

Alma Mater Studiorum – Università di Bologna

**DOTTORATO DI RICERCA IN
ONCOLOGIA, EMATOLOGIA E PATOLOGIA**

Ciclo XXXIII

Settore Concorsuale: 06/A2

Settore Scientifico Disciplinare: MED/04

**In-Depth Analysis Of Metabolic Phenotype By Studying Age-
And Sex-Related Features Of Type 2 Diabetes**

Presentata da: Anna Carbó Meix

Coordinatore Dottorato

Supervisore

Co-supervisore

Prof.ssa

Prof.

Prof.

Manuela Ferracin

Stefano Salvioli

Paolo Garagnani

Esame finale anno 2021

Contents

LIST OF FIGURES	4
LIST OF TABLES	6
LIST OF ABBREVIATIONS	8
ACKNOWLEDGEMENTS	12
ABSTRACT	13
INTRODUCTION	14
1. TYPE 2 DIABETES.....	14
1. Overview	14
2. Epidemiology	14
3. Genetics.....	14
4. Diagnosis.....	15
5. Motivation	15
2. ATHEROGENIC DYSLIPIDEMIA.....	16
6. Lipoproteins (LDL, VLDL, HDL), Cholesterol, Triglycerides and Apolipoproteins.....	16
7. Common risk factors in cardiovascular disease and type 2 diabetes	19
8. Thrombosis and fibrinolysis	20
3. GLYCANS	21
9. Glycans definition.....	21
10. Glycans biosynthesis.....	21
11. Molecular diversity of glycans	23
12. Functions of glycans	23
13. Genetics of protein glycosylation	24
14. Environmental factors affecting protein glycosylation.....	25
MATERIALS & METHODS	26
15. Data description.....	26
16. Preprocessing: Clinical Database and N-glycans.....	31
17. Cohorts: controls, prediabetics, diabetics	36
CHAPTER 1: CHARACTERIZATION OF THE T2D-DYSMETABOLIC-AGING PHENOTYPE	37
4. INTRODUCTION.....	37
18. Intracellular biological roles of glycans: focus on cellular metabolism	38
19. Extracellular biological roles of glycans: focus on inflammation (IgG glycosylation)	44

20.	N-glycans as biomarkers in liver diseases, aging, and age-related diseases	46
5.	MATERIALS AND METHODS.....	54
PART I.....		55
6.	AIMS.....	55
7.	RESULTS.....	56
21.	Demographic characteristics of the cohort	56
22.	Prediction of (non-)diabetic status by HbA1c and by N-glycans.....	59
23.	Visualization of mean values and integrated weights.....	67
PART II.....		85
8.	AIM.....	85
9.	RESULTS.....	86
24.	Sources of variation in Phenomics, T2D, Lipids, Liver, Kidney, Iron, Coagulation, Blood & Glycomics data sets.....	86
10.	DISCUSSION.....	93
 CHAPTER 2: ESTIMATION OF AN INNOVATIVE COMPOSITE AGING CLOCK AS BIOMARKER FOR TYPE 2 DIABETES AND RELATED METABOLIC PHENOTYPES 95		
11.	INTRODUCTION.....	95
25.	Molecular biomarkers of aging.....	96
26.	Phenotypical biomarkers of aging.....	100
27.	Motivation	101
12.	MATERIALS AND METHODS	101
28.	Materials.....	101
29.	Methods.....	101
13.	AIM.....	103
14.	RESULTS.....	104
30.	Univariate regression models: Age ~ Variable	104
31.	Statistical models.....	105
32.	Association between “Biological Age” and “Chronological Age”, separately in sexes:	107
33.	Association between “Biological Age” with disease (T2D) separately in sexes:.....	108
34.	Associations with continuous traits or N-glycans: Trait/N-glycan ~ Phenotypical Age + Age + Sex + Group.....	109
15.	DISCUSSION.....	111
CONCLUSIONS		113

LIST OF FIGURES

Figure 1. Plasma lipid exchange..... 17

Figure 2. Association between BMI and T2D..... 19

Figure 3. Relative risk of CVD in normoglycemia, prediabetes and diabetes..... 20

Figure 4. Examples of glycans (IUPAC Gold Book)..... 21

Figure 5. Protein N-glycosylation biosynthesis..... 22

Figure 6. O-GlcNAcylation in cell metabolism..... 23

Figure 7. Biological roles of glycans proposed by Varki (2017)..... 24

Figure 8. Distribution of N-glycans without vs with clr transformation 32

Figure 9. Distribution of Blood Cells without vs with clr transformation..... 33

Figure 10. Metaflammation intra-organ crosstalk 38

Figure 11. Metaflammation inter-organ crosstalk 39

Figure 12. Metaflammation intracellular crosstalk 39

Figure 13. O-GlcNAcylation in conditions of hyperglycaemia..... 40

Figure 14. Relationship between chronic glycaemia, chronic glycosylation and insulin sensitivity 41

Figure 15. Hyperglycaemia, glycosylation and cardiovascular disease 41

Figure 16. The absence of glycosylation of Glut2 in mice fed with high fat diet impairs insulin secretion..... 42

Figure 17. Inflammasomes activation in glucotoxicity/lipotoxicity 42

Figure 18. Endoplasmic reticulum stress responses..... 43

Figure 19. Antibody glycosylation controls antibody activity..... 44

Figure 20. What comes first in RA, HIV and metabolic syndrome, disease or inflammation? 45

Figure 21. The relationship between AGEs and T2D/obesity 46

Figure 22. Glycosyltransferases enzymes 47

Figure 23. Summary of main findings regarding alteration of glycosylation in liver pathologies..... 47

Figure 24. Two-way orthogonal partial least squares algorithm..... 54

Figure 25. Decomposed parts of the O2PLS integrations between 44 clinical variables and 10 N-glycans in four different age ranges by groups and sexes 60

Figure 26. Joint PC1 loading values of control males aged 20-85 y..... 61

Figure 27. Joint PC1 loading values of control females aged 20-85 y..... 62

Figure 28. Joint PC1 loading values of prediabetic males aged 20-85 y.....	63
Figure 29. Joint PC1 loading values of prediabetic females aged 20-85 y.....	64
Figure 30. Joint PC1 loading values of diabetic males aged 20-85 y.....	65
Figure 31. Joint PC1 loading values of diabetic females aged 20-85 y.....	66
Figure 32. Parallel plots of groups of variables (I)	67
Figure 33. Parallel plots of groups of variables (II).....	68
Figure 34. Means and weights of fasting glucose.....	68
Figure 35. Means and weights of glycated haemoglobin	69
Figure 36. Means and weights of HOMA	69
Figure 37. Means and weights of fasting insulin	70
Figure 38. Means and weights of BMI	70
Figure 39. Means and weights of WHR.....	70
Figure 40. Means and weights of total cholesterol	71
Figure 41. Means and weights of LDL	71
Figure 42. Means and weights of HDL.....	72
Figure 43. Means and weights of triglycerides.....	72
Figure 44. Means and weights of alkaline phosphatase	73
Figure 45. Means and weights of total bilirubin.....	73
Figure 46. Means and weights of AST	74
Figure 47. Means and weights of C reactive protein	74
Figure 48. Means and weights of creatinine.....	75
Figure 49. Means and weights of azotemia.....	75
Figure 50. Means and weights of uric acid	76
Figure 51. Means and weights of ferritin.	76
Figure 52. Means and weights of fibrinogen.....	77
Figure 53. Means and weights of haemoglobin.....	78
Figure 54. Means and weights of white cells	78
Figure 55. Means and weights of platelets	79
Figure 56. Means and weights of telomere length.....	79
Figure 57. Means and weights of GP1	80
Figure 58. Means and weights of GP2.....	80

Figure 59. Means and weights of GP3.....	81
Figure 60. Means and weights of GP4.....	81
Figure 61. Means and weights of GP5.....	82
Figure 62. Means and weights of GP6.....	82
Figure 63. Means and weights of GP7.....	83
Figure 64. Means and weights of GP8.....	83
Figure 65. Means and weights of GP9.....	84
Figure 66. Means and weights of GP10.....	84
Figure 67. Decomposed parts of the O2PLS integrations between 8 different data sets and 10 N-glycans in four different age ranges and the total age range by groups and sexes	86
Figure 68. Pearson inter-correlations among the joint PCs in Ctrl M.....	88
Figure 69. Pearson inter-correlations among the joint PCs in Ctrl F	89
Figure 70. Pearson inter-correlations among the joint PCs in PreDiab M	89
Figure 71. Pearson inter-correlations among the joint PCs in PreDiab F.....	90
Figure 72. Pearson inter-correlations among the joint PCs in Diab M.....	90
Figure 73. Pearson inter-correlations among the joint PCs in Diab F.....	91
Figure 74. Age acceleration among controls, prediabetics, and diabetics in both sexes	108

LIST OF TABLES

Table 1. Database classified variables by biomedical groups.....	26
Table 2. 10 N-glycans obtained with the DSA-FACE (images obtained with the Glycoworkbench software).....	30
Table 3. Summary of the major glycomic signatures of T2D described in literature.....	49
Table 4. Summary of the major glycomic signatures of aging described in literature	52
Table 5. Patients baseline characteristics	56
Table 6. Number and frequency of individuals that are affected by clinical conditions	58
Table 7. Number and frequency of individuals that take medication.....	58
Table 8. Confusion matrix for HbA1c	59
Table 9. Confusion matrix for 10 N-glycans.....	59
Table 10. Sample size of groups by age ranges and in the full age range and frequency of individuals in each age range per group.....	60
Table 11. Top 10 joint PC1 loading values of phenomics and glycomics data sets in Ctrl M.....	61
Table 12. Top 10 joint PC1 loading values of phenomics and glycomics data sets in Ctrl F.....	62

Table 13. Top 10 joint PC1 loading values of phenomics and glycomics data sets in PreDiab M.....	63
Table 14. Top 10 joint PC1 loading values of phenomics and glycomics data sets in PreDiab F	64
Table 15. Top 10 joint PC1 loading values of phenomics and glycomics data sets in Diab M	65
Table 16. Top 10 joint PC1 loading values of phenomics and glycomics data sets in Diab F	66
Table 17. Pearson intra-correlations between pairs of Phenomics/Glycomics joint scores.....	67
Table 18. Pearson intra-correlations of the joint PC scores in each pair of integrated data sets.....	87
Table 19. Top 4 loading values of the joint parts for each pair of integrated data sets	91
Table 20. Univariate regression models. P-value was adjusted for multiple comparisons with Benjamini-Hochberg correction at $\alpha=0.05$	104
Table 21. Best combination of hyperparameters calculated with cross-validation	106
Table 22. Metrics of model performance	106
Table 23. Summary output of the linear model: Predicted Age = Biological Age ~ Chronological Age.....	107
Table 24. Summary output of the multinomial logistic model: Group (Ctrl vs PreDiab/Diab) ~ Biological Age	108
Table 25. Pairwise comparisons using Wilcoxon's rank sum test in KDM. P-value adjustment method: Benjamini-Hochberg. $P<0.05$ (*), $p<0.01$ (**), $p<0.001$ (***) at $\alpha=0.05$	109
Table 26. Association between continuous N-glycans species and Biological Age, Chronological Age, Sex and Disease Status: Trait ~ Biological Age + Chronological Age + Disease Status.....	109
Table 27. Association between continuous clinical traits and Biological Age, Chronological Age, Sex and Disease Status: Trait ~ Biological Age + Chronological Age + Disease Status	110

LIST OF ABBREVIATIONS

AD: Alzheimer's disease

ADA: American Diabetes Association

ADCC: antibody-dependent cellular cytotoxic

ADMA: asymmetric dimethylarginine

AFAR: American Federation for Aging Research

Age acceleration: AgeAccel

AGE: advanced glycated end product

ALD: alcoholic liver disease

ALE: advanced lipoxidation end product

AMPK: AMP-activated protein kinase

Apo: apolipoprotein

ASAH1: lysosomal enzyme

AT: adipose tissue

ATP: adult treatment panel

AUC: area under the curve

BMI: body mass index

CAD: coronary artery disease

CCP: anti-citrullinated antibodies

CDG: congenital disorders of glycosylation

CETP: cholesteryl ester transfer protein

CHD: cardiovascular heart disease

CI: confidence intervals

Clr: center-log ratio

CM: chylomicrons

CRP: C reactive protein

CVD: cardiovascular disorders

DAMP: damage-associated molecular pattern

DDP-4: dipeptidyl peptidase 4

DNA methylation: DNAm

DSA-FACE: DNA sequencer-aided fluorophore-assisted carbohydrate electrophoresis

EC: endocannabinoids

EDA: exploratory data analysis

EGFR: estimated glomerular filtration rate

EN: elastic net regression

ER: endoplasmic reticulum

FEV1: forced expiratory volume in one second

FFA: free fatty acids

Galectin9: Gal9

GFAT: glutamine-fructose-6-phosphate transaminase

GGT: gamma glutamyl transpeptidase

GP: glycan peaks

GWAS: genome-wide association studies

HbA1c: glycated haemoglobin

HBP: hexosamine biosynthetic pathway

HIV: human immunodeficiency virus

HMGB1: high mobility group box 1

HTP: high-throughput

IAPP: Islet Amyloid Polypeptide

IDF: International Diabetes Federation

IDL: intermediate density lipoprotein

IFG: impaired fasting glucose

Ig: immunoglobulin

IGT: impaired glucose tolerance

IKK2: Inhibitor kappa B kinase 2

IL: interleukin

IR: insulin receptor

IR: insulin receptor

ISR: insulin receptor substrate

KDM: Klemra-Doubal method

K-S: Kolmogorov-Smirnov test

LAL: lysosomal acid lipase

LDL: Low density lipoprotein

LDL-C: low density cholesterol

LLFS: Long Life Family Study

LPS: lipopolysaccharide

MAD: mean absolute deviations

MCP-1: monocyte chemoattractant protein

MS: mass spectrometry

MetS: metabolic syndrome

MSE: mean-squared error

mTOR: mammalian target of rapamycin

NA: missing value(s)

NHANES: National Health and Nutrition Examination Survey

NLRP3: NLR family pyrin domain containing 3

O2PLS: two-way orthogonal partial least squares

O-GlcNAc: O-linked β -N-acetylglucosamine

O-GlcNAcylation: O-GlcNAc modification

OOB: out-of-bag

OR: odds ratio

P.LRC: platelet large cell ratio

PC: principal components

PCA: principal component analysis

PD: Parkinson's disease

PI3K: phosphatidylinositol-3-kinase

PKC: protein kinase C

PTP1B: protein tyrosine phosphatase 1B

R²: correlation

RA: rheumatoid arthritis

RDW-CV: red blood cell distribution width – coefficient of variation

RNA: ribonucleic acid
ROS: reactive oxygen species
RR: relative risk
RR: ridge regression
S6K1: S6 kinase 1
SAMPs: self-associated molecular patterns
SDMA: symmetric dimethylarginine
SNP: single nucleotide polymorphism
SOCS3: suppressor of cytokine signaling 3
SVM: support vector machine
SVR: support vector regression
T1D: type 1 diabetes
T2D: Type 2 diabetes
T2DM: Type 2 diabetes mellitus
TAFI: thrombin-activatable fibrinolysis inhibitor
TG: triglycerides
TNF- α : tumor necrosis factor-1
TSC1/2: tuberous sclerosis 1/2
TXNIP: thioredoxin interacting protein
U: transformed units
VLDL: very low density lipoprotein
WC: waist circumference
WHR: waist-to-hip ratio
WTCCC: Wellcome Trust Case Control Consortium

ACKNOWLEDGEMENTS

I would like to express my gratitude to everybody who helped me during the course of the PhD.

I would like to thank Claudia Sala and Cristina Giuliani for having introduced me to statistics and R programming and genetics, respectively. I would like to thank Maria Giulia Bacalini for her mentorship and feedback during my PhD closure and thesis writing. I would also like to thank Stefano Salvioli for the useful comments on PhD reports and thesis.

I would like to specially thank Noémie Génous for her availability and help when I needed, as well as to Francesco Ravaioli, and Kat Kwiatkowska Malgorzata for hanging out some evenings.

Especial thanks to the colleagues from the IMforFUTURE project for sharing both science and fun. At the scientific level, I would like to thank: Shafiq for showing me different software for post-genomic analysis; Maarten, for helping me with code by simulating data; Arianna Landini, for her time and willingness to show me how to perform a GWAS; Tamás, for showing me a software for designing glycans; Iva, for the many discussions and shared beers, as well as for the help with the server; Zhujie, for helping me with the questions coming up, and Azra, for sharing information about glycans normalization methods. I thank Samira, Frania and Annah for the shared enjoyment during meetings.

I also would like to thank Hae-Won Uh, Zhujie Gu and Said el Bouhaddani for welcoming me at UMCU, and for the guidance and feedback for my analysis.

Per suposat, vull agrair el suport incondicional en tot de mons pares, Santi i Maria José, sense el qual no hauria arribat mai fins aquí. Vull donar també les gràcies a Míriam, el millor regal que mai m'han fet els meus pares, pel suport i la confiança durant aquests anys difícils de doctorat. També vull agrair als familiars que van contribuir que la meva estada a Bolonya fos possible: iaies Pepita i Neus, tiets Hilde i Pepi, Jordi i Tere, moltes gràcies. I voldria manifestar també el meu agraïment a les ties, Imma i Maribel, pel seu suport, així com a la resta de la família.

Per ultimo ma non meno importante, voglio ringraziare a Paolo Garagnani il supporto a livello scientifico ed emozionale, per cui senza non avrei finito il dottorato. “La vita è complicata...” La vita è la intensità di luce che scegliamo portare ad ogni posto ad ogni istante. Grazie per esserne stato un grande maestro.

ABSTRACT

Type 2 diabetes mellitus (T2DM) is an age-related disease characterized by chronic hyperglycaemia mainly explained by insulin resistance and impaired insulin secretion and strongly linked to dysregulation of carbohydrate, lipid and protein metabolism. T2DM is a worldwide increasing disabling disease – in 2013, the International Diabetes Federation estimated that 382 million adults suffered from T2DM and that by 2035 there will be 592 people affected. These worrisome numbers challenge biomedical research at identifying new biomarkers for the diagnosis. The purpose of this study was to analyse and integrate different sources of phenomics (clinical) data – clinical history, anthropometrical measurements, biochemical parameters, cell counts and blood differential, and medication – with glycomics data in control, prediabetics and diabetics cohorts, in order to 1) identify the major sources of variation in both data sets, 2) visualize trends or patterns in variables within- and between-omics (e.g. a combination of N-glycans highly correlated with some biochemical parameters), 3) determine whether the identified patterns ‘naturally’ cluster according to known biological sources or conditions (i.e. diagnostic T2D parameters, lipidic profile, liver, kidney, iron, coagulation biomarkers, or blood-related factors), 4) deeply study the fluctuation of clinical variables over age, by sex and by groups (Ctrl/PreDiab/Diab), and 5) estimate an aging clock based on the clinical variables and N-glycans and apply it to assess whether the groups of prediabetic and diabetic patients show an accelerated aging as compared with control, and, within each group, whether there are differences between the two sexes. The analytical methods employed were two-way partial least squares (O2PLS) and regression. Results indicate that 1) the phenomics and glycomics joint components are different among groups, 2) males and females follow a different pattern dynamics over age reflected by the relative changes of clinical variables and N-glycans, 3) intra- and inter-correlations between joint PCs obtained integrating the phenomics data set or sets of endophenotypes with the glycomics data set point to a common N-glycan signature (instead of endophenotype-specific), and 4) a model based on the Klemmera-Doubal method (KDM) estimates that T2DM patients are biologically older than prediabetics and controls, being this effect more evident for male patients. Our main conclusions are that *i)* a combination of N-glycans could be used as complementary tool for the early diagnosis of metabolic dysregulation and/or T2D, as N-glycan changes are already present in prediabetics, *ii)* glycan peaks (GP) 1, GP2, and GP6 are confirmed as markers of aging, while GP8 and GP10 appear associated with dyslipidemia, and *iii)* this is the first time that prediabetics and diabetics have been included in an aging clock, as pure “healthy” controls do not exist: *“the effects of atherosclerosis are superimposed on normal aging of the underlying vessel”* (Wang and Bennett, 2012).

INTRODUCTION

1. TYPE 2 DIABETES

1. Overview

Type 2 Diabetes (T2D) is characterized by dysregulation of carbohydrate, lipid and protein metabolism. Although it is primarily caused by insulin resistance, followed by impaired insulin secretion, T2D is a multifactorial disease involving genetic and environmental factors (DeFronzo et al., 2015).

2. Epidemiology

The epidemic of diabetes mellitus and its complications has already become a global challenge. The International Diabetes Federation (IDF) estimated that 1 in 11 adults aged 20-79 years (415 million adults had diabetes mellitus (including type 1 diabetes (T1D), T2D, gestational diabetes) in 2015. This alarming numbers are calculated to raise to 642 million by 2040, and the largest increments will come from emerging countries transitioning from low-income to middle-income wealth (IDF Diabetes Atlas, 7th Edition). However, these estimates may be, in absolute numbers, underestimated by the current pandemic of coronavirus, which is causing many deaths in this group of individuals, as glycaemic control, body mass index (BMI), T2D and cardiovascular disease are risk factors for COVID-19 mortality (Holman et al., 2020), and by other undergoing rapid demographic transitions (Zimmet, 2017). The grounds for this ascending epidemic of diabetes mellitus are numerous and diverse: population ageing, economic development, urbanization, unhealthy eating habits and sedentary lifestyles (Zheng et al., 2018). Even though the genetic background might partially determine an individual's response to environmental stimuli (Fuchsberger et al., 2016), the main drivers are the raise in obesity, a sedentary lifestyle, energy-dense diets, and population ageing (Chatterjee et al., 2017). In fact, strong evidence points to that many cases of T2D could be prevented by keeping a healthy weight, engaging in a healthy diet, exercising daily for 30 min, avoiding smoking and consuming alcohol in moderation (Schellenberg et al., 2013; Hu et al., 2001).

3. Genetics

T2D runs in families and is heritable. The relative risk (RR) for T2D with 0 affected siblings is ~2-3 compared with siblings of non-T2D families, but, when 2 siblings are affected, the relative risk of other siblings for T2D is 30. Interestingly, the RR for T2D is higher when the mother is affected as compared to when the father is affected. Also, RR for T2D is increased if BMI ≥ 30 or fasting glucose concentrations >5.5 mmol/L (DeFronzo et al., 2015).

Over the last decade, plenty of genome-wide association studies (GWAS) were performed to identify common variants involved in T2D. In particular, in 2007 the Wellcome Trust Case Control Consortium (WTCCC) made a significant breakthrough by reporting single nucleotide polymorphisms (SNPs) associated with T2D (WTCCC, 2007). The strongest signal mapped in the gene *TCF7L2*, that up until now remains the most important SNP associated with T2D (Garagnani et al., 2013). Other genes significantly associated with T2D are *CAPN10* (Horikawa et al., 2000), *KCNQ1* (Yasuda et al., 2008), *KCNJ11* (Gloyn et al., 2003), *ABCC8* (Gloyn et al., 2003), *SLC30A8*

(Sladek et al., 2007), *IDE-KIF11-HHEX* (Sladek et al., 2007) and *EXT2-ALX4* (Sladek et al., 2007). However, the fact that most of the genetic variants map into (unknown) non-coding regions of the genome, the problem of the missing heritability (whereby the genetic loci described explain only a small proportion of the observed heritability, possibly due to disease heterogeneity, exclusion of rare variants, gene-environment interactions, gene-gene interactions and epigenetics), modest effect sizes, and the lack of replication of many genetic variants, inter alia, hamper the comprehension of the specific role each variant may have in increasing the risk for T2D. Moreover, novel mutations cannot explain the diabetes epidemic, which is instead largely explained by the epidemic of obesity. Beyond that, the majority of non-diabetic people carry risk variants for T2D, which roughly account for ~15% of heritability (DeFronzo et al., 2015).

4. Diagnosis

The diagnosis of T2D can be made on the basis of one of the following (American Diabetes Association, 2020):

- Increased casual plasma glucose test (≥ 200 mg/dL) in patients with classic symptoms of hyperglycaemia or hyperglycemic crisis
- Fasting plasma glucose levels (≥ 126 mg/dL)
- 2h postload glucose level (≥ 200 mg/dL after 75g oral glucose)
- Hb1Ac ($\geq 6.5\%$), confirmed by repeat testing

With regard to prediabetes, the diagnosis is established by one of the following criteria (American Diabetes Association, 2020):

- Fasting plasma glucose levels (100-125 mg/dL)
- 2h plasma glucose during 75g oral glucose tolerance test (140-199 mg/dL)
- Hb1Ac (5.7-6.4%)

5. Motivation

T2D is a multifactorial complex and heterogeneous disease – the clinical presentation, underlying pathophysiology and disease progression in patients with diabetes can vary remarkably among individuals and, at times, atypical manifestation of symptoms can make clear-cut classification of prediabetes and diabetes difficult –. In addition, diabetes often overlaps with other complex pathologies, such as obesity or cardiovascular disease, and/or it may evolve to vascular complications. In this doctorate studies, in order to investigate the complexity of the disease, I made use of a rich database comprising up to 55 variables, that could be grouped in clinical and anthropometrical measurements, biochemical parameters (diagnostic parameters of T2D, lipid profile, liver profile, kidney profile, cell counts and blood differential) and N-glycans. Besides, we aimed to take into account the heterogeneity of all individuals, not only prediabetics and diabetics, but also controls, since above a certain age (~60 years) the definition of healthy controls blurs with preclinical age-related diseases. To achieve that, we grouped patients by sex and age ranges.

Accordingly, the following subsections of the general introduction tackle the contextualization of the variables that are used in this research.

2. ATHEROGENIC DYSLIPIDEMIA

6. Lipoproteins (LDL, VLDL, HDL), Cholesterol, Triglycerides and Apolipoproteins

The insolubility of cholesterol and triglycerides (TG) in plasma requires that they are transported in macromolecules named lipoproteins, which are composed of a hydrophobic core containing phospholipid, fat-soluble antioxidants and vitamins, and cholesteryl ester, and a hydrophilic coat that contains free cholesterol, phospholipid and apolipoprotein molecules (reviewed from Hegele, 2009). Lipoproteins can be grouped according to their densities: chylomicrons (CM), very low density lipoprotein (VLDL), low density lipoprotein (LDL), intermediate density lipoprotein (IDL), and high density lipoprotein (HDL). CM and VLDL are the main TG-carrying lipoproteins, while the main cholesterol-carrying lipoproteins are LDL and HDL (reviewed from Hegele, 2009).

The terms “cholesterol”, “LDL”, and “LDL cholesterol (LDL-C)” are often used interchangeably, but they have different meanings. LDL refers to the particles circulating in fasting blood. Instead, the LDL-C term is used in clinical practice as a measure of the total amount of cholesterol contained in LDL particles, since plasma LDL level is generally not measured directly. Cholesterol is an essential component of cell membranes and a precursor of bile acids and steroid hormones; it can be of exogenous and endogenous origin and it is transported to peripheral cells, mostly by Apo B – containing lipoproteins in plasma (reviewed from Ference et al., 2017). Analogous to LDL-C, HDL-C is the measure of the total amount of cholesterol contained in HDL particles.

Patients with T2D often have lipid profiles that seem more benign than those of other high-risk individuals without T2D. Generally, LDL-C levels in diabetic individuals are not higher than in non-diabetic individuals who are matched for age, sex, and body weight. Indeed, the most frequent LDL-C level is “borderline high” (130-159 mg/dL) (The Expert Panel, 2002). Besides, high LDL-C levels (>160 mg/dL) are not found at higher rates in T2D individuals. Nevertheless, LDL-C has an important role in cardiovascular disease (CVD) in both T2D and non-T2D individuals. In fact, LDL-C levels may underestimate CVD risk in T2D (Buse et al., 2007). This is due to the LDL fraction characteristic in T2D individuals: small, dense particles with less cholesterol than normal-sized LDL particles, which are exceptionally atherogenic (Krentz, 2003; Marcovina and Packard, 2006; Goldberg, 2001). Such atherogenic nature of LDL particles in T2D individuals accounts for that the small, dense LDL particles are more readily oxidized and glycosylated and can more easily penetrate the arterial wall than larger LDL particles. As a result, atherosclerosis is enhanced and atherosclerotic plaques can increase migration and apoptosis of vascular smooth muscle cells (reviewed from Nesto, 2008). Accordingly, individuals with T2D have a two- to threefold increased risk for CVD (myocardial infarction, stroke, peripheral vascular disease) compared with non-diabetics, and CVD is responsible for ~80% of mortality in T2D patients (Morrish et al., 2001).

LDL-C not only presents a major role in T2D, but it also interacts with risk factors of the metabolic syndrome that multiply the risk of CVD. As an example, increased small, dense LDL particles and raised triglycerides seem to be tied to insulin sensitivity. Insulin resistance in skeletal muscle fosters hepatic TG synthesis, which in the long-run generates a large amount of atherogenic TG-rich lipoprotein particles, specially VLDL. In turn, many TG molecules from VLDL are exchanged for cholesterol in LDL particles, which leads to the formation of TG-enriched (and cholesterol-depleted) LDL. These LDL particles will become smaller and denser through the TG hydrolysis mediated by the hepatic lipase. Hence, adverse changes in LDL particles are positively correlated

with TG levels. Once TG levels exceed 100 mg/dL, small, dense LDL particles predominate (reviewed from Nesto, 2008). Figure 1 shows that when plasma concentrations of VLDL are high, cholesteryl ester transfer protein (CETP) will exchange VLDL triglyceride (TF) for cholesteryl ester (CE) in the core of LDL and HDL particles. This triglyceride can then be converted to free fatty acids by the actions of plasma lipases, primarily hepatic lipase. The net effect is a decrease in density of both LDL and HDL particles (Goldberg, 2001).

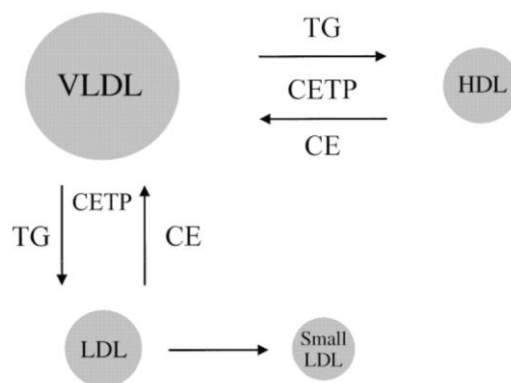


Figure 1. Plasma lipid exchange.

In this line, the Adult Treatment Panel III (ATP) report of the National Cholesterol Education Program considers elevated LDL-C as the primary target of lipid-lowering therapy for reducing CVD risk (NCEP, 2001). Conversely, some *a posteriori* studies claim that apo B and apoA1 may, indeed, be more powerful lipid-related predictors of risk for CVD (Packard, 2003; Sniderman et al., 2003; Walldius and Jungner, 2004).

Apo B-100 is a master protein component of the pro-atherogenic VLDL, IDL and LDL particles, each containing one Apo B molecule. Thus, plasma Apo B levels reflect the total numbers of the pro-atherogenic particles. TG endogenously synthesized in the liver are transported with VLDL particles into plasma, where they undergo lipolysis to IDL by the action of lipoprotein lipase. IDL, in turn, is lipolyzed by hepatic lipase, giving rise to LDL, or taken up by the liver via the LDL receptor. Apo B is also fundamental for the binding of LDL particles to the LDL receptor for cellular uptake and removal of LDL particles. In contrast, Apo A1 is the main apolipoprotein component of the anti-atherogenic HDL, and thus Apo A1 concentrations are strongly associated with HDL-C levels. Apo A1 is essentially involved in removing excess cholesterol from tissues and embedding it into HDL for reverse transport through the macrophage ATP-binding cassette transporter ABCA1 (reviewed from Chan and Watts, 2006). Chan and Watts (2006) compared different observational studies that used either LDL-C and/or apolipoproteins as predictors of coronary/cardiovascular risk. Among them, *Apolipoprotein-related Mortality Risk Study (AMORIS)*, N=98,722 men and N=76,831, *Second Northwick Park Heart Study (NPHSII)*, N=2,505 men, and *Women's Health Study (WHS)*, N=15,632 women) estimated the association of both predictors (LDL-C and the apolipoproteins B and A1) with CVD, and the three of them showed that the odds ratio (OR) of the Apo B / Apo A1 ratio is noticeably higher than the OR of LDL-C (Walldius et al., 2001; Talmud et al., 2002; Ridker et al., 2005; Chan and Watts, 2006).

On the other hand, HDL enhances reverse cholesterol transport and it is an immunomodulator. According to different studies, HDL reduces LDL oxidation (Norata et al., 2006), inhibits oxidized LDL-induced MCP-1 (monocyte chemoattractant protein 1) production and monocyte

transmigration in a co-culture of human aortic endothelial cells and human aortic smooth muscle cells (Van Lenten et al. 1995; Navab et al., 1991), blunts inflammatory response of endothelial cells to tumour necrosis factor-1 (TNF- α) and interleukin (IL)-1 stimuli (Ashby et al., 1998), and exerts anti-thrombotic and anti-apoptotic effects (Vergès, 2015; Femlak et al., 2017).

HDL is composed of proteins that become affected by oxidative stress during an inflammatory response, including Apo A1. Modification of the protein content can turn HDL a proinflammatory particle, a scenario in which HDL becomes “dysfunctional”. Epidemiological research has shown that HDL levels are inversely correlated with the risk of CVD. However, there are individuals who have suffered a clinical event and present normal or high levels of HDL. Low levels of HDL cholesterol are also associated with increased risk for type 2 diabetes. The hypothesis of “dysfunctional” HDL has been studied in T1D. Overall, T1D patients with lowered levels of HDL-C have a higher cardiovascular risk, but T1D patients with high levels of HDL-C may not have a lower cardiovascular risk (Orchard, 1990). Thus, the mechanisms of such relationship are yet to be understood. To date, though, it has been suggested that different HDL subfractions relate to coronary artery disease (CAD) incidence in a different manner (reviewed from Femlak et al., 2017).

1. BMI and WHR

Along with the lipid profile, some anthropometric measures are also taken into account as a function of health in patients. A few well-known parameters are waist circumference (WC), waist-to-hip ratio (WHR), and BMI. There has been controversy in whether BMI or WHR is a better predictor of obesity (Murray, 2006; Qyao and Nyamdorj, 2010). Qyao and Nyamdorj (2010) revised different investigations with the goal of comparing between BMI, WC, and WHR in their relation to the incidence and prevalence of T2D. They concluded that all studies showed that either BMI or WC predicted or was associated with T2D independently and stressed that the nature of the study could condition on results. In other words, prospective studies favoured equally BMI, WC, WHR and the waist-to-stature ratio (though these studies were limited to ethnic groups), while cross-sectional studies evidence pointed to that WC or WHR discriminate better cases with diabetes from those without, as compared with BMI (though these studies provide only possible association). Another investigation aimed to explore both the relative magnitude of association and the discriminative capability of multiple indicators of obesity with CVD mortality risk by conducting a meta-analysis of 9 cohort studies of men and women from the British general population (Czernichow et al., 2014). They measured BMI, WC, and WHR, and, after a mean of 8.1 years of follow-up, 6,641 deaths were recorded from a total of 82,864 individuals. In this research, they estimated the association of CVD risk and CVD mortality with the mentioned variables and reached the conclusion that measures of abdominal adiposity, but not BMI, were related to an increased risk of CVD mortality (Czernichow et al., 2014). In any case, the increasing adiposity is the single most important risk factor for T2D (De Fronzo et al., 2015). **Figure 2**, extracted from De Fronzo et al. (2015), shows the relationship between the age-adjusted RR of T2D and BMI and reflects that the higher the BMI, the higher the age-adjusted RR of T2D is, being greater in women than men.

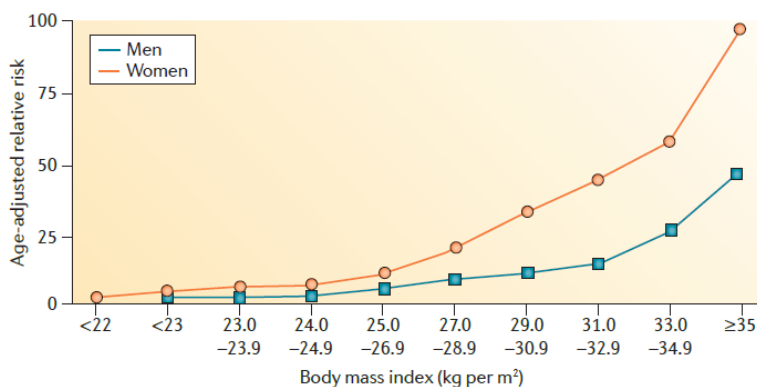


Figure 2. Association between BMI and T2D

7. Common risk factors in cardiovascular disease and type 2 diabetes

Thus far, I reviewed the role of key players in CVD and T2D without clearly distinguishing between the two diseases, which, indeed, share many traditional and non-traditional risk factors. Traditional CVD risk factors in T2D include dyslipidemia (altered lipid profile), increased blood pressure, being overweight or obese ($>25 \text{ kg/m}^2$), having abdominal or central obesity (independent of BMI), reduced physical exercise, and cigarette smoking. Non-traditional CVD risk factors comprise insulin resistance and hyperinsulinemia, postprandial hyperglycaemia, glucose variability, microalbuminuria, haematological factors, thrombogenic factors, increased levels of C-reactive protein (CRP), homocysteine and vitamins, erectile dysfunction, and genetics and epigenetics. There are consistent evidences which underscore that control of the traditional risk factors are necessary for reducing CVD risk in T2D patients. Cardiovascular benefits are acquired if the control of such risk factors start early in subjects with short duration of diabetes and low cardiovascular risk. On the contrary, in elderly subjects with a longer time exposure to hyperglycaemia and high cardiovascular risk, there are no cardiovascular improvements but losses. This positive or negative effect could be attributed to the “metabolic memory” hypothesis, whereby the early glycaemic exposure is imprinted in target organs, resulting in long-term protective or deleterious consequences (reviewed from Martín-Timón et al., 2014). The main gain of reducing concentration of plasma glucose in T2D is prevention of long-lasting microvascular complications and, to lesser extent, of macrovascular complications. However, hyperglycaemia is a loose risk factor for CVD (UKPDS, 1998; Holman et al., 2008), and interventions (Gerstein et al., 2008; Patel et al., 2008; Duckworth et al., 2009) focusing on reducing plasma glucose have unsuccessfully reduced CV risk and mortality. The cluster of CV/metabolic factors (obesity, dyslipidemia, hypertension, endothelial dysfunction, procoagulant state) associated with insulin resistance, that is usually moderate to severe in individuals with T2D, is known as insulin resistance (metabolic) syndrome and is a combinatorial major factor responsible for CV risk in T2D (DeFronzo, 2009; DeFronzo, 2010; Abdul-Ghani et al., 2017). In this line, obese non-T2D individuals but with insulin resistance syndrome exhibit an akin raised risk for CVD compared with T2D individuals, which reinforces the notion that hyperglycaemia is not a paramount risk factor for CVD. In accordance with this, antidiabetic drugs (e.g. insulin, sulfonylureas, and dipeptidyl peptidase 4 (DPP-4) inhibitors), which diminish plasma glucose without affecting insulin resistance, do not lessen CVD risk and mortality in T2D (Abdul-Ghani et al., 2017).

2. Prediabetes is Key for Understanding Cardiovascular Complications in Type 2 Diabetes

Prediabetes is characterized by one of the following (DeFronzo et al., 2015):

- *Impaired fasting glucose (IFG)*: muscle insulin resistance + impaired late insulin secretion.
- *Impaired glucose tolerance (IGT)*: hepatic insulin resistance + impaired early insulin secretion.
- ↑*Glycated haemoglobin (HbA1c)*.

Both IFG and IGT use hyperinsulinemia as a coping mechanism to reduce elevated fasting or postprandial glucose (Laakso, 2010). Some studies, though, have provided evidence that individuals with IGT have a more marked degree of insulin resistance, whereas individuals with IFG are characterized by a more marked β -cell defect (Weyer et al., 1999; Festa et al., 2004). In addition to that, it has been shown that insulin resistance *per se* is correlated with atherosclerosis, even in normoglycemic individuals (Laakso et al., 1991). Furthermore, several investigations have demonstrated that high insulin level is associated with risk of cardiovascular heart disease (CHD) (Laakso, 1996). Prediabetics have a cluster of CVD risk factors that may differ of the cluster of CVD risk factors in diabetics. Nevertheless, diabetics are at least as insulin resistant as prediabetics (Laakso, 1996). Therefore, insulin-resistance RR in both prediabetics and diabetics are likely to explain an important part of the atherosclerotic process (**figure 3**, extracted from Laakso, 2010).

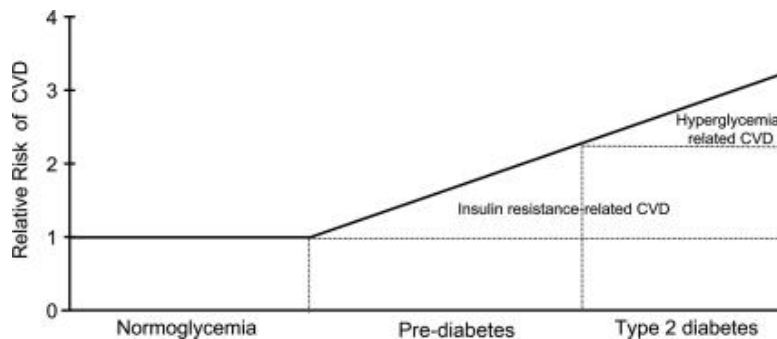


Figure 3. Relative risk of CVD in normoglycemia, prediabetes and diabetes

3. Sex in Cardiovascular Complications in Type 2 Diabetes

Intriguingly, RR for CHD is higher in T2D female patients than in T2D male patients. The reason for this sex difference is not known, but it could be explained by a heavier risk-factor burden and a greater effect of blood pressure and atherogenic dyslipidemia on the risk of CVD in T2D women than in T2D men (Juutilainen et al., 2004).

8. Thrombosis and fibrinolysis

Insulin resistance and diabetes are associated with an increased prothrombotic risk (coagulation factors VII, XII, and fibrinogen) and with suppressed fibrinolysis, due to elevated levels of the fibrinolytic inhibitor PAI-1 and of the thrombin-activatable fibrinolysis inhibitor (TAFI) (Grant, 2007; Hori et al., 2002).

3. GLYCANS

9. Glycans definition

The terms glycan and polysaccharide are defined by IUPAC as synonyms meaning "compounds consisting of a large number of monosaccharides linked glycosidically". In practice, glycan, carbohydrate, saccharide and sugar are generic terms used interchangeably (IUPAC Gold Book – Glycans). **Figure 4** shows some examples of glycans.

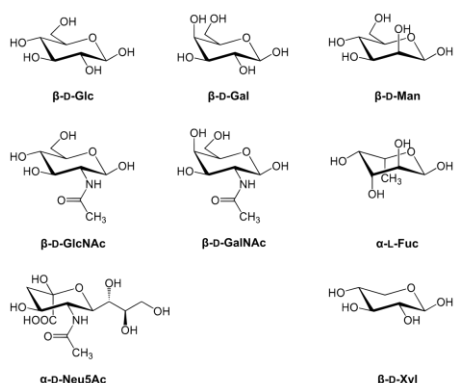


Figure 4. Examples of glycans (IUPAC Gold Book)

Glycans are not only defined by monosaccharide units, but also by the bond position, its anomeric configuration (alpha or beta), the number of branches and the position of branching (Lauc and Zoldoš, 2010). Along with nucleic acids, proteins and lipids, glycans are essential macromolecular and structural components or, in other words, building blocks, of all cells, and they can be found attached to proteins and lipids or as independent macromolecules (Marth and Grewal, 2008).

10. Glycans biosynthesis

In the case of protein glycosylation, glycans synthesis can be split into N-linked, O-linked, and the GlcNAc modification (revised from Zhou et al., 2018):

- **N-glycosylation synthesis:** N-linked glycosylation starts with the transfer of a large precursor oligosaccharide from a membrane-embedded dolichol phosphate lipid to the asparagine of a nascent protein while being synthesized in the rough endoplasmic reticulum (ER). The glycan assists in protein folding by mediating interactions with ER chaperones like calnexin and calreticulin, thereby functioning as a quality control checkpoint. Upon proper folding, the glycans are trimmed to “high mannose” structures prior to moving to the Golgi apparatus, where they undergo further trimming and then rebuilding through the combined action of various glycosyltransferases. This results in both hybrid and complex-type N-glycans, which are common at the plasma membrane and on secreted glycoproteins, including immunoglobulin (Ig) G and essentially every surface protein on a cell (reviewed from Zhou et al., 2018; **figure 5**, extracted from New England Biolabs).

