression in RNA sequencing data using observation weights', *Nucleic Acids Res*, 42 (11), e91.

Zhu, X., Shen, W., Liu, Z., Sheng, S., Xiong, W., He, R., Zhang, X., Ma, L., and Ju, Z. (2020), 'Effect of Metformin on Cardiac Metabolism and Longevity in Aged Female Mice', *Front Cell Dev Biol*, 8, 626011.

Zimmet, P., Alberti, K. G., and Shaw, J. (2001), 'Global and societal implications of the diabetes epidemic', *Nature*, 414 (6865), 782-7.

APPENDICES

APPENDIX I

List of inclusion and exclusion criteria of the study cohorts in metformin studies

Metformin strongly affects transcriptome of peripheral blood cells in healthy individuals
Monta Ustinova, Ivars Silamikelis, Ineta Kalnina, Laura Ansone, Vita Rovite, Ilze Elbere, Ilze Radovica-Spalvina, Davids Fridmanis, Jekaterina Aladyeva, Ilze Konrade, Valdis Pirags, Janis Klovins

S1 Table. Inclusion and exclusion criteria for the cohort of healthy individuals

Principal inclusion criteria
1. Healthy person with no known illnesses at the time of application that could possibly alter the results of the study;
2. Body characteristic parameters (e.g. weight) are within the conventional range;
3. Mental condition allows a person to understand the research process and give a legal consent for the participation in it;
4. Age: 18 – 64 years;
5. European descent;
6. Both women and men with reproductive potential correspond to the contraceptive requirements stated in the study protocol;
7. Prior to the study-related procedures, the consent of a person's participation in the clinical trial is received by submitting a signed and dated informed consent document.
Principal exclusion criteria
1. Allergies to any of *Metforal* 850mg components;
2. Usage of any other medication which is not compatible with *Metforal* 850mg;
3. Pregnancy or lactation;
4. Type 1 or type 2 diabetes, pancreatogenic diabetes, impaired glucose tolerance;
5. Polycystic ovarian syndrome;
6. Chronic diseases of intestinal tract, oncological or autoimmune diseases;
7. Renal failure or renal impairment;
8. Hepatic impairment or alcoholism;
9. Acute conditions with the potential effects on kidney;
10. Acute or chronic disease which may cause tissue hypoxia;
11. Diarrhea during the past week;
12. Long term previous administration of metformin;
13. Use of the following products during the previous two months: a. antibiotics, b. probiotics, c. proton pump inhibitors, d. immunosuppressive agents, e. corticosteroids;
14. Intravascular administration of iodinated contrast agents intendedduring the active period of the clinical trial.
ALAT - alanine aminotransferase; HbA1c - hemoglobin A1c

Whole-blood transcriptome profiling reveals signatures of metformin and its therapeutic response

Monta Ustinova, Laura Ansone, Ivars Silamikelis, Vita Rovite, Ilze Elbere, Laila Silamikele, Ineta Kalnina, Davids Fridmanis, Jelizaveta Sokolovska, Ilze Konrade, Valdis Pirags, Janis Klovins

S2 Table. Inclusion and exclusion criteria for the cohort of type 2 diabetes patients

Principal inclusion criteria
1. (a) Newly diagnosed type 2 diabetes mellitus (ICD-10 code E11) requiring oral antidiabetic therapy, or (b) previously diagnosed type 2 diabetes mellitus but no oral antidiabetic therapy or insulin has been used for the last three years, or (c) newly diagnosed type 2 diabetes mellitus and intensive insulin therapy initiated in a hospital for acute glycemic normalization;
2. The patient is not currently involved and is not planning to enroll in clinical trials during the OPTIMED study;
3. The patient has attained 18 years of age;
4. The patient is not pregnant at the time of application;
5. The patient meets the criteria for the diagnosis of type 2 diabetes mellitus: (a) fasting blood glucose level ≥ 7 mmol/l, (b) a blood glucose level ≥ 11.1 mmol/l for a two-hour glucose tolerance test with 75 g intake;
6. Prior to the study-related procedures, the consent of a person's participation in the study is received by submitting a signed and dated informed consent document.
Principal exclusion criteria
1. The patient is receiving oral antidiabetic therapy on a regular basis or has received the therapy during the last three years;
2. The patient is receiving insulin therapy at the time of application;
3. The patient is pregnant.