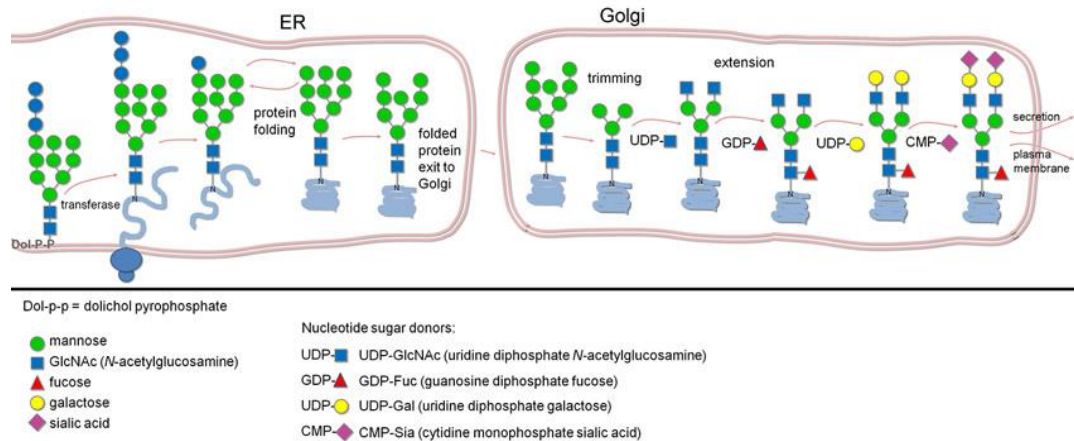


Figure 5. Protein N-glycosylation biosynthesis

- **O-glycosylation synthesis:** O-linked glycosylation is a wide category comprising various modifications named for the glycosidic linkage of oxygen on threonine or serine residues, and includes O-GalNAc, O-fucose, and O-mannose glycans. Unlike the co-translational N-glycosylation, all of these are built progressively from a single initiating sugar post-translationally cell (reviewed from Zhou et al., 2018).
- **O-GlcNAcylation (O-GlcNAc glycosylation):** O-GlcNAc glycosylation corresponds to the addition of N-acetylglucosamine on serine and threonine residues of cytosolic and nuclear proteins. O-GlcNAcylation is a dynamic post-translational modification, analogous to phosphorylation, that regulates the stability, the activity or the subcellular localisation of target proteins (e.g. RNA polymerase II, histones, histone deacetylase complexes and members of the Polycomb and Trithorax group). This reversible modification depends on the availability of glucose – 2-5% of glucose enters the hexosamine biosynthetic pathway (HBP) – and therefore constitutes a powerful mechanism by which cellular activities are regulated according to the nutritional environment of the cell (as above-mentioned GlcNAc is considered to be the metabolic sensor of the cell). O-GlcNAcylation has been implicated in important human pathologies including Alzheimer disease and type-2 diabetes. Only two enzymes, OGT and O-GlcNAcase, control the O-GlcNAc level on proteins (reviewed from Issad et al., 2010; **figure 6**).

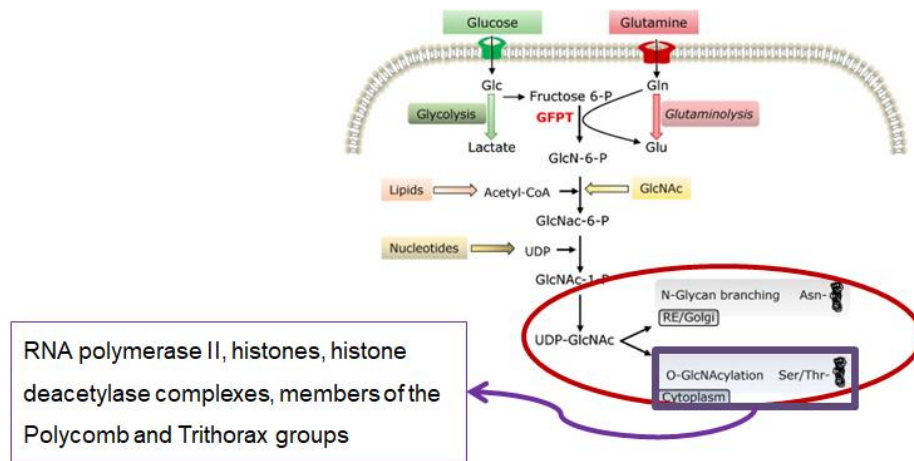


Figure 6. O-GlcNAcylation in cell metabolism

11. Molecular diversity of glycans

Glycosylation is not only the most abundant post-translational modification, but also by far the most structurally diverse. Conservative estimates state that >50% of proteins are glycoproteins, without considering the countless number of nuclear and cytoplasmic proteins modified by O-GlcNAc. It is estimated that between 2-5 glycans are attached to an average glycoprotein, the 10% of which are O-linked, and the other 90% are N-linked or both (Apweiler et al., 1999). In this line, there are at least 13 different monosaccharides and 8 different amino acids involved in glycoprotein linkages, with a total of at least 41 different chemical bonds known to be linking the glycan to the protein (Spiro, 2002). Importantly, each one of these glycan:protein linkages is unique in both structure and function. However, if the structural diversity of different branches composed by different monosaccharides is added to these single-linkages, the molecular diversity increases exponentially. Just the “sialome” rivals or exceeds many other post-translational modifications in abundance and structural/functional diversity. Besides, chemical modifications, such as phosphorylation, sulfation and acetylation, increase the glycan structural/functional diversity even more. Therefore, categorizing glycosylation as a single type of post-translational modification may be misleading and useless (Cohen and Varki, 2010).

12. Functions of glycans

Very conservative estimates indicate that there are over a million different glycan structures in a mammalian cell's glycome, and consequently glycoproteins are involved in a myriad of functions.

Varki (2017) revised the multiple functions of glycans, and classified them into four groups (see **figure** extracted from Varki, 2017):

- i. **Structural and modulatory roles:** as example, the addition of O-GlcNAc residues to histones surrounding chromosomal DNA is key in the histone code that regulates gene expression.
- ii. **Extrinsic (interspecies) recognition of glycans:** as example, glycans, such as polysaccharide A, derived from the mammalian gut microbiome, helps to modulate the host immune system to a more tolerant state (via T-reg engagement).

- iii. **Intrinsic (intraspecies) recognition of glycans:** as example, the conservative glucosylation/deglucosylation cycle of protein folding in the ER.
- iv. **Molecular mimicry of host glycans:** as example, molecular evolution of microorganisms has allowed them the acquisition of host sialoglycans or the direct transfer of host sialic acids by trans-sialidases, thereby acting as self-associated molecular patterns (SAMPs) recognized by Siglecs – receptors of sialoglycans –, limiting complement activation, or masking antibody recognition.

Structural and modulatory roles	Extrinsic (interspecies) recognition of glycans
Physical structure	Bacterial, fungal and parasite adhesins
Physical protection and tissue elasticity	Viral agglutinins
Water solubility of macromolecules	Bacterial and plant toxins
Lubrication	Soluble host proteins that recognize pathogens
Physical expulsion of pathogens	Pathogen glycosidases
Diffusion barriers	Host decoys
Glycoprotein folding	Herd immunity
Protection from proteases	Pathogen-associated molecular patterns
Modulation of membrane receptor signaling	Immune modulation of host by symbiont/parasite
Membrane organization	Antigen recognition, uptake and processing
Modulation of transmembrane receptor spatial organization and function	Bacteriophage recognition of glycan targets
Antiadhesive action	Intrinsic (intraspecies) recognition of glycans
Depot functions	Intracellular glycoprotein folding and degradation
Nutritional storage	Intracellular glycoprotein trafficking
Gradient generation	Triggering of endocytosis and phagocytosis
Extracellular matrix organization	Intercellular signaling
Protection from immune recognition	Intercellular adhesion
Effects of glycan branching on glycoprotein function	Cell–matrix interactions
Cell surface glycan:lectin-based lattices	Fertilization and reproduction
Masking or modification of ligands for glycan-binding proteins	Clearance of damaged glycoconjugates and cells
Tuning a range of function	Glycans as clearance receptors
Molecular functional switching	Danger-associated molecular patterns
Epigenetic histone modifications	Self-associated molecular patterns
	Antigenic epitopes
	Xeno-autoantigens
	Molecular mimicry of host glycans
	Convergent evolution of host-like glycans
	Appropriation of host glycans

Figure 7. Biological roles of glycans proposed by Varki (2017)

13. Genetics of protein glycosylation

The synthesis of glycans, in contrast to genes, is a template-independent process – glycans are encoded in a dynamic network of hundreds of genes that code for enzymes involved in glycan synthesis, e.g. glycosyltransferases (>250), glycosidases, enzymes for sugar nucleotide biosynthesis, etc. In addition to over 600 such proteins, there are various transcription factors, Golgi organizers, proton pumps, etc. that affect their expression and activity (reviewed from Lauc and Zoldoš, 2010).

In genetics, three groups of mutations/SNPs in genes coding for enzymes involved in glycosylation have been described at large:

- a. **Embryologically lethal:** those affecting enzymes involved in the pathway of glycans biosynthesis, and that consequently they impair the proper functioning of thousands of proteins (reviewed from Lauc and Zoldoš, 2010).

- b. **Rare mutations:** those comprising a group of rare genetic disorders (the so-called congenital disorders of glycosylation or CDG) caused by mutations in the core N-glycosylation pathway – these mutations result in residual enzymatic activity, but are nonetheless associated with significant mortality and broad motoric, immunologic, digestive and neurologic symptoms (reviewed from Lauc and Zoldoš, 2010).
- c. **Glycophenotype:** those variations in genes involved in modifications of glycan antennas, which are common in the population and apparently cause a large part of individual phenotypic variations that exist in humans and in other higher organisms. Indeed, the majority of variability originates from SNPs that individually do not have visible glycophenotypes, but if present in specific combinations within the same individual can have significant phenotypic effects (reviewed from Lauc and Zoldoš, 2010).

Regardless of mutations/SNPs, it has been reported that a large part of the observed variability is under genetic control (Lauc and Zoldoš, 2010).

14. Environmental factors affecting protein glycosylation

The complex dynamic network of hundreds of proteins taking part in glycans biosynthesis makes the process of protein glycosylation intrinsically sensitive to all changes occurring within a cell, so that the glycan structures produced at any instant of time reflect the significant past events in the cell. Indeed, glycosylation sites on the same protein can contain different glycan structures that reflect the cell type, developmental stage and metabolic state of the cell in which they are synthesized.

While general glycan structures appear to be mostly defined by the genetic makeup, some specific glycans are sensitive to environmental factors. Some examples of environmental factors that change the glycomic profile are age, smoking and diet.

MATERIALS & METHODS

15. Data description

4. Clinical database

A database consisted of 123 variables (including 10 N-glycans) of N=1185 individuals was provided. The clinical data were classified in groups to facilitate the comprehension (**table 1**).

Table 1. Database classified variables by biomedical groups

GROUP	VARIABLE
Sex – Age	Sex
	Age
Clinical History	Bypass
	Somatic Neuropathy
	Nephropathy
	Chronic Renal Insufficiency
	Retinopathy
	Arteriopathy Obliterans of Lower Limbs
	Arteriopathy Obliterans of Upper Limbs
	Cardiac Ischemia
	Major Adverse CDV Events
Anthropometrical Measurements	Weight
	Height
	Body Mass Index
	Waist
	Hips
	Waist-to-Hip ratio
Biochemical Parameters	Glycaemia
	GLIC MM (unk)
	Hb1Ac
	Azotemia
	Creatinine
	Uric Acid
	Alkaline Phosphatase

	AST
	ALT
	Cholesterol
	HDL
	Triglycerides
	Insulin
	HOMA
	Estimated glomerular filtration rate
	C Reactive Protein
	Fibrinogen
	PAI-1
	Thrombin activatable fibrinolysis inhibitor
	Gamma glutamyl transpeptidase
	Total Bilirubin
	LDL
	Apolipoprotein A1
	Apolipoprotein B
	Total iron
	Transferrin
	Ferritin
	Total Protein
	Telomere Length
	Symmetric Dimethylarginine
	Asymmetric Dimethylarginine
	ASAH1 (lysosomal enzyme)
	ASAH1 (lysosomal enzyme percentage)
	Glucoamylase (lysosomal enzyme)
	Glucoamylase (lysosomal enzyme percentage)
Cell counts	White Cells
	Red Cells
	Hemoglobin
	Hematocrit

	Mean Corpuscular Volume
	Mean Corpuscular Hemoglobin
	Mean Corpuscular Hemoglobin Concentration
	Red blood cell Distribution Volume – SD
	Red blood cell Distribution Width – CV
	Platelets
	Platelets Distribution Width – SD
	Mean Platelets Volume
Blood Differential	P.LCR (Platelet Large Cell Ratio)
	Neutrophils
	Lymphocytes
	Basophils
	Eosinophils
	Monocytes
	Absolute Neutrophils
	Absolute Lymphocytes
	Absolute Basophils
	Absolute Eosinophils
	Absolute Monocytes
Medication	ACE Inhibitors
	Anti-inflammatories
	Bisphosphonates
	Anti-arrhythmics
	Calcium Antagonists
	Beta blockers
	Fibrates
	Statins
	Metformin
	Sulphonylureas
	Glinides
	Insulin administration
	Vasodilators

	Anti-aggregants
	Diuretics
	Gastroprotectors
	Micronutrients
	Antibiotics
	SNC Medication
	Thyroid Medication
	Lipid Therapy
	Glycaemia Therapy
	Hypertension Therapy
	Presence Therapy
	Metabolic Syndrome (ATPIII diagnosis)*
	Metabolic Syndrome (IDF diagnosis)*
Survival-like variables	Smoker
	Smoking Years
	Number Cigarettes
	Stop Smoking Year
	Diabetes
	Diabetes Duration
	Hypertension Medication
	Hypertension Onset Year
	Cholesterol Medication
	Hypercholesterolemia Onset Year
	Acute Myocardial Infarction
	Acute Myocardial Infarction Onset Year
	Ictus
	Ictus Onset Year
Genetic markers	Haplogroup
	RFLP markers
	Pavia
Genetics geographical precedence	Municipality
	Province

Genetics genealogical precedence	Paternal Grandfather
	Paternal Grandmother
	Maternal Grandfather
	Maternal Grandmother

***ATPIII**: adult treatment panel III; ***IDF**: international diabetes foundation; **unk**: unknown; **yellow ochre color**: variables used in all analyses; **green color**: variables used for the global O2PLS; **grey color**: variables never used.

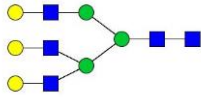
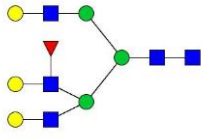
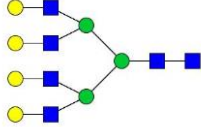
The variables in gray (a total of 32) were removed – except the “Diabetes” variable –, either because they were redundant with other variables or were out of the scope of this research.

5. N-Glycans

10 quantified total plasma N-glycan abundances measured with DSA-FACE were used for the analysis (**table 2**).

Table 2. 10 N-glycans obtained with the DSA-FACE (images obtained with the Glycoworkbench software)

Short Code (DSA-FACE Code)	Structure
GP1 (NGA2F)	
GP2 (NGA2FB)	
GP3 (NG1(6)A2F)	
GP4 (NG1(3)A2F)	
GP5 (NA2)	
GP6 (NA2F)	
GP7 (NA2FB)	

GP8 (NA3)	
GP9 (NA3F(b))	
GP10 (NA4)	

16. Preprocessing: Clinical Database and N-glycans

Summary of the preprocessing – Phenomics data set

The preprocessing was run 3 times for the exploratory data analysis (EDA) visualization plots. On the 1st time 2 impossible values were detected: Height=110cm (BMI=61), Glucoamylase_Lys.Enz=-80.3 (negative value), which were considered as missing values. Plots and summary tables were obtained and a screening was made for extreme outliers (>3/4 times the 0.01/0.99 percentiles), and 13 values were removed. After that, preprocessing was run a second time, in which variables that were skewed were transformed and pairs of variables considered redundant (highly correlated) were filtered (1 out of the pair). The 3rd run contains plots from the whole processed data set.

6. Categorical variables to binary

The variables “Anti-aggregants”, “Vasodilators”, “Anti-arrythmics”, “Anti-inflammatories” and “Statins” from the group “Medication” were all converted to binary, i.e. she takes medication – YES/NO, instead of considering the class of drug.

7. Center log-ratio (clr) transformation on N-glycans and Blood cells %

Compositional data are measures of proportions, percentages, parts per million, etc. which sum up to the unity or 100 constraint. In the case of N-glycans, this means that when they are transformed to percentage, the level of one glycan increases at the expense of the decrease of (another) glycan(s). This might lead to a misinterpretation of the correlation coefficients between the original glycans, commonly referred as the negative bias problem (Houwing-Duistermaat et al., 2017). To deal with compositional glycans, we used the center log-ratio (clr) transformation, mathematically expressed as:

$$\text{clr}(x) = (\log(x_1/g(x)), \dots, \log(x_D/g(x)))$$

where x represents the composition vector, $g(x)$ is the geometric mean of the composition x , and x_D is Euclidean distances between the individual N-glycan variables. The clr was carried out with the *compositions* package from R software. In **figure 8** the glycans distribution is shown for non- clr and clr transformed glycans.

Before applying *clr* transformation, it was noted that 17 rows in the Glycomics data set were completely missing, and thus removed (N=1168). No other missing values were detected in the data set.

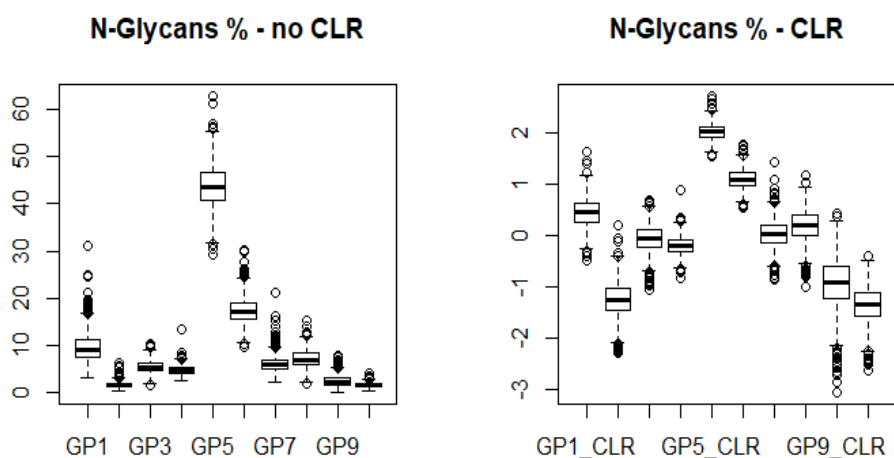


Figure 8. Distribution of N-glycans without vs without clr transformation

The same applies to blood cell percentages, thereby the sum of neutrophils %, lymphocytes %, basophils %, eosinophils % and monocytes % equals 100. Accordingly, the clr too was performed in blood cell counts (**figure 9**). These 5 variables had a total of 6 NA, but it should be noted that the *clr* deals with missing values when performing the transformation.

The *clr* on N-glycans and blood cells was conducted on separate dataframes (df), respectively; afterwards, they were rejoined to the “Phenomics” df for the imputation.

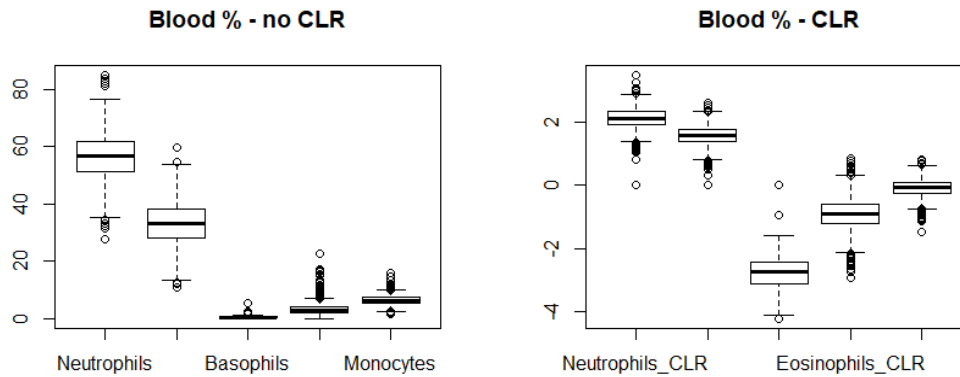


Figure 9. Distribution of Blood Cells without vs without clr transformation

8. Missing values and imputation

1164/1168 rows contained missing values, which were present in 70/122 columns. Briefly, 96 variables had <1% NA, 11 variables had 1-25% NA, 0 variables had between 25-50% NA, and 16 variables >50% NA.

The 16 variables >50% NA were deleted, so that the final number of clinical variables became 106. All 106 ($\leq 25\%$ NA) were imputed with the *mice* package.

9. Outliers detection and treatment

After a first process of data cleaning (following NA deletion or imputation), different plots (histograms, Q-Q plots, boxplots) and tables (summary statistics and quantiles) were produced (data not shown). Cohorts are explained in the next section, but so far N=292 controls (Ctrl), N=313 prediabetics (PreDiab) and N=563 diabetics (Diab).

These EDA served to check for data distribution and to detect whether there could be possible influential outliers that could drastically bias/change the fit of estimates and predictions.

In general, values were not extreme, indicating that the best is to include them over other methods, like trimming based on reference values – up to 130 samples would be lost – or winsorizing, i.e. imputing the most external top and bottom outliers to the 5th and the 95th quantiles, which would introduce bias. Based on the EDA, the following extreme outliers were detected:

- Age: 17-Ctrl
- Creatinine: 6-Ctrl
- HOMA: 47.3-Diab
- Triglycerides: 1528-Diab

- Estimated glomerular filtration rate (EGFR): 364.8-PreDiab, 7.3-Ctrl
- C Reactive Protein: 130.6-PreDiab
- Gamma Glutamyl Transpeptidase: 946-Diab
- Total Bilirubin: 7.5-Diab, 9-Ctrl
- Ferritin: 1500-PreDiab
- Platelets: 965-Diab
- Alkaline Phosphatase: 300-Ctrl, 410-Diab

These values were removed (except for a technical reason the “EGFR=7.3-Ctrl” (ID and value could not be read by R)), giving to: N=288 Ctrl, N=310 PreDiab and N=557 Diab (final N=1155).

10. Descriptive Statistics

After dealing with missing data and extreme outliers, data were visually check for normally distribution or skewness through histograms, Q-Q plots, boxplots and tables of summary statistics (data not shown). From plots and tables it could be inferred that the following variables are:

- **Highly right skewed:** Creatinine, AST, HOMA, ALT, Triglycerides, Insulin, C Reactive Protein, Gamma Glutamyl Transpeptidase, Ferritin, Alkaline Phosphatase, Azotemia, Uric Acid, Telomere Length, Symmetric Dimethylarginine, Asymmetric Dimethylarginine, Red blood cell Distribution Width.
- **Highly left skewed:** Mean Corpuscular Volume.

1. *Kolmogorov-Smirnov test*

For the 17 highly right skewed variables, a Kolmogorov-Smirnov (K-S) normality test was performed to numerically check for normality. “Cholesterol” was used as reference variable, since it follows a fair Gaussian distribution. The null hypothesis that each of the skewed variables follows a normal distribution and, thus, it is not different from the “Cholesterol” distribution was rejected in all skewed variables (p-value < 2.2e-16).

11. Skewed Continuous Variables Transformation

Next, 10 different transformations were tested in each variable – log, square root, raise to the power of coefficients ranging from 0.1 to 0.3 – to decide which transformation suited most to each variable.

2. *Summary*

- **No transformation:** Mean Corpuscular Volume.
- **Log:** Creatinine, AST, ALT, Triglycerides, Insulin, Gamma Glutamyl Transpeptidase, Alkaline Phosphatase, Azotemia, Uric Acid, Telomere Length, Red blood cell Distribution Width.
- **Square root:** Symmetric Dimethylarginine, Asymmetric Dimethylarginine.
- **Raise to the power 0.125:** Ferritin.
- **Raise to the power 0.275:** C Reactive Protein.
- **Raise to the power 0.3:** HOMA.

12. Redundant Correlated Variables Filtering

A high correlation between two variables means they have similar trends and are likely to bring similar information. This can bring down the performance of some models drastically (e.g. linear and logistic regression models (Sharma P (2018))).

A Pearson correlation matrix between numerical variables was calculated, and for variables crossing a certain threshold, one of them was deleted. Since dropping a variable is highly subjective, choosing between one of them was done based on domain knowledge, so that, for example, for blood variables, those already surpassing a threshold of ≥ 0.5 or ≤ -0.5 were removed, while for the rest of variables a threshold of ≥ 0.7 or ≤ -0.7 was set, excluding core variables for the study, such as HbA1c/HOMA. Ten variables (7 blood-related and 3 lysosomal enzymes) were dropped:

- Basophils
- Red blood cell Distribution Volume – SD
- Mean Corpuscular Hemoglobin Concentration
- Hematocrit
- Platelet Large Cell Ratio
- Platelets Distribution Width – SD
- Mean Corpuscular Hemoglobin
- Glucoamylase
- ASAH1 %
- ASAH1

A second Pearson correlation matrix was computed after filtering for those variables, to visually assess for correlations among the remaining variables (not shown).

17. Cohorts: controls, prediabetics, diabetics

As mentioned before, N=288 Ctrl, N=310 PreDiab and N=557 Diab (final N=1155).

These cohorts were established by the diabetologists of Ancona (where the database was produced), who determined if individuals were diabetics or not. However, since there is a time gap between the diagnosis of T2DM – ranging from 70's up to the early 2000, being the 2 most recent diagnosed individuals in 2005, plus incomplete cases – and the data collection – year 2009 –, it was considered appropriate to redefine the groups based on the 2 first diagnostic reference values from table (extracted from De Fronzo et al. 2015), in agreement with the American Diabetes Association (ADA), because they could be more informative than the previous ones. For that, I did not change the “diabetes” status established by the diabetologists from Ancona (to date, diabetes cannot be cured, just controlled), and I just took the individuals with absence of diabetes (according to the diabetologists) and classified them into controls and prediabetics following the ADA criteria (**subsection Diagnosis**).

CHAPTER 1: Characterization of the T2D-Dysmetabolic-Aging Phenotype

4. INTRODUCTION

Aging is a time-dependent multifactorial process that results in a global deterioration of the physiological functions and elevated risk of pathologies, including CVD, neurodegenerative diseases, cancer and diabetes (López-Otín et al., 2013). Over recent years, the new approach termed “Geroscience” has focused the attention on a limited number of highly interconnected “pillars”, shared between aging and age-related diseases (López-Otín, 2013; Kennedy et al., 2014). Among them, inflammation plays a key role, though being indispensable for survival and paradoxically non-detrimental in childhood and adulthood (Furman et al., 2019). A particular feature of the old immunophenotype is the chronic, low-grade, sterile inflammation status, termed inflamm-aging, that can be largely explained by a disequilibrium between inflammatory and anti-inflammatory networks. Within this context, immune cells are major players in eliciting inflammatory responses and constantly adapting their functions at facing internal and external stimuli, but not unique.

Different organs, tissues and cells, such as adipose tissue (AT), liver, muscle, fibroblasts or endothelial cells, are also capable of inducing a proinflammatory status, thus contributing to systemic inflammation and to subsequent chronic diseases (Salvioli, 2013). Of note, adipocytes and hepatocytes are pivotal metabolic cells that, under chronic overnutrition, promote inflammatory local pathways, which, in turn, drive the recruitment of proinflammatory macrophages. In the long run, this may result in obesity, CVD and hepatic steatosis, which constitute intermediary phenotypes that combined raise risk for metabolic syndrome (MetS) (Grundy et al., 2004). Besides, chronic obesity-associated inflammation greatly contributes to insulin resistance and posterior β -cell inflammation and dysfunction, increasing risk for T2DM. Within this framework, a common denominator for these maladies is the low-grade, chronic inflammation state induced by metabolic cells in response to excess nutrients and energy, termed metaflammation (Gregor et al., 2011).

Metaflammation and inflamm-aging share molecular mechanisms of inflammation that characterize aging and age-related diseases. For example, inflammatory cytokines can impair the insulin signalling pathway by inactivating the insulin receptor substrate 1 (IRS1) via phosphorylation, thus leading to the inactivation of the phosphatidylinositol-3-kinase (PI3K)-Akt path. From a clinical point of view, some T2DM patients are non-obese and some obese patients are non-diabetic, but anti-inflammatory drugs do not prevent diabetes and CVD events progression, which points towards inflammation as link between obesity and diabetes development (Franceschi et al., 2000; Franceschi et al., 2017; Prattichizzo et al., 2018). Though, immunometabolic disorders often come up in clusters and nurture aging and age-related diseases (Hotamisligil, 2006; 2017). In point of fact, the World Health Organization estimates that 39% adults worldwide are overweight and 13% are also obese, and that between 2015 and 2050 the world's population over 60 years will increase from 12% to 22%. These worrisome numbers challenge biomedical research at identifying new biomarkers for the diagnosis of age-related diseases. A field that has a lengthy trajectory, but it is only recently being recognized by immunologists, is glycomics (Zhou et al., 2018), upon which this chapter will focus on for reviewing shared links between inflamm-aging and metaflammation.

Glycomics refers to the studies that strive for defining and quantifying the glycome of a cell, tissue or organism (Hart and Copeland, 2010). The glycome is the entire set of glycans, or carbohydrates, and glycoconjugates produced by a cell or organism under specific conditions (Bertozi and

Sasisekharan, 2009). The covalent addition of glycans to proteins and lipids, termed glycosylation, is the most abundant post-translational modification and by far the most structurally richest. In mammalian cells, protein glycosylation can be mainly classified into three classes: the co-translational N-linked glycosylation, the post-translational O-linked glycosylation, and the O-linked β -N-acetylglucosamine (O-GlcNAc) modification (O-GlcNAcylation). N-linked chains are those glycans attached to the nitrogen of asparagine; O-linked glycans are the ones attached to the oxygen of serine or threonine residues, and comprise O-linked N-acetylgalactosamine(O-GalNAc), O-fucose, and O-mannose glycans; while O-GlcNAcylation is an ancient modification whereby GlcNAc is swiftly added or removed, being uridine diphosphate (UDP)-GlcNAc the metabolic sensor of cellular condition (Zhou et al., 2018). Not surprisingly, these post-translational modifications directly or indirectly take part in a myriad of cellular and extracellular functions (Varki, 2017). Importantly, glycans and glycan-binding proteins are core to a properly functioning immune system (Johnson et al., 2013), and the genetic makeup and environmental changes can significantly impact the glyco-immune homeostasis, favouring/triggering different pathological conditions (Lauc and Zoldoš, 2010). These encompass infectious and complex diseases, including cancer, neurological disorders and severe inflammation (Moran et al., 2011, Theodoratou et al., 2014), as well as the aging process. Aberrant changes in glycosylation are particularly relevant for immunoglobulin G (IgG) effector functions, since these antibodies can prompt proinflammatory responses and fuel metaflammation and inflamm-aging. Indeed, glycoforms of the immune system and of circulating proteins can be captured by examining either the whole spectrum of N-linked glycans (N-glycome) or the N-glycans attached to IgGs (IgG glycome) present in serum or plasma. In this regard, recently developed high-throughput (HTP) methods have significantly enhanced investigations on N-glycans.

In this chapter, we briefly introduce the main characteristics of glycosylation and the immunometabolic crosstalk between aging and metabolic syndrome. Then, we describe the pathophysiological roles of glycans and their receptors in inflamm-aging and metaflammation. Finally, we discuss the most relevant N-glycan biomarkers in these inflammatory processes.

18. Intracellular biological roles of glycans: focus on cellular metabolism

13. Metaflammation intra-organ crosstalk

Obesity elicits various triggering events (*step 1*), such as hypoxia, lipotoxicity, ER stress, that initiate activation of inflammatory signaling pathways in parenchymal cells, such as adipocytes and hepatocytes (Lee et al., 2018; **figure 10**).

These cells, in turn, secrete different chemokines, which drive chemotaxis and migration of macrophages and other immune cell types, into

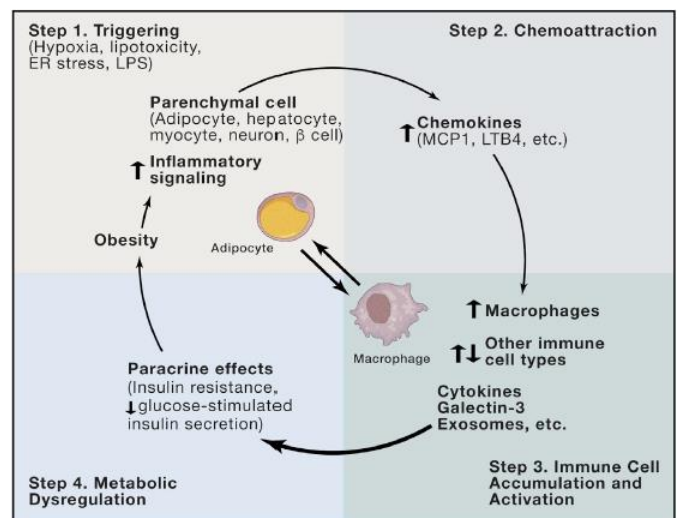


Figure 10. Metaflammation intra-organ crosstalk

the underlying tissue (*step 2*; Lee et al., 2018; **figure 10**).

Overall, these immune cells take on a proinflammatory phenotype and secrete a number of factors (cytokines, galectin-3, exosomes; *step 3*), which exert local paracrine effects to cause insulin resistance in adipocytes, hepatocytes and myocytes, or decreased insulin secretion stimulated by glucose in beta cells (*step 4*; Lee et al., 2018; **figure 10**).

14. Metaflammation inter-organ crosstalk

Obesity gives rise to chronic inflammation in metabolic tissues (liver, adipose tissue, muscle; **figure 11**, extracted from Lee et al., 2018). It also promotes dysbiosis in the gastrointestinal tract and gliosis in the central nervous system. These tissues secrete factors that not only act locally (as paracrine effects), but also enter the bloodstream to cause distal effects on insulin sensitivity and food intake. Each factor has the potential to foster the production and secretion of some other factors, adding to the complexity of the integrated network comprising metabolic inter-organ crosstalk (Lee et al., 2018).

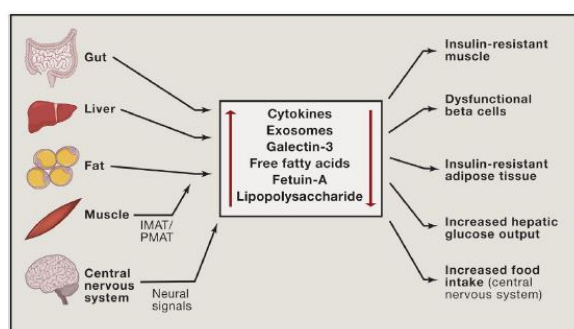


Figure 11. Metaflammation inter-organ crosstalk

15. Metaflammation intracellular crosstalk

In obesity, NF- κ B can be turned on by multiple mechanisms (**figure 12**, extracted from Catrysse and van Loo, 2017): microbiota-derived lipopolysaccharide (LPS), free fatty acids (FFAs), advanced glycation end products (AGEs), inflammatory cytokines, oxidative stress, and ER stress all recruit inflammatory signaling cascades activating NF- κ B kinase subunit 2 (IKK2). On activation, IKK2 will phosphorylate insulin receptor substrate (IRS) proteins on inhibitory Ser sites, prompting insulin resistance. In addition, IKK2 can induce the activation of the inhibitory insulin regulatory proteins mammalian target of rapamycin (mTOR) and ribosomal protein S6 kinase1 (S6K1) by suppressing the

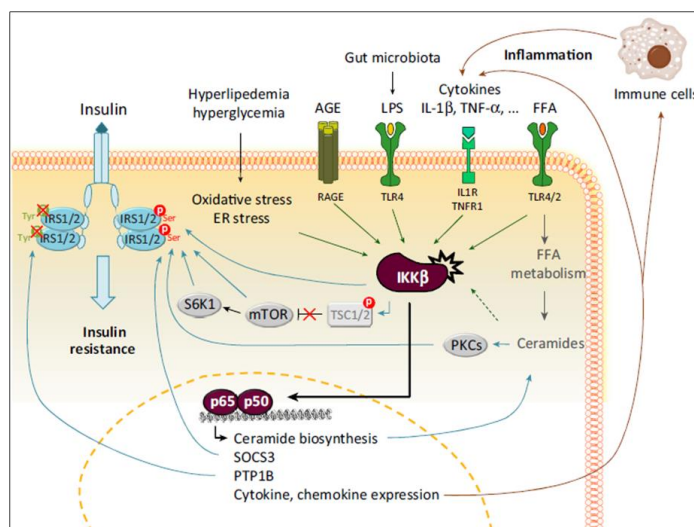


Figure 12. Metaflammation intracellular crosstalk

activity of tuberous sclerosis 1/2 (TSC1/2). NF- κ B also affects insulin signaling by transcriptional induction of protein tyrosine phosphatase 1B (PTP1B) and suppressor of cytokine signaling 3 (SOCS3), which can both interfere with the phosphorylation state of IRS proteins. Also, ceramide biosynthesis genes are upregulated, leading to increased production of ceramide from FFA metabolism and the activation of protein kinase C (PKC), which contribute to the development of insulin resistance. NF- κ B signaling also induces a low-grade inflammatory environment by driving the production of a wide range of inflammatory chemokines and cytokines, which further leads to the recruitment of immune cells, augmenting the inflammatory response (Catrysse and van Loo, 2017).

16. Glucotoxicity in the frame of metaflammation

In this subsection, a mechanism contributing to metaflammation is described. Accordingly, the higher the input concentration of glucose and glutamine to the cell, the higher the concentration of UDP-GlcNAc, and, consequently, more proteins become O-GlcNAcylated (Issad et al., 2010; **figure 13**).

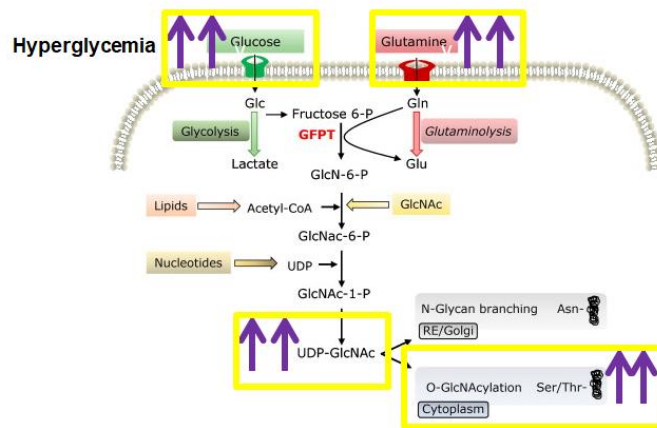


Figure 13. O-GlcNAcylation in conditions of hyperglycaemia

Insulin activates the Akt/PI3K pathway in the cell, which subsequently boosts glucose transport and metabolism. ~2-5% of glucose is directed towards the HBP, which induces the biosynthesis of UDP-GlcNAc, the OGT substrate. OGT goes to the plasma membrane and glycosylates (O-GlcNAcylation) the proximal elements of insulin signaling, resulting in signal attenuation. In chronic hyperglycemia, this chronic glycosylation in signaling proteins may foster the deterioration of insulin sensitivity (Issad et al., 2010; **figure 14**).

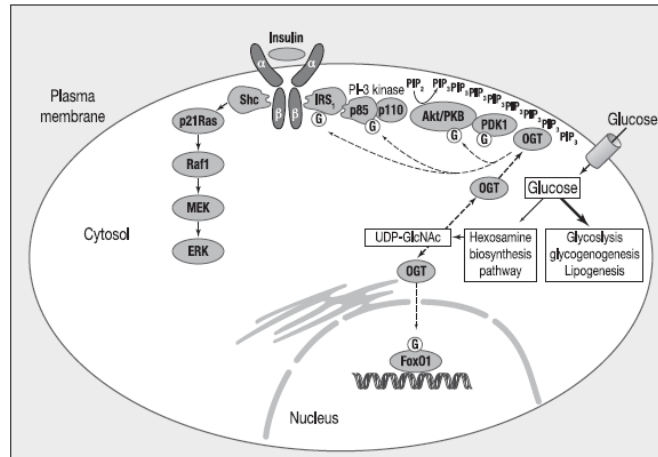


Figure 14. Relationship between chronic glycaemia, chronic glycosylation and insulin sensitivity

However, it has not been established a direct causal relationship between the increase in the UDP-GlcNAc intracellular levels and the development of insulin resistance, since, even though that the raise in the intracellular concentrations of glucose and glutamine results in the increase of UDP-GlcNAc and O-GlcNAc levels, there are three factors accounting against this idea: 1) the increase in oxidative and ER stress, that promotes chronic inflammation and insulin resistance; 2) UDP-GlcNAc inhibits the glutamine-fructose-6-phosphate transaminase (GFAT) enzyme, which is the rate limiting enzyme controlling the HBP flux, and 3) the inhibition of OGT in animal models in hyperglycaemic conditions still results in insulin resistance (reviewed from Copeland et al., 2008).

Hyperglycemia is not only linked to insulin resistance, but also to cardiovascular disease. In conditions of hyperglycemia (muscle/fat), there is an activation of glucose responsive genes that are associated to insulin resistance (**figure 15**). For example, the transcription factor SP1 is glycosylated by OGT and it enhances the expression of PAI-1, which is an inhibitor of fibrinolysis, and TGF beta, a profibrotic factor, which contribute to atherosclerotic lesions in obese and diabetic patients. But there can be involved many transcription factors, genes and proteins (Copeland et al., 2008).

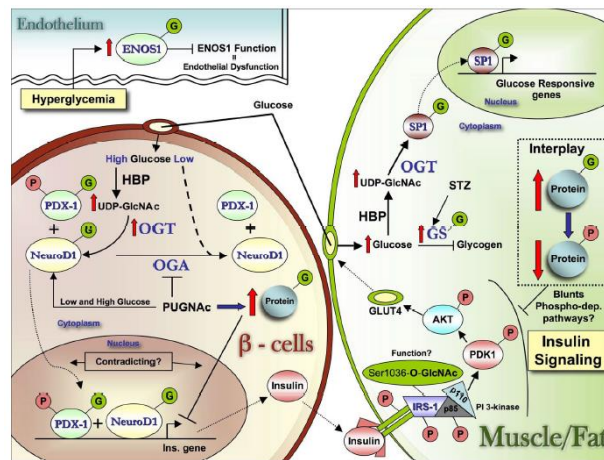


Figure 15. Hyperglycaemia, glycosylation and cardiovascular disease

Another important thing in CVD is that endothelial dysfunction is an abnormality that may accompany insulin resistance and diabetes but not necessarily be a direct cause of it (Copeland et al., 2008).

17. Lipotoxicity in the frame metaflammation

In this subsection, another mechanism involving protein and glycosylation and T2D is summarized. In islets from wild-type mice fed with normal chow, Foxa2 and HNF1a control the expression of Slc2a2 (Glut2) and of Mgat4a (GnT-4a, an N-acetylglucosamine transferase, **figure 16**, extracted from Thorens, 2011). GnT-4a is required for the formation of a N-glycan

structure on Glut2 (and Glut1 in human islets), a process that takes place in the secretory pathway. The N-glycan is required to anchor Glut2 at the cell surface through interaction with a lectin, galectin9 (Gal9). In HFD-fed mice or upon exposure of mouse or human islets to palmitic acid, Foxa2 and HNF1a are excluded from the nucleus, by a mechanism that can be inhibited by the antioxidant N-acetylcysteine (NAC), suggesting that reactive oxygen species (ROS), produced as a result of lipid catabolism, are engaged in this process. Nuclear exclusion of the transcription factors reduces GnT-4a and Glut2 expression, as well as the normal glycosylation of Glut2, leading to transporter internalization, reduced cell-surface expression, and reduced GSIS. This results in glucose intolerance, insulin resistance, and hepatic steatosis. These deregulations can be prevented by transgenic overexpression of Mgat4a and Slc2a2 (Ohtsubo et al., 2011; Thorens, 2011).

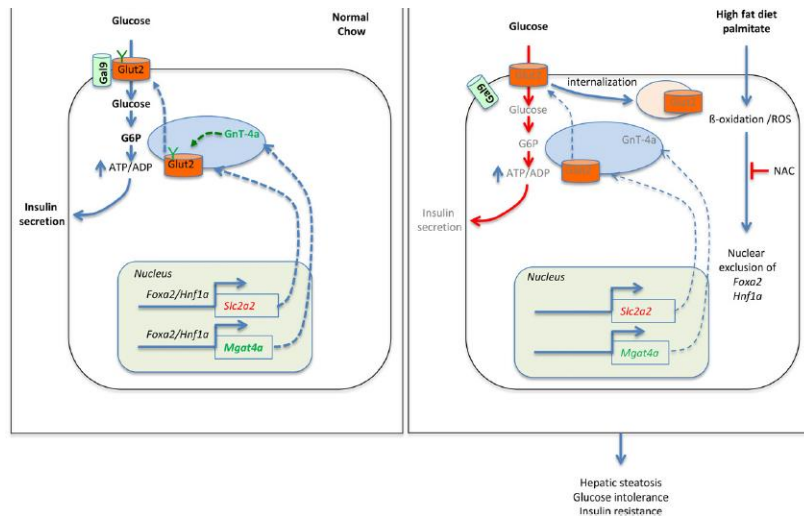


Figure 16. The absence of glycosylation of Glut2 in mice fed with high fat diet impairs insulin secretion

18. Inflammasomes activation following glucotoxicity / lipotoxicity

Another mechanism contributing to inflamm-aging/metaflammation is the activation of the intracellular inflammasomes.

It has been proposed a model for IL-1beta-induced inflammation in response to diabetes-associated DAMPs (Shin et al., 2015). Long-lasting exposure of macrophages or pancreatic cells to elevated levels of glucose and palmitate boosts NLRP3 activation through a mechanism that involves ROS and TXNIP activation, resulting in an increased production of mature IL-1. In

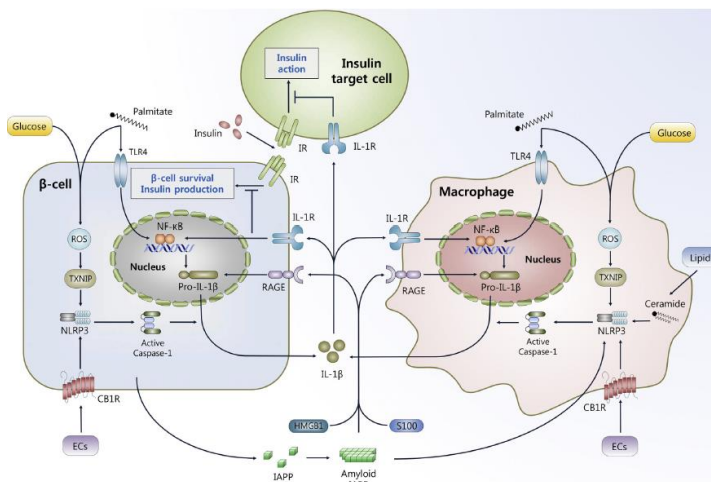


Figure 17. Inflammasomes activation in glucotoxicity/lipotoxicity

addition, palmitate stimulates the expression of IL-1 through NF- κ B activation via TLR4. Ceramide can also trigger the processing of pro-IL-1 into its mature form through NLRP3 activation in macrophages. IAPP secreted from cells aggregates into insoluble islet amyloid deposits and is internalized by macrophages, inducing NLRP3 activation and the production of mature IL-1. Endocannabinoids (ECs) also promote NLRP3 activation and IL-1 production via CB1R. RAGE and its ligands (AGE, HMGB1, S100 protein, and IAPP) can enhance IL-1 expression and beta-cell death. Deleterious levels of IL-1 induce beta-cell death and deregulate insulin receptor (IR) signaling, potentially leading to insulin resistance in insulin target cells (revised from Shin et al., 2015).

19. Endoplasmic reticulum stress response

Endoplasmic reticulum is a well-known target organelle for inducing inflammation (**figure 19**, extracted from Gregor and Hotamisligil, 2011). There have been reported three different ER stress pathways leading to inflammation, deriving from the ER's unfolded protein response (reviewed from Gregor and Hotamisligil, 2011). IRE-1 uses its kinase domain partnering with TRAF2 to activate the inflammatory kinases JNK and IKK; this leads to upregulation of inflammatory mediators via the transcription factors AP-1 and NF- κ B, respectively. IRE-1 splicing of XBP1 mRNA also triggers inflammatory consequences, as XBP1 has been shown to regulate inflammatory cytokine stimulation and immune responses in various cell types, especially macrophages. PERK activation results in decreased translation of I κ B α , a suppressor of NF- κ B signaling, thereby augmenting NF- κ B transcriptional activity. In addition, PERK activation mediates ATF-4 translation, and ATF-4 was shown to regulate inflammatory cytokine induction, although the mechanism remains unknown. PKR is also activated by ER stress and contributes to JNK and IKK activation. Finally, ATF-6 has also been demonstrated to increase NF- κ B transcriptional activity (reviewed from Gregor and Hotamisligil, 2011).

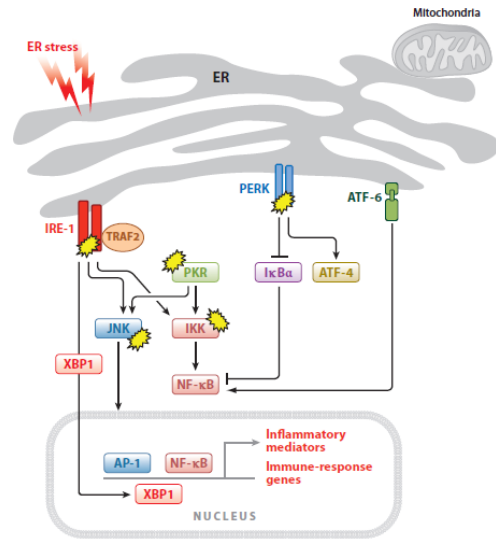
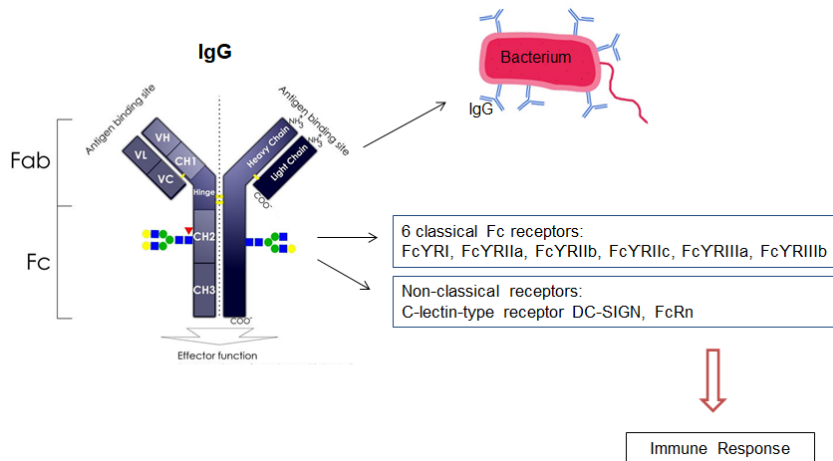


Figure 18. Endoplasmic reticulum stress responses

19. Extracellular biological roles of glycans: focus on inflammation (IgG glycosylation)

20. Antibody glycosylation controls antibody activity



Each IgG has an N-glycan attached at Asn297 in the polypeptide chain of each heavy part (figure 19, modified from Higuel et al., 2016); since the constant Fc region of the heavy chains defines the capacity of the antibody to interact with Fc receptors, the IgG immune response can be measured as a function of the interaction between

IgG N-glycan composition and Fc receptor type

Figure 19. Antibody glycosylation controls antibody activity

(Higuel et al., 2016). There are two major groups of Fc receptors: classical (Fc γ RI, Fc γ RIIa, Fc γ RIIb, Fc γ RIIc, Fc γ RIIIa, Fc γ RIIIb) and non-classical (C-lectin-type receptor DC-SIGN, FcRn) (Winkel and Anderson, 1991). It is interesting to note that 70% of human IgGs are asymmetrically glycosylated (being both Fc chains always glycosylated, while the Fab regions are glycosylated only in 20% of IgGs), and that the addition and removal of sugars (galactose, fucose, b-GlcNAc, or sialic acid) are directly linked to altered antibody functionality (Higuel et al., 2016; Jennewein and Alter, 2017). In this line, it has been described that: removal of fucose enhances IgG affinity for Fc γ RIIIa and thus it boosts antibody-dependent cellular cytotoxicity (ADCC) (Seeling et al., 2017), exposed bisecting GlcNAc residues (“G0”) also promote ADCC (Seeling et al., 2017), fucosylation and bisection are largely mutually exclusive (Seeling et al., 2017), removal of galactose either increases or decreases affinity for Fc γ R (depending on specific Fc γ R receptor and IgG subclass), and presence of sialic acids has been linked to anti-inflammatory effects as they decrease binding to Fc γ RIII receptors (Seeling et al., 2017).

21. Glycovariation in physiological and pathological conditions

Glycosylation is a highly ordered and conserved process. However, disease-associated changes (e.g autoimmune or infectious diseases) suggest that it is actively modulated during inflammatory responses. The specific mechanisms leading to this variation in glycosylation, though, are incompletely understood. Nevertheless, some factors that shape antibody glycosylation have been proposed (reviewed from Zhou et al., 2018):

- Glycosyltransferase/glycosidase expression
- Shifts in monosaccharide availability
- ER stress

- Golgi pH and organization
- Kinetics of protein production
- Availability of vesicular transport machinery

22. Aberrant antibody glycosylation in ageing and disease: the egg or the chicken?

In this subsection, I modestly disclose the most important research questions related to my investigation.

What comes first: metaflammation or inflamm-aging? (**Figure 20**) Do they become an unavoidable unbreakable dichotomy in the long run? In other words, are older patients (>60 years old), already exhibiting a low-grade pro-inflammatory phenotype and with more abundance of agalactosylated glycans species, more prone to metaflammation and other pathophysiological changes in protein glycosylation? Do N-glycan species related to inflamm-aging differ from N-glycan species related to metaflammation? There is some evidence in literature that this might be the case (see next subsection). Although more studies are needed, considering this starting point, another question comes up: are glycans the cause or the consequence of metaflammation and/or inflamm-aging?

Rheumatoid arthritis (RA) and the human immunodeficiency virus (HIV) are two examples that aid to provide a better insight to these questions. In the case of RA, loss of galactosylation on IgG1 appears in circulation before onset, shortly after the development of anti-citrullinated (CCP) antibodies (Ercan et al., 2010); therefore, in RA, the sequence of events apparently is 1) auto-antibodies, 2) loss of galactosylation, 3) onset of disease, pointing to that glycans are the consequence of RA disease. In HIV, it has been demonstrated that spontaneous controllers (low to undetectable viral replication) without retroviral therapy, maintain the highest levels of agalactosylated antibodies (reviewed from Zhou et al., 2018), which suggests that in HIV glycans are also the consequence of the disease.

In the case of metaflammation and glycosylation, the answer to what is first is intuitive, as the question becomes too simplistic, due to the complexity and heterogeneity of the disease. Thus, it should be reformulated: do N-glycans change over the course of preclinical (that is, insulin resistance and impaired insulin secretion) to clinical diagnoses (obesity and/or T2D, T2D complications, metabolic syndrome)?

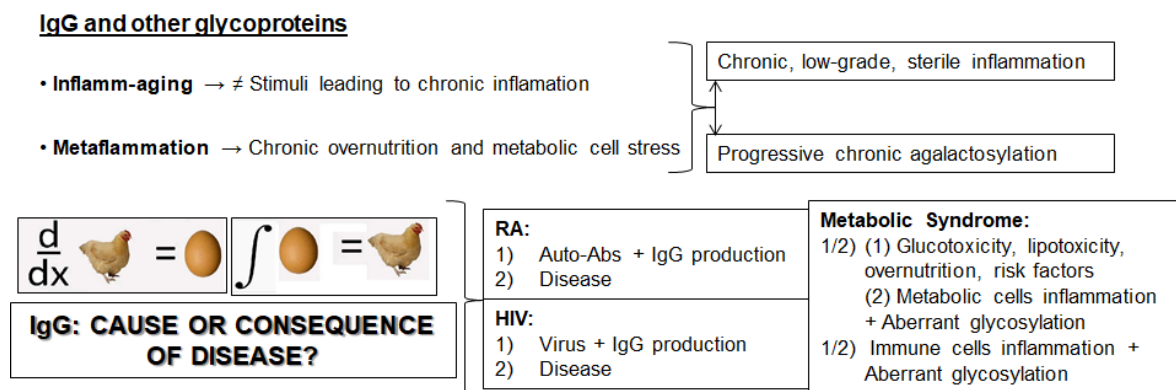


Figure 20. What comes first in RA, HIV and metabolic syndrome, disease or inflammation?

23. Advanced glycation end products / HbA1c: surrogate biomarkers of N-glycans in T2D?

Though further research is needed to answer the posed research questions concerning N-glycans species as disease trackers, specifically in T2D, it has been shown that the so-called advanced glycation end products (AGEs) do precede diabetes mellitus (Vlassara and Striker, 2011). AGEs are a heterogeneous family of sugar-amino acid adducts formed by a nonenzymatic covalent binding of reducing sugars to protein amino groups.

Vlassara and Striker postulated that current processed food contains appetite-enhancing AGEs, that prompt food consumption and overnutrition, which in the long-term lead to increased BMI, obesity and diabetes mellitus, as well as oxidant overload. Steady influx of nutrient AGEs and advanced lipoxidation end products (ALEs) leads to suppression of innate host defenses and an overflow of intracellular ROS, which increases the basal oxidant stress and inflammation (reviewed from Vlassara and Striker, 2011). The interaction of these processes can simultaneously cause β -cell dysfunction, impaired insulin secretion and insulin resistance, as well as diabetic complications. Restriction of food-derived AGEs reduces oxidative stress and prevents or improves type 1 and type 2 diabetes mellitus in mice (figure 21, extracted from Vlassara and Striker, 2011).

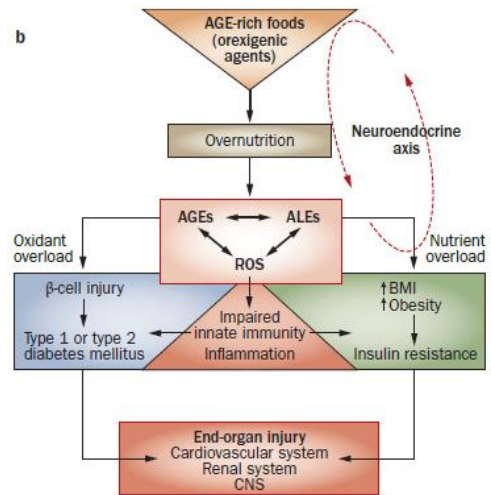


Figure 21. The relationship between AGEs and T2D/obesity

The most studied glycosylated protein is glycosylated hemoglobin (or HbA1c), which, indeed, has been shown to display a strong association with AGEs. As mentioned in the general introduction, HbA1c serves as indicator of long-term glycaemia (average lifespan of red cells is of 4 months). In 2010, the American Diabetes Association (ADA) added HbA1c as a further diagnostic criterion for diabetes and prediabetes in an attempt to shorten the diagnosis time of T2D (the average time between onset and diagnosis of T2D is 7 years) and as a marker of glycaemic control in established patients – in T2D individuals, the rate of formation of HbA1c is a direct function of the average blood glucose concentration – (Gillett, 2009; American Diabetes Association 2010). In this regard, we wondered whether a combination of N-glycans would predict the T2D status better than HbA1c alone (see aims of thesis below), considering *a*) the nature of the bond type (non-enzymatic in N-glycans vs enzymatic in HbA1c), and *b*) the origin of bond formation (mostly intracellular in N-glycans vs extracellular in HbA1c).

20. N-glycans as biomarkers in liver diseases, aging, and age-related diseases

In this section, I review major studies conducted in aging and (the underlying diseases that give rise to) metabolic syndrome which use N-glycans as candidate biomarkers.

24. Liver diseases

Most of plasma proteins, and thus N-glycans (Apweiler et al., 1999), are synthesized in the liver and pancreas (Uhlén et al., 2015), whereas immunoglobulins are synthesized specifically in cells of the immune system (Rhoades and Pflanzner, 2002). Particularly, liver is not only an important source of glycoproteins generation, but also a metabolic organ. Thus, it is worthy to focus on liver diseases to better understand N-glycan profiling of age-related diseases (e.g. T2D or obesity). In this line, a bunch of literature has extensively reported the characteristic derived glycan traits (i.e. N-glycan features: sialylation, galactosylation, bisection, fucosylation) and the implicated enzymes in liver pathologies (**figure 22**, extracted from Takahashi et al. (2016)).

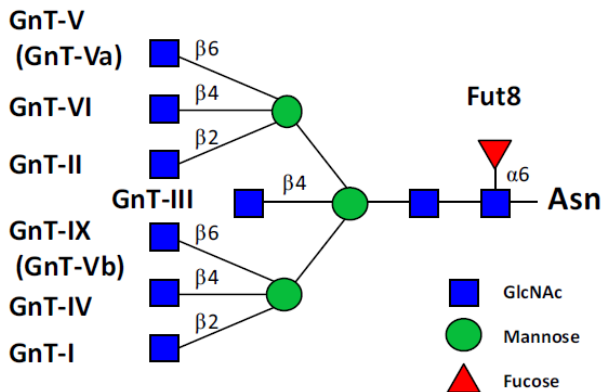


Figure 22. Glycosyltransferases enzymes

- (Liver) diseases: **hyperfucosylation**, **↑branching**, **↑bisection N-GlcNAc**
- **Alcoholic liver disease (ALD)**: ↓dolichol, desialylation (↓ST6GalI ↑sialidase in hepatocytes), ↓mannosylation (ApoE), ↓galactosylation, ↑liver weight (hepatic steatosis)
 - Biomarkers of abnormal glycosylation: transferrin, ↑haptoglobin (↑branching, ↑fucosylation)
- **Fatty liver diseases:**
 - GnT-III (↑bisection GlcNAc) expression (hepatocytes)
 - α1,6-fucosyltransferase expression (hepatocytes)
- **Hepatocellular carcinoma (HCC) and cirrhosis:**
 - GnT-III (↑bisection GlcNAc) expression (hepatocytes)
 - ↑ST6GalI (in contrast to ALD)
 - Biomarker: ↑ fucosylated α-fetoprotein (AFP)

Figure 23. Summary of main findings regarding alteration of glycosylation in liver pathologies

3. Alcoholic liver disease

In alcoholic liver disease (ALD), haptoglobin, α1-acid glycoprotein, α2-HS glycoprotein, and transferrin are known to exhibit an increased branching. The Golgi apparatus plays an essential role in the alteration of glycosylation patterns in all liver diseases. In ALD, there is a characteristic significant accumulation of hepatic protein caused by impaired glycosylation and glycoprotein

trafficking. A proposed explanation for the diminished Golgi functioning is the deficient polymerization of microtubular protein as a downstream consequence of hepatic acetaldehyde due to ethanol oxidation. In an experimental rat model of ALD by Gosh et al., it was described a decreased activity of the mannosyltransferase and galactosyltransferase enzymes, a lowered intracellular dolichol concentration, and a strong decreased synthesis and activity of ST6GalI. The authors also reported an increase of 30% in liver weight as compared to the body weight of that rat model, which was attributed to the accumulation of hepatic lipids and proteins leading to fatty liver and steatosis. In this research, they also studied O-glycosylation in Apo E, which showed, as in N-glycosylation, a decrease in mannosylation and sialylation. The relative ratio of labeled sugar to leucine incorporation (glycosylation index) revealed a 50% reduction in relative mannosylation of Apo E at both the microsomal and the Golgi level. The impairment in the glycan structures of Apo E was hypothesized to be responsible for the defective clearance of HDL and VLDL, resulting in a defective cholesterol transport to the liver and the subsequent hepatic accumulation (reviewed from Blomme et al., 2009; Callewaert et al., 2004).

4. *Fatty liver diseases*

At least two different studies investigated the impact of aberrant glycosylation by overexpressing N-acetylglucosaminyltransferase (GnT) III in transgenic mice. An increased GnT-III activity results in an increased level of bisecting GlcNAc in Apo B, which impairs the glycoprotein functionality by preventing its release in bloodstream, and thus its accumulation in the liver. Not only Apo B concentrations were significantly elevated in the liver, but also Apo A1. These studies point to that N-glycans can have an important impact on lipid metabolism (reviewed from Blomme et al., 2009; Callewaert et al., 2004).

Not only GnT-III, but also the ectopic expression of α 1,6-fucosyltransferase (α 1,6-FT) causes steatosis in the liver and the kidney (reviewed from Blomme et al., 2009; Callewaert et al., 2004). The increased expression of this enzyme lowered the activity of the lysosomal acid lipase (LAL), the accumulation of which might contribute to the lipid accumulation observed in the lysosomes of the liver.

5. *Hepatocellular carcinoma and cirrhosis*

Alteration of glycosylation has also been documented in malignant cellular transformation. In hepatocellular carcinoma (HCC), three glycosyltransferases are considered crucial: GnT-V, GnT-III and α 1,6-FT. GnT-V is coded on the *MGAT5* gene, which is regulated by the Ras signaling pathway (commonly up-regulated in cancerous cells). Though GnT-III and GnT-V compete for the same substrate, the bisecting GlcNAc, and the activity of both enzymes is increased in HCC, the activity of GnT-III is more prominent; on the other hand, GnT-V is directly associated with metastasis.

In HCC, the activity of ST6GalI is increased, in contrast to ALD, in which is decreased. But the best known biomarker in HCC is the elevated serum concentration of fucosylated α -fetoprotein (AFP) (reviewed from Blomme et al., 2009; Callewaert et al., 2004).

25. Type 2 diabetes mellitus

Table 3 shows a summary of collected literature about major glycomic biomarkers findings in T2D. As a whole, the studies report increased levels of N-glycan branching (Testa et al., 2015; Keser et al., 2017; Adua et al., 2018) in whole plasma glycome. Regarding sialylation, there is controversy: two observational studies show a reduction in levels of sialylation (Lemmers et al., 2017; Dotz et al., 2018) in the IgG plasma N-glycome of T2D patients, measured with MALDI-TOF, and an experimental analysis by Tanigaki et al. (2018) demonstrated that IgG hyposialylation in mice promotes obesity-induced insulin resistance. Conversely, Keser et al. (2017) revealed an increase in trisialylated glycans in whole plasma N-glycome of patients with hyperglycaemia, Liu et al. (2019) also evinced increased levels of sialylation whole plasma N-glycome in an Uyghur population of T2D patients. With respect to digalactosylated structures, Adua et al. (2018) and Testa et al. (2015) report decreased levels, while Liu et al. (2019) shows an increase in bigalactosylation. There are also divergent results with regard to bisecting GlcNAc structures: Lemmers et al. (2017) shows an increase while Liu et al. (2019) shows a reduction.

The causes of divergencies between studies should be deeply analyzed in each case, but possible sources are: technology used for the collection of N-glycans, IgG or whole N-glycome, small sample size, lack of randomization, unbalanced number of cases vs controls, poor variables matching in case-control studies, confounders not considered (thus, structure present in the data that explain the results), possible batch effects, different normalization techniques, models used for the analyses and the covariates employed.

Table 3. Summary of the major glycomic signatures of T2D described in literature

Set up	Analysis Overview	Main results	Reference
MALDI-TOF 70 N-glycans, 96 derived traits IgG plasma N-glycome <i>Discovery cohort:</i> 1583 cases, 728 controls. <i>Replication cohort:</i> 232 cases, 108 controls.	Logistic regression models: Model1 = Group ~ N-glycans + Sex + Age + Sex*Age Model2 = Model1 + BMI Model3 = Model2 + HDLc + nonHDLc + smoking	↓alpha2,3-linked sialylation	Dotz et al., 2018
MALDI-TOF 70 N-glycans, 96 derived traits, IgG plasma N-glycome <i>Discovery cohort:</i> DiaGene: N=1886 cases, N=854 controls <i>3 Replication cohorts:</i> CROATIA-Korcula	1. Logistic regression models (in each of the 4 cohorts): Model1 = Group ~ N-glycans + Sex + Age + Sex*Age Model2 = Model1 + BMI Model3 = Model2 + HDLc + nonHDLc + smoking 2. Meta-analysis of the 4 cohorts: weighted z-transform	IgG in T2D: ↓galactosylation ↓sialylation ↑bisection of fucosylated structures	Lemmers et al., 2017

<p>CROATIA-Vis ORCADES N=162 cases, N=3162 controls</p>	<p>method, Bonferroni correction for multiple testing 3. Logistic regression models for prediction analyses</p>		
<p>HILIC-HPLC 46 N-glycans, 12 derived traits, Whole plasma N-glycome <i>Case/control study:</i> Cases: hyperglycaemic Controls: normoglycaemic Cases or controls with critical illness are considered to have higher risk for T2D. <i>Discovery cohort:</i> N=59 cases, N=49 controls <i>Test cohort:</i> N=52 cases, N=14 controls <i>3 Replication cohorts:</i> FinRisk population: N=37 T2D, N=37 controls ORCADES: N=94 cases with ↑[HbA1c], N=658 controls SABRE: N=307 cases with ↑[HbA1c], N=307 controls</p>	<p>Analyses of associations between clinical trait of interest and glycan measurements were performed using a regression model with age and sex included as additional covariates.</p>	<p>Individuals who developed hyperglycaemia had: ↓low-branching ↑high-branching ↑trigalactosylated ↑tetragalactosylated ↓neutral glycans ↑trisialylated glycans</p>	<p>Keser et al., 2017</p>
<p>UPLC 39 N-glycans Total plasma N-glycome Case cohorts from the Prospective EPIC - Potsdam cohort (n=27548): Incident cases of T2D (N=5820; median follow-up time 6.5 years) Cases of CVD (N=5508; median follow-up time 8.2 years)</p>	<p>1. N-glycan data were adjusted for age by applying multiple fractional polynomial (MFP) regression separately in men and women. The age-adjusted residuals were used to construct an N-glycan score to predict type 2 diabetes with machine learning techniques. 2. The score was validated in an independent cohort and used in 6 different models to assess the ability of different predictors to</p>	<p>The N-glycan-based type 2 diabetes score was strongly predictive for diabetes risk (weighted C-index 0.83, 95% CI 0.78–0.88). N-glycans were moderately predictive for CVD incidence (weighted C indices 0.66, 95% CI 0.60–0.72, for men; 0.64, 95% CI 0.55–0.73, for women).</p>	<p>Wittenbecher et al., 2020</p>

	<p>predict type 2 diabetes.</p> <p>4. Selection of potential CVD-specific predictors was conducted analogous to the selection of diabetes-related N-glycans with machine learning techniques, deriving a weighted score.</p>		
<p>DSA-FACE</p> <p>10 Whole serum N-glycans</p> <p><i>Italian cohort:</i></p> <p>N=562 T2D, N=599 healthy controls</p>	<p>1) Comparison of serum N-glycan profiles in T2DM, without and with complications,</p> <p>2) association of N-glycans with diabetic complications,</p> <p>3) associations with MetS,</p> <p>4) correlation of N-glycans profiling with MetS parameters</p>	<p>↓Monogalactosylated, core-fucosylated diantennary N-glycans in T2D compared with controls</p>	<p>Testa et al., 2015</p>
<p>HILIC-UPLC</p> <p>Total plasma N-glycome</p> <p>39 N-glycans and 22 derived traits</p> <p><i>Cohort:</i></p> <p>Ghana population</p> <p>N=232 cases, N=219 controls</p>	<p>Association between N-glycans and age in both males and females for cases and controls, determined by linear regression, and multiple testing adjustment with Benjamini-Hochberg method to control FDR</p>	<p>T2DM compared with controls:</p> <p>↑high branching</p> <p>↑trigalactosylation</p> <p>↑antennary fucosylated</p> <p>↑triantennary</p> <p>↓low branching</p> <p>↓non-sialylated</p> <p>↓biantennary galactosylation</p> <p>↓biantennary structures</p>	<p>Adua et al., 2018</p>
<p>HILIC-UPLC</p> <p>IgG plasma N-glycome</p> <p>24 N-glycans</p> <p>Case-control Australian population cohort:</p> <p>N=217 cases, N=632 controls</p>	<p>1) logistic models: disease status ~ IgG glycan traits + covariates (age, sex, BMI, WHR, SBP, and DBP), multiple testing with Benjamini-Hochberg procedure;</p> <p>2) classification model based on the significantly altered glycans to evaluate the potential of IgG N-glycans as T2D biomarkers</p>	<p>Two directly measured and four derived glycan peaks were significantly associated with T2DM.</p>	<p>Li et al., 2019</p>
<p>HILIC-UPLC</p> <p>IgG subclass Fc N-glycopeptidome plasma/serum (unknown)</p>	<p>1. Associations between IgG subclass-specific Fc N-glycopeptide profiles and disease status were performed using logistic regression, adjusting for risk factors for</p>	<p>27 directly measured and 4 derived glycan traits of the IgG subclass-specific N-glycopeptides were significantly associated</p>	<p>Liu et al., 2019</p>

24 N-glycans Case-control Uygur population cohort: N=115 cases, N=122 controls	T2DM covariates; 2. LASSO to select IgG subclass-specific Fc N-glycans; 3. Model evaluation	with T2DM. IgG2 in T2DM: ↓bisecting GlcNAc ↑digalactosylation IgG4 in T2DM: ↓agalactosylation ↑sialylation	
Human and transgenic C57BL/6J mouse pancreatic beta cells	Experimental analyses in human and mice beta pancreatic cells to: induction of molecular events by exposure of elevated levels of free fatty acids (FFA) or by administration of high-fat diet associated with obesity	↑[FFA] caused nuclear exclusion and ↓expression of the transcription factors FOXA2 and HNF1A in beta cells; this resulted in a deficit of GnT-4a glycosyltransferase expression in beta cells that produced signs of metabolic disease, including hyperglycemia	Ohtsubo et al., 2011
Male C57BL/6 WT, FcγRIIB ^{-/-} , Fcγ ^{-/-} , B ^{-/-} , SAP ^{-/-} , FcγRIIB ^{fl/fl} and VECad-Cre mice, or in offspring from their mating	Experimental analyses	↓IgG sialylation in mice is implicated in obesity-induced insulin resistance	Tanigaki et al., 2018

26. Aging research

Table 4 shows a summary of collected literature about major glycomic biomarkers findings in aging. In general, studies agree in the fact that over age there is a progressive decrease in bigalactosylated species that goes along an increase of agalactosylated species.

Table 4. Summary of the major glycomic signatures of aging described in literature

Set up	Analysis Overview	Main results	Reference
HILIC and WAS HPLC 33 IgG plasma N-glycome population-based study of N=1914 individuals from Vis and Korkula	Multiple linear regression used to analyze the association of selected predictor variables and each glycan feature	F2, G0, G2, S2, A2 glycan traits: age-dependent ↑galactosylation correlate with lipid status	Knezevic et al., 2010
DSA-FACE	GlycoAge test definition	GlycoAgeTest: -Significantly ↑in dementia patients	Vanhooren et al., 2010

<p>10 whole serum N-glycan</p> <p>Participants:</p> <p>N=425 healthy Belgian adults</p> <p>N=84 Italian centenarians</p> <p>N=79 with dementia</p> <p>N=6 Cockayne Syndrome (CS)</p>		<p>compared with age-matched controls.</p> <p>-Highly heterogeneous in CS, but even the lowest value is much higher than in the controls</p>	
<p>HILIC-HPLC</p> <p>26 whole plasma N-glycans</p> <p>N=2396 middle aged participants in the Leiden Longevity Study (LLS)</p>	<p>Regression strategies applied to evaluate associations between glycan patterns, familial longevity, and healthy aging.</p>	<p>↑A2G2: associated with healthy ageing (longevity)</p> <p>↓A2G1S1: associated with occurrence of myocardial infarction</p>	<p>Ruhaak et al., 2011</p>
<p>UPLC</p> <p>24 IgG plasma N-glycans</p> <p>Participants:</p> <p>N=906 Vis island</p> <p>N=915 Korčula island</p> <p>N=2035 Orkney islands</p> <p>N=1261 TwinsUK</p>	<p>A predictive model of age was built ("GlycanAge")</p>	<p>IgG N-glycans markers of age: FA2B, FA2G2, and FA2BG2, explaining 58% variance</p>	<p>Kristic et al., 2014</p>
<p>MALDI-TOF-MS</p> <p>12 tryptic IgG1 and IgG2 Fc glycopeptides</p> <p>N=1967 participants in the LLS</p>	<p>Several regression strategies were applied to evaluate the association of IgG glycosylation with age, sex, and longevity</p>	<p>Galactosylated glycoforms: ↑bisecting GlcNAc with age.</p> <p>Sex-related differences below 60 years: younger females had higher galactosylation, which ↓stronger with age</p>	<p>Ruhaak et al., 2010</p>

5. MATERIALS AND METHODS

This section, following the general description of materials and methods, describes the algorithm used in the analyses, that is two-way orthogonal partial least squares (O2PLS) (**figure 24**):

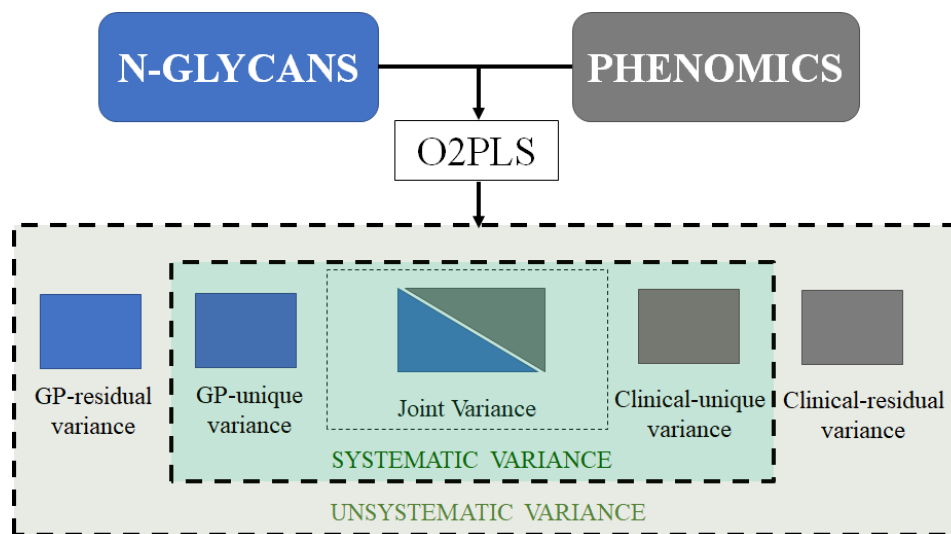


Figure 24. Two-way orthogonal partial least squares algorithm

O2PLS is a dimension reduction technique that decomposes the variance of 2 data sets in 3 parts (Bouhaddani et al., 2016):

- Joint part: biologic variance common to both data sets (clinical and N-glycans).
- Single-omics specific part: the variance specific to each data set.

A critical point for identifying the joint and the specific parts is to choose the number of joint and specific principal components (PCs) through 1) the elbow technique by the visualization of the eigenvalues plot, and 2) cross-validation. The analyses were done with the *OmicPLS* package. Since in this thesis dissertation, a high number of O2PLS models were performed, the number of PCs was chosen only by cross-validation, by selecting the combination of joint and specific PCs with lower mean-squared error (MSE).

PART I

6. AIMS

The general aims of this study are:

1. To assess whether N-glycans predict diabetic or non-diabetic status better than the HbA1c.
2. To identify the major sources of variation both in Phenomics and Glycomics in four age ranges in six groups of individuals (Ctrl F, Ctrl M, PreDiab F, PreDiab M, Diab F, Diab M) and to identify whether such sources of variation correspond to major biological conditions (e.g. prediabetic or diabetic status, metabolic syndrome).
3. To visualize the integrated weights (i.e. loading values) and mean values of both data sets.

The specific for this study are:

1. To perform a logistic model by modelling a binary outcome composed of controls (N=285) and diabetics (N=555) as function of HbA1c or as function of N-glycans.
2. To identify covariating N-glycan species and endophenotypes characteristic across different age ranges and per sex by considering N=288 controls (FIRB cohort) – individuals with normal ranges of HbA1c, insulin, HOMA and fasting glucose –, N=310 prediabetics and N=557 diabetics – individuals with abnormal ranges of HbA1c, insulin, HOMA and fasting glucose –. This will be achieved by performing O2PLS models on two data sets, one composed of 44 clinical variables (excluding sex, as it does not follow a normal distribution) and another one composed of 10 N-glycans, in each of 4 predefined age ranges (20-44, 45-55, 56-70, 71-85 years) and in the full age range (20-85 years) in Ctrl and, due to smaller sample sizes, 3 age ranges in PreDiab and in Diab (45-55, 56-70, 71-85 years), separately in males and females (total of 23 O2PLS models).
3. To generate line plots of the integrated weights (i.e. loading values) and mean values for the most relevant clinical variables and N-glycans to detect patterns taking place at specific age ranges and between sexes either in “healthy” controls, prediabetics, or diabetics.

7. RESULTS

21. Demographic characteristics of the cohort

The characteristics of the participants included in the study are described in **Table 5** and **Table 6**. **Table 7** illustrates the number and the frequency of individuals that take medication.

Table 5. Patients baseline characteristics

Variable	Ctrl F (n=177)		Ctrl M (n=109)		PreDiab F (n=187)		PreDiab M (n=122)		Diab F (n=259)		Diab M (n=298)	
	<i>mean</i>	<i>Sd</i>	<i>mean</i>	<i>sd</i>	<i>mean</i>	<i>sd</i>	<i>mean</i>	<i>Sd</i>	<i>mean</i>	<i>Sd</i>	<i>mean</i>	<i>sd</i>
Azotemia (log U)	36.38	8.97	40.83	12.56	38.19	9.67	39.22	8.4	40.26	13.89	40.25	10.27
Age (years)	56.4	13.21	65.14	8.27	60.44	11.52	60.24	10.14	66.6	7.73	56.39	12.44
BMI (kg/m ²)	25.75	4.81	28.32	4.29	27.75	5	27.71	3.5	29.66	5.16	27.17	4.23
WHR (cm)	0.83	0.09	0.97	0.06	0.85	0.07	0.94	0.05	0.9	0.06	0.92	0.06
Fasting glucose (mg/dL)	87.2	6.28	165.6	48.75	96.16	10.85	100.9	8.97	160.83	48.72	89.22	5.59
HbA1c (%)	5.43	0.36	7.41	1.3	5.93	0.3	5.82	0.31	7.5	1.22	5.35	0.21
Cholesterol (mg/dL)	218.26	38.8	196.73	35.27	221.53	35.35	210.46	36.32	218.88	37.52	204.33	36.45
HDL (mg/dL)	64.24	13.92	48.72	11.88	60.19	13.92	51.61	13.55	56.68	16.41	50.18	11.81
EGFR (mL/min)	87.14	20.35	80.38	19.86	83.53	17.96	90.77	25.8	76.47	21.25	95.16	26.02
Fibrinogen (mg/dL)	289.5	73.05	296.41	81.01	306.91	77.08	269.5	67.28	315.33	79.98	267.19	84.21
PAI1 (ng/mL)	19.28	10.23	19.56	8.99	22.3	11.14	23.45	12.4	22.41	10.3	25.01	12.24
TAFI (%)	132.39	17.24	126.11	17.3	134.17	21.25	130.18	17.37	133.9	17.26	133.91	18.7
Total Bilirubin (mg/dL)	0.67	0.26	0.73	0.31	0.61	0.25	0.79	0.37	0.62	0.24	0.79	0.35
LDL (mg/dL)	125.04	34.34	110.8	28.13	126.15	29.79	122.78	30.08	123.12	32.67	120.98	30.73
Apo A1 (mg/dL)	186.22	29.68	159.01	31.39	183.99	29.05	168.04	31.68	178.24	34.88	167.64	27.1
Apo B (mg/dL)	97.47	27.9	98.24	26.15	103.05	27.77	103.92	27.67	105.57	26.58	100.38	28.87
Total Iron (mcg/dL)	79.64	27.77	86.45	27.94	78.47	24.15	82.74	31.38	78.63	25.15	84.05	33.5
Transferrin (mg/dL)	255.44	44.75	259.24	37.46	261.49	42.71	258.55	38.09	270.58	47.05	262.99	48.11
Total Protein (g/dL)	7.12	0.43	7.25	0.52	7.03	0.49	7.09	0.45	7.15	0.49	7.2	0.47
White Cells (n/mm ³)	5.97	1.59	6.58	1.57	6.06	1.34	6.45	1.56	6.8	1.65	6.3	1.69
Red Cells (n/mm ³)	4.56	0.39	4.83	0.44	4.53	0.36	4.97	0.36	4.57	0.42	4.96	0.44
Haemoglobin	13.52	0.99	14.89	1.21	13.53	0.99	14.99	0.91	13.6	1.14	14.95	1.11

(g/dL)												
MCV (fL)	88.12	5.31	88.84	5.13	88.82	5.22	88.5	5.06	87.84	5.45	88.77	5.1
Platelets (n/mm ³)	237.81	51.24	199.59	54.53	241.07	58.34	215.73	50.99	239.07	63.57	213.78	57.77
MPV (fL)	11.12	0.98	11.21	0.96	11.11	1.01	11.1	0.85	11.11	0.96	11.17	0.98
Neutrophils (CLR U)	2.1	0.4	2.12	0.33	2.06	0.37	2.05	0.35	2.17	0.33	2.09	0.41
Lymphocytes (CLR U)	1.65	0.32	1.49	0.31	1.57	0.32	1.54	0.33	1.6	0.31	1.54	0.36
Eosinophils (CLR U)	-0.98	0.52	-0.86	0.53	-0.93	0.5	-0.84	0.52	-0.93	0.49	-1	0.52
Monocytes (CLR U)	-0.07	0.26	-0.1	0.3	-0.09	0.3	-0.03	0.27	-0.14	0.28	0	0.33
Creatinine (log U)	0.73	0.14	1.01	0.32	0.74	0.14	0.91	0.22	0.82	0.28	0.89	0.21
Uric Acid (log U)	3.9	1.03	4.95	1.25	4.42	1.19	5.37	1.12	4.68	1.24	5.28	0.98
Alkaline Phosphatase (log U)	70.98	19.66	72.58	21.74	73.4	22.1	71.43	20.77	82.51	26.97	74	18.68
AST (log U)	20.35	5.33	22.55	10.02	21.4	7.45	23.25	10.35	21.61	12.29	25.43	13.21
ALT (log U)	35.79	8.47	44.7	17.14	39.3	15.06	42.63	14.83	41.66	15.04	42.17	12.19
Triglycerides (log U)	82.11	40.64	134.12	96.14	101.97	56.04	127.06	95.86	141.73	86.35	112.29	82.26
Fasting insulin (log U)	5.7	3.61	6.94	6	6.98	5.65	7.76	7.36	7.22	4.84	6.48	4.18
HOMA (pow. 0.3 U)	1.22	0.79	2.95	3.12	1.7	1.56	1.87	2.04	2.9	2.41	1.42	0.88
CRP (pow. 0.275 U)	2.94	3.67	4.46	8.85	4.17	6.71	3.06	4.29	5.07	7.02	3.71	9.9
GGT (log U)	45.16	15.71	60.22	37.14	47.64	16.63	53.7	17.34	55.27	40.86	51.73	19.59
Ferritin (pow. 0.125 U)	69.69	70.39	166.97	162.33	79.96	62.77	125.34	105.88	88.9	78.3	119.48	123.58
Telomere Length (log U)	0.5	0.2	0.43	0.22	0.45	0.18	0.46	0.19	0.45	0.19	0.48	0.16
SDMA (sqrt U)	1	0.57	0.9	0.61	0.96	0.5	1.11	0.67	0.91	0.65	1.1	0.65
ADMA (sqrt U)	1.01	0.6	1.04	0.56	1.04	0.51	1.07	0.67	1.09	0.61	0.92	0.5
RDW-CV (log U)	13.4	1	13.17	0.73	13.48	0.84	13.34	0.8	13.52	1.03	13.3	0.94

Log U: log transformed units; pow. 0.125/0.275/0.3 U: raise to the power of 0.125/0.275/0.3 transformed units; sqrt U: square root transformed units; CLR U: centre-log ratio transformed units; BMI: body mass index; WHR: waist to hip ratio; HbA1c: glycated haemoglobin; HDL: high-density lipoprotein; EGFR: estimated glomerular filtration rate; PAI1: plasminogen activator 1; TAFI: thrombin activatable fibrinolysis inhibitor; LDL: low-density lipoprotein; Apo A1/B: apolipoprotein A1/B; MCV: mean corpuscular volume; MPV: mean platelets volume; CRP: C reactive protein; GGT: gamma glutamyl transpeptidase; SDMA: symmetric dimethylarginine; ADMA: asymmetric dimethylarginine; RDW-CV: red blood cell distribution width – coefficient of variation.

Table 6. Number and frequency of individuals that are affected by clinical conditions

Clinical History	Ctrl F (n=177)	Ctrl M (n=109)	PreDiab F (n=187)	PreDiab M (n=122)	Diab F (n=259)	Diab M (n=298)
Metabolic Syndrome (ATPIII)	12 (7%)	7 (6%)	37 (20%)	14 (11%)	190 (73%)	140 (47%)
Metabolic Syndrome (IDF)	15 (8%)	15 (14%)	59 (32%)	33 (27%)	201 (78%)	180 (60%)
Bypass	2 (1%)	2 (2%)	0	1 (1%)	10 (4%)	39 (13%)
Somatic Neuropathy	0	0	0	0	35 (14%)	67 (22%)
Nephropathy	0	0	0	0	24 (9%)	48 (16%)
Chronic Renal Insufficiency	0	0	0	0	5 (2%)	15 (5%)
Retinopathy	0	0	0	0	72 (28%)	82 (28%)
Arteriopathy Obliterans of Lower Limbs	0	0	0	0	14 (5%)	21 (7%)
Arteriopathy Obliterans of Upper Limbs	0	0	0	0	10 (4%)	16 (5%)
Cardiac Ischemia	0	0	0	0	37 (14%)	62 (21%)
Major Adverse CDV Events	6 (3%)	1 (1%)	3 (2%)	6 (5%)	26 (10%)	58 (19%)

Table 7. Number and frequency of individuals that take medication

Medication	Ctrl F (n=177)	Ctrl M (n=109)	PreDiab F (n=187)	PreDiab M (n=122)	Diab F (n=259)	Diab M (n=298)
ACE Inhibitors	29 (16%)	10 (9%)	37 (20%)	19 (16%)	120 (46%)	113 (38%)
Anti-inflammatories	7 (4%)	4 (4%)	12 (6%)	12 (10%)	55 (21%)	90 (30%)
Bisphosphonates	9 (5%)	0	6 (3%)	0	5 (2%)	3 (1%)
Anti-arrythmics	2 (1%)	1 (1%)	1 (1%)	4 (3%)	18 (7%)	18 (6%)
Calcium-Antagonists	8 (5%)	6 (6%)	14 (7%)	11 (9%)	49 (19%)	86 (29%)
Beta-blockers	12 (7%)	7 (6%)	15 (8%)	9 (7%)	48 (19%)	56 (19%)
Fibrates	0	1 (1%)	1 (1%)	1 (1%)	9 (3%)	4 (1%)
Statins	9 (5%)	8 (7%)	17 (9%)	8 (7%)	55 (21%)	53 (18%)
Metformin	0	0	0	0	108 (42%)	96 (32%)
Sulphonylureas	0	0	0	0	122 (47%)	146 (49%)
Glinides	0	0	0	0	5 (2%)	8 (3%)
Insulin Administration	0	0	0	0	52 (20%)	48 (16%)
Vasodilators	1 (1%)	2 (2%)	2 (1%)	2 (2%)	21 (8%)	30 (10%)
Anti-aggregants	5 (3%)	11 (10%)	8 (4%)	6 (5%)	43 (17%)	55 (18%)
Diuretics	11 (6%)	1 (1%)	10 (5%)	4 (3%)	35 (14%)	30 (10%)
Gastroprotectors	15 (8%)	6 (6%)	12 (6%)	4 (3%)	22 (8%)	23 (8%)
Micronutrients	12 (7%)	3 (3%)	12 (6%)	1 (1%)	22 (8%)	13 (4%)
Antibiotics	3 (2%)	1 (1%)	1 (1%)	0	4 (2%)	4 (1%)
SNC Medication	23 (13%)	2 (2%)	24 (13%)	2 (2%)	32 (12%)	22 (7%)
Thyroid Medication	14 (8%)	2 (2%)	13 (7%)	5 (4%)	14 (5%)	4 (1%)
Lipid Therapy	9 (5%)	11 (10%)	18 (10%)	9 (7%)	65 (25%)	58 (19%)
Glycaemia Therapy	0	0	0	0	195 (75%)	224 (75%)

Hypertension Therapy	44 (25%)	21 (19%)	63 (34%)	30 (25%)	162 (63%)	174 (58%)
Presence Therapy	91 (51%)	34 (31%)	106 (57%)	43 (35%)	240 (93%)	273 (92%)

22. Prediction of (non-)diabetic status by HbA1c and by N-glycans

May N-glycans predict T2D status than the classic HbA1c? To answer that I performed two logistic models, one using HbA1c as predictor and N-glycans for the other model. The total N was of 840 individuals. The areas under the curve (AUC) and the confusion matrixes of the following models evidence that HbA1c predicts diabetic or non-diabetic status far better than the 10 N-glycans. Confidence intervals (CI) for the AUC were generated by bootstrapping, using an $\alpha=10\%$.