Novel Susceptibility Loci Identified In A Genome-Wide Association Study Of Type 2 Diabetes Complications In Population of Latvia

Monta Ustinova, Raitis Peculis, Raimonds Rescenko, Vita Rovite, Linda Zaharenko, Ilze Elbere, Laila Silamikele, Ilze Konrade, Jelizaveta Sokolovska, Valdis Pirags, Janis Klovins

S3 Table. Inclusion and exclusion criteria of the Genome-wide association study

Principal inclusion criteria
1. Clinically confirmed diagnosis of type 2 diabetes mellitus (E11 diagnosis code, ICD-10);
2. Information on age at diagnosis, sex, weight, and height available;
3. Records of national diabetes registry and Latvian hospital inpatient discharges available;

APPENDIX II

List of candidate variants selected for the targeted association analysis

Novel Susceptibility Loci Identified In A Genome-Wide Association Study Of Type 2 Diabetes Complications In Population of Latvia

Monta Ustinova, Raitis Peculis, Vita Rovite, Linda Zaharenko, Ilze Elbere, Laila Silamikele, Ilze Konrade, Jelizaveta Sokolovska, Valdis Pirags, Janis Klovins

S4. Table. List of candidate variants selected for the targeted association analysis

Candidate variants associated with ophthalmic complications					
SNP	**Chromosome**	**Position (Grch 38)**	**Closest gene**	**Reported trait**	**Reference**
rs12602486	17	44164561	*C17orf53, ASB16*	diabetic retinopathy	Chen P et al. 2013
rs61741249	7	157516897	*AC006372.2*	diabetic retinopathy	Pollack S et al. 2018
rs74705672	8	132042380	*AC100868.1, OC90*	proliferative diabetic retinopathy	Pollack S et al. 2018
rs2064196	6	144266047	*UTRN, AL024474.1*	proliferative diabetic retinopathy	Pollack S et al. 2018
rs61811867	1	154802768	*KCNN3*	proliferative diabetic retinopathy	Pollack S et al. 2018
rs11488711	1	229105486	*LINC02814*	proliferative diabetic retinopathy	Pollack S et al. 2018
rs1414474	1	34197810	*C1orf94*	proliferative diabetic retinopathy	Pollack S et al. 2018
rs11018670	11	89623460	*FOLH1B, AP003400.1*	diabetic retinopathy, type II diabetes mellitus	Meng W et al. 2018
rs10927101	1	244010570	*LINC02774*	diabetic retinopathy	Grassi MA et al. 2011
rs10403021	19	29588697	*VSTM2B, POP4*	diabetic retinopathy	Grassi MA et al. 2011
rs7139352	12	60951968	*AC090017.1, AC090022.2*	diabetic retinopathy	Pollack S et al. 2018
rs9882204	3	167848495	*LRRC77P, SERPINI1*	diabetic retinopathy	Pollack S et al. 2018
rs10878791	12	68215266	*IFNG-AS1, IL26*	diabetic retinopathy	Pollack S et al. 2018
rs11575234	12	56350492	*STAT2*	proliferative diabetic retinopathy	Pollack S et al. 2018
rs1566115	6	99184447	*FAXC, BDH2P1*	proliferative diabetic retinopathy	Pollack S et al. 2018
rs2037601	2	195417334	*AC010983.1, AC104823.1*	proliferative diabetic retinopathy	Pollack S et al. 2018
rs1046896	17	82727657	*FN3KRP*	diabetic retinopathy	Chen P et al. 2013
rs4838605	10	48491914	*ARHGAP22*	diabetic retinopathy	Huang YC et al. 2011
rs17376456	5	94221997	*KIAA0825*	diabetic retinopathy	Huang YC et al. 2011
rs2696835	16	86331965	*LINC00917*	diabetic retinopathy	Grassi MA et al. 2011
rs7772697	6	149113975	*AL603766.1, AL031056.1*	diabetic retinopathy	Grassi MA et al. 2011
rs10199521	2	2515741	*MYT1L, AC018685.2*	diabetic retinopathy	Grassi MA et al. 2011
rs1571942	10	20253705	*PLXDC2*	diabetic retinopathy	Huang YC et al. 2011
rs74152685	1	247647989	*AL390860.1*	diabetic retinopathy	Pollack S et al. 2018
rs7533141	1	217543535	*GPATCH2*	proliferative diabetic retinopathy	Pollack S et al. 2018
rs1342038	1	173332377	*AL645568.3, TNFSF4*	diabetic retinopathy	Grassi MA et al. 2011
rs6702784	1	36439119	*OSCP1*	diabetic retinopathy	Grassi MA et al. 2011
rs6763376	3	167886331	*LRRC77P, AC026353.1*	diabetic retinopathy	Pollack S et al. 2018
rs1990145	2	75650524	*MRPL19*	diabetic macular edema, type II diabetes mellitus	Graham PS et al. 2018
rs1144964	12	68944857	*CPM*	proliferative diabetic retinopathy	Pollack S et al. 2018