Model 1: Ctrl/Diab ~ HbA1c

AUC=0.979 (0.970-0.988)

Table 8. Confusion matrix for HbA1c

	Expected Ctrl	Expected Diab
Observed Ctrl	277	8
Observed Diab	31	524

Model 2: Ctrl/Diab ~ 10 N-glycans

AUC=0.706 (0.782-0.843)

Table 9. Confusion matrix for 10 N-glycans

	Expected Ctrl	Expected Diab
Observed Ctrl	159	126
Observed Diab	81	474

2. Sources of variation in Phenomics and Glycomics data sets

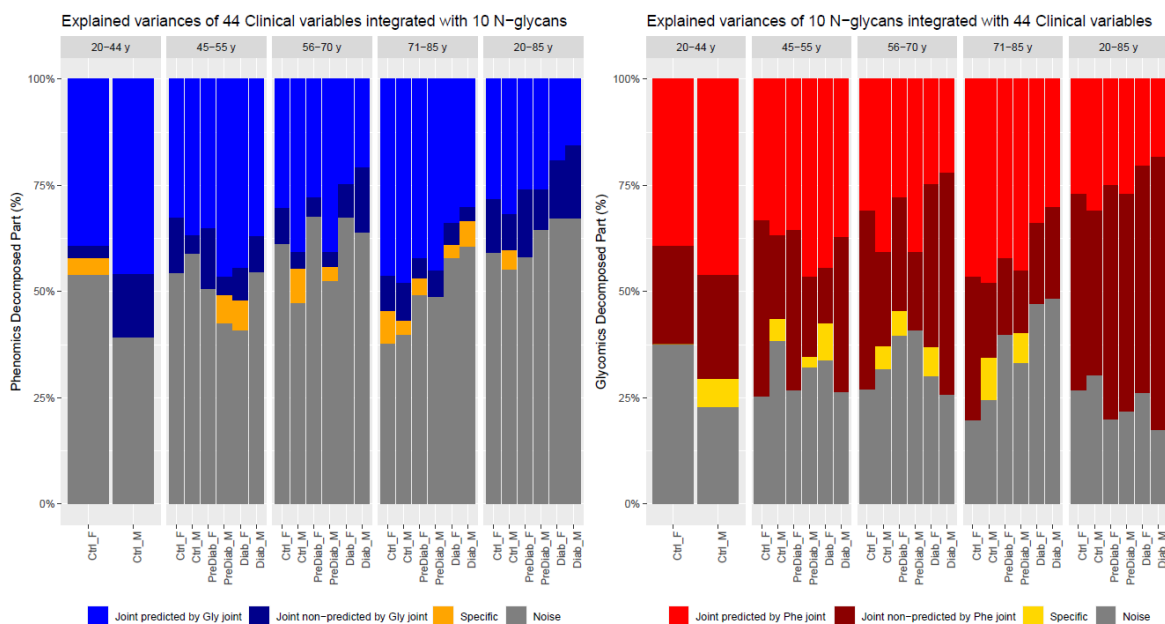


Figure 25. Decomposed parts of the O2PLS integrations between 44 clinical variables and 10 N-glycans in four different age ranges by groups and sexes

Figure 25 illustrates the decomposed parts of the two integrated data sets in all groups by age ranges (on the left, the phenomics joint part; on the right, the glycomics joint part). In general, the joint phenomics data set has lower signal/noise ratio, as compared with the joint glycomics data set; the latter also exhibits a higher percentage of joint part non-predicted by phenomics (thus predicted but something unknown but correlating with this part), when compared with the non-predicted by glycomics in the phenomics data set. These non-predicted parts become more evident as the sample size increases (range 20-85 years); this is especially true for the glycomics data set, in which the noise slightly decreases, in contrast to that in phenomics, which slightly heightens.

In order to make results of joint loadings more comprehensive, the description that follows comprises the six groups (Ctrl F, Ctrl M, PreDiab F, PreDiab M, Diab F, Diab M) considering the full age range (20-85 years), instead of examining the groups in the four age ranges previously shown (**Table 10**). However, in the next section, the loading values for all age ranges will be presented.

Table 10. Sample size of groups by age ranges and in the full age range and frequency of individuals in each age range per group

Age Range	Group					
	Ctrl F	Ctrl M	PreDiab F	PreDiab M	Diab F	Diab M
20-44 y	37 (20.9%)	14 (13%)	18 (9.7%)	9 (7.4%)	2 (0.78%)	5 (1.7%)
45-55 y	44 (24.9%)	39 (36.1%)	45 (24.3%)	28 (23.1%)	19 (7.4%)	32 (10.7%)
56-70 y	74 (41.8%)	39 (36.1%)	88 (47.6%)	67 (55.4%)	157 (61%)	183 (61.4%)
71-85 y	22 (12.4%)	16 (14.8%)	34 (18.3%)	17 (14%)	79 (30.7%)	78 (26.1%)
20-85 y	177	108	185	121	257	298

Total N=1146

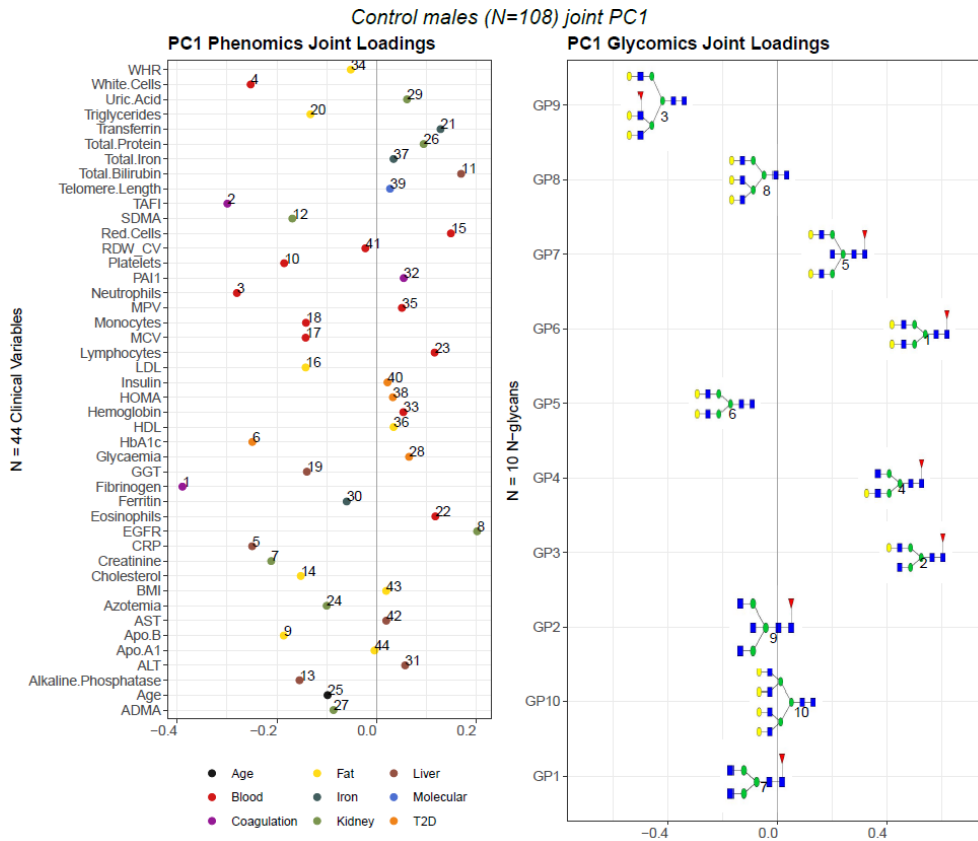


Figure 26. Joint PC1 loading values of control males aged 20-85 y.

Table 11. Top 10 joint PC1 loading values of phenomics and glycomics data sets in Ctrl M

	Phe_load	Phenomics_var	Gly_load	Glycan_var
1	-0.38972	Fibrinogen	0.519007	GP6
2	-0.2994	TAFI	0.504762	GP3
3	-0.28026	Neutrophils	-0.4416	GP9
4	-0.25263	White.Cells	0.426182	GP4
5	-0.24958	CRP	0.219393	GP7
6	-0.2493	HbA1c	-0.19362	GP5
7	-0.21139	Creatinine	-0.07801	GP1
8	0.202448	EGFR	-0.06804	GP8
9	-0.18609	Apo.B	-0.04437	GP2
10	-0.18532	Platelets	0.03005	GP10

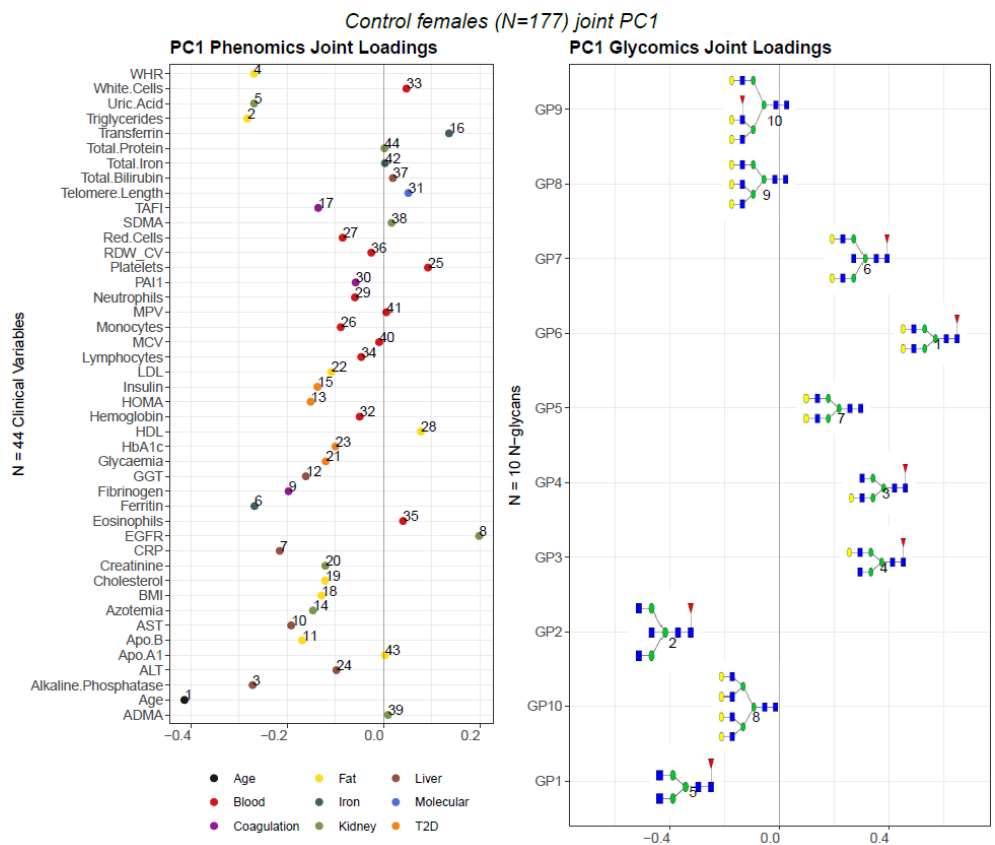


Figure 27. Joint PC1 loading values of control females aged 20-85 y.

Table 12. Top 10 joint PC1 loading values of phenomics and glycomics data sets in Ctrl F

	Phe_load	Phenomics_var	Gly_load	Glycan_var
1	-0.41373	Age	0.54965	GP6
2	-0.28302	Triglycerides	-0.41975	GP2
3	-0.27213	Alkaline.Phosphatase	0.36011	GP4
4	-0.26922	WHR	0.352816	GP3
5	-0.26905	Uric.Acid	-0.34491	GP1
6	-0.26828	Ferritin	0.292037	GP7
7	-0.21569	CRP	0.196503	GP5
8	0.198648	EGFR	-0.11466	GP10
9	-0.19739	Fibrinogen	-0.07649	GP8
10	-0.19192	AST	-0.07536	GP9

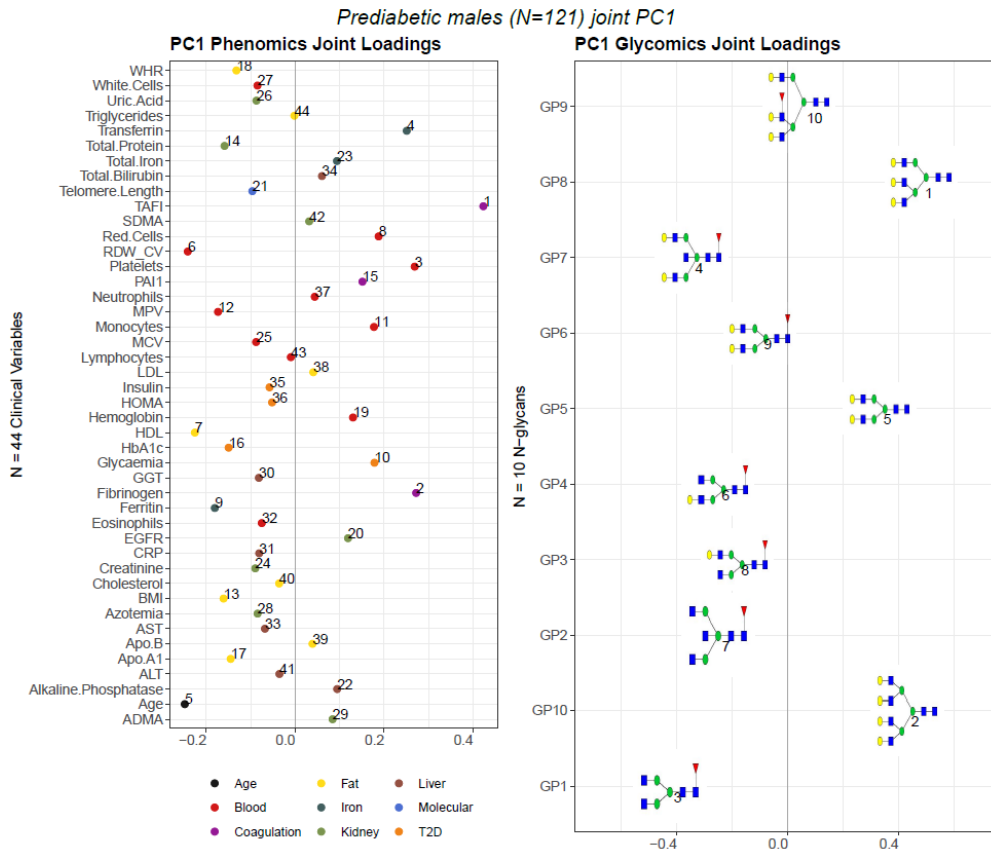


Figure 28. Joint PC1 loading values of prediabetic males aged 20-85 y.

Table 13. Top 10 joint PC1 loading values of phenomics and glycomics data sets in PreDiab M

	Phe_load	Phenomics_var	Gly_load	Glycan_var
1	0.423193	TAFI	0.480792	GP8
2	0.27211	Fibrinogen	0.430459	GP10
3	0.268923	Platelets	-0.42573	GP1
4	0.251033	Transferrin	-0.34782	GP7
5	-0.24822	Age	0.330355	GP5
6	-0.24152	RDW_CV	-0.2525	GP4
7	-0.22559	HDL	-0.25155	GP2
8	0.187685	Red.Cells	-0.18286	GP3
9	-0.18065	Ferritin	-0.10071	GP6
10	0.178323	Glycaemia	0.039649	GP9

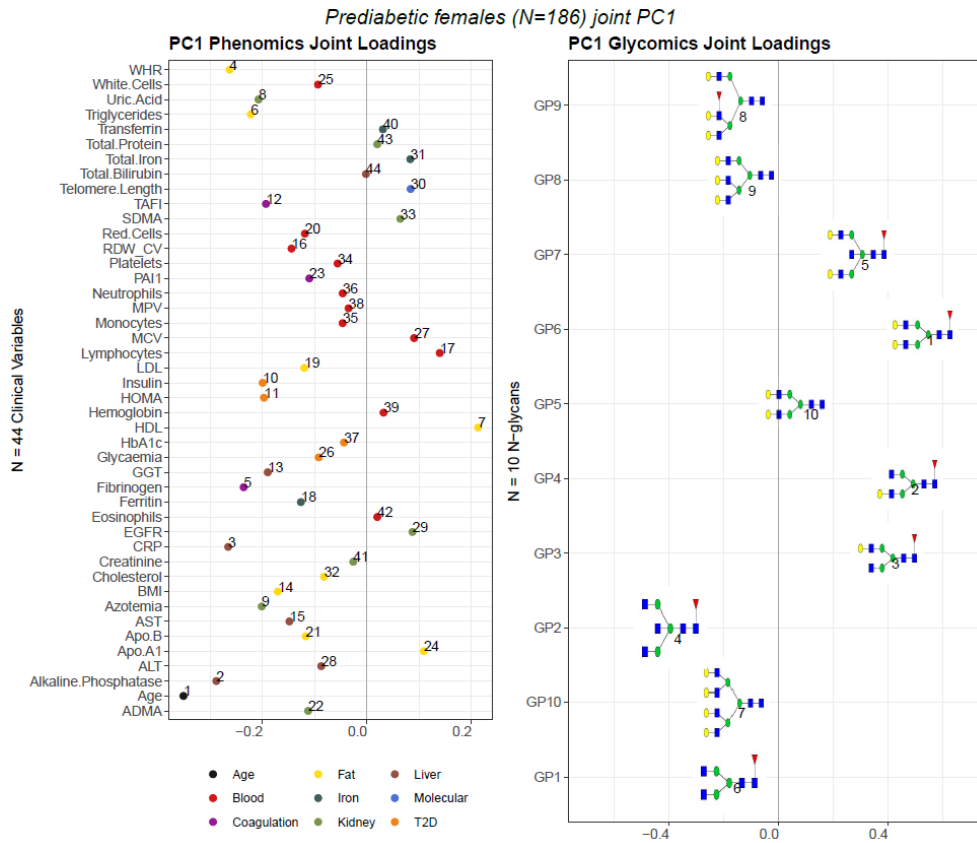


Figure 29. Joint PC1 loading values of prediabetic females aged 20-85 y.

Table 14. Top 10 joint PC1 loading values of phenomics and glycomics data sets in PreDiab F

	Phe_load	Phenomics_var	Gly_load	Glycan_var
1	-0.35149	Age	0.525998	GP6
2	-0.28861	Alkaline.Phosphatase	0.470829	GP4
3	-0.26585	CRP	0.398601	GP3
4	-0.26315	WHR	-0.39574	GP2
5	-0.23627	Fibrinogen	0.286845	GP7
6	-0.22258	Triglycerides	-0.17994	GP1
7	0.212599	HDL	-0.16396	GP10
8	-0.20779	Uric.Acids	-0.15836	GP9
9	-0.20166	Azotemia	-0.12574	GP8
10	-0.19987	Insulin	0.061027	GP5

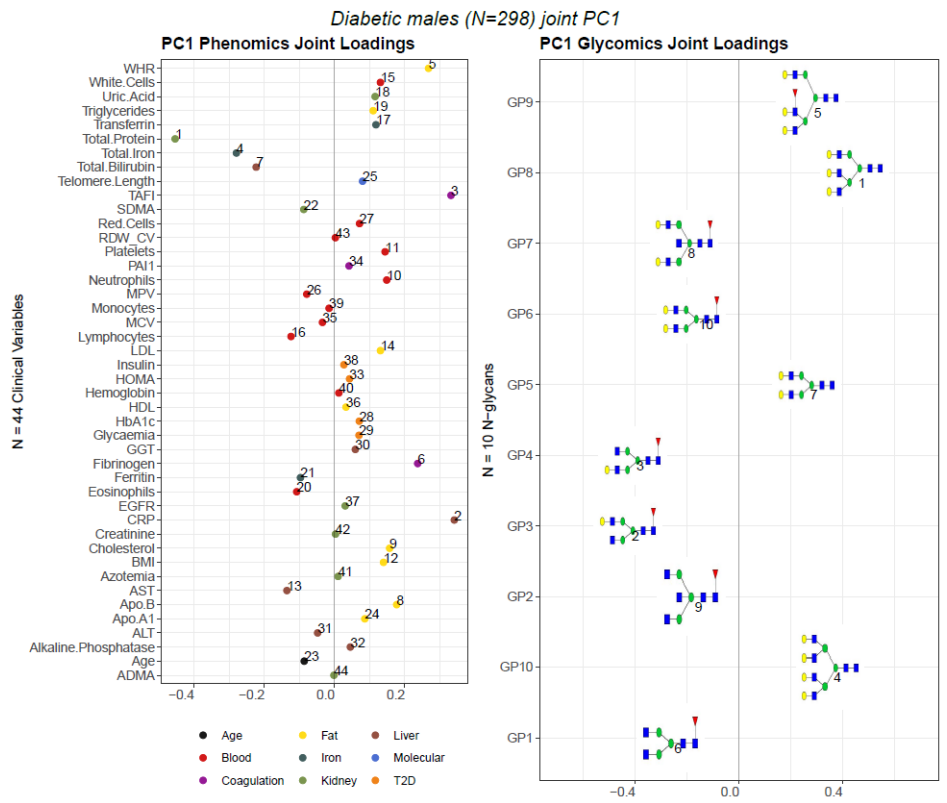


Figure 30. Joint PC1 loading values of diabetic males aged 20-85 y.

Table 15. Top 10 joint PC1 loading values of phenomics and glycomics data sets in Diab M

	Phe_load	Phenomics_var	Gly_load	Glycan_var
1	-0.45435	Total.Protein	0.446021	GP8
2	0.340855	CRP	-0.42861	GP3
3	0.33144	TAFI	-0.40985	GP4
4	-0.2797	Total.Iron	0.351605	GP10
5	0.267464	WHR	0.274701	GP9
6	0.236585	Fibrinogen	-0.26355	GP1
7	-0.22355	Total.Bilirubin	0.260263	GP5
8	0.176752	Apo.B	-0.21204	GP7
9	0.156645	Cholesterol	-0.18503	GP2
10	0.148456	Neutrophils	-0.18413	GP6

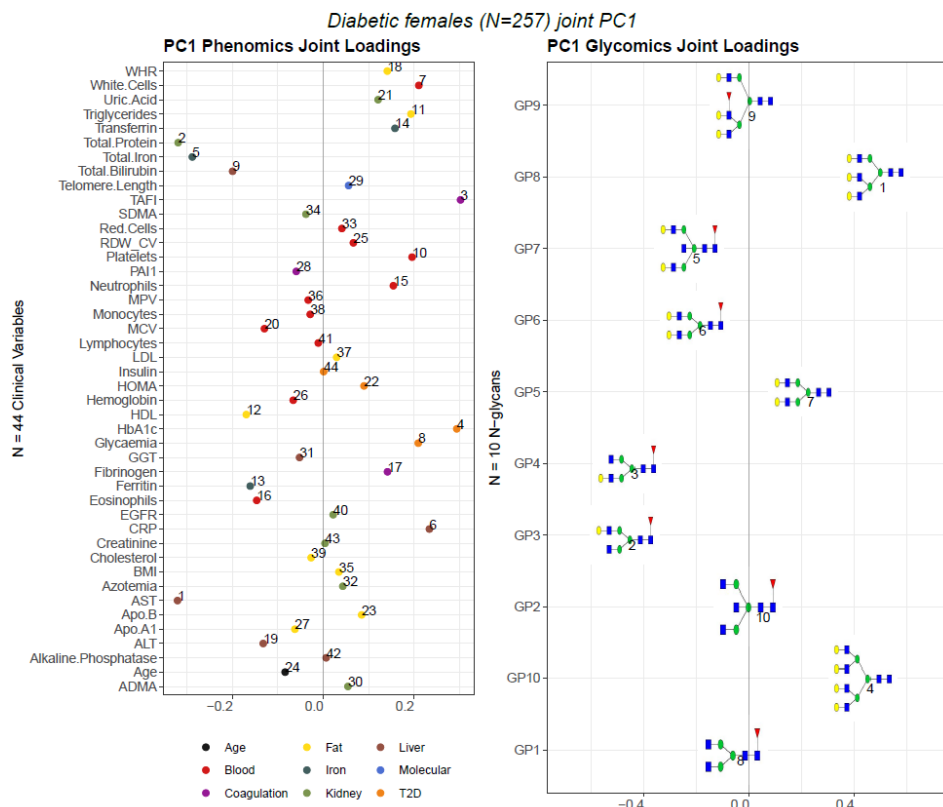


Figure 31. Joint PC1 loading values of diabetic females aged 20-85 y.

Table 16. Top 10 joint PC1 loading values of phenomics and glycomics data sets in Diab F

	Phe_load	Phenomics_var	Gly_load	Glycan_var
1	-0.3216	AST	0.479851	GP8
2	-0.32025	Total.Protein	-0.47455	GP3
3	0.303065	TAFI	-0.46463	GP4
4	0.294732	HbA1c	0.432795	GP10
5	-0.28893	Total.Iron	-0.22885	GP7
6	0.234337	CRP	-0.20625	GP6
7	0.211017	White.Cells	0.205671	GP5
8	0.209755	Glycaemia	-0.06177	GP1
9	-0.20004	Total.Bilirubin	-0.01761	GP9
10	0.196148	Platelets	-0.00388	GP2

Figures 26-31 and **Tables 11-16** show the loading variables in each joint set for each group. It can be noticed that three of the top 4 joint glycans in control males, control females and prediabetic females are GP6 (digalactosylated, fucosylated glycan, already described in the literature to be negatively associated with aging), GP3 and GP4 (monogalactosylated, fucosylated glycans), while GP2 (agalactosylated, bisecting GlcNAc, fucosylated, described too in the literature as typical of the aging phenotype) is also common in the top 4 in control females and prediabetic females (**Figure 27, Table 12**). GP6, GP3 and GP4 are negatively correlated mainly with coagulation and blood

parameters in control males and with age (**Figure 26, Table 11**), alkaline phosphatase and lipid parameters in control females (**Figure 27, Table 12**). Conversely, the top 4 glycans in diabetic males and females are GP8, GP10, GP3 and GP4 (**Figures 30-31, Tables 15-16**). Prediabetic males also show GP8 and GP10 as top 2, and they are mainly positively correlated with three coagulation parameters (**Figure 28, Table 13**). In diabetics, both sexes show that GP8 and GP10 are positively correlated with diverse parameters from the phenomics data set, including lipid, coagulation, iron, blood parameters and factors used for the T2D diagnosis, while GP3 and GP4 are negatively correlated with those (**Figures 30-31, Tables 15-16**).

When testing the correlation between phenomics and glycomics joint scores in each group (**Table 17**), controls display joint PC1 correlations of the order of .7, prediabetics, of the order of .6, and diabetics of the order of .5. This observation highlights that the inter-variation phenomics/glycomics decreases with disease (from controls to prediabetics and to diabetics), thus that the intra-variation of each data set increases with disease, which, in turn, underlines the heterogeneity and/or complexity of disease and age.

Table 17. Pearson intra-correlations between pairs of Phenomics/Glycomics joint scores

Group	Correlation between Phenomics/Glycomics joint scores
Control males	Joint PC1: $R^2=0.71$ Joint PC2: $R^2=0.61$
Control females	Joint PC1: $R^2=0.70$ Joint PC2: $R^2=0.53$
Prediabetic males	Joint PC1: $R^2=0.64$ Joint PC2: $R^2=0.59$
Prediabetic females	Joint PC1: $R^2=0.64$ Joint PC2: $R^2=0.55$
Diabetic males	Joint PC1: $R^2=0.53$ Joint PC2: $R^2=0.36$
Diabetic females	Joint PC1: $R^2=0.52$ Joint PC2: $R^2=0.48$

23. Visualization of mean values and integrated weights

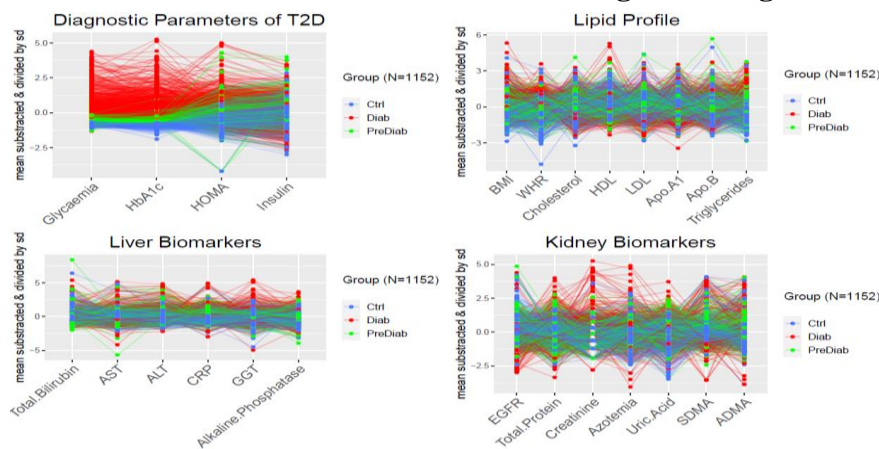


Figure 32. Parallel plots of groups of variables (I)

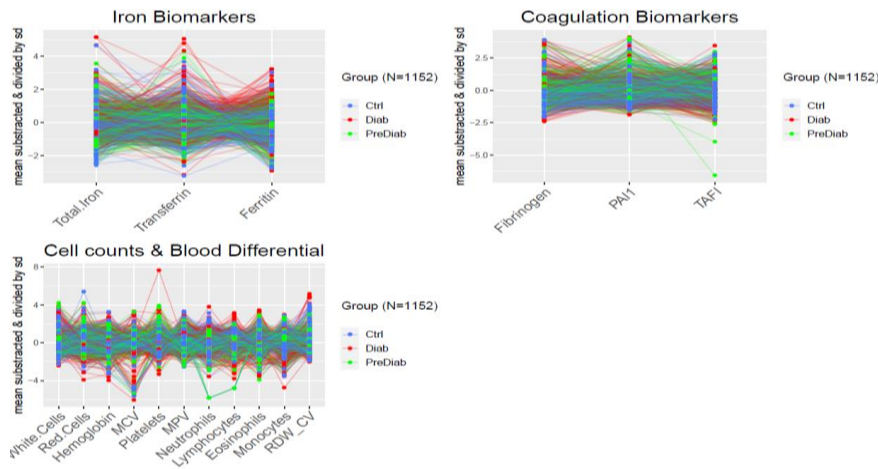


Figure 33. Parallel plots of groups of variables (II)

Figures 32 and 33 visually describes the differences among the groups of individuals in groups of variables – 42 phenotypical variables (excluding sex, age, telomere length) were grouped according to the tissue of production, their functionality or their diagnostic value: diagnostic parameters of T2D, lipid profile, liver, kidney, iron, coagulation, and cell counts & blood differential, and visualized with parallel plots for EDA – N=1152, because individuals >85 years old had also been included.

In T2D parameters, differences are clear, especially for glycaemia (fasting glucose) and HbA1c. In the remaining groups of variables, overall diabetics stand out over controls and prediabetics in WHR, LDL, Apo A1, triglycerides, GGT, ferritin and TAFI. There many variables that illustrate the heterogeneity of diabetics, by peaking up and falling down at the same time, especially for kidney, coagulation and blood parameters: EGFR, creatinine, azotemia, SDMA, ADMA, fibrinogen, TAFI, red cells, MCV, lymphocytes, eosinophils, monocytes.

The next subsection shows in more detail the mean values per variable (on the left) as well as the loading values obtained with O2PLS (on the right) grouped by the same groups of variables as the parallel plots; only the most relevant variables within each group are shown.

Diagnostic parameters of T2D

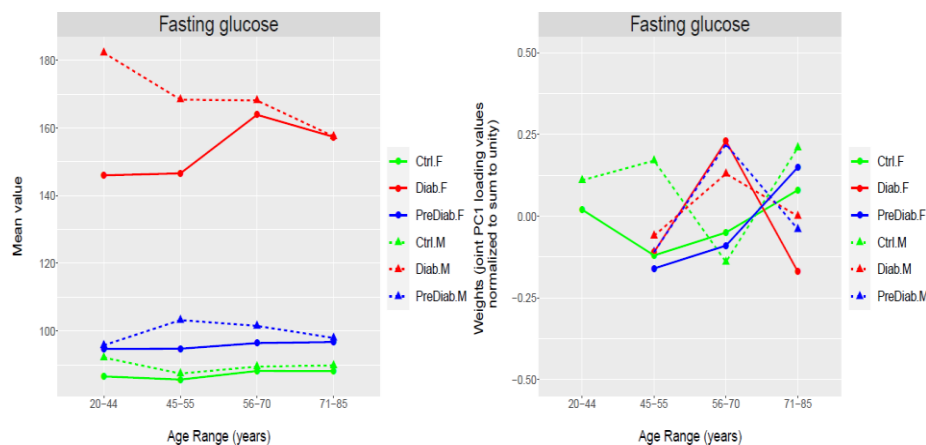


Figure 34. Means and weights of fasting glucose

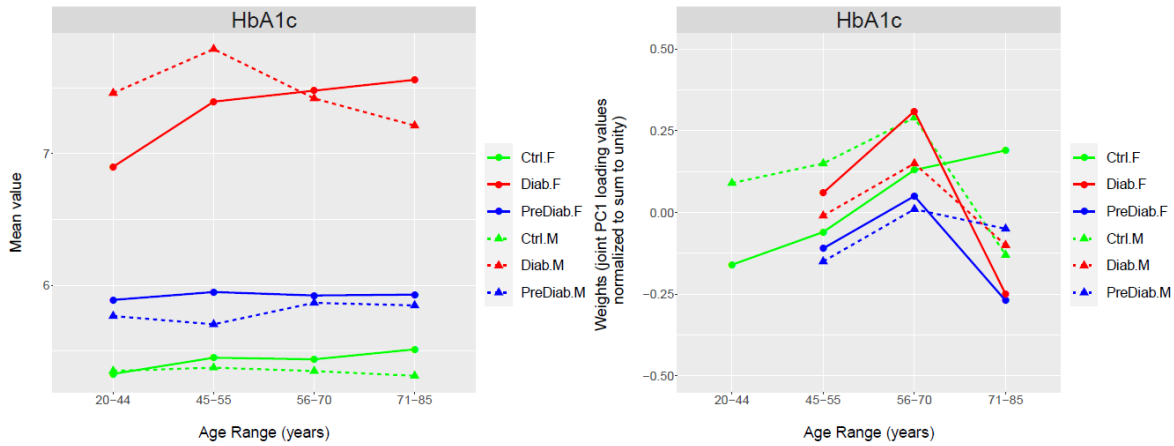


Figure 35. Means and weights of glycated haemoglobin

Fasting glucose and HbA1c. A clear segregation of groups is observed in the mean values plot. In addition, females present lower concentrations of fasting glucose across all age ranges than males, but higher concentrations of HbA1c (an indicator of long-term glucose), except at the age ranges 20-44 and 45-55. Weights reach a maximum at 56-70 years (positive for all of them) and diminish again at 71-85 years, forming a triangle-like shape across the age ranges 45-55, 56-70, and 71-85.

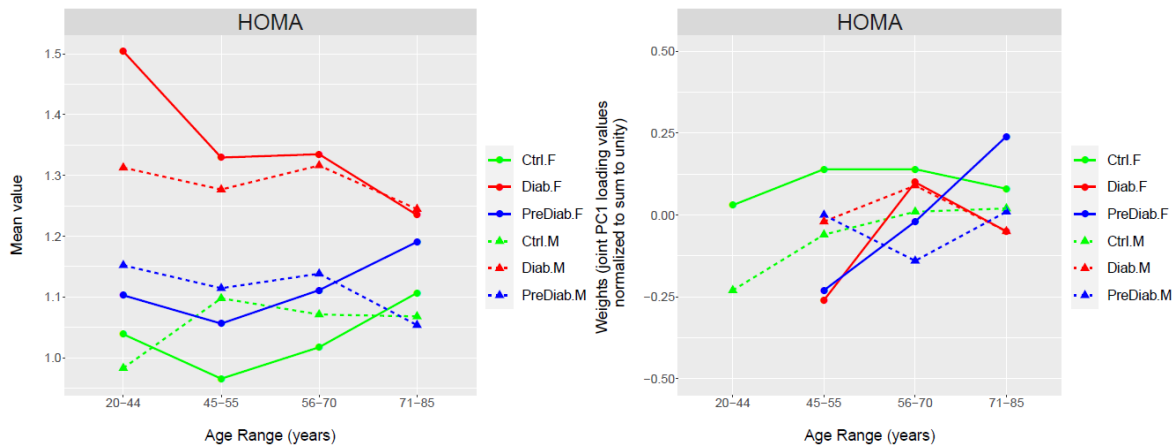


Figure 36. Means and weights of HOMA

HOMA and fasting insulin. The means show that in the case of HOMA, groups are segregated, though at the age ranges 45-55 and 71-85, controls and prediabetics mix up. In the case of insulin, control and prediabetic females follow a similar curve to that of HOMA, that is at 45-55 years of age, they shrink, and at the next age range they raise again. Diabetic females present the highest concentrations of insulin in all age ranges (specially at 20-44y), except at 71-85 years of age. In males, prediabetics have higher levels of insulin in the first three age ranges, while diabetic males present constant concentrations across ages, and control males aged 20-44 years present the lowest levels. Weights do not follow any particular pattern, though in fasting insulin, groups tend to aggregate at 56-70 and 71-85 years of age.

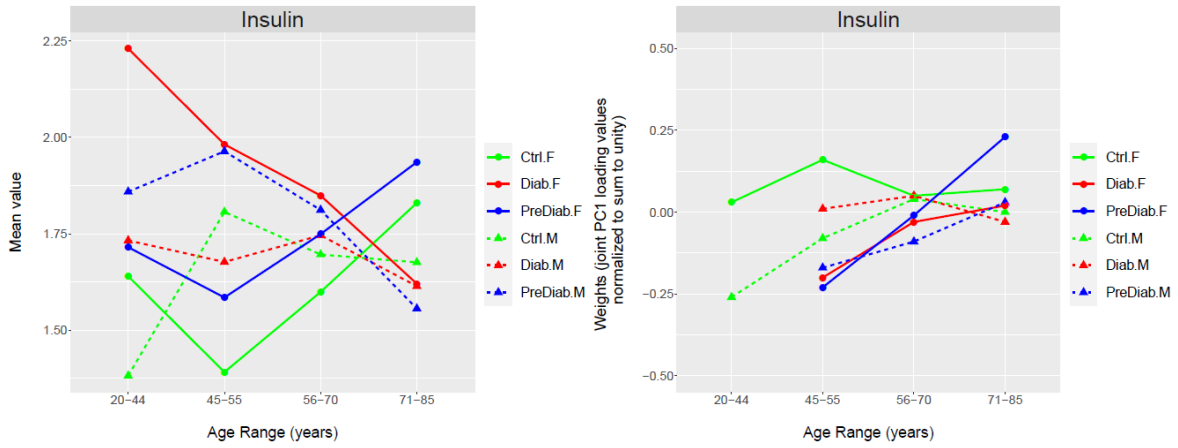


Figure 37. Means and weights of fasting insulin

Lipid profile

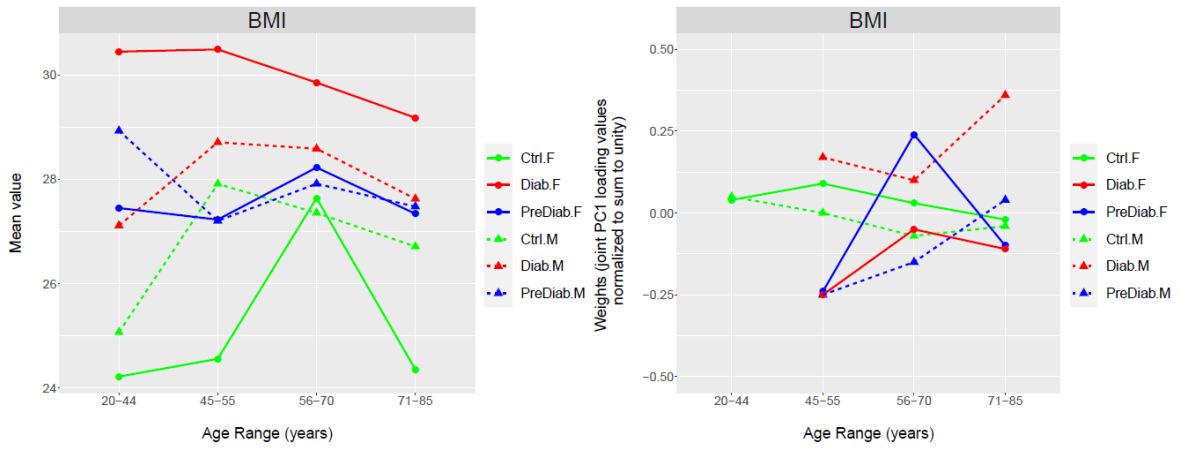


Figure 38. Means and weights of BMI

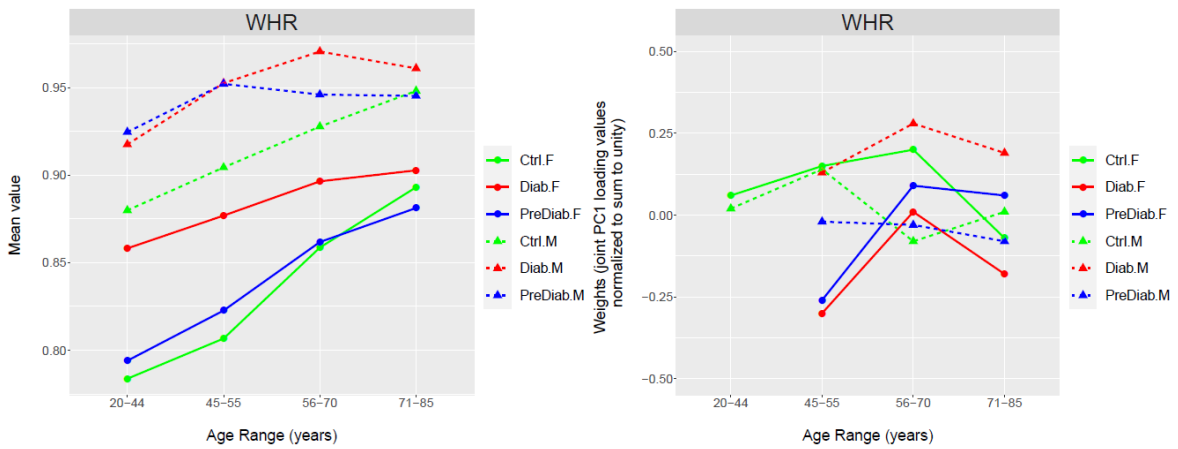


Figure 39. Means and weights of WHR

BMI and WHR. Mean values show that diabetic females have the highest BMIs (29 to 32, according to the age range), while control females, the lowest (24 to 25), though at the age 56-70, it raises above 27. Diabetic males and prediabetic females and males present a BMI of 27-29 across all age ranges. With respect to WHR, males display much higher WHR than females in the three groups across all age ranges. Loading values do not follow a specific pattern.

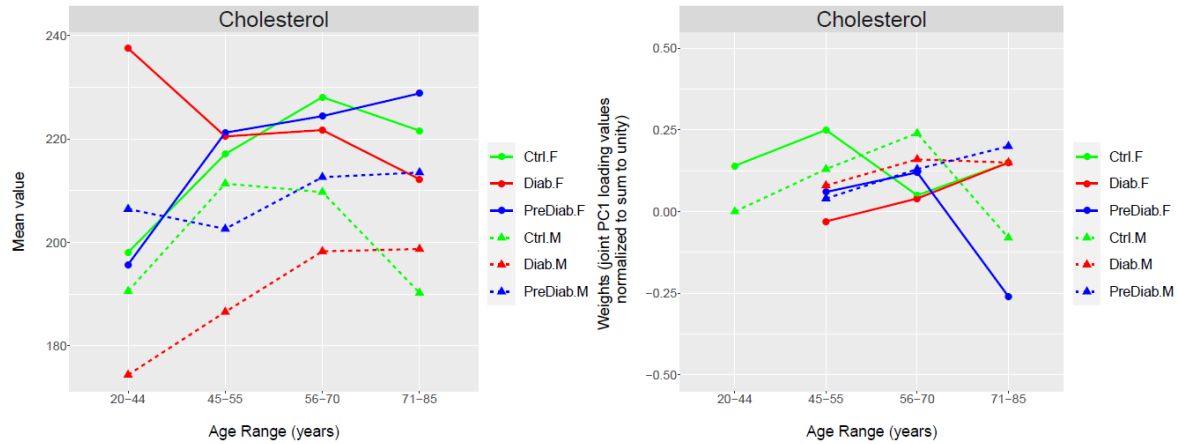


Figure 40. Means and weights of total cholesterol

Cholesterol and LDL. Mean values illustrate that, in general, females have more elevated cholesterol levels than males; in particular, diabetic males have the lowest concentrations of cholesterol and diabetic females, the highest in the age range 20-44 years old. In relation to LDL, diabetic males follow the same pattern as cholesterol levels across ages, and roughly the same applies in control and prediabetic females. Weights of the groups tend to converge at 56-70 years, while before and after they look more scattered.

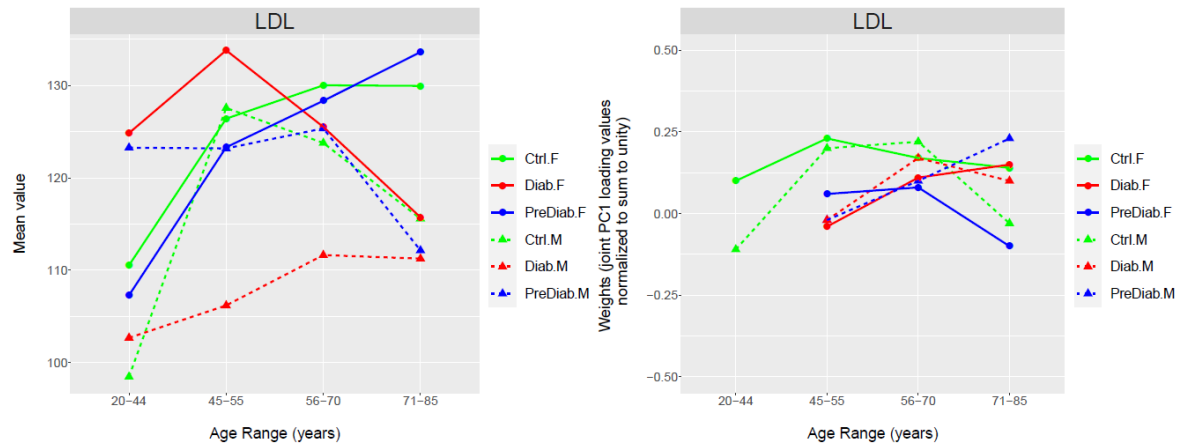


Figure 41. Means and weights of LDL

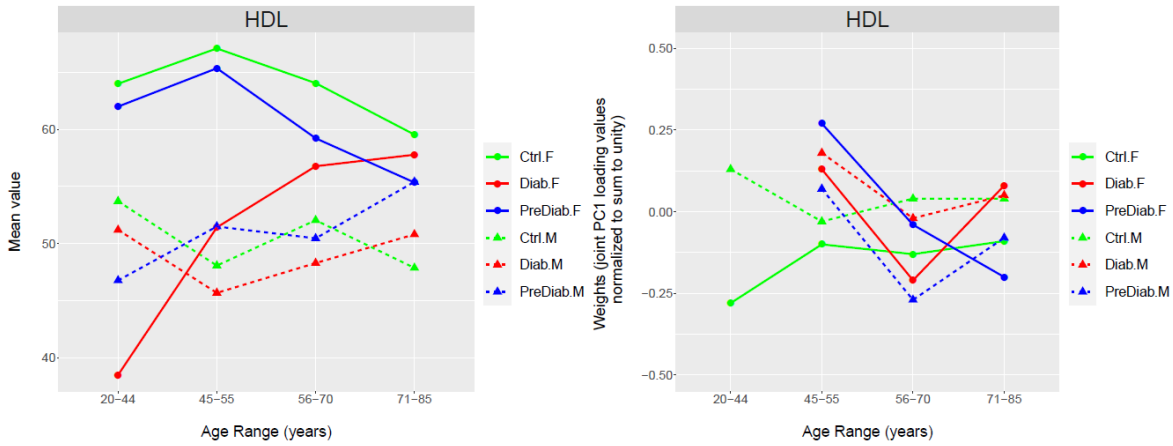


Figure 42. Means and weights of HDL

HDL and Triglycerides. Mean values confirm that diabetic females aged 20-44 years have very low levels of HDL and, at the same time, very high levels of triglycerides as compared with males, while prediabetic and control females have the highest levels of HDL and the lowest levels of triglycerides. Weights in HDL in prediabetics and diabetics shrink at 56-70 years, while in controls they remain constant from 45-55 to 71-85 years; in triglycerides, prediabetics and diabetics have loading values clustered between 0 and -0.25, and controls of +0.25, in 56-70 years, the six groups group together to dissipate at 71-85 years.

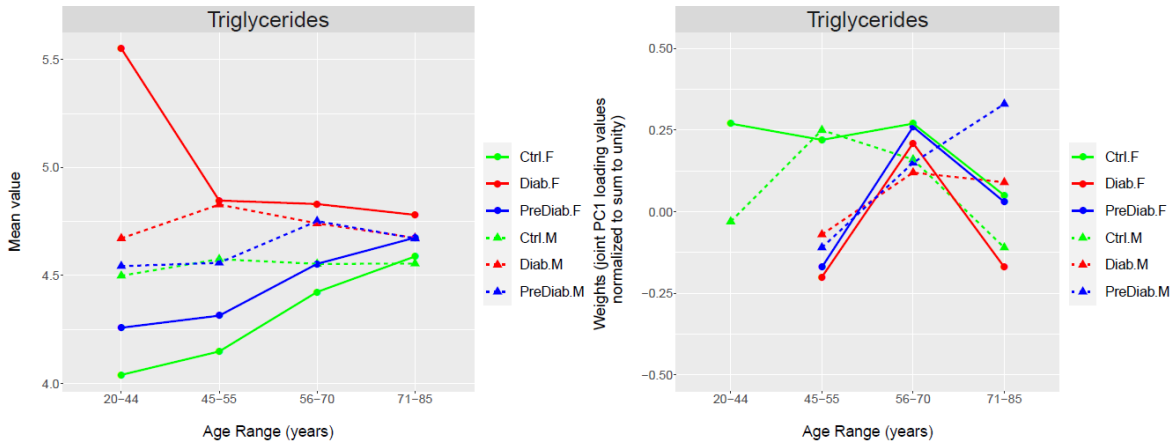


Figure 43. Means and weights of triglycerides

Liver biomarkers

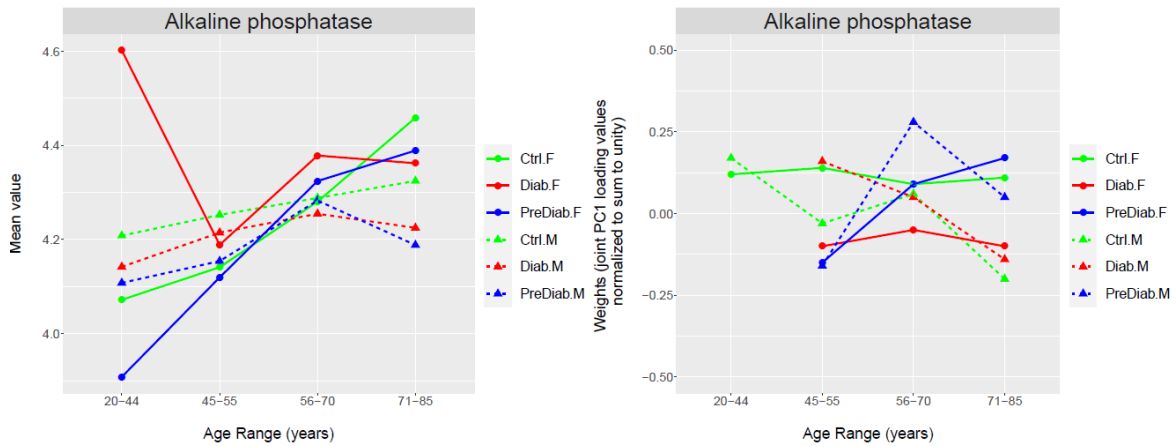


Figure 44. Means and weights of alkaline phosphatase

*Alkaline phosphatase**. Overall, mean values show an upward trend across the four groups of age, with the exception of diabetic females, in which at 20-44 y, the mean value is exceptionally high, and conversely, very low in prediabetic females in the same group of age (*enzyme that comes primarily from the liver, but also from the bones; higher levels may indicate a problem in the liver or a bone disorder). Weights do not follow a specific pattern.

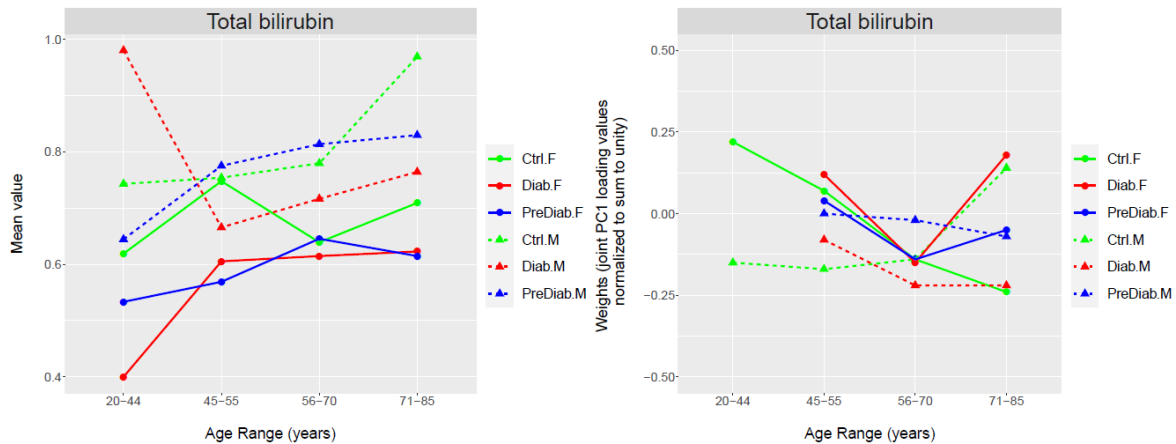


Figure 45. Means and weights of total bilirubin

*Total bilirubin**. In this parameter, there is a general separation between males and females in the means, except at 45-55, in which values for control females raise and for diabetic males fall (*higher bilirubin levels indicate that either red cells are breaking down at an unusual rate or that liver isn't breaking down waste properly and clearing it from blood). Loading values evolve at once for females with an inverted triangle-shaped pattern (fall at 45-55 y – raise at 56-70 y), while males do not follow a specific pattern.

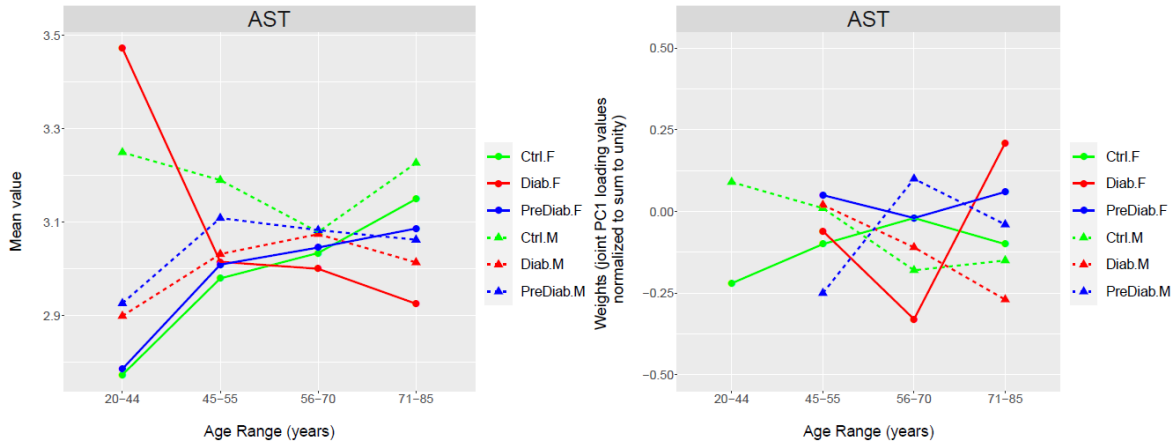


Figure 46. Means and weights of AST

*AST**. In mean values, except for control males and diabetic females, there is a rise in ALT levels from 20-44 to 45-55 years, followed by a general stabilization across ages (*high levels may indicate hepatitis, cirrhosis, mononucleosis or other liver diseases). Weights do not follow a specific pattern.

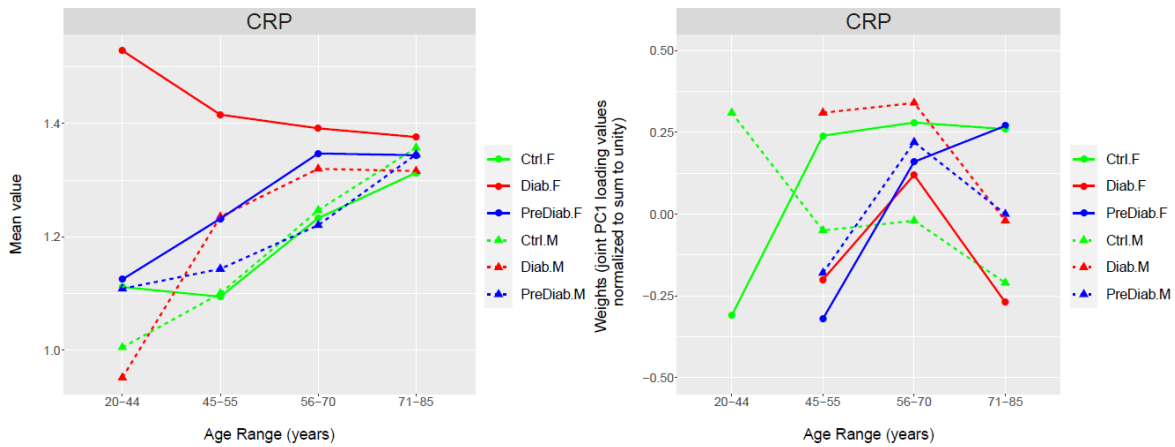


Figure 47. Means and weights of C reactive protein

*CRP**. Regarding mean values, except for diabetic females, who display very high levels of CRP at 20-44, which slightly fall over ages, there is an upward trend in all groups from 20-44 years up to 71-85 years (*it is a marker of inflammation, and high levels may indicate infection, cancer, cardiovascular disease, aging). Weights are similar for prediabetic males and females and diabetic females at 45-55 and 56-70 years, while the other groups do not follow a specific pattern.

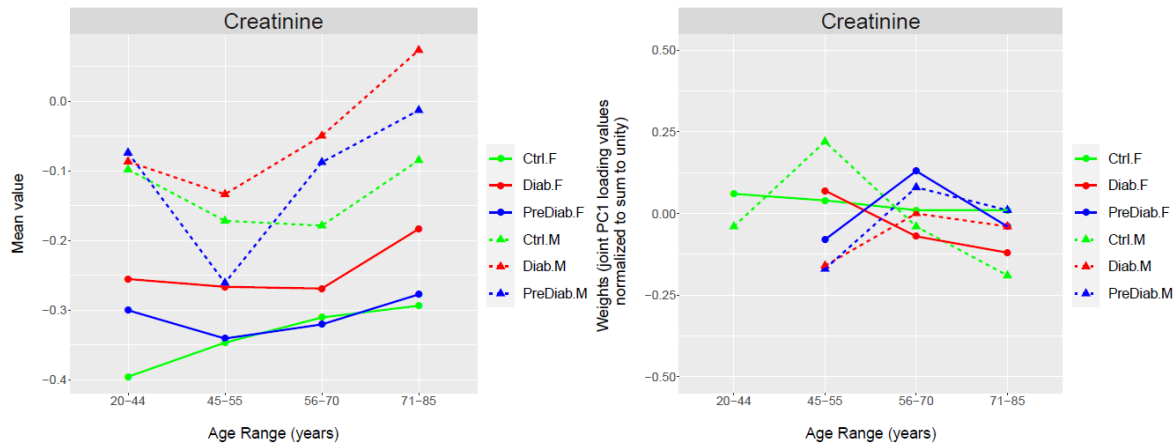


Figure 48. Means and weights of creatinine

*Creatinine**. In mean values, there is a clear separation between males and females, the formers presenting higher mean values, and except for control/prediabetic males at 45-55 years of age, controls, prediabetics and diabetics are escalated in each sex (*high creatinine levels indicate that kidneys are not working properly). Weights follow a triangle-shaped pattern, except in control males and diabetic females, in which from 45-55 to 71-85 years they fall down.

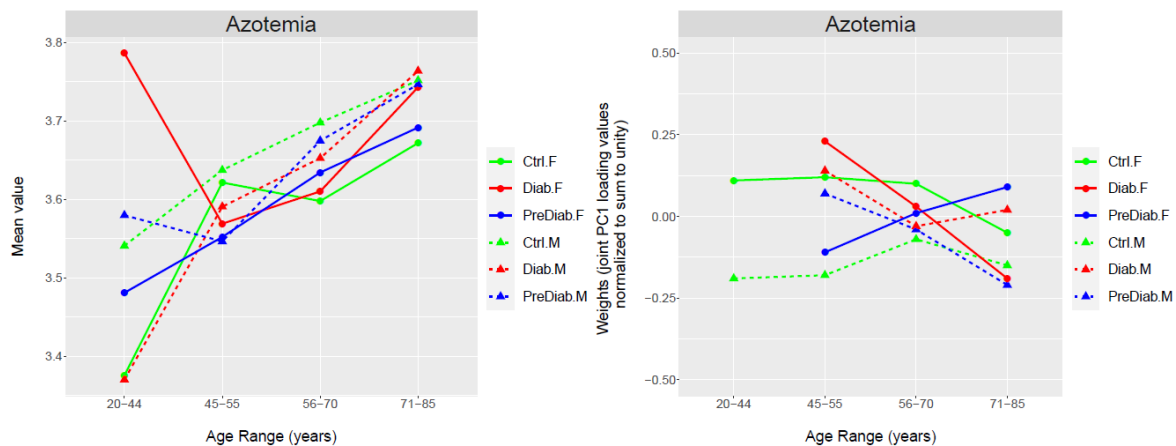


Figure 49. Means and weights of azotemia

*Azotemia**. Overall, mean values raise across age ranges, except in diabetic females (*high levels indicate abnormally high levels of nitrogen-containing compounds, such as urea, creatinine, waste compounds...). Weights tend to cluster at 56-70 years, though they do not follow a specific pattern.

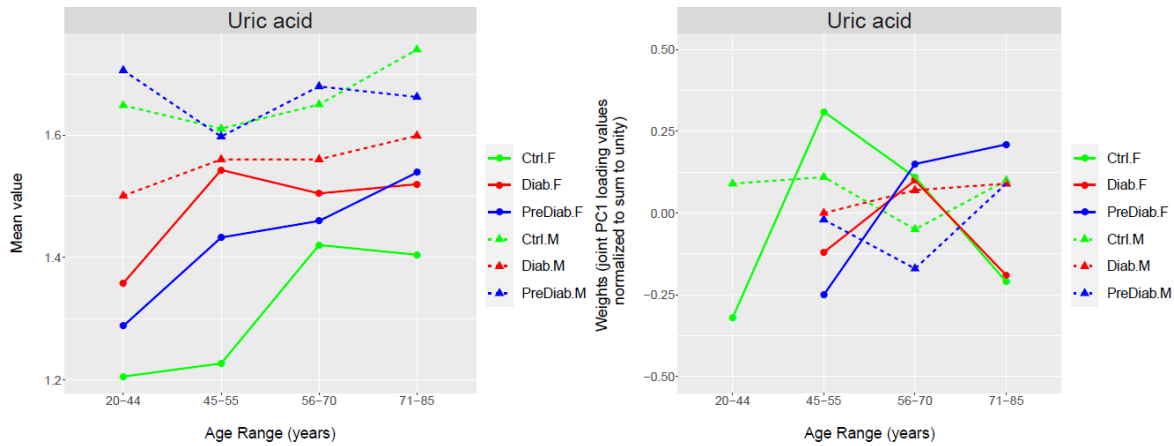


Figure 50. Means and weights of uric acid

Uric acid. Males means show higher levels of uric acid across all age ranges, and in the case of females, groups are escalated (diabetics presenting the highest and controls, the lowest). Loading values do not follow a specific pattern, though at 56-70 years, females aggregate. Weights do not follow a specific pattern.

Iron biomarkers

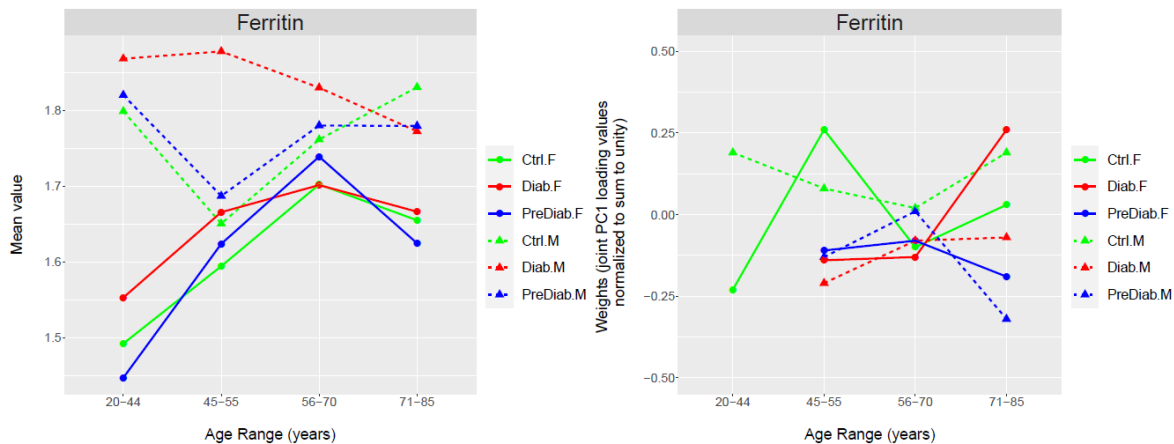


Figure 51. Means and weights of ferritin.

*Ferritin**. Males and females follow a diverse dynamics: diabetic males present higher mean values of ferritin that peak up at 45-55, followed by a reduction across ages; prediabetic and control males present their lowest levels at 45-55; females present the lowest levels at 20-44 years, which rise up until 56-70 years to diminish again at 71-85 (*it measures the amount of iron stored in the liver). Weights do not follow a particular pattern. Weights segregate negatively in prediabetics and diabetics and positively in controls at 45-55 years, at 56-70 years they cluster around 0 and minus 0, and at 71-85 years dissipate again.

Coagulation biomarkers

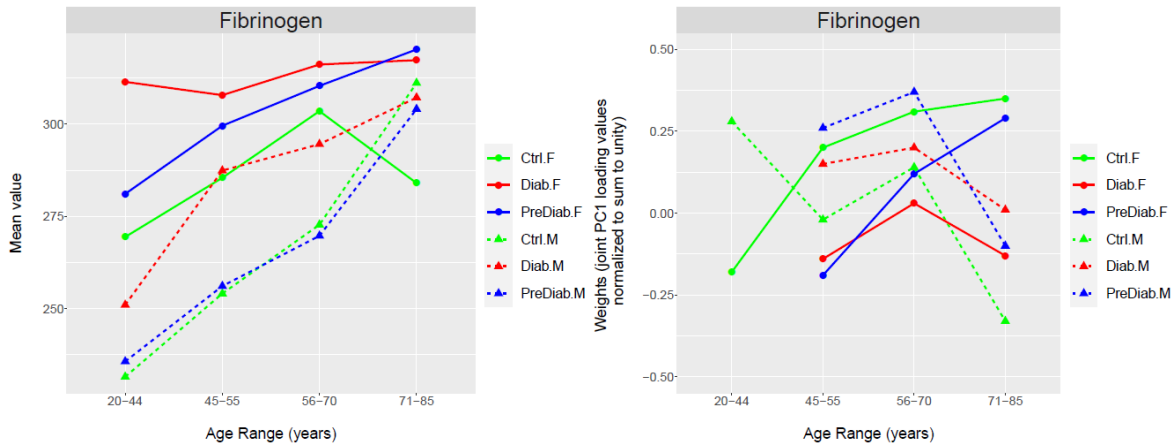
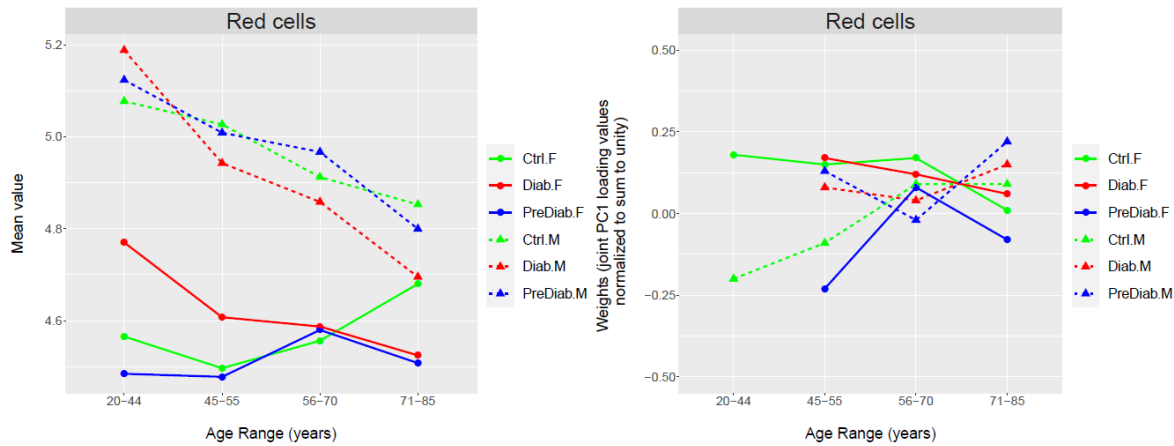


Figure 52. Means and weights of fibrinogen

Fibrinogen. It clearly separates both groups and sexes, being males and controls the ones with lowest mean values. At 20-44 years, they are at their lowest levels and increase steadily until 71-85 years, except for control females, who diminish fibrinogen concentrations at 71-85 years. Weights follow a triangle-shaped pattern (raise at 45-55 y – fall at 56-70 y), except in control and prediabetic females, that augment progressively from 20-44 y and 45-55, respectively. Weights follow roughly the same pattern than in CRP.

Cell counts and blood differential



Red cells. Mean values separate males and females, and the dynamics between sexes follows a different pattern: it diminishes across ages in males, while in females, red cells concentration starts raising at 45-55 years (except diabetics females at 20-44 years, who have high levels), and in 71-85 years, that diminish in prediabetics and diabetics, while in female controls still increases. Weights follow roughly the same pattern than creatinine and azotemia.

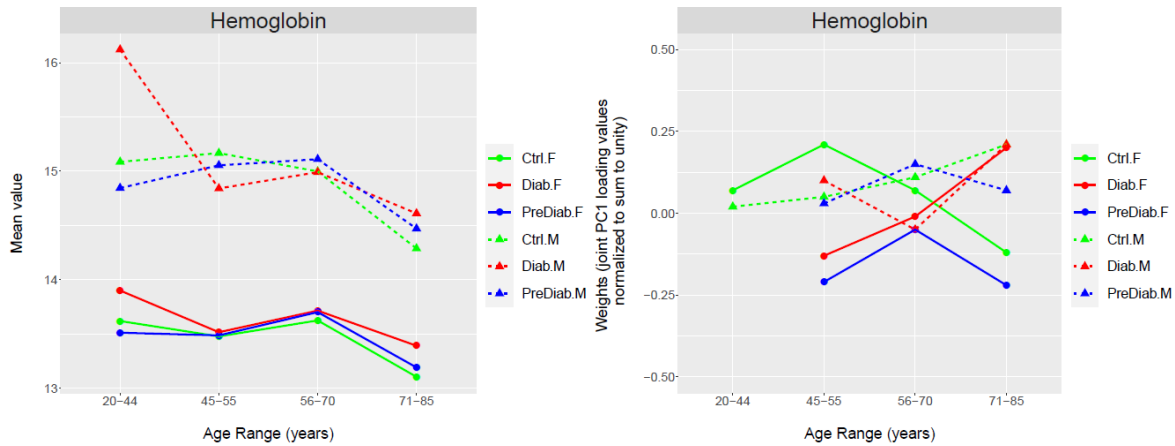


Figure 53. Means and weights of haemoglobin

Haemoglobin. Mean values clearly show much higher concentration of hemoglobin in males than females, and diabetic males show much higher levels at 20-44 years. Weights do not follow a specific pattern.

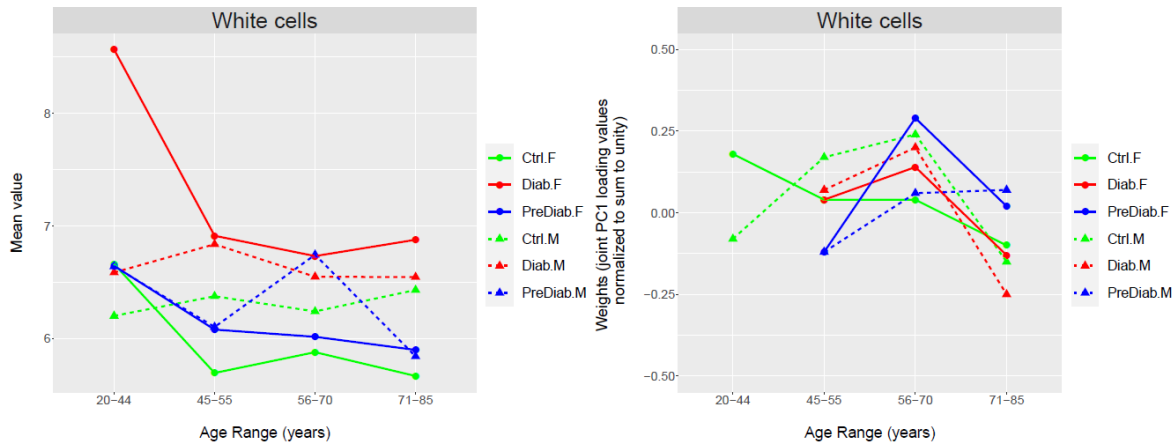


Figure 54. Means and weights of white cells

White cells. Diabetic males and females means are the highest, the latter group displaying a very high value at 20-44 years; prediabetic males and females have the same mean values, except a raise in males at 56-70, which are lower than diabetics, and control females present the lowest values. Weights follow a triangle-shaped pattern (raise at 45-55 y – fall at 56-70 y), except in control and prediabetic males.

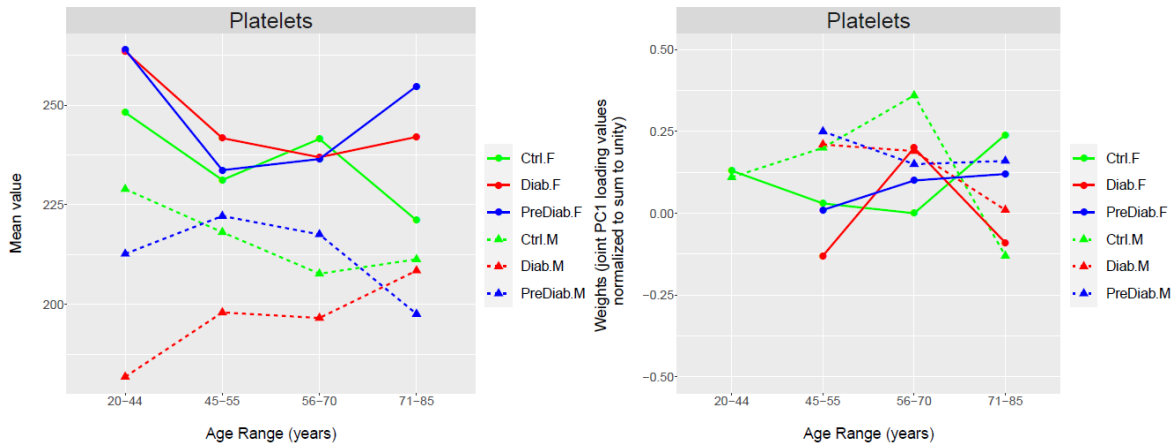


Figure 55. Means and weights of platelets

Platelets. In this case, the sexes cluster, being females the group with higher mean values, which are at its highest at 20-44 years of age. Weights do not follow a specific pattern.

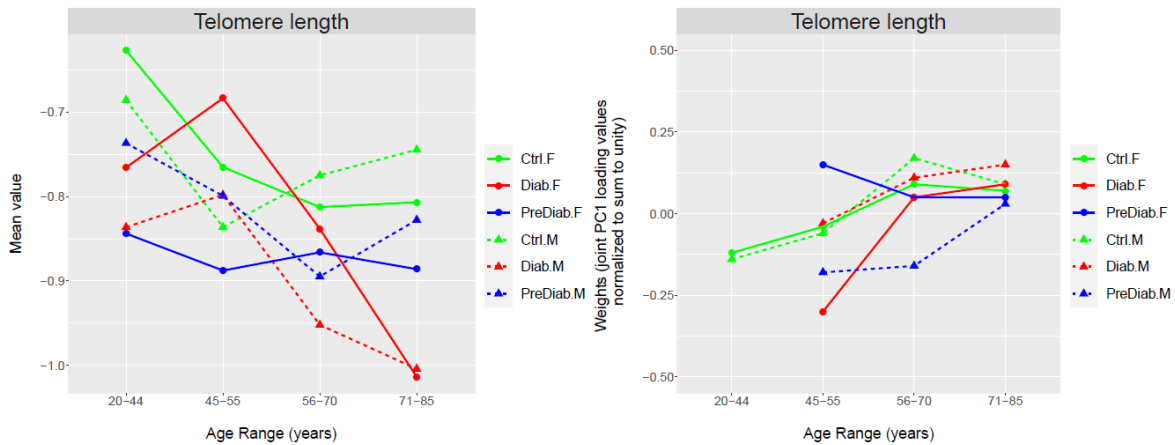


Figure 56. Means and weights of telomere length

Telomere length. Mean values vary according to the cohort (defined by age range, sex, and group), but it becomes evident that at 71-85 years diabetics present the lowest levels, followed by prediabetics and diabetics. Weights group at 56-70 years, except in prediabetic males, and at 71-85 years.

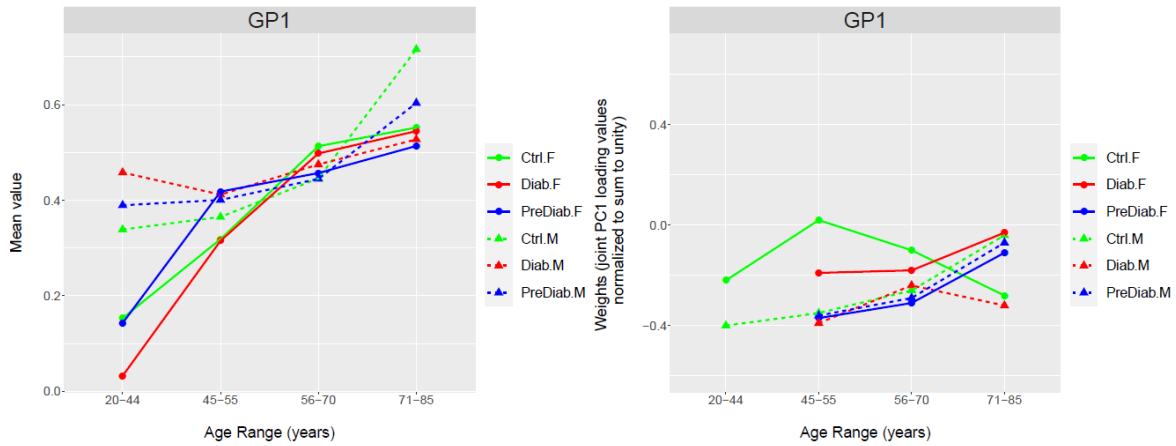


Figure 57. Means and weights of GP1

GP1. Mean values increase over age and reach the maximum values at 71-85 years age; there are differences between sexes at 20-44 years, by which females show much lower values, though at the same time, they display a steeper raise until 45-55 years – for males, from the youngest age range until the 2nd age range, mean values remain roughly constant. Weights do not follow a specific pattern.

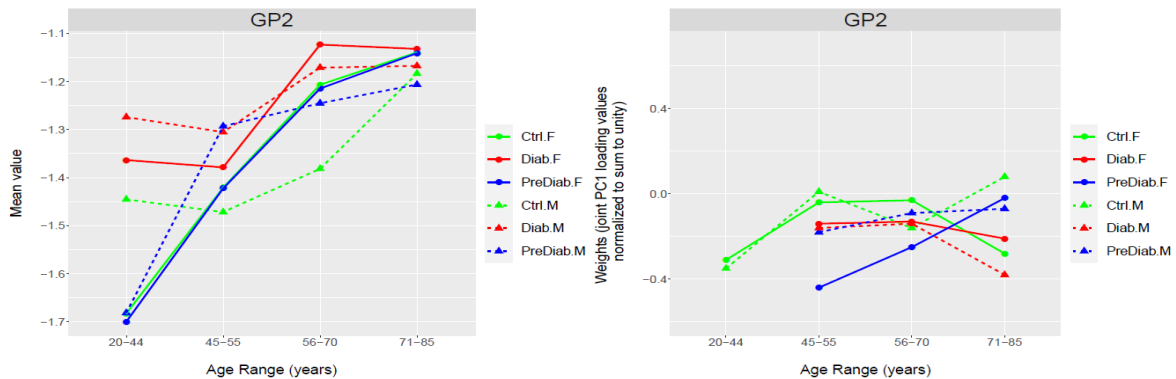


Figure 58. Means and weights of GP2

GP2. Mean values present a pronounced increase from younger ages until 56-70 or 71-85 years; in the case of diabetic males, diabetic females and control males (with this order) they present much higher mean values already at 20-44 years, in contrast to the other three groups, and start raising their values at 45-55 years. Weights do not follow a general pattern – they remain roughly constant between around -0.4 and 0.

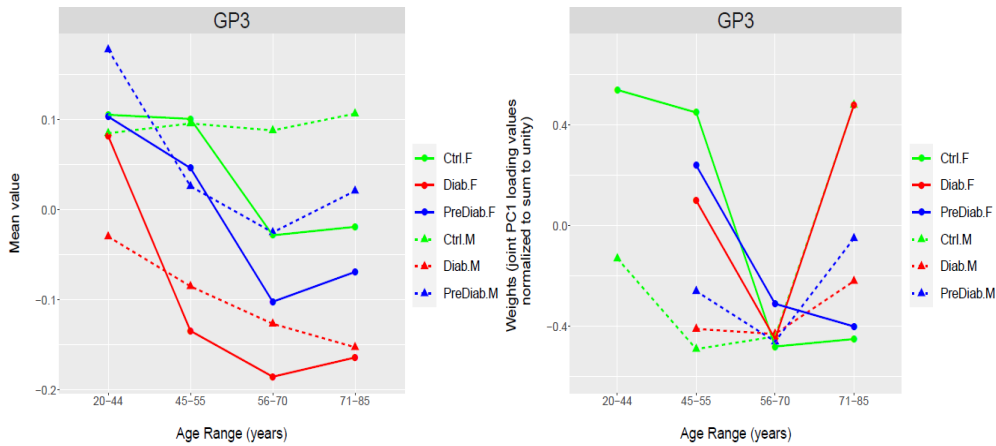


Figure 59. Means and weights of GP3

GP3. Means of *GP3* depart from high positive values and fall down up until 56-70 years, except in diabetic males, and slightly raise again at 71-85 years. It is surprising that controls, prediabetics and diabetics are quite escalated, being controls generally the ones to have higher mean values. Weights separate males and females at 45-55 years (negative and positive loading values, respectively), converge at 56-70 years (female loadings become negative) to scatter again at 71-85 years.

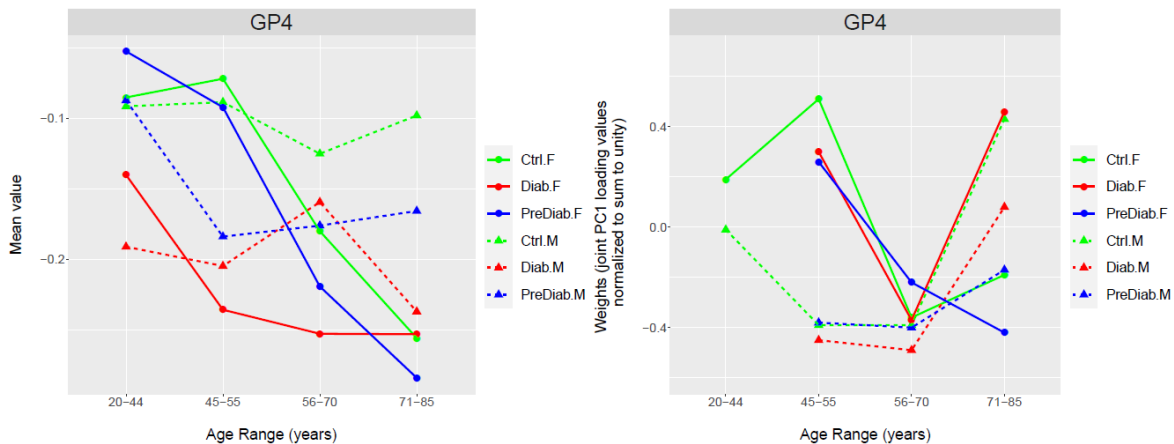


Figure 60. Means and weights of GP4

GP4. Mean values follow a step decrease from 20-44 years until 71-85 years for females, being controls the ones with higher values, followed by prediabetics, while in males do not decrease or decrease less from 20-44 years until 45-55 years, and increase/decrease or remain constant until 71-85 years. Weights follow roughly the same pattern as in *GP3*.

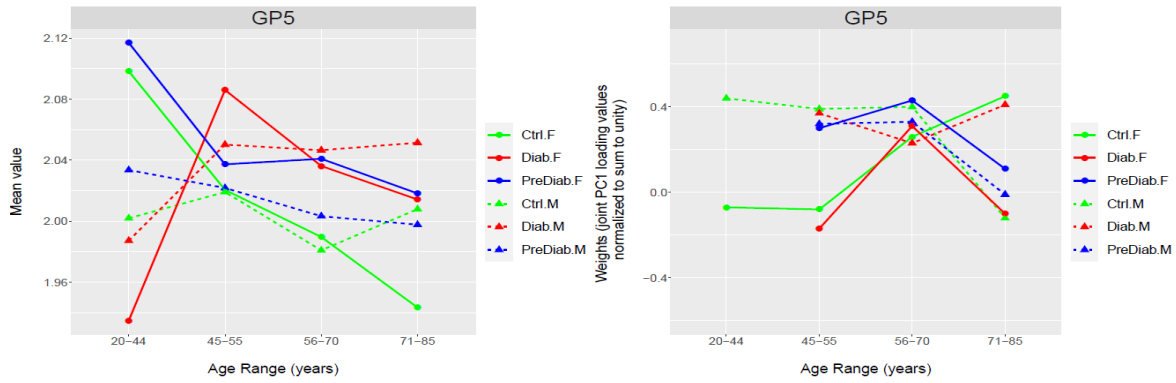


Figure 61. Means and weights of GP5

GP5. Mean values in control and prediabetic females are positive high at 20-44 years and sharply decrease until 45-55 years, while in diabetic females the contrary applies; from 45-55 to 71-85 years mean values slightly decrease, though the fall is more important in controls. Instead, males follow an approximate constant pattern from 20-44 until 71-85 years. Weights, except in control and diabetic males, present a triangle-shaped pattern (raise at 45-55 years – fall at 56-70 years), which is sharper in females.

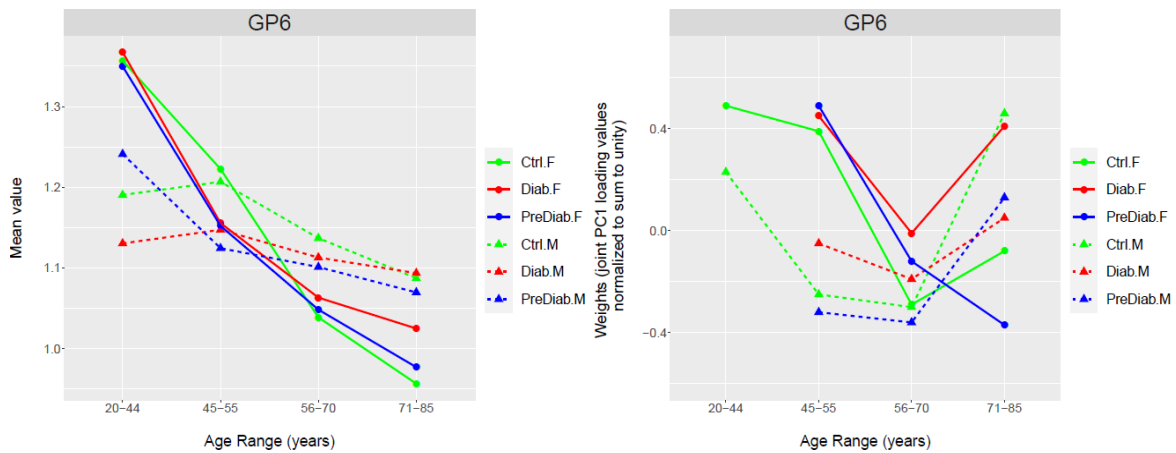


Figure 62. Means and weights of GP6

GP6. Mean values are very high at 20-44 years for females, and they markedly decline, almost lineally, until 71-85 years old; in males, from 20-44 to 45-55 years remain constant for controls and diabetics while diminishing in prediabetics, and from 45-55 until 71-85 years they steadily lessen. Weights, except in prediabetic females, follow an inverted triangle-shaped pattern (fall at 45-55 years – raise at 56-70 years), more closed ($<90^\circ$) in females and more opened ($>90^\circ$) in males.

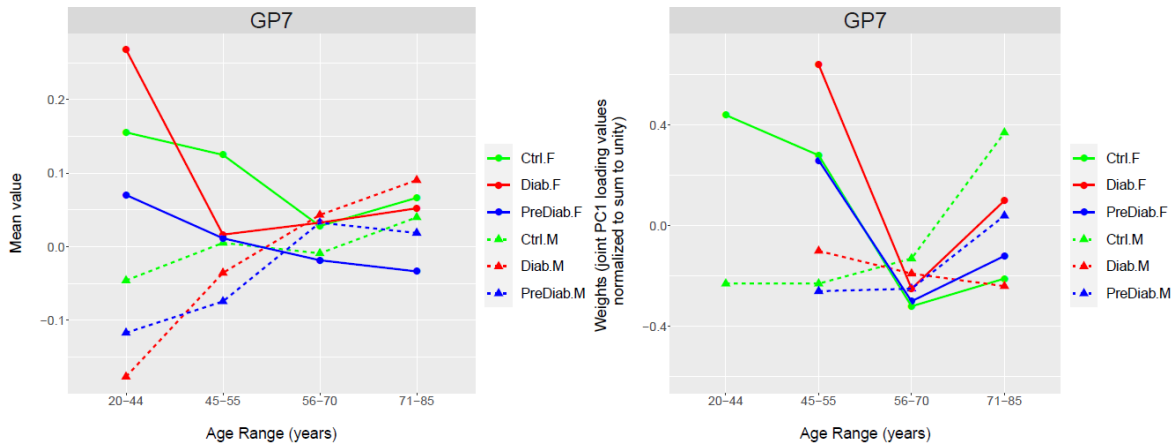


Figure 63. Means and weights of GP7

GP7. Mean values are quite low for males at 20-44 years and steadily increase until 71-85 years; for females is quite the opposite, mean values are high at 20-44 years and diminish until 71-85 years, so that sexes mix. Weights follow a similar pattern to that of GP3.