rs2038823	13	:96299179	*HS6ST3*	diabetic retinopathy	Huang YC et al. 2011
rs3007729	1	18468761	*AL591896.1, KLHDC7A*	diabetic retinopathy	Grassi MA et al. 2011
rs2300993	5	109702928	*MAN2A1*	diabetic retinopathy	Pollack S et al. 2018
rs73050171	3	174801688	*NAALADL2*	proliferative diabetic retinopathy	Pollack S et al. 2018
rs1326934	10	95524324	*SORBS1*	age-related macular degeneration, type II diabetes	Chen W et al. 2010, Scott et al. 2007
rs1799884	7	44189469	*GCK*	diabetic retinopathy	Chen P et al. 2013
rs11765845	7	28351523	*CREB5*	diabetic retinopathy	Grassi MA et al. 2011
rs9543976	13	75562512	*AL137782.1, UCHL3*	diabetic retinopathy	Sheu WH et al. 2013
rs13064954	3	157136953	*CCNL1, LINC00881*	diabetic retinopathy	Grassi MA et al. 2011
rs1000708	12	59689707	*SLC16A7*	proliferative diabetic retinopathy	Pollack S et al. 2018
rs71354195	19	36385416	*ZFP82*	proliferative diabetic retinopathy	Pollack S et al. 2018
rs11201335	10	85018443	*AL358787.2, RPS3AP5*	proliferative diabetic retinopathy	Pollack S et al. 2018
rs1073203	5	125983763	*AC116362.1*	diabetic retinopathy	Grassi MA et al. 2011
rs12219125	10	20304158	*AMD1P1, PLXDC2*	diabetic retinopathy	Huang YC et al. 2011
rs10509061	10	56782127	*ZWINT, MIR3924*	diabetic retinopathy	Fu YP et al. 2010
rs2296041	13	43412635	*ENOX1*	diabetic retinopathy	Fu YP et al. 2010
rs3790568	67	67370377	*IL12RB2*	diabetic retinopathy	Fu YP et al. 2010
rs1463075	11	41610312	*LRRC4C, LOC100507205*	diabetic retinopathy	Fu YP et al. 2010
rs7184426	16	82788303	*CDH13*	diabetic retinopathy	Fu YP et al. 2010
rs10501855	11	96510025	*JRKL, NONE*	diabetic retinopathy	Fu YP et al. 2010
rs899036	11	41661360	*LRRC4C, LOC100507205*	diabetic retinopathy	Fu YP et al. 2010
rs1197310	3	133409380	*BFSP2*	diabetic retinopathy	Fu YP et al. 2010
rs10502019	11	103842804	*DYNC2H1, MIR4693*	diabetic retinopathy	Fu YP et al. 2010
rs2202802	6	54302883	*MLIP, TINAG*	diabetic retinopathy	Fu YP et al. 2010
rs10499143	6	128896035	*LAMA2*	diabetic retinopathy	Fu YP et al. 2010
rs3098224	8	103434468	*DCAF13*	diabetic retinopathy	Fu YP et al. 2010
rs7777578	7	145979789	*TPK1, CNTNAP2*	diabetic retinopathy	Fu YP et al. 2010
rs436563	9	10100851	*PTPRD*	diabetic retinopathy	Fu YP et al. 2010
rs10497168	2	156179203	*KCNJ3, NR4A2*	diabetic retinopathy	Fu YP et al. 2010
rs7923214	10	2251795	*LOC282980, LOC399708*	diabetic retinopathy	Fu YP et al. 2010
rs2320761	13	32909156	*PDS5B, KL*	diabetic retinopathy	Fu YP et al. 2010
rs1351253	11	80343490	*ODZ4, MIR4300*	diabetic retinopathy	Fu YP et al. 2010