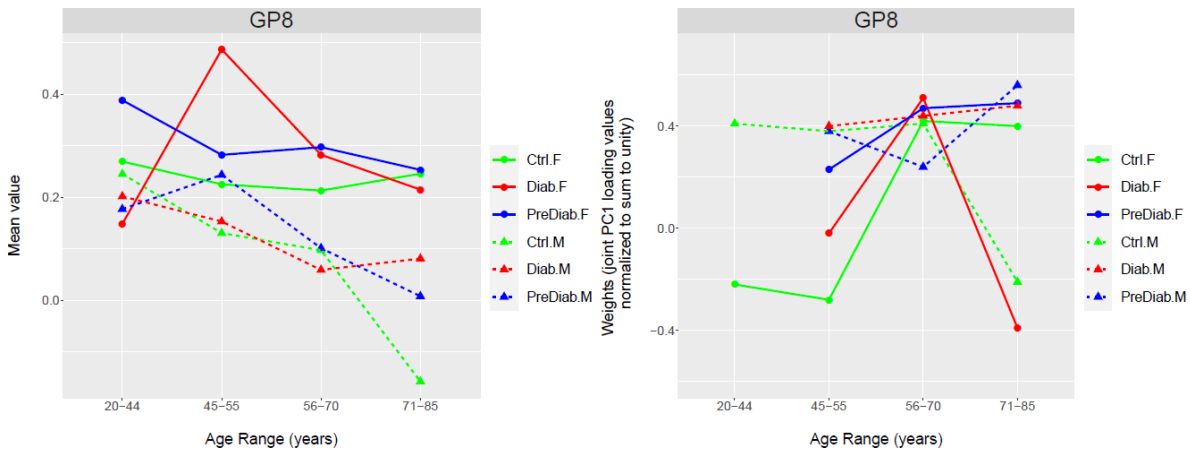


Figure 64. Means and weights of GP8

GP8. Mean values show that groups are differentiated, that is females have higher values from 20-45 years and keep roughly constant until 71-85 years, except for diabetic females at 20-44 years, which sharply rise until 45-55 years to diminish again. Weights follow a pattern from 45-55 to 56-70 years, in which males have high positive values from 20-44 years until 45-55 years, while females exhibit less positive, neutral or negative loadings at 45-55 years, and increase until 45-55 years, where all the groups converge.

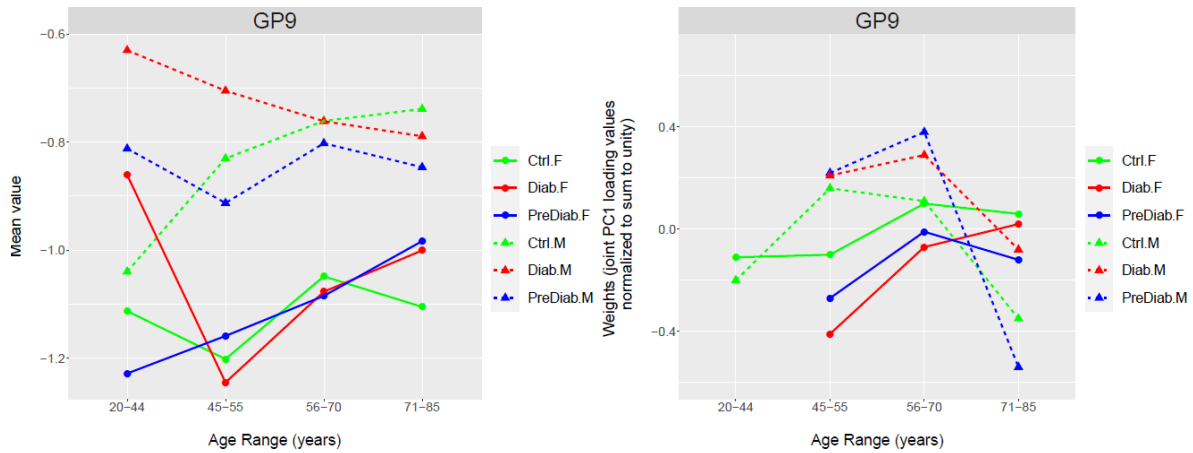


Figure 65. Means and weights of GP9

GP9. Mean values are lower for females than males across ages, except control males, which present a lower loading value at 20-44 years than the rest of the groups, and diabetic females have a much higher value than the other groups; from 45-55 years, means increase until 56-70 or 71-85 years. Weights are positive from 45-55 years in males and negative in females, though in males they steeply decrease from 56-70 years, while the decrease in females is less important.

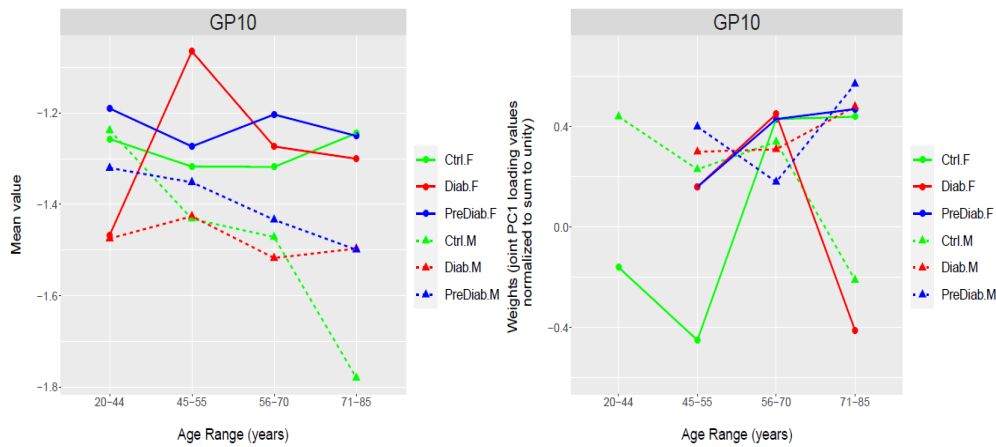


Figure 66. Means and weights of GP10

GP10. Mean values and weights follow the same pattern as in *GP8*.

PART II

8. AIM

The general aim of this study was:

- To identify the major sources of Phenomics variation (when integrated with N-glycans) by integrating endophenotypes/groups of biological variables, composed by biologically similar variables, that is variables diagnostic of T2D, lipid profile, liver biomarkers, kidney biomarkers, iron and coagulation parameters and blood factors, again in the six groups of individuals in four age ranges and the full age range, and to identify N-glycans changes according to these endophenotypes and to the group.

The specific aim of this study was:

- To follow the procedure described in the 2n objective of the aims in Part I, but in addition to integrate phenomics with glycomics, to apply O2PLS in seven additional sets of endophenotypes with N-glycans. Unlike in Part I, for the integration Phenomics/Glycomics, the variable age was not included.

9. RESULTS

24. Sources of variation in Phenomics, T2D, Lipids, Liver, Kidney, Iron, Coagulation, Blood & Glycomics data sets

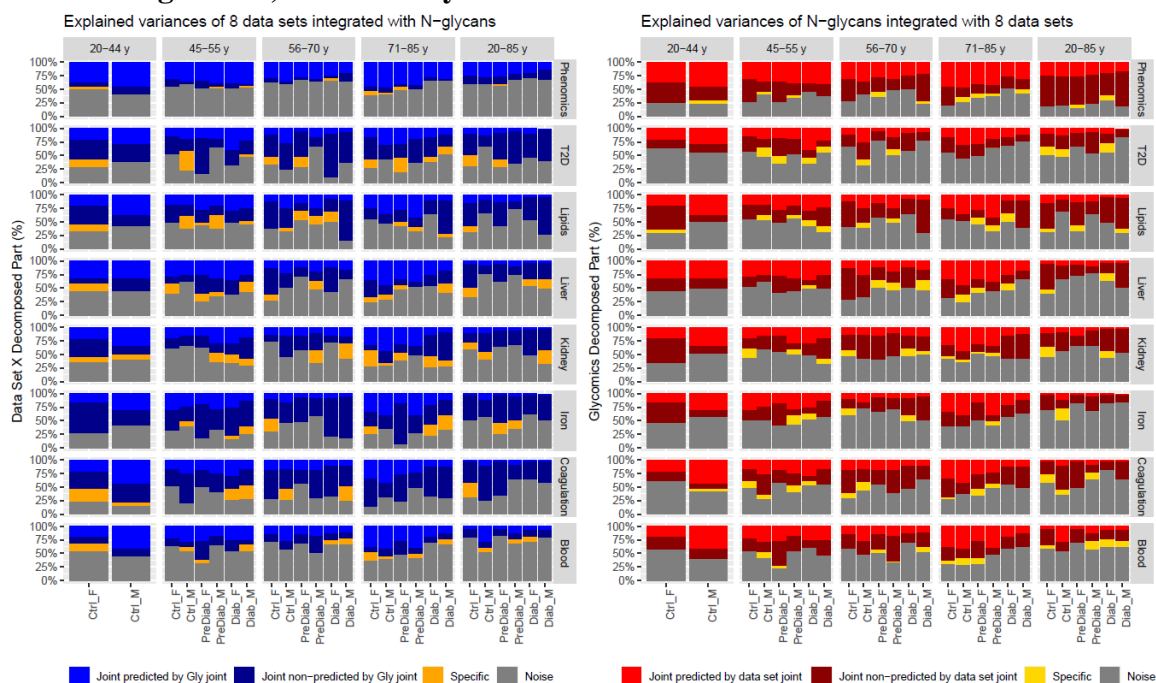


Figure 67. Decomposed parts of the O2PLS integrations between 8 different data sets and 10 N-glycans in four different age ranges and the total age range by groups and sexes

Figure 67 illustrates the decomposed parts of the eight integrated data sets with N-glycans in all groups by age ranges (“T2D” refers to the four variables used for diagnosing or quantifying T2D, that is fasting glucose, fasting insulin, HbA1c and HOMA). In general, the joint parts of the endophenotypes sets have higher signal/noise ratio, as compared with the joint phenomics data set. Conversely, in the glycomics joint part, while the covariating Phenomics exhibits a high signal/noise ratio, the covariating endophenotypes show a much lower signal/noise ratio. The same applies to the non-predictive parts, which are in higher percentage in joint endophenotypes parts but in lower percentage in the joint phenomics part, when compared with the non-predicted by glycomics in the phenomics data set. Instead, in the glycomics joint part, the covariating phenomics set presents a high percentage of non-predicted by phenomics part, while this percentage falls down in the glycomics joint of the covariating endophenotypes.

In order to make results of joint loadings more comprehensive, the description that follows comprises the six groups (Ctrl F, Ctrl M, PreDiab F, PreDiab M, Diab F, Diab M) considering the full age range (20-85 years) in each of the 8 integrated data sets, instead of examining the groups in the four age ranges previously shown.

The intra-correlations between pairs of joint PC scores show that in prediabetics, in both sexes, N-glycans scores correlate less with T2D diagnostic parameters and lipid profile scores, while they display a high correlation with liver and coagulation scores (**Table 18**). Conversely, in diabetics, in both sexes, the correlation of joint scores between N-glycans and T2D diagnostic parameters and

lipid profile is surprisingly low. In turn, the correlation with liver and coagulation is higher, yet much lower than the same correlations observed in prediabetics, which may be explained by medication (**Table 7**). By contrast, in comparison to controls and prediabetics, the correlation of the glycomics scores with kidney scores increases (**Table 18**). Yet, the expected tri-/tetragalactosylated glycans GP8 and GP10 are not in the top 4 of glycomics loadings integrated with liver, except for diabetic females, though they are in the coagulation data set in prediabetic males, diabetic females, and diabetic males. Instead, the top 3 N-glycans common in both prediabetic males and females are GP6 (digalactosylated, fucosylated), GP3 (monogalactosylated, fucosylated) and GP4 (monogalactosylated, fucosylated) (**Table 19**). Regarding controls, the highest intra-correlations between joint PCs are found in T2D, lipids, kidney and blood data sets pairing with N-glycans, and iron and coagulation for males too (**Table 18**).

Table 18. Pearson intra-correlations of the joint PC scores in each pair of integrated data sets

Group	Phenomics/Gly	T2D/Gly	Lipids/Gly	Liver/Gly	Kidney/Gly	Iron/Gly	Coagulation/Gly	Blood/Gly
Ctrl F	Joint PC1: R ² =0.63 Joint PC2: R ² =0.53 Joint PC3: R ² =0.44	Joint PC1: R ² =0.35	Joint PC1: R ² =0.29 Joint PC2: R ² =0.24	Joint PC1: R ² =0.21 Joint PC2: R ² =0.22	Joint PC1: R ² =0.29	Joint PC1: R ² =0.22	Joint PC1: R ² =0.25	Joint PC1: R ² =0.29
Ctrl M	Joint PC1: R ² =0.69 Joint PC2: R ² =0.55 Joint PC3: R ² =0.62	Joint PC1: R ² =0.39	Joint PC1: R ² =0.30	Joint PC1: R ² =0.27	Joint PC1: R ² =0.34 Joint PC2: R ² =0.30	Joint PC1: R ² =0.33	Joint PC1: R ² =0.29 Joint PC2: R ² =0.20	Joint PC1: R ² =0.30
PreDiab F	Joint PC1: R ² =0.69 Joint PC2: R ² =0.55 Joint PC3: R ² =0.51	Joint PC1: R ² =0.32	Joint PC1: R ² =0.46 Joint PC2: R ² =0.32	Joint PC1: R ² =0.52	Joint PC1: R ² =0.41	Joint PC1: R ² =0.33	Joint PC1: R ² =0.44 Joint PC2: R ² =0.23	Joint PC1: R ² =0.39
PreDiab M	Joint PC1: R ² =0.60 Joint PC2: R ² =0.54 Joint PC3: R ² =0.53	Joint PC1: R ² =0.32 Joint PC2: R ² =0.18	Joint PC1: R ² =0.39	Joint PC1: R ² =0.45	Joint PC1: R ² =0.25	Joint PC1: R ² =0.35	Joint PC1: R ² =0.48	Joint PC1: R ² =0.38
Diab F	Joint PC1: R ² =0.52 Joint PC2: R ² =0.47	Joint PC1: R ² =0.35	Joint PC1: R ² =0.21 Joint PC2: R ² =0.23	Joint PC1: R ² =0.41	Joint PC1: R ² =0.37 Joint PC2: R ² =0.20	Joint PC1: R ² =0.30	Joint PC1: R ² =0.33	Joint PC1: R ² =0.35
Diab M	Joint PC1: R ² =0.51 Joint PC2: R ² =0.33 Joint PC3: R ² =0.42 Joint PC4: R ² =0.29	Joint PC1: R ² =0.18	Joint PC1: R ² =0.26 Joint PC2: R ² =0.23 Joint PC3: R ² =0.22	Joint PC1: R ² =0.34 Joint PC2: R ² =0.32	Joint PC1: R ² =0.40 Joint PC2: R ² =0.21	Joint PC1: R ² =0.23	Joint PC1: R ² =0.35	Joint PC1: R ² =0.25

In this line, the inter-correlations among joint PC scores, for control males and females are mid-low: in males, the highest correlations after the phenomics/glycomics joint parts are 1) between the glycomics joint PC1 integrated with T2D and the coagulation joint PC1 ($R^2=-0.46$), and 2) between the blood joint PC1 with its glycomics joint PC1 counterpart ($R^2=0.41$) (**Figure 68**); while in females, the highest correlations after the phenomics/glycomics joint parts are 1) the glycomics joint PC1 [integrated with T2D] and lipids joint PC1 ($R^2=0.5$) and 2) T2D joint PC1 with lipids joint PC1 ($R^2=0.45$) and the glycomics joint PC1 [integrated with liver] with the glycomics joint PC2 [integrated with liver] ($R^2=0.45$) (**Figure 69**). Interestingly, in prediabetics and diabetics (**Figure 70-73**), there can be observed very high inter-correlations: the glycomics joint PC1 scores [integrated with phenomics] are strongly associated with glycomics joint PC1 [integrated with T2D] ($R^2=0.96$ PreDiab M, $R^2=0.99$ PreDiab F, $R^2=0.7$ Diab M, $R^2=-0.82$ Diab F), glycomics joint PC1 [integrated with lipids] ($R^2=0.83$ PreDiab M, $R^2=0.99$ PreDiab F, $R^2=0.92$ Diab M, $R^2=0.9$ Diab F), glycomics joint PC1 [integrated with liver] ($R^2=-0.14$ PreDiab M, $R^2=-0.98$ PreDiab F, $R^2=0.98$ Diab M, $R^2=0.92$ Diab F), glycomics joint PC1 [integrated with kidney] ($R^2=-0.89$ PreDiab M, $R^2=0.95$ PreDiab F, $R^2=0.95$ Diab M, $R^2=0.78$ Diab F), glycomics joint PC1 [integrated with iron] ($R^2=0.88$ PreDiab M, $R^2=-0.1$ PreDiab F, $R^2=0.2$ Diab M, $R^2=-0.97$ Diab F), glycomics joint PC1 [integrated with coagulation] ($R^2=-0.80$ PreDiab M, $R^2=0.99$ PreDiab F, $R^2=-0.96$ Diab M, $R^2=-0.93$ Diab F), and with glycomics joint PC1 [integrated with blood] ($R^2=0.89$ PreDiab M, $R^2=0.79$ PreDiab F, $R^2=-0.84$ Diab M, $R^2=0.12$ Diab M). These observations highlight that N-glycans are not specific for the endophenotypes but rather indicative of a biological condition. On the other hand, the highest inter-correlation of phenomics joint PC1 is with liver joint PC1 for prediabetic females ($R^2=-0.74$), diabetic males ($R^2=0.67$) and diabetic females ($R^2=0.69$), while for prediabetic males, the highest is with glycomics joint PC1 [integrated with blood] ($R^2=0.64$). This indicates that liver is one of the primarily organs to be affected in T2D disease.

It is to note the inter-correlations that are equal 0; they are attributed to an intrinsic assumption of O2PLS, thereby the reciprocal joint is uncorrelated with the following joint PC (for example, the glycomics joint PC1 [integrated with phenomics] has a correlation of 0 with phenomics joint PC2).

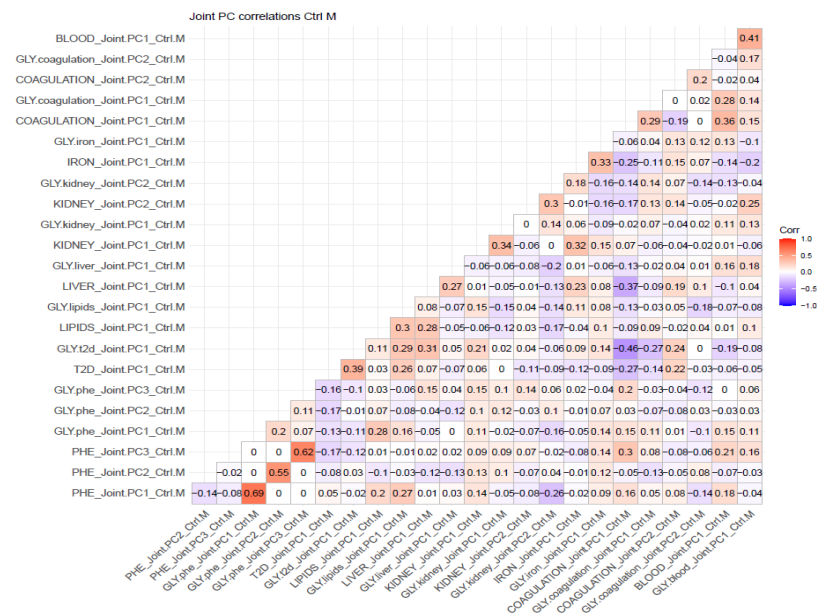


Figure 68. Pearson inter-correlations among the joint PCs in Ctrl M

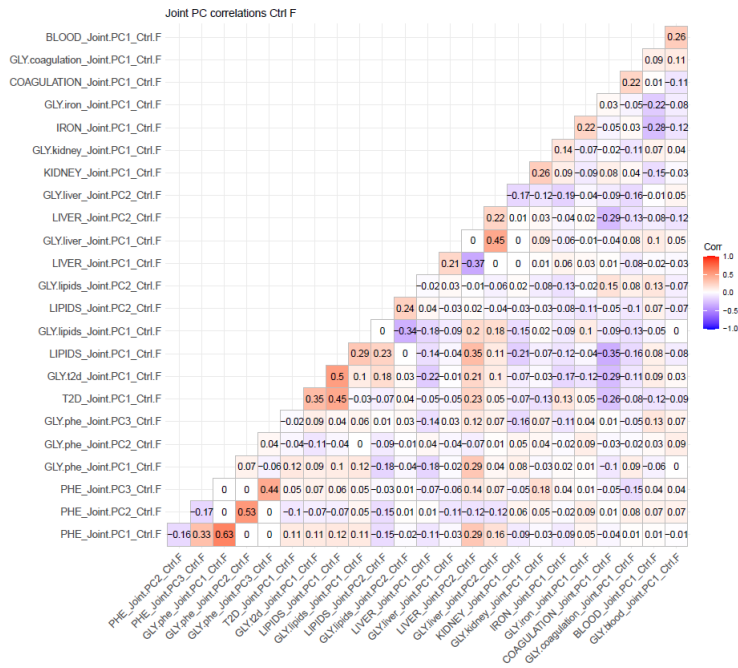


Figure 69. Pearson inter-correlations among the joint PCs in Ctrl F

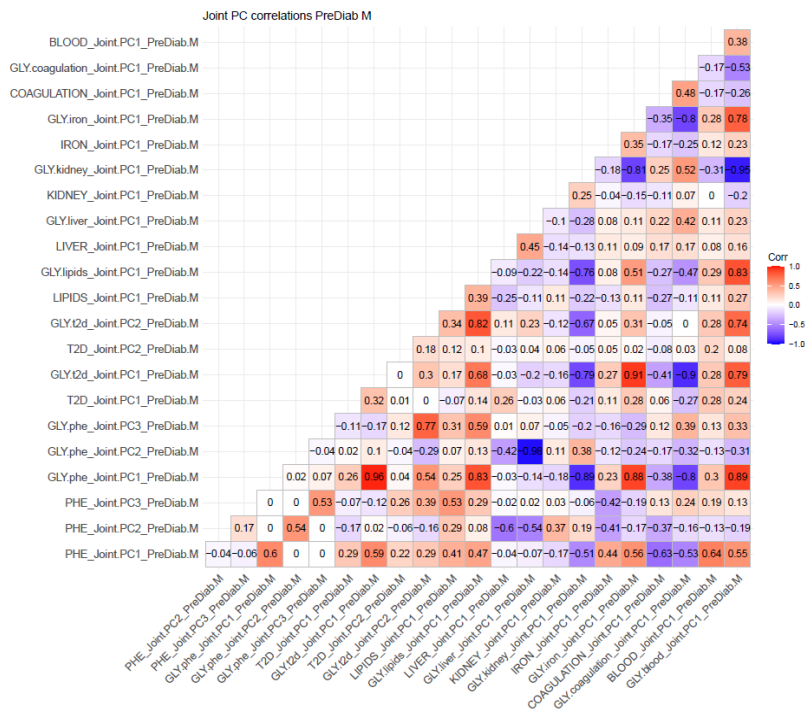


Figure 70. Pearson inter-correlations among the joint PCs in PreDiab M

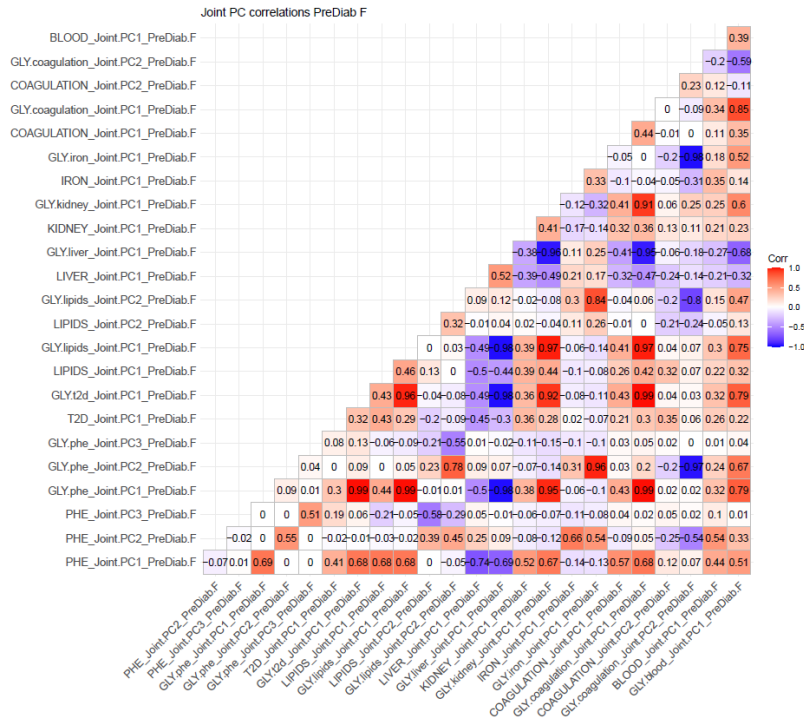


Figure 71. Pearson inter-correlations among the joint PCs in PreDiab F

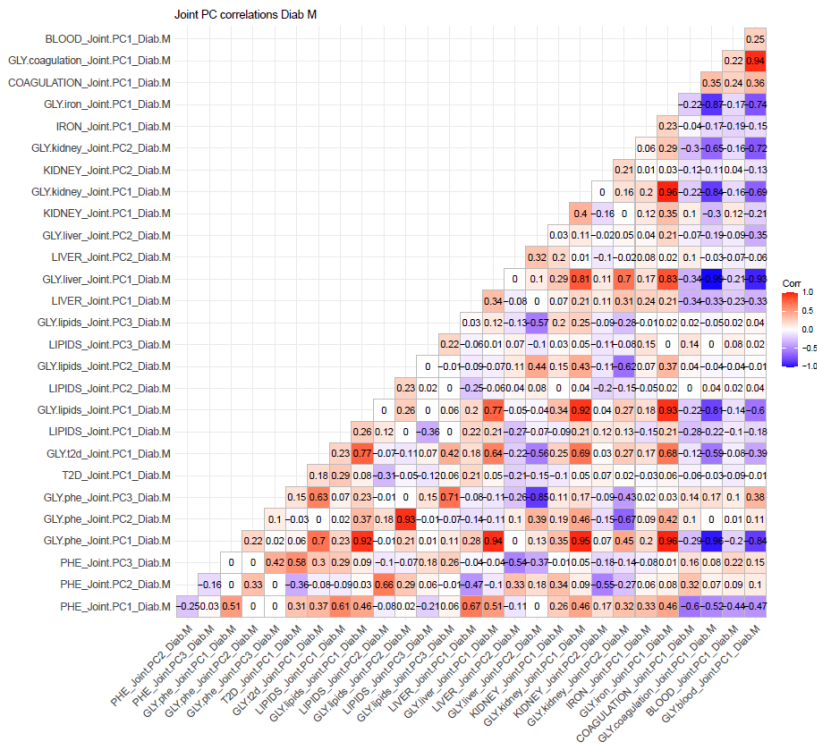


Figure 72. Pearson inter-correlations among the joint PCs in Diab M

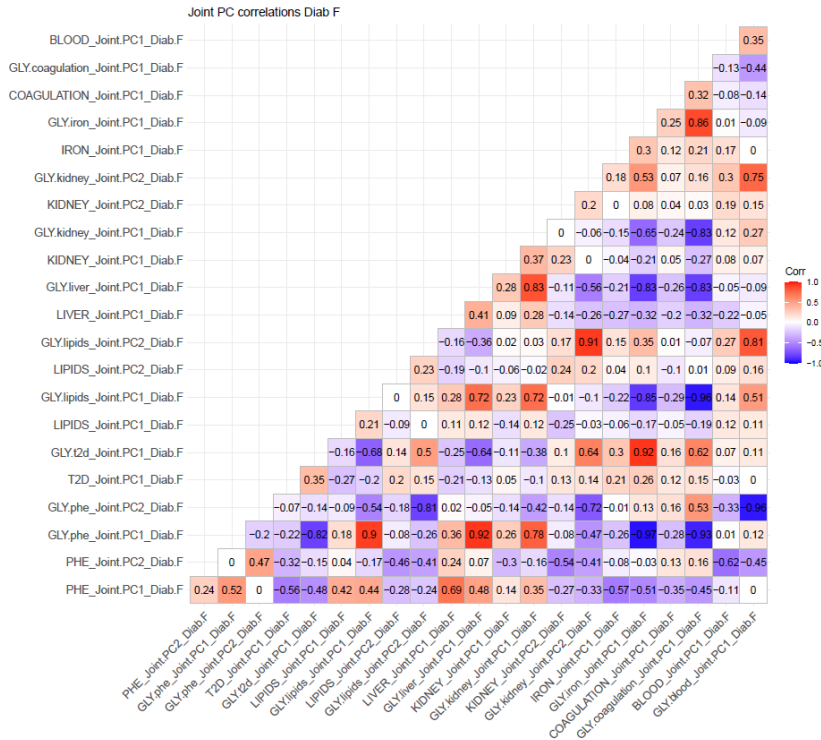


Figure 73. Pearson inter-correlations among the joint PCs in Diab F

Intra- and inter-correlations between pairs of joint PCs and individual joint PCs, respectively, therefore, point to one underlying condition having a functional impact in multiple organs and tissues, rather than different medical entities, in prediabetics and diabetics. **Table 19** reinforces this hypothesis, as the most common N-glycans covarying with the endophenotypes sets are GP3, GP4, GP6 for PreDiab F, GP5, GP8, GP10, GP1, GP7 for PreDiab M, GP3, GP4, GP8, GP10 for Diab F, and GP3, GP4, GP7, GP8, GP10 for Diab M; all of them fairly constant across glycomics scores of the same group, thus confirming the inter-correlations results.

Table 19. Top 4 loading values of the joint parts for each pair of integrated data sets

Group	Phenomics/Gly	T2D/Gly	Lipids/Gly	Liver/Gly	Kidney/Gly	Iron/Gly	Coagulation/Gly	Blood/Gly
Ctrl F	<i>Phenomics:</i> Triglycerides=-0.32 AP=-0.29 Uric Acid=-0.29 WHR=-0.29	<i>T2D:</i> HOMA=-0.59 Insulin=-0.55 FG=-0.45 HbA1c=-0.38	<i>Lipids:</i> WHR=-0.65 Triglycerides=-0.53 BMI=-0.34 HDL=0.26	<i>Liver:</i> AP=0.62 AST=0.59 TB=-0.37 ALT=0.32	<i>Kidney:</i> ADMA=0.56 Creatinine=0.48 EGFR=-0.47 TP=0.37	<i>Iron:</i> TI=0.69 Ferritin=0.53 Transferrin=-0.49	<i>Coagulation:</i> PAII=0.87 TAFI=0.36 Fibrinogen=0.34	<i>Blood:</i> Neutrophils=-0.48 Hemoglobin=-0.43 Lymphocytes=-0.38 Monocytes=-0.38
	<i>Glycomics:</i> GP6=0.53 GP2=-0.40 GP3=0.38 GP4=0.37	<i>Glycomics:</i> GP7=0.51 GP6=0.41 GP3=0.40 GP1=-0.38	<i>Glycomics:</i> GP4=0.60 GP2=-0.46 GP6=0.39 GP5=0.38	<i>Glycomics:</i> GP3=0.51 GP7=0.48 GP5=-0.37 GP8=-0.36	<i>Glycomics:</i> GP7=0.68 GP2=-0.34 GP1=-0.33 GP10=0.32	<i>Glycomics:</i> GP8=0.51 GP10=0.41 GP5=0.41 GP4=-0.35	<i>Glycomics:</i> GP4=0.44 GP10=-0.38 GP9=0.38 GP6=0.34	<i>Glycomics:</i> GP7=0.45 GP6=0.45 GP1=-0.42 GP4=0.38

Ctrl M	<i>Phenomics:</i> Fibrinogen=-0.38 TAFI=-0.31 Neutrophils=-0.27 White cells=-0.26	<i>T2D:</i> HbA1c=0.90 FG=-0.37 HOMA=-0.19 Insulin=-0.14	<i>Lipids:</i> HDL=-0.72 Apo A1=-0.38 Triglycerides=0.38 BMI=0.35	<i>Liver:</i> CRP=0.52 ALT=-0.47 TB=-0.44 AST=-0.38	<i>Kidney:</i> TP=-0.81 Uric acid=-0.48 Creatinine=0.21 SDMA=-0.19	<i>Iron:</i> Transferrin=-0.89 TI=-0.46 Ferritin=-0.07	<i>Coagulation:</i> Fibrinogen=-0.93 PAI1=-0.29 TAFI=-0.22	<i>Blood:</i> Platelets=-0.62 Eosinophils=-0.43 Lymphocytes=0.36 White cells=-0.32
	<i>Glycomics:</i> GP3=0.52 GP6=0.50 GP4=0.45 GP9=-0.42	<i>Glycomics:</i> GP3=-0.43 GP5=0.38 GP4=-0.36 GP9=0.36	<i>Glycomics:</i> GP6=-0.64 GP2=0.54 GP3=-0.39 GP7=-0.28	<i>Glycomics:</i> GP3=0.61 GP6=0.46 GP9=-0.45 GP4=0.31	<i>Glycomics:</i> GP7=-0.51 GP10=0.49 GP8=0.47 GP9=-0.31	<i>Glycomics:</i> GP2=-0.64 GP10=0.42 GP1=-0.38 GP6=0.28	<i>Glycomics:</i> GP6=0.55 GP9=-0.44 GP7=0.32 GP2=-0.32	<i>Glycomics:</i> GP6=0.47 GP7=0.46 GP3=0.40 GP5=-0.39
PreDiab F	<i>Phenomics:</i> AP=-0.31 CRP=-0.29 WHR=-0.29 Triglycerides=-0.28	<i>T2D:</i> Insulin=-0.68 HOMA=-0.67 FG=-0.29 HbA1c=-0.10	<i>Lipids:</i> WHR: -0.51 Triglycerides=-0.49 HDL=0.43 BMI=-0.33	<i>Liver:</i> AP=0.62 CRP=0.55 GGT=0.40 AST=0.33	<i>Kidney:</i> Azotemia=-0.65 Uric acid=-0.62 EGFR=0.27 SDMA=0.18	<i>Iron:</i> Transferrin=-0.74 Ferritin=0.63 TI=0.22	<i>Coagulation:</i> Fibrinogen=-0.72 TAFI=-0.64 PAI1=-0.26	<i>Blood:</i> RDW-CV=-0.57 White cells=-0.42 Lymphocytes=0.41 Platelets=-0.31
	<i>Glycomics:</i> GP6=0.53 GP4=0.46 GP3=0.42 GP2=-0.34	<i>Glycomics:</i> GP4=0.52 GP6=0.43 GP7=0.38 GP3=0.36	<i>Glycomics:</i> GP4=0.48 GP6=0.47 GP2=-0.45 GP3=0.36	<i>Glycomics:</i> GP6=-0.52 GP3=-0.46 GP4=-0.46 GP2=0.43	<i>Glycomics:</i> GP6=0.58 GP2=-0.42 GP1=-0.35 GP3=0.33	<i>Glycomics:</i> GP5=-0.46 GP8=-0.44 GP1=0.40 GP10=-0.40	<i>Glycomics:</i> GP6=0.58 GP4=0.43 GP3=0.36 GP7=0.32	<i>Glycomics:</i> GP10=-0.49 GP8=-0.41 GP3=0.40 GP4=0.39
PreDiab M	<i>Phenomics:</i> TAFI=0.48 RDW-CV=-0.30 Transferrin=0.27 HDL=0.26	<i>T2D:</i> FG=0.69 HbA1c=-0.52 Insulin=-0.36 HOMA=-0.34	<i>Lipids:</i> HDL=-0.59 BMI=-0.57 Apo A1=-0.35 WHR=-0.29	<i>Liver:</i> CRP=-0.71 AP=-0.56 TB=0.29 ALT=0.28	<i>Kidney:</i> TP=0.70 SDMA=0.42 Azotemia=0.32 EGFR=-0.27	<i>Iron:</i> Transferrin=0.71 Ferritin=-0.66 TI=0.24	<i>Coagulation:</i> TAFI=-0.81 Fibrinogen=-0.48 PAI1=-0.32	<i>Blood:</i> RDW-CV=-0.61 Monocytes=0.52 Hemoglobin=0.36 White cells=-0.25
	<i>Glycomics:</i> GP8=0.50 GP10=0.43 GP1=-0.41 GP5=0.35	<i>Glycomics:</i> GP7=-0.55 GP10=0.47 GP5=0.36 GP8=0.34	<i>Glycomics:</i> GP1=-0.54 GP5=0.43 GP4=-0.39 GP9=0.39	<i>Glycomics:</i> GP6=0.63 GP4=0.46 GP9=-0.38 GP3=0.37	<i>Glycomics:</i> GP5=-0.54 GP2=0.47 GP8=-0.39 GP10=-0.38	<i>Glycomics:</i> GP8=0.53 GP10=0.50 GP7=-0.44 GP1=-0.37	<i>Glycomics:</i> GP8=-0.42 GP10=0.41 GP6=0.39 GP7=0.38	<i>Glycomics:</i> GP1=-0.54 GP5=0.42 GP8=0.42 GP10=0.35
Diab F	<i>Phenomics:</i> AST=-0.32 TP=-0.31 TI=-0.30 TAFI=0.30	<i>T2D:</i> HbA1c=-0.71 FG=-0.60 HOMA=-0.36 Insulin=-0.11	<i>Lipids:</i> HDL=-0.61 Triglycerides=0.61 Apo B=0.37 WHR=0.23	<i>Liver:</i> AST=-0.62 CRP=0.52 TB=-0.45 ALT=-0.34	<i>Kidney:</i> TP=-0.97 EGFR=0.16 ADMA=0.12 Creatinine=-0.10	<i>Iron:</i> TI=-0.72 Ferritin=0.50 Transferrin=-0.49	<i>Coagulation:</i> TAFI=-0.96 Fibrinogen=-0.24 PAI1=0.13	<i>Blood:</i> Lymphocytes=0.62 Red cells=0.38 Hemoglobin=0.35 Monocytes=0.34
	<i>Glycomics:</i> GP3=-0.48 GP4=-0.48 GP8=0.45 GP10=0.42	<i>Glycomics:</i> GP4=0.57 GP2=0.56 GP2=-0.35 GP8=-0.34	<i>Glycomics:</i> GP8=0.61 GP10=0.50 GP4=-0.40 GP3=-0.30	<i>Glycomics:</i> GP3=-0.48 GP4=-0.43 GP8=0.39 GP10=0.34	<i>Glycomics:</i> GP5=0.54 GP1=-0.43 GP8=0.35 GP7=-0.32	<i>Glycomics:</i> GP4=0.46 GP3=0.44 GP10=-0.42 GP8=-0.40	<i>Glycomics:</i> GP8=-0.61 GP10=-0.53 GP7=0.36 GP3=0.29	<i>Glycomics:</i> GP6=0.52 GP8=0.42 GP2=-0.27 GP1=-0.35
Diab M	<i>Phenomics:</i> TP=-0.45 TAFI=0.34 CRP=0.28 WHR=0.28	<i>T2D:</i> HOMA=0.54 FG=0.52 HbA1c=0.49 Insulin=0.45	<i>Lipids:</i> WHR=0.58 Apo B=0.42 Cholesterol=0.39 LDL=0.31	<i>Liver:</i> CRP=0.73 TB=-0.59 AST=-0.27 AP=0.15	<i>Kidney:</i> TP=-0.96 EGFR=0.19 SDMA=-0.16 Creatinine=-0.12	<i>Iron:</i> TI=-0.72 Transferrin=0.54 Ferritin=-0.44	<i>Coagulation:</i> TAFI=-0.78 Fibrinogen=-0.61 PAI1=-0.10	<i>Blood:</i> Neutrophils=-0.41 MPV=0.41 Platelets=-0.39 White cells=-0.35
	<i>Glycomics:</i> GP8=0.45 GP3=-0.44 GP4=-0.41 GP10=0.35	<i>Glycomics:</i> GP3=-0.59 GP4=-0.54 GP4=0.33 GP7=0.29	<i>Glycomics:</i> GP8=0.55 GP3=-0.49 GP4=-0.45 GP10=0.32	<i>Glycomics:</i> GP4=-0.44 GP6=-0.41 GP3=-0.38 GP7=-0.36	<i>Glycomics:</i> GP1=-0.51 GP5=0.46 GP4=-0.33 GP3=-0.32	<i>Glycomics:</i> GP5=0.52 GP3=-0.43 GP8=0.42 GP10=0.41	<i>Glycomics:</i> GP8=-0.43 GP6=0.39 GP4=0.37 GP7=0.37	<i>Glycomics:</i> GP8=-0.47 GP3=0.44 GP10=-0.36 GP7=0.32

10. DISCUSSION

N-glycans species change in conditions of inflammation and metaflammation. In the frame of T2D and metabolic syndrome, it is not clear whether these alterations in N-glycans profile are the cause or the consequence of the disease. The clinical diagnostic of T2D is achieved through the criteria established by ADA (2020); however, it has two potential drawbacks: it takes an average of 7 years, and, when T2D is confirmed, the individual already presents an overt clinical symptomatology that rarely ameliorates and, instead, it progresses towards vascular complications. Thus, when the clinical diagnostic of T2D is achieved, the quality of life of an individual and the lifespan will likely have already been conditioned. As a whole, the increasing number of geriatric population, along with the late diagnosis of T2D, will lead to a socio-economical burden is going to be difficult to manage, unless systematic changes in the health system are carried out and/or more biomarkers of T2D/dysmetabolism/aging (which often go hand in hand) are found.

Here, we firstly address whether N-glycans could predict better the diabetic or non-diabetic status than HbA1c, used in clinical routine as parameter to diagnose T2D (ADA, 2020). Results confirmed that HbA1c is better for the diagnosis than N-glycans ($AUC_{HbA1c} = 0.97$, $AUC_{N-glycans} = 0.71$). Thereupon, we aimed to integrate 44 clinical variables of different sources (including T2D diagnostic parameters, biochemical parameters from the lipid profile, liver, kidney, and cell counts and blood differential factors) with 10 N-glycans obtained with DSA-FACE in six groups of individuals, that is, control males, control females, prediabetic males, prediabetic females, diabetic males, and diabetic females, in four age ranges, that is, 20-44, 45-55, 56-70, 71-85 years of age, and the full age range. We made these subgroups for different reasons: i) males and females present a different biology, ii) clinical parameters and N-glycans change with age, and iii) controls, with age, tend to resemble prediabetics/diabetics, especially from 60 years on, since as healthy individuals get older tend to become less healthy and start presenting subclinical alterations. Results of the six groups, focusing on the full age range, evidenced from the loading values of the integrated variables, in hand with their corresponding mean values, that N-glycans GP3, GP4, GP6 are negatively associated with age in the six groups, while GP8 and GP10 are positively associated with diverse clinical parameters, especially from coagulation and lipidic profile, in diabetic males and females, and in prediabetic males. These signals are suggestive of being some of the team players mirroring a metabolic dysfunction. In addition, the decreasing intra-correlations between joint phenomics and joint glycomics from controls to diabetics supports the heterogeneity and/or complexity of T2D and/or metabolic syndrome.

To the best of our knowledge, there are currently no previous studies integrating clinical variables and N-glycans in the context of metaflammation and T2D. Notwithstanding, we confirmed GP6 (digalactosylated, fucosylated) as an N-glycan that negatively correlates with aging (Vanhooren et al., 2008; 2010; Kristic et al., 2014), and GP8 and GP10 (multi-branched, galactosylated N-glycans) as N-glycans associated with lipidic profile and T2D diagnostic parameters (Testa et al., 2015, Keser et al., 2017).

We further aimed to better understand whether these N-glycans changes arise from a common medical condition or there are specific signatures of N-glycans according to the endophenotypes to which they are integrated. Results of the intra-correlations between pairs of joint PCs scores showed that, in prediabetics, N-glycans scores correlate less with T2D diagnostic parameters and lipid profile scores, while they displayed a high correlation with liver and coagulation scores. Conversely, in diabetics the correlation of joint scores between N-glycans and T2D diagnostic

parameters and lipid profile was surprisingly low; in turn, the correlation with liver and coagulation was higher, yet much lower than the same correlations observed in prediabetics. Overall, this could be indicative of the fact that medication changes N-glycan profile, as diabetics take more drugs than prediabetics (and controls) (**Table 7**), especially for lowering the glycaemic and lipid indexes, or anti-aggregants for vascular disease-related conditions. On the other side, the highest intra-correlation in controls was found in the T2D set, which may be attributed to the fact that these control individuals are not the prototype of healthy controls, as the mean BMI is 28.32 for males and 25.75 for females, and thus it is suggestive of an onset of biological dysregulation. Secondly, inter-correlations between individual joint PCs scores show that in prediabetics and diabetics N-glycans are generally the same when integrated with different endophenotypes, instead of revealing a specific signature for each endophenotype, suggesting one biomedical condition; whereas in controls, inter-correlations are too low for interpretation.

Together, this extensive characterization of the phenotype by integrating clinical and N-glycans data has shed light into the subclinical window of T2D/metabolic syndrome, obtained from healthy younger controls and less healthy older controls and prediabetics, and into the clinical window, obtained from younger and older diabetic patients. Overall, individuals should not only be studied on the basis of T2D diagnostic parameters, but also considering a combination of parameters characteristic of insulin resistance (metabolic) syndrome, which starts before the clinical onset and diagnosis of T2D, thus in control and prediabetic individuals. Further research is needed to refine and elaborate this design. Here we identify several future directions that can build on our work for investigating the joint variation between clinical and glycomic variables or other metabolites that provide further insight into the pathology.

First, our analysis was limited to a single cohort, and one that lacked ethnic minority populations. Replication in other cohorts is needed, and replication using a larger number of younger and older healthy controls, younger and older less healthy controls, prediabetics and diabetics can also provide additional perspectives. Larger samples can also help understanding the less common phenotypes.