rs1920241	3	189576370	*TPRG1, TPRG1*	diabetic retinopathy	Fu YP et al. 2010
rs2893603	8	127067712	*FAM84B, POU5F1B*	diabetic retinopathy	Fu YP et al. 2010
rs4074672	3	184012507	*ABCC5*	diabetic retinopathy	Fu YP et al. 2010
rs1347637	3	21267223	*SGOL1, VENTXP7*	diabetic retinopathy	Fu YP et al. 2010
rs2542930	2	210843606	*CPS1, ERBB4*	diabetic retinopathy	Fu YP et al. 2010
rs1436302	2	125416354	*CNTNAP5, GYPC*	diabetic retinopathy	Fu YP et al. 2010
rs9290870	3	188138614	*BCL6, LOC339929*	diabetic retinopathy	Fu YP et al. 2010
rs1370005	8	103432781	*DCAF13*	diabetic retinopathy	Fu YP et al. 2010
rs10488574	7	76294675	*SRRM3,HSPB1*	diabetic retinopathy	Fu YP et al. 2010
rs1889031	10	36558325	*ANKRD30A,FZD8*	diabetic retinopathy	Fu YP et al. 2010
rs2505585	10	36560065	*ANKRD30A,FZD8*	diabetic retinopathy	Fu YP et al. 2010
rs3761296	20	16729017	*SNRPB2*	diabetic retinopathy	Fu YP et al. 2010
rs2300782	5	111453087	*CAMK4*	diabetic retinopathy	Fu YP et al. 2010
rs1537328	10	36560664	*ANKRD30A,FZD8*	diabetic retinopathy	Fu YP et al. 2010
rs2483021	10	36558204	*ANKRD30A,FZD8*	diabetic retinopathy	Fu YP et al. 2010
rs10495528	2	5917715	*LOC150622, SOX11*	diabetic retinopathy	Fu YP et al. 2010
rs1558068	7	30463747	*NOD1*	diabetic retinopathy	Fu YP et al. 2010
rs1317318	15	52840866	*ONECUT1, WDR72*	diabetic retinopathy	Fu YP et al. 2010
rs2023139	8	117388472	*MED30, SLC30A8*	diabetic retinopathy	Fu YP et al. 2010
rs724333	3	55924031	*ERC2*	diabetic retinopathy	Fu YP et al. 2010
rs1424548	7	137025013	*CHRM2, PTN*	diabetic retinopathy	Fu YP et al. 2010
rs7921283	10	56790127	*ZWINT, MIR3924*	diabetic retinopathy	Fu YP et al. 2010
rs2226859	18	73376963	*LOC100505817, FBXO15*	diabetic retinopathy	Fu YP et al. 2010
rs10504455	8	69558367	*SULF1*	diabetic retinopathy	Fu YP et al. 2010
rs6544021	2	22530353	*LOC645949, KLHL29*	diabetic retinopathy	Fu YP et al. 2010
rs10504454	8	69558119	*SULF1*	diabetic retinopathy	Fu YP et al. 2010
rs957940	13	53727505	*MIR1297, OLFM4*	diabetic retinopathy	Fu YP et al. 2010
rs764637	2	210776279	*ERBB4, CPS1*	diabetic retinopathy	Fu YP et al. 2010
rs2197713	3	74120197	*CNTN3, PDZRN3*	diabetic retinopathy	Fu YP et al. 2010
rs10492603	13	58297687	*PCDH17, DIAPH3*	diabetic retinopathy	Fu YP et al. 2010
rs1428445	5	158568795	*EBF1, CLINT1*	diabetic retinopathy	Fu YP et al. 2010
rs10520201	4	171183705	*LOC100506122, GALNTL6*	diabetic retinopathy	Fu YP et al. 2010
rs1445754	5	84279813	*EDIL3*	diabetic retinopathy	Fu YP et al. 2010