Second, DSA-FACE N-glycans are low-dimensional and have the disadvantage that sialic acids are cut-off from the sugar, which blurs the interpretation of results. Further studies are needed that use N-glycans obtained with other technologies which include sialic acids and produce more N-glycan species, and not only from whole plasma glycome, but also from IgG glycome, which is especially relevant for understanding the role of inflamm-aging and metaflammation played by N-glycans.

Third, the nature of the cross-sectional study hinders the ability to answer the question “*do N-glycans and other biochemical parameters change over the course of preclinical (that is, insulin resistance and impaired insulin secretion) to clinical diagnoses (obesity and/or T2D, T2D complications, metabolic syndrome)?*”

Our findings suggest that future studies of T2D/dysmetabolism incorporate longitudinal repeated measures of biomarkers to track change. They also suggest that these investigations of T2D/dysmetabolism incorporate multiple biomarkers to track change across different organ systems.

Within the bounds of these limitations, the implication of the present work is that it is possible to identify subclinical and clinical signatures of T2D/dysmetabolism by integrating clinical and N-glycans data.

CHAPTER 2: Estimation Of An Innovative Composite Aging Clock As Biomarker For Type 2 Diabetes And Related Metabolic Phenotypes

11. INTRODUCTION

Aging is a time-dependent multifactorial process that results in a global deterioration of the physiological functions and elevated risk of pathologies, including cardiovascular disorders, neurodegenerative diseases, cancer and diabetes (López-Otín et al., 2013). It is well-known that age is a major risk factor for functional impairments, chronic diseases and mortality; however, the aging rate is not universal for humans, as it depends on the individual exposure/resilience trade-off. Consequently, chronological age may not be a reliable indicator of the body's physiological decline, but rather a proxy of the aging growth rate. In this line, the unprecedented growth rate of world's aging population is highlighting the need for better understanding the aging process and the determinants of healthy and unhealthy aging.

Traditionally, the aging rate has been estimated through the mortality curves of populations, i.e. individuals are tracked until their death to estimate their "biological age" at time points when they were alive (Moskalev, 2020). Thus, to measure the functional decline, and to predict the morbidity onset and the life expectancy with this method are not possible.

An alternative to mortality risk estimation for measuring the biological age is the identification of a set of endophenotypes that change with aging, called "biomarkers of age", that best predict chronological age. This is where the "paradox of biomarkers", reformulated by Hochschild (1994), comes into play: "A hypothetical biomarker that approaches perfect correlation with chronological age could be replaced by chronological age and would be insensitive to differences in aging among individuals." The core of the paradox is that biological age is estimated by multiple linear regression, which uses chronological age as dependent variable and a set of biological predictors as independent variables, thus resulting in biological age equal to chronological age (Klemra and Doubal, 2006).

According to the American Federation for Aging Research (AFAR), a biomarker of age has to fulfil the following criteria (Johnson, 2006):

1. it must predict the rate of aging (it should tell exactly where a person is in their total lifespan and it must be a better predictor of lifespan than chronological age);
2. it must monitor a basic process that underlies the aging process, not the effects of disease;
3. it must be able to be tested repeatedly without harming the person (for example a blood test or an imaging technique);
4. it must be something that works in humans and in laboratory animals, such as mice (so that it can be tested in laboratory animals before being validated in humans).

However, the existence of biomarkers meeting all AFAR criteria has been called into question, as organs and tissues age at different rates (Franceschi et al., 2018), which makes the estimation of the overall individual aging rate difficult and, in turn, the estimation of the overall population aging rates, too. Examples of aging rate types between individuals are: persons with Down syndrome (a

progeroid syndrome), who show accelerated aging from a young age; individuals with T2D (an age-related disease), who have an increased risk for CVD, which make them more likely to age at faster rates than matched individuals without T2D; aged individuals with history of disease; aged individuals with no history of disease; and centenarians, who have avoided or postponed age-related diseases in their 60s, 70s and 80s and show decelerated aging. Thus, the phenotype in the elderly depends on the genetic background and the interaction with environment and lifestyle.

Further impediments to identify biomarkers of aging are the challenge to disentangle the intrinsic processes underlying healthy aging from the ones of unhealthy aging or of onset of age-related diseases and to capture independent and overlapping mechanistic processes driving the aging process. To date, the biological mechanisms underlying aging that have been proposed are the “hallmarks of aging” (López-Otín et al., 2013) and the “seven pillars of aging” (Kennedy et al., 2014).

Many research studies have described biomarkers of aging, which constitute tiny pieces encompassing the complex puzzle of the aging process. These can be molecular – DNA methylation, transcriptome-based, proteome-based, metabolome-based, N-glycans, telomere length – and phenotypical – blood measures, neuroimaging, frailty, musculoskeletal functions –.

25. Molecular biomarkers of aging

27. DNA methylation

A promising biomarker of age is DNA methylation age (DNAm), also referred as epigenetic clock, because it applies to all sources of DNA (cells, tissues and organs) and to the whole age spectrum (from prenatal tissue to tissues of centenarians) (Horvath and Raj, 2018).

So far, at least nine robust state-of-the-art estimators have been described: seven epigenetic clocks based on DNAm (Horvath, 2013; Horvath et al., 2018; Hannum et al., 2013), Weidner et al., 2014, Levine et al., 2018; Lu et al., 2019), and CpG sites (genes islands *ELOVL2* and *FHL2*, Garagnani et al., 2012).

Horvath and Hannum epigenetic clocks show high correlations with chronological age ($r=0.96$ for Horvath and $r=0.91$ for Hannum) and small, mean absolute deviations (MAD) from calendar age (3.6 and 4.9 years, respectively) in their corresponding validation cohorts. Weidner reported a $MAD=3.34$ years and $R^2=0.98$ in the discovery cohort, and a MAD of 5.79, 5.52 and 4.02 years in the validation cohorts. Sample size for the three algorithms was of $N=8000$ for Horvath, $N=656$ for Hannum and $N=151$ for Weidner. The Horvath clock is a multi-tissue predictor based on methylation levels of 353 CpG sites on the Illumina 27k array, Hannum clock uses only 71 CpG sites from the Illumina 450k array and performs best using whole blood samples, and Weidner identifies 3 age-related CpGs by using bisulfite pyrosequencing (Horvath, 2013; Hannum et al., 2013; Weidner et al., 2014; Jylhävä et al., 2017). Selection of CpG sites was done using elastic net, a penalized regression multivariate model, in Horvath and Hannum (yet, they only have 6 CpG sites in common), and a recursive feature elimination in Weidner. The most important characteristics about Horvath and Hannum clocks is their ability to predict all-cause mortality independent of classic risk factors (Chen et al., 2016). When it comes to aging phenotypes, Marioni and colleagues reported that age and sex-adjusted Horvath’s clock was associated with different measures of fitness, that is, epigenetic age acceleration was associated to a poorer cognitive performance, lower

grip strength, and poorer lung function at baseline (Marioni et al., 2015), though the baseline DNAm at 70 years did not predict the rate of change of these fitness measures. Other associations have been performed in blood, in which the Horvath and Hannum clocks correlate with certain blood cell types that also show age-related changes (Marioni et al., 2015; Chen et al., 2016). The fourth clock, Levine's clock, named DNAm PhenoAge, represents an improvement of the previous epigenetic biomarkers for several reasons: i) 41/513 CpGs in DNAm PhenoAge were shared with the Horvath clock, 5 of which were also shared with the Hannum's clock), ii) unlike the previous epigenetic clocks, which aimed to optimize the correlation of the CpGs predictors with chronological age, the CpGs in DNAm PhenoAge were tuned to predict a multi-system proxy of physiological dysregulation (phenotypic age). As a result, the CpGs with the highest effect sizes in the new clock did not correlate with chronological age, but instead were related to the difference between phenotypic and chronological age (i.e. divergence in the rate of aging).

Several investigations have also analysed the association between the epigenetic age and diseases of aging: in Alzheimer's disease (AD) patients, prefrontal cortex was associated with the presence of plaques, amyloid load and a decline in global cognitive functioning, episodic memory and working memory (Levine et al., 2015); in Parkinson's disease (PD), DNAm is higher for PD patients when compared with controls, after adjusting for cell blood composition (Horvath and Ritz, 2015); in cancer, the Hannum clock presents increased epigenetic age in the tested tumor tissues (Hannum et al., 2013), whereas the Horvath's clock shows increased DNAm only in certain cancer types (Horvath, 2013; Horvath, 2015); in osteoarthritis, the Horvath's clock was associated with an increased DNAm in the joint of the affected cartilage, but not in the nearby bone or blood (Vidal-Bralo et al., 2016).

The two CpG sites considered epigenetic markers map in the gene islands of *ELOVL2* and *FHL2*. Both CpG sites display high correlations with chronological age ($r=0.92$ for *ELOVL2* and $r=0.80$ for *FHL2*) that correspond to a progressive increase in methylation with age. The discovery was made on 64 subjects with Illumina 450k array on whole blood DNA and validated with a cohort of 501 subjects analysed with Sequenom's EpiTYPER (Garagnani et al., 2012). A subsequent investigation examined the relationship between DNA methylation and the prospective development of breast cancer and colorectal cancer in 845 subjects (Durso et al., 2017). The authors found that the two most used epigenetic clocks (Horvath and Hannum) were unable to detect age acceleration effects in females blood that were later diagnosed of breast cancer, while *ELOVL2* did; conversely, Horvath's clock and *FHL2* predictor were associated with colorectal cancer in males (Durso et al., 2017).

28. Transcriptomics-based

Ferrucci and colleagues conducted a transcriptome-wide gene expression analysis in peripheral blood leukocyte samples of individuals aged 30-104 years (N=698), to determine which transcripts were most associated with advancing age in the InCHIANTI study (Ferrucci et al., 2000), and they found that using expression levels of only 6 genes (*LRRN3*, *CD27*, *GRAP*, *CCR6*, *VAMP5* and *CD248*), they were able to distinguish between younger (age<65) and older subjects (age≥75) with high accuracy (Ferrucci et al., 2000). Later on, they studied the performance of a modified model in three populations and determined whether individuals predicted to be biologically younger than their chronological age had biochemical and functional measures consistent with a younger biological age: individuals with younger gene expression patterns had higher muscle strength and serum albumin, and lower interleukin (IL)-6 and blood urea concentrations relative to "biologically

older” individuals (OR=1.05, 1.13, 0.61, 0.98; $p=3.2\times 10^{-2}$, 2.7×10^{-4} , 1.1×10^{-2} , 2.6×10^{-2} respectively) (Holly et al., 2013).

Another large study performed a whole-blood gene expression meta-analysis in 14,983 individuals and identified 1,497 genes that were differentially expressed with chronological age; from these significant genes, the age-associated ones were enriched for the presence of potentially functional CpG methylation sites in enhancer and insulator regions. The authors calculated the “transcriptomic age” from the gene expression profiles and showed that differences between transcriptomic age and chronological age were associated with processes characteristics of aging, such as blood pressure, cholesterol levels, fasting glucose, and BMI (Peters et al., 2015).

In another study, Fleischer and colleagues generated a data set of genome-wide RNA-seq profiles of human dermal fibroblasts from people aged 1 to 94 years old to test whether signatures of aging are encoded within the transcriptome (Fleischer et al., 2018). With this purpose, they made use of an ensemble machine learning method that predicts age to a median error of 4 years, thus outperforming algorithms proposed by prior studies that predicted age from DNA methylation (Horvath, 2013; Hannum et al., 2013; Peters et al., 2015; Xu et al., 2015).

A further investigation aimed to account for the inter-individual variability by performing gene expression analysis in T lymphocytes from 27 healthy monozygotic twins aged between 22 and 98 years (Remondini et al., 2017). The authors obtained a transcriptomic signature of 125 genes, from which chronological age could be estimated. It should be noted the relevance of this work as this signature represents genes involved in normative aging, which expression changes independently of environmental factors.

29. Proteome-based

A recent study quantified 2,925 plasma proteins from 4,263 individuals of 18 to 95 years old and developed a new bioinformatics approach that uncovered marked waves of changes in the proteome in the 4th, 7th and 8th decades of life (Lehallier et al., 2019). Specifically, to determine whether the plasma proteome could predict biological age and serve as a “proteomic clock”, the authors used 2,817 randomly selected subjects to fine-tune a predictive model that was tested on the remaining 1,446 subjects. Indeed, they identified a sex-independent plasma proteomic clock consisting of 373 proteins, thereby individuals who were predicted to be younger than their chronological age performed better on cognitive and physical tests. The work is promising, because the panel of 373 proteins can be used to assess the relative health of an individual and to measure health span, and thus, despite it needs more large-scale validation, it can be considered analogous to epigenetic clocks based on DNA methylation patterns (Lehallier et al., 2019).

A subsequent research, by Johnson and colleagues, systematically reviewed 36 different proteomics studies of proteomics that significantly changed with age (Johnson et al., 2020). They discovered 1,128 proteins that had been reported by ≥ 2 or more analyses and 32 proteins that had been reported by ≥ 5 analyses, and proposed two proteomic aging clocks, a smaller panel and a larger panel comprised of proteins that were reported to change with age in plasma in ≥ 4 and ≥ 3 studies, respectively (Johnson et al., 2020). They achieved so by validating the plasma proteins with an online aging plasma proteome interface created by Lehallier and colleagues and by demonstrating that these protein panels are bona fide aging clocks that can accurately predict patient age in a large cohort of 3,301 individuals. The authors reported that 2/3 of 1,128 proteins changed with age in ≥ 2 tissues/cell types, by which they suggested that, similar to the Horvath’s clock, a single proteomic

clock might be estimated with the use of machine learning algorithms in a multi-tissue human dataset (Johnson et al., 2020).

30. Metabolome-based

The clinical manifestations of many age-related diseases, such as T2D, AD, PD are well known. However, the complex interconnection of biomolecular pathways contributing to the aging process are just beginning to be unveiled. In other words, the few pathways described so far, such as DNA damage or accumulation of ROS, have been seen as independent processes, but most of the links among them are missing – one exception is the evidenced interconnection between DNA damage and metabolic control (Bai et al., 2011; Sahin et al., 2011). Examples of metabolic systems known to be implicated in aging are the insulin/IGF1 pathway, the mTOR, sirtuin and AMP-activated protein kinase (AMPK) pathways (Houtkooper et al., 2011).

Several studies have aimed to identify metabolic signatures of aging. A key investigation that shed light on the functioning of metabolic processes in longevity combined ¹H-NMR profiling and targeted mass spectrometry (MS) approaches in a human aging cohort comprised mostly of female centenarians, elderly, and young individuals, and showed that, with increasing age, targeted MS profiling of blood serum markedly decreases in tryptophan concentration, while a unique alteration of specific glycerophospholipids and sphingolipids are seen in the longevity phenotype (Collino et al., 2013). The authors proposed that the overall lipidome changes reflect the centenarians unique capacity to adapt to the accumulative oxidative and chronic inflammatory conditions of their extreme aging phenotype. Yet, the observational nature of the study made unable to establish causal links between the inflammatory status, aging, and modulation of lipid metabolism (Collino et al., 2013).

In this line, as in other omics studies, the lack of multilevel integration of different data sets hinders the understanding of the metabolic decline during aging that predisposes to age-related diseases. A novel work systemically integrated *in vivo* phenotyping with gene expression, biochemical analysis and metabolomics in young and 2 year old mice. They analysed muscle and liver affected pathways, and found that alterations in glucose and fatty acid metabolism and redox homeostasis translated in decreased long chain acylcarnitines and increased fatty acid levels and a marked reduction in various amino acids in the plasma of aged mice. The authors concluded that such metabolites serve as biomarkers for aging and health span (Houtkooper et al., 2011).

31. Glycomics

As mentioned before, glycans are product of a complex pathway that involves hundreds of different proteins and are encoded in a complex dynamic network of hundreds of genes (Lauc and Zoldoš, 2010). Unlike the genome, which remains constant in life, glycans are stable molecules that change with physiological and pathological conditions; thus, their relative quantification and analysis can contribute to the understanding of the underpinning processes of aging.

A pioneering study conducted by Krištić and colleagues confirmed that glycans, indeed, are a novel biomarker of chronological and biological ages (Krištić et al., 2014). The authors analysed 24 immunoglobulin (Ig)G plasma N-glycans using the UPLC technology in N=906 subjects from Vis island, N=915 subjects from Korčula island, N=2035 subjects from Orkney islands, and N=1261 subjects from TwinsUK cohort. The researchers built a predictive model of age, termed

“GlycanAge”, by which 3 IgG N-glycans, FA2B, FA2G2, and FA2BG2, explaining 58% of variance, were found to be markers of age (Krištić et al., 2014). Equally important was the work performed by Vanhooren et al. (2008; 2010) with the use of the DSA-FACE technology in total blood serum, in which they described that the GlycoAge Test, obtained from the log ratio of P1/P6, is better than chronological age at estimating physiological age, and is higher in dementia and Cockayne syndrome than in healthy individuals.

It is worthy to note that, as Levine et al. (2018) highlighted, the biological age should not be estimated by optimizing the correlation of the predictors with chronological age, and instead it should focus on time-dependent weights of these predictors. In this line, Krištić et al. (2014), similar to Levine and colleagues with DNAm PhenoAge, defined the GlycanAge = intercept + $\beta_1 \cdot GP6 - \beta_2 \cdot GP6^2 - \beta_3 \cdot GP14 + \beta_4 \cdot GP15$, separately in males and females, but estimated the difference in the biological age with chronological age by computing the delta median of the GlycanAge calculated both in 2003 and in 2013 in the same subjects – in the case of females (N=20, aged 38-50) the median delta GlycanAge was of 9.6 years, while for the males (N=6, aged 43-49) was equal to 0.6.

26. Phenotypical biomarkers of aging

32. Composite-based

Other investigations have searched for multiple biomarkers that combined produce a single biological age predictor, which can be advantageous in biological research, public health and clinical practice.

An important study was conducted by Levine et al. (2013), in which he compared the predictive ability of 5 different statistical methods for estimating Biological Age using the cohort National Health and Nutrition Examination Survey III (NHANES III). The study comprised N=9,389 subjects, aged 30-75 years, that within 18-year follow-up, 1,843 deaths were recorded. Results showed that the Klemmera and Doubal method (KDM) (Klemmera and Doubal, 2006) was the most reliable predictor of mortality and performed significantly better than chronological age; thus, it may represent a useful algorithm for future research studies (Levine, 2013). In this line, other research works have also employed this algorithm to estimate BA (Belsky et al., 2015).

Belsky and colleagues used a dual approach to study Biological Age and the Pace of Aging in N=954 young adults from the Dunedin Study birth cohort. Biological Age was obtained cross-sectionally using the KDM, while the Pace of Aging was measured longitudinally at 3 different points of time. Results evidenced that study members with older Biological Age had faster Pace of Aging over the preceding 12 years (Belsky et al., 2015). Authors suggested as strategy for future studies to use longitudinal repeated measures of biomarkers to track change of multiple tissues and organs, and that these studies will require new statistical methods to measure the Pace of Aging, likewise KDM currently calculates Biological Age.

A totally different approach was used by Sebastiani and colleagues, thereby they measured 19 blood biomarkers, comprised of hematological measures, lipids profile, and markers of inflammation and frailty in 4,704 participants of the Long Life Family Study (LLFS), aged 30-110 years. Briefly, they used an agglomerative algorithm to group LLFS subjects into clusters, which yielded 26 biomarker signatures, and they subsequently correlated them with longitudinal changes in physiological

functions and incident risk of cancer, CVD, T2D, and mortality, with the end goal of testing whether these signatures were associated with differences in biological aging. One of those signatures was associated with lower mortality, morbidity and more healthier aging, while nine other signatures were associated with higher risks for mortality, morbidity and less successful aging. Seven of the signatures were replicated in an independent cohort (Sebastiani et al., 2017).

To conclude with the overview of the current described molecular and phenotypical clocks, future studies need to take into consideration: i) longitudinal repeated measures to measure the Pace of Aging, ii) new statistical methods for estimating both Biological Age and the Pace of Aging, iii) young and old subjects, i.e. an age spectrum as wide as possible, iv) multiple biomarkers accounting for different organs and tissues of molecular and phenotypical nature.

27. Motivation

In this study, not only controls are used, but also prediabetics and diabetics. It can be noted that the inclusion of prediabetics and diabetics in the experiment does not fulfill the AFAR criteria for a biomarker to calculate the biological age, because “it must monitor a basic process that underlies the aging process, *not the effects of disease*”. Notwithstanding, many studies reflect that blood vessel damage, atherosclerosis, CVD and CVD-associated mortality affect a great majority of the elders (e.g. Yazdanyar and Newman, 2009). In accordance with this, Wang and Bennett (2012) recently proposed that “*the effects of atherosclerosis are superimposed on normal aging of the underlying vessel*”. We, therefore, considered a different strategy with regard to previous studies, that is, to test whether a mix biomarker that includes determinants of healthy and unhealthy aging (i.e. variables of atherosclerosis that overlap with aging) can estimate biological age, considering i) that pure controls do not exist and ii) that, in the elder population, the concept of healthy control is cumbersome, since when progressively approaching the age of the average life expectancy, even in absence of overt diseases, the likelihood that subclinical pathological processes are active is high. At the same time, the fact of not only considering controls, but also prediabetics and diabetics, increases the sample size, which, in turn, increases the power of the study.

12. MATERIALS AND METHODS

28. Materials

The study used a total of 1146 individuals, comprising 177 Ctrl F, 108 Ctrl M, 185 PreDiab F, 121 PreDiab M, 257 Diab F, and 298 Diab M. 39 phenotypical variables and 10 N-glycans were used, while HbA1c, Fasting Glucose, HOMA, and Insulin, which are variables that define the pre-T2D and T2D status, were excluded from the study, along with sex and age. The inclusion criteria considered individuals aged 20-85 years old, group- and sex-matched, and with an age balanced distribution among groups.

29. Methods

In this section, the methods utilized are as follows:

6. Univariate linear regression

It determines the relationship between one independent variable (X) and one dependent variable (Y). It is handy when the relationship between 2 variables is not obvious to the naked eye. The disadvantage of univariate regression is that it leads to high bias error because the model is too simple (Oby Tayo, 2019). In this study, univariate regression models were used to test the association of age with each variable: Age ~ Variable. P-values were adjusted with Bonferroni correction for multiple testing at an $\alpha=0.05$ and n =number of variables ($n=55$).

7. Multiple linear regression

It determines the relationship between several independent variables ($X_1 + X_2 + X_j$) and one dependent variable (Y). In this study, we tested the association of each phenotypical variable as function of sex, age and diabetes (Variable~Sex+Age+T2D status) in order to check for the partial effect of these 3 variables on the phenotypical variables.

8. Penalization or shrinkage methods

Unlike univariate regression, where there is high bias and low variance, multivariate regression models often have large variance, especially in situations of collinearity (note: the square of the bias and the variance are the 2 components of the MSE). In such models, the variance is large and the MSE is mainly determined by this variance. Penalization models impose a bias by applying a penalty to the regression coefficients (revised from Archarjee, 2013).

1. Ridge regression

It shrinks the regression coefficients by imposing a penalty on the sum of squares (L2 norm) of regression coefficients. In ridge, none of the estimated regression coefficients becomes exactly zero; therefore, all variables stay in the model (Archarjee, 2013).

2. Lasso regression:

It shrinks the regression coefficients by imposing a penalty on the sum of the absolute values of the regression coefficients (L1 norm) of regression coefficients. In LASSO, some of the estimated regression coefficients become exactly zero; therefore, only the selected variables stay in the model (Archarjee, 2013).

3. Elastic net regression:

It is a combination of ridge and LASSO because it shrinks the regression coefficients by imposing a penalty on the sum of squares (L2 norm) and a penalty on the sum of the absolute values of the regression coefficients (L1 norm) of regression coefficients (Archarjee, 2013).

9. Machine learning methods:

The goal of machine learning is to build a computer system that can learn from experience:

4. Random forest regression:

It is made up of decision trees, each of which is built from bootstrap samples of the data set. In general, 2/3 parts of the samples (in this study 9/10) are used in the training set, and 1/3 (here 1/10) is left out (the so-called out-of-bag (OOB) samples). Variables used which decrease the prediction error obtain higher variable importance (Archarjee, 2013).

5. Support vector regression:

The goal of SVR is to find a function $f(x)=wx+b$ that has at most ε deviation from the actually obtained response (Age) for all the predictors, and at the same time, minimizes the distance between predicted and observed values (Archarjee, 2013).

10. Dimension-reduction methods:

The goal is to transform the data from a high-dimensional space into a low-dimensional space, so that the latter retains meaningful properties of the original data:

6. Principal component analysis (PCA):

It is carried out on all original regressors, and each component (latent variable) is represented by a linear combination of the original variables (Archarjee, 2013).

7. Two-way orthogonal partial least squares (O2PLS):

See materials and methods of chapter 1.

11. Klemera-Doubal method (KDM):

Briefly, it takes information from m number of regression lines of chronological age regressed on m number of biomarkers (Klemera and Doubal, 2006).

13. AIM

As posed before, the research problem is based in that age is a risk factor for functional impairments, chronic diseases and mortality, and that the world's population is aging at a fast speed, thus generating a major healthcare and socio-economic burden. Accordingly, the research questions to be addressed in this study are: 1) *How can we better understand the underlying mechanisms of the aging process?* 2) *Which are the determinants of healthy and unhealthy aging?* To shed light on this topic, the aim of this research is to identify a composite biomarker capable of estimating biological age.

14. RESULTS

30. Univariate regression models: Age ~ Variable

First, to understand which variables may be importantly related to the Age variable, we conducted univariate models for all clinical variables and N-glycans (**Table 20**).

Table 20. Univariate regression models. P-value was adjusted for multiple comparisons with Benjamini-Hochberg correction at $\alpha=0.05$

Variable	β coefficient	p-value	adj. p-value	Pearson cor (R^2)
GP2	0.01204	2.99E-46	1.58378E-44	0.404099
GP6	-0.00659	3.37E-45	8.92652E-44	-0.3997
EGFR	-0.68998	2.92E-32	5.15406E-31	-0.33925
GP1	0.008343	2.08E-29	2.75043E-28	0.323951
WHR	0.002344	2.22E-26	2.34839E-25	0.306679
Azotemia	0.006782	1.09E-21	9.65249E-21	0.277368
HbA1c	0.030458	1.23E-18	9.32151E-18	0.256222
GP3	-0.00573	9E-18	5.9604E-17	-0.24987
GP4	-0.004	1.14E-17	6.70608E-17	-0.24911
Fasting glucose	1.090439	1.88E-16	9.95026E-16	0.239809
Alkaline phosphatase	0.005926	2.28E-13	1.09884E-12	0.214277
Creatinine	0.004929	1.63E-12	7.21594E-12	0.206599
Triglycerides	0.010435	2.38E-12	9.7003E-12	0.205099
Fibrinogen	1.460053	9.37E-12	3.54891E-11	0.199527
CRP	0.007166	5.7E-10	2.01502E-09	0.181771
Uric acid	0.004198	6.1E-09	2.02152E-08	0.170667
RDW-CV	0.000984	1.63E-08	5.0929E-08	0.165833
Ferritin	0.003007	7.35E-08	2.16303E-07	0.158169
Telomere length	-0.00591	1.99E-07	5.55367E-07	-0.15288
GP8	-0.00373	3.07E-06	8.14523E-06	-0.13732
HOMA	0.003507	4.45E-06	1.12292E-05	0.13509
Lymphocytes	-0.00333	0.000147	0.000355207	-0.11188
Transferrin	-0.44121	0.000163	0.000376039	-0.11114
GGT	0.003923	0.000334	0.000736812	0.105797
MCV	0.049961	0.000459	0.000973306	0.10333
GP10	-0.00326	0.000527	0.001074489	-0.10225
BMI	0.040743	0.001718	0.00337147	0.092515
Apo B	0.223262	0.002726	0.005159625	0.088456
Red cells	-0.00346	0.003931	0.007184274	-0.08512
GP9	0.003783	0.004745	0.008383449	0.083361
MPV	0.006914	0.007817	0.013364789	0.078536
GP5	-0.00102	0.014079	0.023318579	-0.07251
Hemoglobin	-0.0085	0.015148	0.024328113	-0.07173

Cholesterol	0.227759	0.027009	0.042102768	0.065325
HDL	-0.08464	0.035829	0.054255352	-0.06201
Neutrophils	0.001929	0.045682	0.067254532	0.059043
ALT	-0.00149	0.052744	0.07555229	-0.05723
Platelets	-0.30588	0.058009	0.080906828	-0.05601
Eosinophils	0.002526	0.071087	0.09660524	0.053337
AST	0.001594	0.084948	0.109811313	0.05091
SDMA	-0.00126	0.083844	0.109811313	-0.05109
Total bilirubin	0.001185	0.147456	0.186074961	0.042818
PAI	0.032517	0.265923	0.327765146	0.03289
LDL	0.081479	0.342078	0.412048548	0.02809
Total protein	0.001047	0.43205	0.508859405	0.023231
TAFI	0.034632	0.490267	0.564872475	0.020399
Monocytes	0.000538	0.501263	0.565253754	0.019885
White cells	-0.00273	0.530401	0.573699376	-0.01855
ADMA	0.000446	0.521458	0.573699376	0.018957
Apo A1	-0.0527	0.554935	0.588230573	-0.01746
Insulin	0.000512	0.757608	0.787317775	0.009126
GP7	0.00016	0.81052	0.826106591	0.00709
Total iron	-0.00044	0.995338	0.995338448	-0.00017

Table 20 shows that variables most significantly correlated with age ($R^2 \geq 0.1$, $p < 0.05$) are GP2, GP6, EGFR, GP1, WHR, Azotemia, HbA1c, GP3, GP4, Fasting glucose, Alkaline Phosphatase, Creatinine, Triglycerides, Fibrinogen, CRP, Uric Acid, RDW-CVM, Ferritin, Telomere Length, GP8, HOMA, Lymphocytes, Transferrin, GGT, MCV, GP10; and variables with larger effect sizes are Fibrinogen ($\beta=1.46$), Fasting glucose ($\beta=1.09$), EGFR ($\beta=-0.69$) and Transferrin ($\beta=-0.44$).

31. Statistical models

Firstly, different statistical methods were tested, one based on a score, and five other methods based on 49 standardized predictors (39 phenotypical, 10 N-glycans).

12. *Biological Age I: Score-Based*

On the one hand, 7 O2PLS models were conducted, in which we integrated Glycomics (10 variables) with Phenomics (39), Lipids (8), Liver (8), Kidney (8), Iron (3), Coagulation (3), and Blood (8). For each model, the number of joint and specific PCs were estimated with cross-validation. Then, the 46 PCs obtained in the 7 models were regressed to chronological age using an elastic net model, which selected 34 PCs.

On the other hand, a PCA was performed on the 49 predictors. Then, they were regressed to chronological age using an elastic net model, which selected all 49 PCs.

13. *Biological Age II: Based on 49 predictors*

Secondly, three penalized models were employed, namely ridge regression (RR), lasso (LASSO), and elastic net (EN), and two machine learning algorithms, that is random forest (RF) and support vector machine (SVM). Each one of the regression models were applied on 49 independent variables and Age as response variable on the training set (using 9/10 folds). For that, a grid search was used to estimate the hyperparameters of the model with cross-validation (**Table 21**), using *glmnet* and *caret* packages from R. The models were computed using the hyperparameters and compared through metrics of performance (**Table 22**).

Table 21. Best combination of hyperparameters calculated with cross-validation

Algorithm	Hyperparameters
O2PLS EN	$\alpha=0.9$ $\lambda_{(1SD)}=0$
PCA EN	$\alpha=0$ $\lambda_{(1SD)}=0.7$
RR	$\alpha=0$ $\lambda_{(1SD)}=2.98$
LASSO	$\alpha=1$ $\lambda_{(1SD)}=0.22$
EN	$\alpha=0.4$ $\lambda_{(1SD)}=0.1$
RF	mtry=49 (OOB=56.44) ntree=500
SVM	$\gamma=0.001$ cost=100 (polynomial kernel)

Overall, the models did not generalize well (they were fitted to non-predictive noise in the training set), specially the machine learning algorithms RF and SVM, which overfit, as evidenced by the high R^2 in the training set and the low R^2 (low accuracy) in the test set. The methods that best performed were O2PLS+EN and PCA+EN, specially the former displays the highest accuracy and lowest error in the test set; therefore, the two of them were chosen over the other methods for the downstream analysis.

Table 22. Metrics of model performance

Method	Metrics of model performance		
	R^2 training set	R^2 test set	MSEP test set
O2PLS+EN	0.44	0.31	93.85
PCA+EN	0.52	0.17	35662.2
RR	0.50	0.08	25535.62
LASSO	0.52	0.07	47782.64
EN	0.54	0.07	49126.24
RF	0.96	0.20	134.62
SVM	0.74	#N/D (sd=0)	842.20

32. Association between “Biological Age” and “Chronological Age”, separately in sexes:

Thirdly, the association between the biological ages obtained with elastic net of O2PLS and elastic net of PCA (the models displaying the best performances) and chronological age was tested, and the KDM on them. Additionally, to have a control, the KDM was applied on the original 49 predictor variables. Eventually, the association between biological and chronological ages of the 3 KDM models (i.e. O2PLS+EN+KDM; PCA+EN+KDM; KDM) was compared with the O2PLS+EN and PCA+EN models. The association was tested separately in males and females and separately in groups (**Table 23**). In the three groups (controls, prediabetics, and diabetics), the raw KDM on the 49 predictors best predicted chronological age; thus, KDM was taken for the downstream analysis.

Table 23. Summary output of the linear model: Predicted Age = Biological Age ~ Chronological Age

Group	Model	Sex	beta	p-value	R ²	
Controls	O2PLS+EN	Female	0.49	<2e-16	0.58	
		Male	0.38	5.318e-14	0.44	
	PCA+EN	Female	0.56	<2e-16	0.65	
		Male	0.47	6.84e-16	0.49	
	KDM	Female	1.07	<2e-16	0.80	
		Male	0.87	<2e-16	0.68	
	O2PLS+EN+KDM	Female	1.11	<2e-16	0.70	
		Male	0.83	<2e-16	0.56	
	PCA+EN+KDM	Female	1.53	2.67e-07	0.15	
		Male	0.76	0.07309	0.03	
	Prediabetics	O2PLS+EN	Female	0.44	<2e-16	0.44
			Male	0.42	6.38e-14	0.42
PCA+EN		Female	0.53	<2e-16	0.58	
		Male	0.47	<2e-16	0.48	
KDM		Female	1.03	<2e-16	0.73	
		Male	0.91	<2e-16	0.70	
O2PLS+EN+KDM		Female	1.07	<2e-16	0.62	
		Male	0.83	<2e-16	0.53	
PCA+EN+KDM		Female	1.20	0.0001	0.08	
		Male	0.75	0.07694	0.02	
Diabetics		O2PLS+EN	Female	0.32	3.855e-10	0.16
			Male	0.32	2.264e-15	0.21
	PCA+EN	Female	0.38	3.701e-16	0.25	
		Male	0.38	<2e-16	0.30	
	KDM	Female	0.81	<2e-16	0.40	
		Male	0.82	<2e-16	0.48	
	O2PLS+EN+KDM	Female	0.69	1.416e-13	0.21	
		Male	0.69	<2e-16	0.28	
	PCA+EN+KDM	Female	1.19	0.0001082	0.08	
		Male	0.75	0.076	0.02	

33. Association between “Biological Age” with disease (T2D) separately in sexes:

A multinomial logistic model was calculated to test the association of the Biological Age with T2D. Ctrl was the reference group to which “PreDiab” and “Diab” levels were compared in each sex (Table 24).

The models output indicates that Biological Age of the KDM can significantly discriminate Ctrl F from PreDiab F and Diab F, and Ctrl M from PreDiab M and Diab M.

Table 24. Summary output of the multinomial logistic model: Group (Ctrl vs PreDiab/Diab) ~ Biological Age

Model	Group	OR	p-value	CI (95%)
KDM	Diab F	1.09	5.71e-19	1.07-1.12
	PreDiab F	1.03	3.60e-4	1.01-1.05
	Diab M	1.10	5.45e-14	1.07-1.13
	PreDiab M	1.05	1.02e-3	1.02-1.07

1. Age acceleration in respect to controls calculated for both prediabetics and diabetics, separately in males and females, using the training data

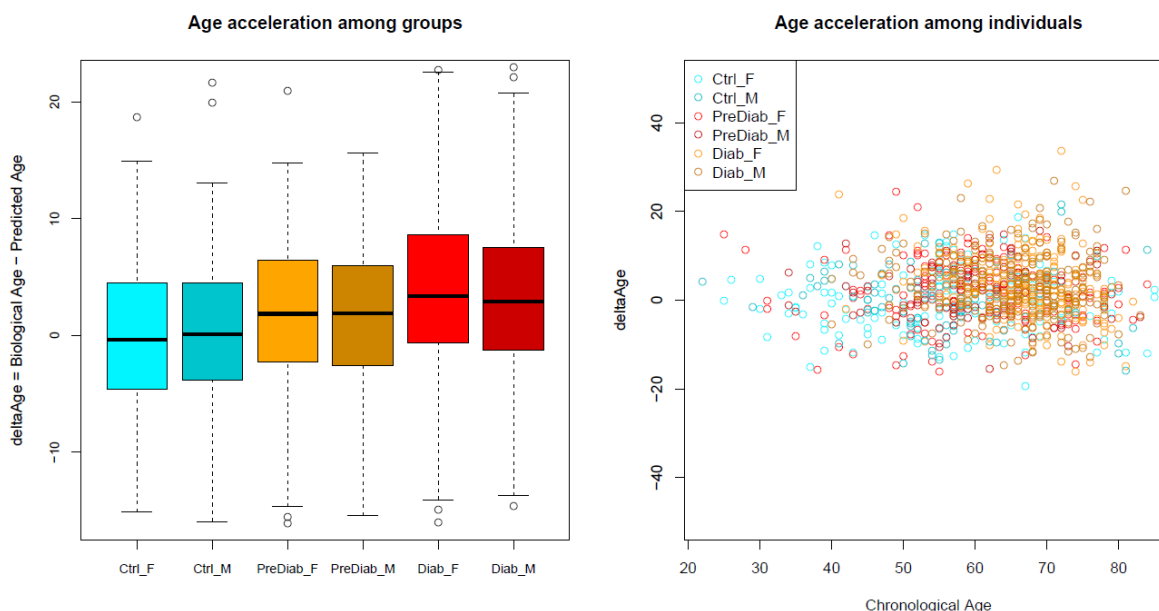


Figure 74. Age acceleration among controls, prediabetics, and diabetics in both sexes

In order to establish whether prediabetics and diabetics had an accelerated aging compared with controls, the delta age (also called Age Acceleration (AgeAccel)) was estimated by calculating the difference between “Biological Age” and “Predicted Age” (defined as the association between “Biological Age” and chronological age) and plotted (Figure 74). The mean values were as follows: 0 Ctrl F, 0 Ctrl M, 1.97 PreDiab F, 1.67 PreDiab M, 3.79 Diab F, 3.25 Diab M.

Pairwise comparisons using Wilcoxon’s rank sum test revealed no statistical differences between sexes in controls, prediabetics and diabetics (Table 25). It is interesting to note that, while control females show a younger trend than control males, this trend is reversed in prediabetics and

diabetics, so that despite there are no statistical differences, females are slightly biologically older than males, as the above-mentioned means highlight. It can also be observed that there are statistical differences, though small, between prediabetic females and control females and males, while prediabetic males showed no differences with controls.

Table 25. Pairwise comparisons using Wilcoxon’s rank sum test in KDM. P-value adjustment method: Benjamini-Hochberg. P<0.05 (*), p<0.01 (), p<0.001 (***) at $\alpha=0.05$.**

	Ctrl F	Ctrl M	PreDiab F	PreDiab M	Diab F
Ctrl M	0.95				
PreDiab F	0.03(*)	0.04 (*)			
PreDiab M	0.06	0.08	0.67		
Diab F	5.3e-05 (***)	0.0002 (***)	0.08	0.036 (*)	
Diab M	9.2e-05 (***)	0.0005 (***)	0.19	0.08	0.57

34. Associations with continuous traits or N-glycans: Trait/N-glycan ~ Phenotypical Age + Age + Sex + Group

Finally, we aimed to measure the association between continuous clinical traits or N-glycans with Biological Age, Age, Sex, and Group with multiple linear regression. Beta values and p-values for N-glycans and clinical traits are shown in **Table 26** and **Table 27**. It can be noted that only GP1, monocytes and RDW-CV were significantly associated with biological age, independently of chronological age, sex and disease status ($p=0$, 0.007, 5.31E-07, respectively). No variables were significantly associated with disease status independently of biological and chronological ages and sex. Platelets and Lymphocytes were significantly associated with Sex, independently of biological and chronological ages and disease status ($p=1.37E-36$ and 1.69E-17, respectively). No variables were significantly associated with chronological age, independently of biological age, sex and disease status.

Table 26. Association between continuous N-glycans species and Biological Age, Chronological Age, Sex and Disease Status: Trait ~ Biological Age + Chronological Age + Disease Status

Variable	β BioAge	p.value BioAge	β Age	p.value Age	β Sex	p.value Sex	β Group	p.value Group
GP1	1	0	-9.2E-17	0.159165	3.96E-16	0.62077	5.91E-16	0.263415
GP2	1.72E-17	2.61E-10	5.02E-18	0.116252	1.68E-16	2.03E-05	1	0
GP3	6.8E-16	9.4E-124	1	0	2.64E-16	0.467184	3.94E-16	0.09964
GP4	0.016255	3.46E-49	-0.00679	5.04E-08	-0.0095	0.532243	-0.03354	0.000852
GP5	0.021727	1.67E-73	-0.00992	6.89E-14	-0.05735	0.000374	0.012241	0.247958
GP6	-0.00938	2.32E-21	0.005436	2.32E-06	0.06289	8.88E-06	-0.07249	1.42E-14
GP7	-0.00789	6.48E-29	0.004164	3.42E-07	0.056447	2.04E-08	-0.01538	0.019611
GP8	-0.00529	7.16E-16	0.003622	2.43E-06	-0.00042	0.964004	0.034274	4.07E-08
GP9	-0.01466	5.8E-117	0.007028	3.19E-25	0.057524	2.53E-12	0.027785	2.48E-07
GP10	-0.00812	2.38E-14	0.007481	2.31E-09	0.000314	0.983589	0.01116	0.267888

Table 27. Association between continuous clinical traits and Biological Age, Chronological Age, Sex and Disease Status: Trait ~ Biological Age + Chronological Age + Disease Status

Variable	β BioAge	p.value BioAge	β Age	p.value Age	β Sex	p.value Sex	β Group	p.value Group
BMI	-0.00255	0.036798	-0.00178	0.215943	-0.18477	3.23E-24	0.036119	0.002062
WHR	0.012832	2.98E-10	-0.0091	0.000147	0.288206	8.78E-22	-0.02197	0.256802
Fasting Glucose	-0.00293	0.041663	-0.00015	0.930601	-0.21333	2.21E-23	0.021797	0.113734
HbA1c	0.193287	1.84E-22	-0.18753	8.05E-16	-0.55755	0.048287	0.929867	6.64E-07
Total Cholesterol	0.003127	6.3E-30	-0.00126	7.2E-05	0.074678	1.13E-70	0.015053	5.42E-09
HDL	0.665669	2.5E-05	-0.65931	0.000413	3.521481	0.124387	38.97394	1.1E-113
EGFR	0.014592	0.000471	-0.01388	0.004892	-0.10381	0.086912	1.061765	6.3E-119
Fibrinogen	0.66555	3.58E-05	-0.27081	0.153788	-18.7322	2.82E-15	-5.27745	0.00063
PAII	-0.47351	1.11E-15	0.438241	2.88E-10	-8.99828	4.35E-25	-1.64194	0.003341
TAFI	-1.40451	6.09E-63	0.694136	1.36E-13	7.794943	1.31E-11	-1.17456	0.117929
Total Bilirubin	3.836766	5.52E-32	-2.33836	5.18E-10	-31.6828	8.65E-12	2.431957	0.421341
LDL	0.15311	0.000843	-0.09494	0.07966	0.10186	0.878472	-1.03515	0.018553
Apo A1	0.393555	5.62E-07	-0.26769	0.003881	-5.00611	1.21E-05	-2.54188	0.000731
Apo B	-0.00586	3.67E-06	0.007381	8.49E-07	0.149039	1.24E-15	-0.03338	0.005847
Total Iron	0.295841	0.029586	-0.10339	0.52012	-9.25305	3.27E-06	-4.33327	0.000918
Transferrin	-0.65591	1.45E-06	0.715181	8.89E-06	-18.1319	1.97E-19	-3.94606	0.002448
Total Protein	0.617433	2.45E-07	-0.39015	0.005629	-4.40632	0.011037	-0.54345	0.633859
White Cells	-0.4812	6.42E-05	0.402626	0.004638	6.573219	0.000177	1.36713	0.235204
Red Cells	-0.85973	2.72E-06	0.162311	0.451898	-3.44471	0.194664	7.921404	6.66E-06
Hemoglobin	0.00825	9.28E-05	-0.00806	0.00123	0.085982	0.005079	0.017416	0.388358
MCV	0.043556	2.11E-10	-0.0547	1.66E-11	-0.00165	0.986661	0.292566	7.92E-06
Platelets	-0.00338	0.060643	-0.00068	0.749749	0.344655	1.37E-36	-0.00544	0.752802
MPV	-0.01324	0.005447	0.00132	0.814402	1.386223	1.37E-75	0.074143	0.104381
Neutrophils	0.01328	0.557997	0.05391	0.044669	0.619151	0.060944	-0.4569	0.035945
Lymphocytes	0.041548	0.866223	-0.1939	0.506538	-31.1282	1.69E-17	-2.1346	0.367369
Eosinophils	0.004856	0.242841	0.002843	0.563229	0.059328	0.327139	-0.03846	0.335169
Monocytes	0.004172	0.007286	-0.00269	0.142703	-0.0217	0.33705	0.010162	0.495213
Creatinine	-0.00538	9.37E-05	0.002004	0.21718	-0.07518	0.000177	-0.00078	0.95251
AST	0.0007	0.75515	0.000777	0.769745	0.059802	0.067417	0.029153	0.175994
HOMA	-0.00047	0.713474	0.002689	0.074996	0.067154	0.000312	-0.05542	6.62E-06
ALT	0.016365	1.35E-69	-0.01139	1.32E-27	0.178272	3.21E-42	0.008913	0.27968
Triglycerides	0.004887	0.00064	-0.00214	0.205616	0.089646	1.76E-05	-0.05405	8.49E-05
Insulin	0.009862	5.77E-19	-0.01012	9.25E-15	-0.00753	0.634585	0.118851	2.18E-28
CRP	0.003668	0.001948	-0.00718	3.31E-07	0.07932	4.49E-06	0.050828	8.15E-06
GGT	0.030034	3.74E-42	-0.02223	2.28E-18	0.023936	0.435827	0.094908	3.11E-06
Ferritin	0.021727	1.11E-16	-0.023	9.94E-14	-0.04607	0.219681	0.019387	0.433187
Alkaline Phosphatase	0.025082	1.27E-50	-0.01801	5.82E-21	-0.11135	1.63E-06	0.007909	0.603337
Azotemia	0.012378	7.71E-13	-0.00951	2.77E-06	0.080524	0.001225	0.024529	0.134333

Uric Acid	0.005185	3.64E-10	-0.00272	0.00516	0.11699	9.12E-22	0.01544	0.049822
Telomere Length	0.014925	4.11E-33	-0.00833	6.35E-09	-0.07923	6.65E-06	-0.02107	0.067996
SDMA	0.011547	2.13E-26	-0.00393	0.001714	0.033996	0.02731	-0.02745	0.00688
ADMA	0.013767	4.69E-39	-0.00948	5.35E-15	0.137505	5.12E-20	-0.0137	0.157582
RDW-CV	-0.00904	5.31E-07	0.003371	0.111945	-0.02992	0.251613	-0.02516	0.143508

15. DISCUSSION

The increase in geriatric population around the globe highlights the need for better understanding the processes underlying healthy and unhealthy aging, as age is a major risk factor for functional impairments, chronic diseases and mortality. In this context, the biomarkers of aging have come into play. There are two main ways of estimating biological age: 1) the traditional estimation of mortality curves (Moskalev, 2020) and 2) the estimation of chronological age (Horvath, 2013). The former has the disadvantage that it cannot be used to measure functional decline, predict morbidity onset and life expectancy, while the latter evidences the unsolved “paradox of biomarkers”, which uses a set of biological predictors to predict chronological age with linear regression, therefore resulting in biological age equal to chronological age (Hochschild, 1994). In addition, so far none of the studies carried out with the goal of defining new biomarkers of age has fulfilled the AFAR criteria (Johnson, 2006).