rs10509060	10	56781676	*ZWINT, MIR3924*	diabetic retinopathy	Fu YP et al. 2010
rs10485129	6	78661466	*IRAK1BP1, HTR1B*	diabetic retinopathy	Fu YP et al. 2010
rs6427247	1	170411339	*GORAB, LOC284688*	diabetic retinopathy	Fu YP et al. 2010

Candidate variants associated with macrovascular complications

SNP	Chromosome	Position (Grch 38)	Closest gene	Reported trait	Reference
rs113805659	7	78663475	*MAGI2*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs16879003	6	16745008	*ATXN1*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs6506897	18	31215320	*DSG1, DSC1*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs9973676	2	230954091	*GPR55*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs91	7	24409992	*AC003044.1*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs59490629	6	107256268	*PDSS2*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs7804216	7	149941248	*AC092681.1, AC092681.2*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs74570061	4	56517605	*ARL9*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs139034117	3	39656373	*NFU1P1, MYRIP*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs162211	3	7781368	*AC077690.1*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs1583674	7	47139764	*AC087175.1, AC004870.3*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs7158214	14	96176310	*BDKRB2, C14orf132*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs7632954	3	8486196	*LMCD1-AS1*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs7806345	7	145029551	*RPL7P59, AC073310.1*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs11794666	9	31665127	*RNA5SP281, HMGB3P23*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs7906829	10	116588374	*PNLIPRP1, PNLIPP1*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs9575182	13	82832889	*AL445255.1*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs17865343	4	117351922	*LINC02262*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs2546338	5	133866276	*WSPAR, AC005178.1*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs7901181	10	3600363	*AL357833.1, AL450322.2*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs10978777	9	107268309	*RAD23B, AL445487.1*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs543554	13	32127501	*FRY*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs17214144	7	24846128	*OSBPL3*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs7689240	4	35313869	*AC020589.1, SEC63P2*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017

rs846178	4	2483888	*RNF4*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs277939	5	71560551	*BDP1*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs6446601	4	4298683	*ZBTB49*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs140365914	7	155314860	*BLACE, INSIG1*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs73307775	5	159299527	*AC008691.1, UBLCP1*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs6993194	8	71117967	*AC015687.1*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs10793968	9	130754074	*ABL1*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs111915800	4	56521209	*ARL9*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs9715521	4	58958769	*LINC02619, AC017091.1*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs597539	11	68834506	*CPT1A*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs75107833	2	174100883	*OLA1*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs17295603	4	174540366	*AC096751.2*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs6445525	3	66007429	*MAGI1*	coronary artery calcification, type II diabetes mellitus	Divers J et al. 2017
rs12526453	6	12927312	*PHACTR1*	myocardial infarction, coronary heart disease	Kathiresan S et al. 2009, Schunkert H et al. 2011
rs646776	1	109275908	*PSRC1, CELSR2*	coronary heart disease, myocardial infarction	Reilly MP et al. 2011, Kathiresan S et al. 2009
rs2259816	12	120997784	*HNF1A*	coronary heart disease, C-reactive protein measurement	Erdmann J et al. 2009
rs11206510	1	55030366	*PCSK9, BSND*	myocardial infarction, coronary heart disease, coronary artery disease	Kathiresan S et al. 2009, Schunkert H et al. 2011
rs10911021	1	182112825	*LINC01344*	cardiovascular disease, type II diabetes mellitus	Qi L et al. 2013
rs266729	3	186841685	*ADIPOQ*	ischemic heart disease and stroke, type II diabetes mellitus	Montesanto et al. 2018

Candidate variants associated with diabetic nephropathy

SNP	Chromosome	Position (Grch 38)	Closest gene	Reported trait	Reference
rs11107616	12	77971000	*NAV3*	diabetic nephropathy	Iyengar SK et al. 2015
rs4667466	2	162832637	*KCNH7*	diabetic nephropathy	Iyengar SK et al. 2015
rs73206603	13	59495931	*DIAPH3, RNU7-88P*	chronic kidney disease, type II diabetes mellitus, diabetic nephropathy	van Zuydam NR et al. 2018
rs7222331	17	40995605	*KRTAP3-2, KRTAP3-3*	diabetes mellitus, diabetic nephropathy	van Zuydam NR et al. 2018
rs955333	6	154626274	*AL591419.1, AL591499.1*	diabetic nephropathy	Iyengar SK et al. 2015
rs7769051	6	132825657	*RPS12, HMGB1P13*	diabetic nephropathy	McDonough CW et al. 2010
rs136161	22	36261386	*APOL1*	diabetic nephropathy	Iyengar SK et al. 2015
rs590884	6	160957160	*AL139393.1*	diabetic nephropathy	Iyengar SK et al. 2015

rs6943931	7	22375027	*RAPGEF5, STEAP1B*	diabetic nephropathy	Iyengar SK et al. 2015
rs10778560	12	107407624	*BTBD11*	diabetic nephropathy	Iyengar SK et al. 2015
rs11961816	6	139003738	*REPS1, ABRACL*	chronic kidney disease, diabetic nephropathy	van Zuydam NR et al. 2018
rs76262407	19	34600451	*SCGB2B2*	chronic kidney disease, diabetic nephropathy	van Zuydam NR et al. 2018
rs10404821	19	50657831	*C19orf81*	chronic kidney disease, diabetic nephropathy	van Zuydam NR et al. 2018
rs16977473	15	57404428	*CGNL1*	Microalbuminuria, type 2 diabetes nephropathy	Taira M et al. 2018
rs11582609	1	82754973	*AL157944.1*	diabetic nephropathy	Iyengar SK et al. 2015
rs13421350	2	172453843	*ITGA6*	diabetic nephropathy	Iyengar SK et al. 2015
rs12497655	3	117045163	*LSAMP*	diabetic nephropathy	Iyengar SK et al. 2015
rs2780902	1	64863417	*JAK1*	diabetic nephropathy	Iyengar SK et al. 2015
rs12764441	10	70931894	*AC073176.1, AC073176.2*	microalbuminuria	Teumer A et al. 2015
rs7145202	14	22623202	*OR6J1, ABHD4*	microalbuminuria	Teumer A et al. 2015
rs6513791	20	42429482	*PTPRT*	microalbuminuria	Teumer A et al. 2015
rs1997066	10	105003805	*SORCS3*	diabetic nephropathy	Iyengar SK et al. 2015
rs731565	7	147709170	*CNTNAP2*	diabetic nephropathy	Iyengar SK et al. 2015
rs768920	11	93008790	*MTNR1B, RPL26P31*	diabetic nephropathy, type II diabetes mellitus, diabetes mellitus	van Zuydam NR et al. 2018
rs12472051	2	113974664	*LINC01191*	chronic kidney disease, diabetic nephropathy	van Zuydam NR et al. 2018
rs10144968	14	68684168	*RAD51B, AL121820.1*	type 2 diabetes nephropathy	Taira M et al. 2018
rs1077216	3	46850671	*MYL3*	microalbuminuria	Teumer A et al. 2015
rs6432852	2	165898043	*TTC21B*	diabetic nephropathy	Iyengar SK et al. 2015
rs10766496	11	18717171	*AC103974.1, IGSF22*	diabetic nephropathy	Iyengar SK et al. 2015
rs1424609	12	115725627	*AC009387.1, RN7SL865P*	diabetic nephropathy	Iyengar SK et al. 2015
rs6962291	7	139971318	*TBXAS1*	diabetic nephropathy	Iyengar SK et al. 2015
rs10911184	1	182996532	*RNU6-41P, LAMC1*	diabetic nephropathy	Iyengar SK et al. 2015
rs73048515	5	4741874	*AC106799.2*	microalbuminuria, type 2 diabetes nephropathy	Taira M et al. 2018
rs10019835	4	155712034	*GUCY1A1*	diabetic nephropathy	Iyengar SK et al. 2015
rs12719264	5	119848245	*AC008574.1, AC008550.1*	microalbuminuria	Teumer A et al. 2015
rs773506	9	91213189	*AUH, AL158071.3*	diabetic nephropathy	McDonough CW et al. 2010
rs4849965	2	4561337	*NPM1P48, LINC01249*	diabetic nephropathy	Iyengar SK et al. 2015
rs11105956	12	91053746	*KERA*	diabetic nephropathy	Iyengar SK et al. 2015
rs7916840	10	25424152	*GPR158*	diabetic nephropathy, type II diabetes mellitus	van Zuydam NR et al. 2018
rs13254600	8	123077286	*TBC1D31*	diabetic nephropathy	Iyengar SK et al. 2015
rs73017308	3	12213890	*SYN2, GSTM5P1*	chronic kidney disease, diabetic nephropathy	van Zuydam NR et al. 2018
rs6930576	6	148383818	*SASH1*	diabetic nephropathy	McDonough CW et al. 2010
rs56094641	16	53772541	*FTO*	type 2 diabetes nephropathy, microalbuminuria	Taira M et al. 2018