In this study, we made use of a novel approach: 39 clinical predictors and 10 N-glycans were used as biological predictors for chronological age, considering three groups of individuals, that is, controls, prediabetics, and diabetics, and both sexes. We tested different models, and the Klemera-Doubal one was the best in terms of performance. We applied this method on the 49 predictors, from which we derived our “biological age”. The calculated biological age was then associated with chronological age, T2D status, and with continuous clinical and N-glycan traits. Eventually, age acceleration was computed for each one of the six groups. Results showed that biological age was significantly associated with chronological age in the three groups in both sexes, that biological age can significantly predict (non-)prediabetic and (non-)diabetic status, and that biological age is significantly associated with GP1, monocytes and RDW-CV, independently of chronological age, sex and (non-)pre-/T2D status. Furthermore, results confirmed that prediabetic females are significantly older than control males and control females, and that diabetics are not significantly different than prediabetics, but they are significantly biologically older than controls, in both sexes.

Summarizing, we have identified the biological age in individuals with diverse phenotypes, that is, healthy, dysmetabolic, prediabetic, and diabetic, by using the chronological ages and clinical and N-glycan predictors in all of them at once, therefore without fulfilling the AFAR criterium by which a biomarker of age “must monitor a basic process that underlies the aging process, *not the effects of disease*”. The underlying reason for such design is that controls, with age, become inevitably less healthy, and especially around 60 years of age and above, it is difficult to find totally healthy controls, or, in other words, without subclinical or clinical symptoms of age-related diseases. Indeed, at least two review articles support this hypothesis. The first, from Wang and Bennett (2012), who stated that “*the effects of atherosclerosis are superimposed on normal aging of the underlying vessel*”. The second, from Franceschi et al. (2018), claims that precise boundaries

between aging and age-related diseases or geriatric syndromes do not exist, but instead they are a continuum.

Unfortunately, our design makes it incomparable with findings of previous studies. Besides, most of the aging biomarkers identified are epigenetic clocks (Jylhävä et al., 2017). Nonetheless, at least two previous studies also estimated a composite-based biomarker of age, and another one who estimated a glycomics-based biomarker. Levine (2013) focused on the methodology and compared different statistical methods, reaching the conclusion that KDM was the best, the same that in our study, though here we compared other statistical methods. In turn, Belsky et al. (2015) used 10 biomarkers for estimating biological age (Glycated haemoglobin, Forced expiratory volume in one second (FEV1), Blood pressure (systolic), Total cholesterol, C-reactive protein, Creatinine, Urea nitrogen, Albumin, Alkaline phosphatase, and Cytomegalovirus IgG.) of each Dunedin Study member's at age 38, that is, 7 of our 49 variables, with the KDM, and one of the conclusions they draw is that biological age can provide a summary of accumulated aging in cases when only cross-sectional data are available (Belsky et al. (2015)), a conclusion that we also draw in our study. Eventually, Krištić et al. (2014) used 24 UPLC N-glycans and estimated both chronological and biological ages by building predictive models. They demonstrated that 3 N-glycans (FA2B, FA2G2, FA2BG2, which correspond to GP2, GP6, and GP7 in DSA-FACE glycans, respectively) changed considerably with age; from these, GP2 and GP6 are the top 2 out of 49 variables most significantly associated with chronological age (**Table 20**) and biological age (**Table 27**), while GP7, in our study, has a Pearson's correlation of 0.007 and $p=0.8$ (**Table 20**).

Together, our findings constitute proof of principle for the measures of biological age in individuals with age-related diseases. Here we identify several future directions that can build on this proof-of-principle for estimating accelerated age from younger to older ages.

First, our analysis was limited to a single cohort, and one that lacked ethnic minority populations. Replication in other cohorts is needed, and replication using a larger number of younger and older healthy controls, younger and older less healthy controls, prediabetics and diabetics can also provide additional perspectives. Larger samples can also help understanding the less common phenotypes.

Second, DSA-FACE N-glycans are low-dimensional and have the disadvantage that sialic acids are cut-off from the sugar, which blurs the interpretation of results. Further studies are needed that use N-glycans obtained with other technologies which include sialic acids and produce more N-glycan species, and not only from whole plasma glycome, but also from IgG glycome, which is especially relevant for understanding the role of inflamm-aging and metaflammation played by N-glycans.

Third, the nature of the cross-sectional makes available just a snapshot of the phenotype. Therefore, further studies comprising multiple biomarkers sampled at different points of time are required to measure the pace of aging, since summarizing the data obtained at one single point cannot measure the functional decline nor predict the morbidity onset and life expectancy.

Within the bounds of these limitations, the implication of the present work is that it is possible to identify a composite based biomarker of aging that not only considers pure healthy controls, but also prediabetics and diabetics.

CONCLUSIONS

In this thesis, we aimed to conduct an in-depth analysis of metabolic phenotype by studying age- and sex-related features of type 2 diabetes. On the one hand, it was first tested whether 10 N-glycans could predict better the diabetic status than HbA1c, and results showed that N-glycans cannot compete with the gold standard HbA1c. Secondly, an extensive characterization of the T2D/(dys)metabolic/aging phenotype was performed. Based on the integration of a set or subsets (part I and part II, respectively) of 44 clinical variables and 10 N-glycans, it can be concluded that:

- i. N-glycans GP3, GP4, GP6 are negatively associated with age in the three groups of individuals, that is controls, prediabetics, and diabetics, in both sexes.
- ii. N-glycans GP8 and GP10 are positively associated with a metabolism dysregulation, as revealed by prediabetic males and diabetics.
- iii. The observed N-glycans in control males are mainly correlated with coagulation and blood parameters; in control and prediabetic females, with age, alkaline phosphatase and lipid parameters; in prediabetic males, with coagulation parameters; and in diabetics, with diverse parameters from the phenomics data set, including lipid, coagulation, iron, blood parameters and factors used for the T2D diagnosis.
- iv. The decreasing intra-correlations between pairs of joint PCs from controls ($R^2=0.7\%$), prediabetics ($R^2=0.6\%$), and diabetics ($R^2=0.5\%$) in the integration phenomics/glycomics (part I) underlines the heterogeneity and/or complexity of disease and age, or, in other words, the higher homogeneity of controls.
- v. Males and females present different biological dynamics, as noted by the biochemical and glycomics data, especially at 20-44 and at 45-55 years of age, while at 56-70 and 71-85 tend to progressively converge.
- vi. Control, prediabetic, and diabetic individuals, generally, present different quantities of biochemical and glycomics parameters.
- vii. Inter-correlations between individual joint PCs scores (part II) show that in prediabetics and diabetics N-glycans are generally the same when integrated with different endophenotypes, instead of revealing a specific signature for each endophenotype.

On the other hand, we computed a composite biomarker of aging, by building on a predictive model of chronological age composed of 49 predictors, based on the Klemra-Doubal method. Over recent years, this method has become increasingly popular for measuring biological age. In this study, this approach was used on 1146 individuals, including controls, prediabetics, and diabetics of both sexes. On the whole, it demonstrated that biological age: a) is significantly associated with chronological age in the three groups in both sexes, b) can significantly predict (non-)prediabetic and (non-)diabetic status, c) is significantly associated with GPI, monocytes and RDW-CV, independently of chronological age, sex and (non-)pre-/T2D status. In addition, age acceleration, computed as the difference between biological age and predicted age, confirmed that prediabetic females are significantly older than control males and control females, and that diabetics are not significantly different than prediabetics, but they are significantly biologically older than controls, in both sexes.

Based on these conclusions, to better understand the implications of these results, future studies should include longitudinal sampling of clinical parameters and N-glycans species, glycomics technologies that are higher dimensional and that include sialic acids, larger cohorts, and cohorts of ethnic minorities. In particular, repetitive measures as well as mortality data could contribute to answer the question of whether N-glycans change over the course of preclinical (that is, insulin resistance and impaired insulin secretion) to clinical diagnoses (obesity and/or T2D, T2D complications, metabolic syndrome). Further research will be needed to address this question.

It is important to emphasize that this work has provided a deeper understanding of N-glycan biomarkers and has confirmed the most important clinical factors involved both in aging and T2D/metabolism.

BIBLIOGRAPHY

2015. *IDF Diabetes Atlas*. 1st ed. Brussels: International Diabetes Federation, pp.p. 350, 362-367.
- Abdul-Ghani M, DeFronzo RA, Del Prato S, Chilton R, Singh R, Ryder REJ. Cardiovascular Disease and Type 2 Diabetes: Has the Dawn of a New Era Arrived? [published correction appears in *Diabetes Care*. 2017 Sep 8;:]. *Diabetes Care*. 2017;40(7):813-820. doi:10.2337/dc16-2736
- Acharjee et al., *Metabolomics* 2013, 3:3. <http://dx.doi.org/10.4172/2153-0769.1000126>
- Action to Control Cardiovascular Risk in Diabetes Study Group, Gerstein HC, Miller ME, et al. Effects of intensive glucose lowering in type 2 diabetes. *N Engl J Med*. 2008;358(24):2545-2559. doi:10.1056/NEJMoa0802743
- Adua E, Memarian E, Russell A, et al. High throughput profiling of whole plasma N-glycans in type II diabetes mellitus patients and healthy individuals: A perspective from a Ghanaian population. *Arch Biochem Biophys*. 2019;661:10-21. doi:10.1016/j.abb.2018.10.015
- ADVANCE Collaborative Group, Patel A, MacMahon S, et al. Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes. *N Engl J Med*. 2008;358(24):2560-2572. doi:10.1056/NEJMoa0802987
- American Diabetes Association. Standards of medical care in diabetes--2008. *Diabetes Care*. 2008;31 Suppl 1:S12-S54. doi:10.2337/dc08-S012
- American Diabetes Association. Standards of medical care in diabetes--2010 [published correction appears in *Diabetes Care*. 2010 Mar;33(3):692]. *Diabetes Care*. 2010;33 Suppl 1(Suppl 1):S11-S61. doi:10.2337/dc10-S011
- Apweiler R, Hermjakob H, Sharon N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta*. 1999;1473(1):4-8. doi:10.1016/s0304-4165(99)00165-8
- Ashby DT, Rye KA, Clay MA, Vadas MA, Gamble JR, Barter PJ. Factors influencing the ability of HDL to inhibit expression of vascular cell adhesion molecule-1 in endothelial cells. *Arterioscler Thromb Vasc Biol*. 1998;18(9):1450-1455. doi:10.1161/01.atv.18.9.1450
- Bai J, Wang MX, Chowbay B, Ching CB, Chen WN. Metabolic profiling of HepG2 cells incubated with S(-) and R(+) enantiomers of anti-coagulating drug warfarin. *Metabolomics*. 2011;7(3):353-362. doi:10.1007/s11306-010-0262-3
- Belsky DW, Caspi A, Houts R, et al. Quantification of biological aging in young adults. *Proc Natl Acad Sci U S A*. 2015;112(30):E4104-E4110. doi:10.1073/pnas.1506264112
- Bertozi CR, Sasisekharan R. Glycomics. In: Varki A, Cummings RD, Esko JD, et al., eds. *Essentials of Glycobiology*. 2nd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009.
- Bouhaddani SE, Houwing-Duistermaat J, Salo P, Perola M, Jongbloed G, Uh HW. Evaluation of O2PLS in Omics data integration. *BMC Bioinformatics*. 2016;17 Suppl 2(Suppl 2):11. Published 2016 Jan 20. doi:10.1186/s12859-015-0854-z

- Buse JB, Ginsberg HN, Bakris GL, et al. Primary prevention of cardiovascular diseases in people with diabetes mellitus: a scientific statement from the American Heart Association and the American Diabetes Association. *Diabetes Care*. 2007;30(1):162-172. doi:10.2337/dc07-9917
- Catrysse L, van Loo G. Inflammation and the Metabolic Syndrome: The Tissue-Specific Functions of NF- κ B. *Trends Cell Biol*. 2017;27(6):417-429. doi:10.1016/j.tcb.2017.01.006
- Chan DC, Watts GF. Apolipoproteins as markers and managers of coronary risk [published correction appears in *QJM*. 2006 Nov;99(11):807]. *QJM*. 2006;99(5):277-287. doi:10.1093/qjmed/hcl027
- Chatterjee S, Khunti K, Davies MJ. Type 2 diabetes [published correction appears in *Lancet*. 2017 Jun 3;389(10085):2192]. *Lancet*. 2017;389(10085):2239-2251. doi:10.1016/S0140-6736(17)30058-2
- Chen BH, Marioni RE, Colicino E, et al. DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging (Albany NY)*. 2016;8(9):1844-1865. doi:10.18632/aging.101020
- Cohen M, Varki A. The sialome--far more than the sum of its parts. *OMICS*. 2010;14(4):455-464. doi:10.1089/omi.2009.0148
- Collino S, Montoliu I, Martin FP, et al. Metabolic signatures of extreme longevity in northern Italian centenarians reveal a complex remodeling of lipids, amino acids, and gut microbiota metabolism [published correction appears in *PLoS One*. 2013;8(8). doi:10.1371/annotation/5fb9fa6f-4889-4407-8430-6dfc7ecdcbdd]. *PLoS One*. 2013;8(3):e56564. doi:10.1371/journal.pone.0056564
- Copeland RJ, Bullen JW, Hart GW. Cross-talk between GlcNAcylation and phosphorylation: roles in insulin resistance and glucose toxicity. *Am J Physiol Endocrinol Metab*. 2008;295(1):E17-E28. doi:10.1152/ajpendo.90281.2008
- Czernichow S, Kengne AP, Stamatakis E, Hamer M, Batty GD. Body mass index, waist circumference and waist-hip ratio: which is the better discriminator of cardiovascular disease mortality risk?: evidence from an individual-participant meta-analysis of 82 864 participants from nine cohort studies. *Obes Rev*. 2011;12(9):680-687. doi:10.1111/j.1467-789X.2011.00879.x
- DeFronzo RA, Ferrannini E, Groop L, et al. Type 2 diabetes mellitus. *Nat Rev Dis Primers*. 2015;1:15019. Published 2015 Jul 23. doi:10.1038/nrdp.2015.19
- DeFronzo RA. Banting Lecture. From the triumvirate to the ominous octet: a new paradigm for the treatment of type 2 diabetes mellitus. *Diabetes*. 2009;58(4):773-795. doi:10.2337/db09-9028
- DeFronzo RA. Insulin resistance, lipotoxicity, type 2 diabetes and atherosclerosis: the missing links. The Claude Bernard Lecture 2009. *Diabetologia*. 2010;53(7):1270-1287. doi:10.1007/s00125-010-1684-1
- Diabetes Association. 2. Classification and Diagnosis of Diabetes: *Standards of Medical Care in Diabetes-2020*. *Diabetes Care*. 2020;43(Suppl 1):S14-S31. doi:10.2337/dc20-S002

Dotz V, Lemmers RFH, Reiding KR, et al. Plasma protein N-glycan signatures of type 2 diabetes. *Biochim Biophys Acta Gen Subj*. 2018;1862(12):2613-2622. doi:10.1016/j.bbagen.2018.08.005

Duckworth W, Abraira C, Moritz T, et al. Glucose control and vascular complications in veterans with type 2 diabetes [published correction appears in *N Engl J Med*. 2009 Sep 3;361(10):1028] [published correction appears in *N Engl J Med*. 2009 Sep 3;361(10):1024-5]. *N Engl J Med*. 2009;360(2):129-139. doi:10.1056/NEJMoa0808431

Durso DF, Bacalini MG, Sala C, et al. Acceleration of leukocytes' epigenetic age as an early tumor and sex-specific marker of breast and colorectal cancer. *Oncotarget*. 2017;8(14):23237-23245. doi:10.18632/oncotarget.15573

Ercan A, Cui J, Chatterton DE, et al. Aberrant IgG galactosylation precedes disease onset, correlates with disease activity, and is prevalent in autoantibodies in rheumatoid arthritis. *Arthritis Rheum*. 2010;62(8):2239-2248. doi:10.1002/art.27533

Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, And Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III). *JAMA*. 2001;285(19):2486-2497. doi:10.1001/jama.285.19.2486

Femlak M, Gluba-Brzózka A, Ciałkowska-Rysz A, Rysz J. The role and function of HDL in patients with diabetes mellitus and the related cardiovascular risk. *Lipids Health Dis*. 2017;16(1):207. Published 2017 Oct 30. doi:10.1186/s12944-017-0594-3

Ference BA, Ginsberg HN, Graham I, et al. Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement from the European Atherosclerosis Society Consensus Panel. *Eur Heart J*. 2017;38(32):2459-2472. doi:10.1093/eurheartj/ehx144

Ferrucci L, Bandinelli S, Benvenuti E, et al. Subsystems contributing to the decline in ability to walk: bridging the gap between epidemiology and geriatric practice in the InCHIANTI study. *J Am Geriatr Soc*. 2000;48(12):1618-1625. doi:10.1111/j.1532-5415.2000.tb03873.x

Festa A, D'Agostino R Jr, Hanley AJ, Karter AJ, Saad MF, Haffner SM. Differences in insulin resistance in nondiabetic subjects with isolated impaired glucose tolerance or isolated impaired fasting glucose. *Diabetes*. 2004;53(6):1549-1555. doi:10.2337/diabetes.53.6.1549

Fleischer JG, Schulte R, Tsai HH, et al. Predicting age from the transcriptome of human dermal fibroblasts. *Genome Biol*. 2018;19(1):221. Published 2018 Dec 20. doi:10.1186/s13059-018-1599-6

Franceschi C, Bonafè M, Valensin S, et al. Inflamm-aging. An evolutionary perspective on immunosenescence. *Ann N Y Acad Sci*. 2000;908:244-254. doi:10.1111/j.1749-6632.2000.tb06651.x

Franceschi C, Garagnani P, Vitale G, Capri M, Salvioli S. Inflammaging and 'Garb-aging'. *Trends Endocrinol Metab*. 2017;28(3):199-212. doi:10.1016/j.tem.2016.09.005

Fuchsberger C, Flannick J, Teslovich TM, et al. The genetic architecture of type 2 diabetes. *Nature*. 2016;536(7614):41-47. doi:10.1038/nature18642

- Furman D, Campisi J, Verdin E, et al. Chronic inflammation in the etiology of disease across the life span. *Nat Med*. 2019;25(12):1822-1832. doi:10.1038/s41591-019-0675-0
- Garagnani P, Bacalini MG, Pirazzini C, et al. Methylation of ELOVL2 gene as a new epigenetic marker of age. *Aging Cell*. 2012;11(6):1132-1134. doi:10.1111/accel.12005
- Garagnani P, Giuliani C, Pirazzini C, et al. Centenarians as super-controls to assess the biological relevance of genetic risk factors for common age-related diseases: a proof of principle on type 2 diabetes. *Aging (Albany NY)*. 2013;5(5):373-385. doi:10.18632/aging.100562
- Gillett MJ. International Expert Committee report on the role of the A1c assay in the diagnosis of diabetes: Diabetes Care 2009; 32(7): 1327-1334. *Clin Biochem Rev*. 2009;30(4):197-200.
- Gloyn AL. The search for type 2 diabetes genes. *Ageing Res Rev*. 2003;2(2):111-127. doi:10.1016/s1568-1637(02)00061-2
- Goldberg IJ. Clinical review 124: Diabetic dyslipidemia: causes and consequences. *J Clin Endocrinol Metab*. 2001;86(3):965-971. doi:10.1210/jcem.86.3.7304
- Grant PJ. Diabetes mellitus as a prothrombotic condition. *J Intern Med*. 2007;262(2):157-172. doi:10.1111/j.1365-2796.2007.01824.x
- Gregor MF, Hotamisligil GS. Inflammatory mechanisms in obesity. *Annu Rev Immunol*. 2011;29:415-445. doi:10.1146/annurev-immunol-031210-101322
- Gregor MF, Hotamisligil GS. Inflammatory mechanisms in obesity. *Annu Rev Immunol*. 2011;29:415-445. doi:10.1146/annurev-immunol-031210-101322
- Grundey SM, Brewer HB Jr, Cleeman JI, et al. Definition of metabolic syndrome: Report of the National Heart, Lung, and Blood Institute/American Heart Association conference on scientific issues related to definition. *Circulation*. 2004;109(3):433-438. doi:10.1161/01.CIR.0000111245.75752.C6
- Hannum G, Guinney J, Zhao L, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell*. 2013;49(2):359-367. doi:10.1016/j.molcel.2012.10.016
- Hart GW, Copeland RJ. Glycomics hits the big time. *Cell*. 2010;143(5):672-676. doi:10.1016/j.cell.2010.11.008
- Hegele RA. Plasma lipoproteins: genetic influences and clinical implications. *Nat Rev Genet*. 2009;10(2):109-121. doi:10.1038/nrg2481
- Higel F, Seidl A, Sörgel F, Friess W. N-glycosylation heterogeneity and the influence on structure, function and pharmacokinetics of monoclonal antibodies and Fc fusion proteins. *Eur J Pharm Biopharm*. 2016;100:94-100. doi:10.1016/j.ejpb.2016.01.005
- Holly AC, Melzer D, Pilling LC, et al. Towards a gene expression biomarker set for human biological age. *Aging Cell*. 2013;12(2):324-326. doi:10.1111/accel.12044
- Holman N, Knighton P, Kar P, et al. Risk factors for COVID-19-related mortality in people with type 1 and type 2 diabetes in England: a population-based cohort study. *Lancet Diabetes Endocrinol*. 2020;8(10):823-833. doi:10.1016/S2213-8587(20)30271-0

Holman RR, Paul SK, Bethel MA, Matthews DR, Neil HA. 10-year follow-up of intensive glucose control in type 2 diabetes. *N Engl J Med*. 2008;359(15):1577-1589. doi:10.1056/NEJMoa0806470

Hori Y, Gabazza EC, Yano Y, et al. Insulin resistance is associated with increased circulating level of thrombin-activatable fibrinolysis inhibitor in type 2 diabetic patients. *J Clin Endocrinol Metab*. 2002;87(2):660-665. doi:10.1210/jcem.87.2.8214

Horikawa Y, Oda N, Cox NJ, et al. Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus [published correction appears in *Nat Genet* 2000 Dec;26(4):502]. *Nat Genet*. 2000;26(2):163-175. doi:10.1038/79876

Horvath S, Oshima J, Martin GM, et al. Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and *ex vivo* studies. *Aging (Albany NY)*. 2018;10(7):1758-1775. doi:10.18632/aging.101508

Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat Rev Genet*. 2018;19(6):371-384. doi:10.1038/s41576-018-0004-3

Horvath S, Ritz BR. Increased epigenetic age and granulocyte counts in the blood of Parkinson's disease patients. *Aging (Albany NY)*. 2015;7(12):1130-1142. doi:10.18632/aging.100859

Horvath S. DNA methylation age of human tissues and cell types [published correction appears in *Genome Biol*. 2015;16:96]. *Genome Biol*. 2013;14(10):R115. doi:10.1186/gb-2013-14-10-r115

Horvath S. Erratum to: DNA methylation age of human tissues and cell types. *Genome Biol*. 2015;16(1):96. Published 2015 May 13. doi:10.1186/s13059-015-0649-6

Hotamisligil GS. Inflammation and metabolic disorders. *Nature*. 2006;444(7121):860-867. doi:10.1038/nature05485

Hotamisligil GS. Inflammation, metaflammation and immunometabolic disorders. *Nature*. 2017;542(7640):177-185. doi:10.1038/nature21363

Houtkooper RH, Argmann C, Houten SM, et al. The metabolic footprint of aging in mice. *Sci Rep*. 2011;1:134. doi:10.1038/srep00134

Hu FB, Manson JE, Stampfer MJ, et al. Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. *N Engl J Med*. 2001;345(11):790-797. doi:10.1056/NEJMoa010492

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 [published correction appears in *Lancet* 1999 Aug 14;354(9178):602]. *Lancet*. 1998;352(9131):837-853.

Issad T, Masson E, Pagesy P. O-GlcNAc modification, insulin signaling and diabetic complications. *Diabetes Metab*. 2010;36(6 Pt 1):423-435. doi:10.1016/j.diabet.2010.09.001

IUPAC. *Compendium of Chemical Terminology, 2nd ed. (the "Gold Book")*. Compiled by A. D. McNaught and A. Wilkinson. Blackwell Scientific Publications, Oxford (1997). Online version (2019-) created by S. J. Chalk. ISBN 0-9678550-9-8. <https://doi.org/10.1351/goldbook>.

Jeanine J Houwing-Duistermaat, Hae Won Uh, Arief Gusnanto. Discussion on the paper 'Statistical Houwing-Duistermaat JJ, Uh HW, Gusnanto A. Discussion on the paper 'Statistical contributions to

bioinformatics: Design, modelling, structure learning and integration' by Jeffrey S. Morris and Veerabhadran Baladandayuthapani. *Statistical Modelling*. 2017;17(4-5):319-326. doi:[10.1177/1471082X17706135](https://doi.org/10.1177/1471082X17706135)

Jennewein MF, Alter G. The Immunoregulatory Roles of Antibody Glycosylation. *Trends Immunol*. 2017;38(5):358-372. doi:10.1016/j.it.2017.02.004

Johnson AA, Shokhirev MN, Wyss-Coray T, Lehallier B. Systematic review and analysis of human proteomics aging studies unveils a novel proteomic aging clock and identifies key processes that change with age. *Ageing Res Rev*. 2020;60:101070. doi:10.1016/j.arr.2020.101070

Johnson JL, Jones MB, Ryan SO, Cobb BA. The regulatory power of glycans and their binding partners in immunity. *Trends Immunol*. 2013;34(6):290-298. doi:10.1016/j.it.2013.01.006

Johnson TE. Recent results: biomarkers of aging. *Exp Gerontol*. 2006 Dec;41(12):1243-6. doi:10.1016/j.exger.2006.09.006. Epub 2006 Oct 30. PMID: 17071038.

Juutilainen A, Kortelainen S, Lehto S, Rönnemaa T, Pyörälä K, Laakso M. Gender difference in the impact of type 2 diabetes on coronary heart disease risk. *Diabetes Care*. 2004;27(12):2898-2904. doi:10.2337/diacare.27.12.2898

Jylhävä J, Pedersen NL, Hägg S. Biological Age Predictors. *EBioMedicine*. 2017;21:29-36. doi:10.1016/j.ebiom.2017.03.046

Kennedy BK, Berger SL, Brunet A, et al. Geroscience: linking aging to chronic disease. *Cell*. 2014;159(4):709-713. doi:10.1016/j.cell.2014.10.039

Keser T, Gornik I, Vučković F, et al. Increased plasma N-glycome complexity is associated with higher risk of type 2 diabetes [published correction appears in *Diabetologia*. 2017 Nov 29;:]. *Diabetologia*. 2017;60(12):2352-2360. doi:10.1007/s00125-017-4426-9

Klemra P, Doubal S. A new approach to the concept and computation of biological age. *Mech Ageing Dev*. 2006;127(3):240-248. doi:10.1016/j.mad.2005.10.004

Knezevic A, Gornik O, Polasek O, et al. Effects of aging, body mass index, plasma lipid profiles, and smoking on human plasma N-glycans. *Glycobiology*. 2010;20(8):959-969. doi:10.1093/glycob/cwq051

Krentz AJ. Lipoprotein abnormalities and their consequences for patients with type 2 diabetes. *Diabetes Obes Metab*. 2003;5 Suppl 1:S19-S27. doi:10.1046/j.1462-8902.2003.0310.x

Krištić J, Vučković F, Menni C, et al. Glycans are a novel biomarker of chronological and biological ages. *J Gerontol A Biol Sci Med Sci*. 2014;69(7):779-789. doi:10.1093/gerona/glt190

Laakso M, Sarlund H, Salonen R, et al. Asymptomatic atherosclerosis and insulin resistance. *Arterioscler Thromb*. 1991;11(4):1068-1076. doi:10.1161/01.atv.11.4.1068

Laakso M. Cardiovascular disease in type 2 diabetes from population to man to mechanisms: the Kelly West Award Lecture 2008. *Diabetes Care*. 2010;33(2):442-449. doi:10.2337/dc09-0749

Laakso M. Insulin resistance and coronary heart disease. *Curr Opin Lipidol*. 1996;7(4):217-226. doi:10.1097/00041433-199608000-00008

- Lauc G, Zoldoš V. Protein glycosylation--an evolutionary crossroad between genes and environment. *Mol Biosyst.* 2010;6(12):2373-2379. doi:10.1039/c0mb00067a
- Lee YS, Wollam J, Olefsky JM. An Integrated View of Immunometabolism. *Cell.* 2018;172(1-2):22-40. doi:10.1016/j.cell.2017.12.025
- Lehallier B, Gate D, Schaum N, et al. Undulating changes in human plasma proteome profiles across the lifespan. *Nat Med.* 2019;25(12):1843-1850. doi:10.1038/s41591-019-0673-2
- Lemmers RFH, Vilaj M, Urda D, et al. IgG glycan patterns are associated with type 2 diabetes in independent European populations. *Biochim Biophys Acta Gen Subj.* 2017;1861(9):2240-2249. doi:10.1016/j.bbagen.2017.06.020
- Levine ME, Lu AT, Bennett DA, Horvath S. Epigenetic age of the pre-frontal cortex is associated with neuritic plaques, amyloid load, and Alzheimer's disease related cognitive functioning. *Aging (Albany NY).* 2015;7(12):1198-1211. doi:10.18632/aging.100864
- Levine ME, Lu AT, Quach A, et al. An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY).* 2018;10(4):573-591. doi:10.18632/aging.101414
- Levine ME. Modeling the rate of senescence: can estimated biological age predict mortality more accurately than chronological age?. *J Gerontol A Biol Sci Med Sci.* 2013;68(6):667-674. doi:10.1093/gerona/gls233
- Li X, Wang H, Russell A, et al. Type 2 Diabetes Mellitus is Associated with the Immunoglobulin G N-Glycome through Putative Proinflammatory Mechanisms in an Australian Population. *OMICS.* 2019;23(12):631-639. doi:10.1089/omi.2019.0075
- Liu J, Dolikun M, Štambuk J, et al. Glycomics for Type 2 Diabetes Biomarker Discovery: Promise of Immunoglobulin G Subclass-Specific Fragment Crystallizable N-glycosylation in the Uyghur Population. *OMICS.* 2019;23(12):640-648. doi:10.1089/omi.2019.0052
- López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. *Cell.* 2013;153(6):1194-1217. doi:10.1016/j.cell.2013.05.039
- Marcovina S, Packard CJ. Measurement and meaning of apolipoprotein AI and apolipoprotein B plasma levels. *J Intern Med.* 2006;259(5):437-446. doi:10.1111/j.1365-2796.2006.01648.x
- Marioni RE, Shah S, McRae AF, et al. The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936. *Int J Epidemiol.* 2015;44(4):1388-1396. doi:10.1093/ije/dyu277
- Marth JD, Grewal PK. Mammalian glycosylation in immunity. *Nat Rev Immunol.* 2008;8(11):874-887. doi:10.1038/nri2417
- Martín-Timón I, Sevillano-Collantes C, Segura-Galindo A, Del Cañizo-Gómez FJ. Type 2 diabetes and cardiovascular disease: Have all risk factors the same strength?. *World J Diabetes.* 2014;5(4):444-470. doi:10.4239/wjd.v5.i4.444
- Moran AP, Gupta A, Joshi L. Sweet-talk: role of host glycosylation in bacterial pathogenesis of the gastrointestinal tract. *Gut.* 2011;60(10):1412-1425. doi:10.1136/gut.2010.212704

Morrish NJ, Wang SL, Stevens LK, Fuller JH, Keen H. Mortality and causes of death in the WHO Multinational Study of Vascular Disease in Diabetes. *Diabetologia*. 2001;44 Suppl 2:S14-S21. doi:10.1007/p100002934

Moskalev A. The challenges of estimating biological age. *Elife*. 2020;9:e54969. Published 2020 Feb 11. doi:10.7554/eLife.54969

Murray S. Is waist-to-hip ratio a better marker of cardiovascular risk than body mass index?. *CMAJ*. 2006;174(3):308. doi:10.1503/cmaj.051561

National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. *Circulation*. 2002;106(25):3143-3421.

Navab M, Imes SS, Hama SY, et al. Monocyte transmigration induced by modification of low density lipoprotein in cocultures of human aortic wall cells is due to induction of monocyte chemotactic protein 1 synthesis and is abolished by high density lipoprotein. *J Clin Invest*. 1991;88(6):2039-2046. doi:10.1172/JCI115532

Norata GD, Pirillo A, Catapano AL. Modified HDL: biological and physiopathological consequences. *Nutr Metab Cardiovasc Dis*. 2006;16(5):371-386. doi:10.1016/j.numecd.2006.01.012

Obi Tayo B. What is the point of univariate regression before multivariate regression? *Quora*. 2019. <https://www.quora.com/What-is-the-point-of-univariate-regression-before-multivariate-regression> Accessed February 15th, 2020

Ohtsubo K, Chen MZ, Olefsky JM, Marth JD. Pathway to diabetes through attenuation of pancreatic beta cell glycosylation and glucose transport. *Nat Med*. 2011;17(9):1067-1075. Published 2011 Aug 14. doi:10.1038/nm.2414

Orchard TJ, Dorman JS, Maser RE, et al. Factors associated with avoidance of severe complications after 25 yr of IDDM. Pittsburgh Epidemiology of Diabetes Complications Study I. *Diabetes Care*. 1990;13(7):741-747. doi:10.2337/diacare.13.7.741

Packard CJ. Triacylglycerol-rich lipoproteins and the generation of small, dense low-density lipoprotein. *Biochem Soc Trans*. 2003;31(Pt 5):1066-1069. doi:10.1042/bst0311066

Peters MJ, Joehanes R, Pilling LC, et al. The transcriptional landscape of age in human peripheral blood. *Nat Commun*. 2015;6:8570. Published 2015 Oct 22. doi:10.1038/ncomms9570

Prattichizzo F, De Nigris V, Spiga R, et al. Inflammageing and metaflammation: The yin and yang of type 2 diabetes. *Ageing Res Rev*. 2018;41:1-17. doi:10.1016/j.arr.2017.10.003

Qiao Q, Nyamdorj R. Is the association of type II diabetes with waist circumference or waist-to-hip ratio stronger than that with body mass index?. *Eur J Clin Nutr*. 2010;64(1):30-34. doi:10.1038/ejcn.2009.93

Remondini D, Intrator N, Sala C, et al. Identification of a T cell gene expression clock obtained by exploiting a MZ twin design. *Sci Rep.* 2017;7(1):6005. Published 2017 Jul 20. doi:10.1038/s41598-017-05694-2

Rhoades, R. A. & Pflanzner, R. G. *Human Physiology 4th* (Thomson Learning, 2002).

Ridker PM, Rifai N, Cook NR, Bradwin G, Buring JE. Non-HDL cholesterol, apolipoproteins A-I and B100, standard lipid measures, lipid ratios, and CRP as risk factors for cardiovascular disease in women. *JAMA.* 2005;294(3):326-333. doi:10.1001/jama.294.3.326

Ruhaak LR, Uh HW, Beekman M, et al. Decreased levels of bisecting GlcNAc glycoforms of IgG are associated with human longevity. *PLoS One.* 2010;5(9):e12566. Published 2010 Sep 7. doi:10.1371/journal.pone.0012566

Ruhaak LR, Uh HW, Beekman M, et al. Plasma protein N-glycan profiles are associated with calendar age, familial longevity and health. *J Proteome Res.* 2011;10(4):1667-1674. doi:10.1021/pr1009959

Sahin E, Colla S, Liesa M, et al. Telomere dysfunction induces metabolic and mitochondrial compromise [published correction appears in *Nature.* 2011 Jul 14;475(7355):254]. *Nature.* 2011;470(7334):359-365. doi:10.1038/nature09787

Salvioli S, Monti D, Lanzarini C, et al. Immune system, cell senescence, aging and longevity--inflamm-aging reappraised. *Curr Pharm Des.* 2013;19(9):1675-1679.

Schellenberg ES, Dryden DM, Vandermeer B, Ha C, Korownyk C. Lifestyle interventions for patients with and at risk for type 2 diabetes: a systematic review and meta-analysis. *Ann Intern Med.* 2013;159(8):543-551. doi:10.7326/0003-4819-159-8-201310150-00007

Sebastiani P, Thyagarajan B, Sun F, et al. Biomarker signatures of aging. *Aging Cell.* 2017;16(2):329-338. doi:10.1111/acel.12557

Seeling M, Brückner C, Nimmerjahn F. Differential antibody glycosylation in autoimmunity: sweet biomarker or modulator of disease activity?. *Nat Rev Rheumatol.* 2017;13(10):621-630. doi:10.1038/nrrheum.2017.146

Sharma P. The Ultimate Guide to 12 Dimensionality Reduction Techniques (with Python codes). *Analytics Vidya.* 2018. Available at: <https://www.analyticsvidhya.com/blog/2018/08/dimensionality-reduction-techniques-python/> Accessed January 20th, 2020.

Shin JJ, Lee EK, Park TJ, Kim W. Damage-associated molecular patterns and their pathological relevance in diabetes mellitus. *Ageing Res Rev.* 2015;24(Pt A):66-76. doi:10.1016/j.arr.2015.06.004

Sladek R, Rocheleau G, Rung J, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature.* 2007;445(7130):881-885. doi:10.1038/nature05616

Sniderman AD, Furberg CD, Keech A, et al. Apolipoproteins versus lipids as indices of coronary risk and as targets for statin treatment. *Lancet.* 2003;361(9359):777-780. doi:10.1016/s0140-6736(03)12663-3

Spiro RG. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology*. 2002;12(4):43R-56R. doi:10.1093/glycob/12.4.43r

Takahashi S, Sugiyama T, Shimomura M, et al. Site-specific and linkage analyses of fucosylated N-glycans on haptoglobin in sera of patients with various types of cancer: possible implication for the differential diagnosis of cancer. *Glycoconj J*. 2016;33(3):471-482. doi:10.1007/s10719-016-9653-7

Talmud PJ, Hawe E, Robertson K, Miller GJ, Miller NE, Humphries SE. Genetic and environmental determinants of plasma high density lipoprotein cholesterol and apolipoprotein AI concentrations in healthy middle-aged men. *Ann Hum Genet*. 2002;66(Pt 2):111-124. doi:10.1017/S0003480002001057

Tanigaki K, Sacharidou A, Peng J, et al. Hyposialylated IgG activates endothelial IgG receptor FcγRIIB to promote obesity-induced insulin resistance. *J Clin Invest*. 2018;128(1):309-322. doi:10.1172/JCI89333

Testa R, Vanhooren V, Bonfigli AR, et al. N-glycomic changes in serum proteins in type 2 diabetes mellitus correlate with complications and with metabolic syndrome parameters. *PLoS One*. 2015;10(3):e0119983. Published 2015 Mar 20. doi:10.1371/journal.pone.0119983

Theodoratou E, Campbell H, Ventham NT, et al. The role of glycosylation in IBD. *Nat Rev Gastroenterol Hepatol*. 2014;11(10):588-600. doi:10.1038/nrgastro.2014.78

Thorens B. Of fat, β cells, and diabetes. *Cell Metab*. 2011;14(4):439-440. doi:10.1016/j.cmet.2011.09.006

Uhlén M, Fagerberg L, Hallström BM, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015;347(6220):1260419. doi:10.1126/science.1260419

van de Winkel JG, Anderson CL. Biology of human immunoglobulin G Fc receptors. *J Leukoc Biol*. 1991;49(5):511-524. doi:10.1002/jlb.49.5.511

Van Lenten BJ, Hama SY, de Beer FC, et al. Anti-inflammatory HDL becomes pro-inflammatory during the acute phase response. Loss of protective effect of HDL against LDL oxidation in aortic wall cell cocultures. *J Clin Invest*. 1995;96(6):2758-2767. doi:10.1172/JCI118345

Vanhooren V, Dewaele S, Libert C, et al. Serum N-glycan profile shift during human ageing. *Exp Gerontol*. 2010;45(10):738-743. doi:10.1016/j.exger.2010.08.009

Vanhooren V, Laroy W, Libert C, Chen C. N-glycan profiling in the study of human aging. *Biogerontology*. 2008;9(5):351-356. doi:10.1007/s10522-008-9140-z

Varki A. Biological roles of glycans. *Glycobiology*. 2017;27(1):3-49. doi:10.1093/glycob/cww086

Vergès B. Pathophysiology of diabetic dyslipidaemia: where are we?. *Diabetologia*. 2015;58(5):886-899. doi:10.1007/s00125-015-3525-8

Vidal-Bralo L, Lopez-Golan Y, Gonzalez A. Simplified Assay for Epigenetic Age Estimation in Whole Blood of Adults [published correction appears in *Front Genet*. 2017 Apr 21;8:51]. *Front Genet*. 2016;7:126. Published 2016 Jul 14. doi:10.3389/fgene.2016.00126

Vlassara H, Striker GE. AGE restriction in diabetes mellitus: a paradigm shift. *Nat Rev Endocrinol*. 2011;7(9):526-539. Published 2011 May 24. doi:10.1038/nrendo.2011.74

- Walldius G, Jungner I, Holme I, Aastveit AH, Kolar W, Steiner E. High apolipoprotein B, low apolipoprotein A-I, and improvement in the prediction of fatal myocardial infarction (AMORIS study): a prospective study. *Lancet*. 2001;358(9298):2026-2033. doi:10.1016/S0140-6736(01)07098-2
- Walldius G, Jungner I. Apolipoprotein B and apolipoprotein A-I: risk indicators of coronary heart disease and targets for lipid-modifying therapy. *J Intern Med*. 2004;255(2):188-205. doi:10.1046/j.1365-2796.2003.01276.x
- Wang JC, Bennett M. Aging and atherosclerosis: mechanisms, functional consequences, and potential therapeutics for cellular senescence. *Circ Res*. 2012;111(2):245-259. doi:10.1161/CIRCRESAHA.111.261388
- Weidner CI, Lin Q, Koch CM, et al. Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biol*. 2014;15(2):R24. Published 2014 Feb 3. doi:10.1186/gb-2014-15-2-r24
- Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447(7145):661-678. doi:10.1038/nature05911
- Weyer C, Bogardus C, Pratley RE. Metabolic characteristics of individuals with impaired fasting glucose and/or impaired glucose tolerance. *Diabetes*. 1999;48(11):2197-2203. doi:10.2337/diabetes.48.11.2197
- Wittenbecher C, Štambuk T, Kuxhaus O, et al. Plasma N-Glycans as Emerging Biomarkers of Cardiometabolic Risk: A Prospective Investigation in the EPIC-Potsdam Cohort Study. *Diabetes Care*. 2020;43(3):661-668. doi:10.2337/dc19-1507
- Xu G, Huang J, Yang Y, Yao YA. Transcriptome Analysis of Flower Sex Differentiation in *Jatropha curcas* L. Using RNA Sequencing. *PLoS One*. 2016;11(2):e0145613. Published 2016 Feb 5. doi:10.1371/journal.pone.0145613
- Yasuda K, Miyake K, Horikawa Y, et al. Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nat Genet*. 2008;40(9):1092-1097. doi:10.1038/ng.207
- Yazdanyar A, Newman AB. The burden of cardiovascular disease in the elderly: morbidity, mortality, and costs. *Clin Geriatr Med*. 2009;25(4):563-vii. doi:10.1016/j.cger.2009.07.007
- Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. *Nat Rev Endocrinol*. 2018;14(2):88-98. doi:10.1038/nrendo.2017.151
- Zhou JY, Oswald DM, Oliva KD, Kreisman LSC, Cobb BA. The Glycoscience of Immunity. *Trends Immunol*. 2018;39(7):523-535. doi:10.1016/j.it.2018.04.004
- Zimmet PZ. Diabetes and its drivers: the largest epidemic in human history?. *Clin Diabetes Endocrinol*. 2017;3:1. Published 2017 Jan 18. doi:10.1186/s40842-016-0039-3