rs10004231	4	42855556	*AC096734.1, RN7SKP82*	diabetic nephropathy	Iyengar SK et al. 2015
rs2243480	7	66134209	*CRCP, AC068533.4, AC068533.3*	diabetic nephropathy	Iyengar SK et al. 2015
rs6910061	6	11101685	*SMIM13*	diabetic nephropathy	Iyengar SK et al. 2015
rs8020941	14	97067025	*AL049833.2, AL049833.4*	diabetic nephropathy	Iyengar SK et al. 2015
rs1677894	12	77944156	*NAV3*	diabetic nephropathy	Iyengar SK et al. 2015
rs12251637	10	124549008	*LHPP*	chronic kidney disease in type 2 diabetes	van Zuydam NR et al. 2018

Candidate variants associated with diabetic neuropathy

SNP	Chromosome	Position (Grch 38)	Closest gene	Reported trait	Reference
rs11615866	12	5284163	*LOC105369617*	neuropathic pain, type 2 diabetes mellitus	Meng W et al. 2014
rs28485846	8	35265058	*UNC5D*	type I diabetes mellitus, diabetic foot, type II diabetes mellitus, neuropathy	Meng W et al. 2017
rs80028505	6	36030611	*MAPK14*	type I diabetes mellitus, diabetic foot, type II diabetes mellitus, neuropathy	Meng W et al. 2017
rs3761980	6	36026129	*SLC26A8, MAPK14*	type I diabetes mellitus, diabetic foot, type II diabetes mellitus, neuropathy	Meng W et al. 2017

Chen P, Ong RT, Tay WT, Sim X, Ali M, Xu H, et al. A study assessing the association of glycated hemoglobin A1C (HbA1C) associated variants with HbA1C, chronic kidney disease and diabetic retinopathy in populations of Asian ancestry. PLoS One. 2013;8(11):e79767.

Chen W, Stambolian D, Edwards AO, Branham KE, Othman M, Jakobsdottir J, et al. Genetic variants near TIMP3 and high-density lipoprotein-associated loci influence susceptibility to age-related macular degeneration. Proc Natl Acad Sci U S A. 2010;107(16):7401-6.

Divers J, Palmer ND, Langefeld CD, Brown WM, Lu L, Hicks PJ, et al. Genome-wide association study of coronary artery calcified atherosclerotic plaque in African Americans with type 2 diabetes. BMC Genet. 2017;18(1):105.

Erdmann J, Grosshennig A, Braund PS, Konig IR, Hengstenberg C, Hall AS, et al. New susceptibility locus for coronary artery disease on chromosome 3q22.3. Nat Genet. 2009;41(3):280-2.

Fu YP, Hallman DM, Gonzalez VH, Klein BE, Klein R, Hayes MG, et al. Identification of Diabetic Retinopathy Genes through a Genome-Wide Association Study among Mexican-Americans from Starr County, Texas. J Ophthalmol. 2010;2010.

Graham PS, Kaidonis G, Abhary S, Gillies MC, Daniell M, Essex RW, et al. Genome-wide association studies for diabetic macular edema and proliferative diabetic retinopathy. BMC Med Genet. 2018;19(1):71.

Grassi MA, Tikhomirov A, Ramalingam S, Below JE, Cox NJ, Nicolae DL. Genome-wide meta-analysis for severe diabetic retinopathy. Hum Mol Genet. 2011;20(12):2472-81.

Huang YC, Lin JM, Lin HJ, Chen CC, Chen SY, Tsai CH, et al. Genome-wide association study of diabetic retinopathy in a Taiwanese population. Ophthalmology. 2011;118(4):642-8.

Iyengar SK, Sedor JR, Freedman BI, Kao WH, Kretzler M, Keller BJ, et al. Genome-Wide Association and Trans-ethnic Meta-Analysis for Advanced Diabetic Kidney Disease: Family Investigation of Nephropathy and Diabetes (FIND). PLoS Genet. 2015;11(8):e1005352.

Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, et al. Complement factor H polymorphism in age-related macular degeneration. Science. 2005;308(5720):385-9.

McDonough CW, Palmer ND, Hicks PJ, Roh BH, An SS, Cooke JN, et al. A genome-wide association study for diabetic nephropathy genes in African Americans. Kidney Int. 2011;79(5):563-72.

Meng W, Shah KP, Pollack S, Toppila I, Hebert HL, McCarthy MI, et al. A genome-wide association study suggests new evidence for an association of the NADPH Oxidase 4 (NOX4) gene with severe diabetic retinopathy in type 2 diabetes. Acta Ophthalmol. 2018;96(7):e811-e9.

Meng W, Veluchamy A, Hebert HL, Campbell A, Colhoun HM, Palmer CNA. A genome-wide association study suggests that MAPK14 is associated with diabetic foot ulcers. Br J Dermatol. 2017;177(6):1664-70.

Myocardial Infarction Genetics C, Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissino D, et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. Nat Genet. 2009;41(3):334-41.

Montesanto A, Bonfigli AR, Crocco P, Garagnani P, De Luca M, Boemi M, et al. Genes associated with Type 2 Diabetes and vascular complications. Aging (Albany NY). 2018;10(2):178-96.

Pollack S, Igo RP, Jr., Jensen RA, Christiansen M, Li X, Cheng CY, et al. Multiethnic Genome-Wide Association Study of Diabetic Retinopathy Using Liability Threshold Modeling of Duration of Diabetes and Glycemic Control. Diabetes. 2019;68(2):441-56.

Qi L, Qi Q, Prudente S, Mendonca C, Andreozzi F, di Pietro N, et al. Association between a genetic variant related to glutamic acid metabolism and coronary heart disease in individuals with type 2 diabetes. JAMA. 2013;310(8):821-8.

Reilly MP, Li M, He J, Ferguson JF, Stylianou IM, Mehta NN, et al. Identification of ADAMTS7 as a novel locus for coronary atherosclerosis and association of ABO with myocardial infarction in the presence of coronary atherosclerosis: two genome-wide association studies. Lancet. 2011;377(9763):383-92.

Schunkert H, Konig IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nat Genet. 2011;43(4):333-8.

Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science. 2007;316(5829):1341-5.

Taira M, Imamura M, Takahashi A, Kamatani Y, Yamauchi T, Araki SI, et al. A variant within the FTO confers susceptibility to diabetic nephropathy in Japanese patients with type 2 diabetes. PLoS One. 2018;13(12):e0208654.

Teumer A, Tin A, Sorice R, Gorski M, Yeo NC, Chu AY, et al. Genome-wide Association Studies Identify Genetic Loci Associated With Albuminuria in Diabetes. Diabetes. 2016;65(3):803-17.

van Zuydam NR, Ahlqvist E, Sandholm N, Deshmukh H, Rayner NW, Abdalla M, et al. A Genome-Wide Association Study of Diabetic Kidney Disease in Subjects With Type 2 Diabetes. Diabetes. 2018;67(7):1414-27.