Blood pathway analyses reveal differences between prediabetic subjects with or without dyslipidaemia. The Cardiovascular Risk in Young Finns Study

Jaakko Laaksonen
Syventävien opintojen kirjallinen työ
Tampereen yliopisto
Lääketieteen ja biotieteiden tiedekunta
Kliinisen kemian tutkimusryhmä
Elokuu 2017

Tampereen yliopisto
Lääketieteen ja biotieteiden tiedekunta
Kliinisen kemian tutkimusryhmä

LAAKSONEN JAAKKO: VERISOLUISSA ILMENTYVIEN GEENIEN SIGNALOINTIREITTIEN EROAVAISUUDET PREDIABEETIKOILLA DYSLIPIDEMIASTATUKSESTA RIIPPUEN

Kirjallinen työ
Ohjaaja: prof. Terho Lehtimäki
Elokuu 2017

Avainsanat: geeniekspressio, geenien ilmentyminen, kolesteroli, paastoglukoosi

Tämän opinnäytteen alkuperäisyys on tarkastettu Turnitin OriginalityCheck-ohjelmalla Tampereen yliopiston laatujärjestelmän mukaisesti.

Tavoitteet:
Tämän tutkimuksen tavoitteena oli selvittää, miten veressä ilmentyvien geenien sekä niiden muodostamien signalointireittien aktiivisuus eroaa niillä prediabeetikoilla, joilla on samanaikaisesti dyslipidemia verrattuna niihin prediabeetikoihin, joilla dyslipidemiaa ei ole, sekä verrata näitä ryhmiä normaalit sokeri- ja rasva-arvot omaavaan kontrolliryhmään.

Menetelmät ja tulokset:
Tutkimusaineistoon kuului 1 240 Lasten sepelvaltimotaudin riskitekijät -tutkimukseen vuonna 2011 osallistunutta henkilöä, jotka olivat tuolloin iältään 34–49-vuotta. Signalointireittien aktiivisuus analysoitiin GSEA (Gene set enrichment analysis) -ohjelman avulla. Niillä henkilöillä, joilla oli prediabetes mutta ei dyslipidemiaa, kolesterolisynteesin sekä tiettyjen CD8-lymfosyytti- ja interleukiini-12-välitteisten reittien aktiivisuus oli kontrolliryhmään verrattuna tilastollisesti merkittävästi lisääntynyt. Vastaavia muutoksia ei havaittu niillä tutkittavilla, joilla oli samanaikaisesti sekä prediabetes että dyslipidemia. Verrattaessa näitä kahta prediabeetikkoryhmää keskenään reittien säätelyssä ei ollut tilastollisesti merkittäviä eroja. Yksittäisten geenien ilmentymisessä oli vain lieviä eroja.

Johtopäätökset:
Sekä prediabetes että dyslipidemia vaikuttavat verisolujen geeniekspressioon. Tulosten kliinistä merkitystä esimerkiksi sydän- ja verenkiertoelimistön sairauksiin liittyen voi olla tarpeellista selvittää seurantatutkimuksissa.

Sisältö

1 Abstract **6**

2 Introduction **8**

3 Materials and methods **10**

3.1 Population 10

3.2 Clinical and biochemical measurements 10

3.3 RNA isolation, microarrays and data processing 11

3.4 Definition of prediabetes and dyslipidemia 11

3.5 Statistical analysis 11

4 Results **13**

5 Discussion **15**

6 Acknowledgements **20**

7 Funding **21**

8 Disclosure statement **21**

9 Contribution statement **21**

10 References **22**

11 Tables **30**

12 Figures **35**

This is the peer reviewed version of the following article:

Laaksonen J, Taipale T, Seppälä I, et al. Blood pathway analyses reveal differences between prediabetic subjects with or without dyslipidaemia. The Cardiovascular Risk in Young Finns Study. Diabetes Metab Res Rev. 2017;e2914,

which has been published in final form at http://dx.doi.org/10.1002/dmrr.2914. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

Blood pathway analyses reveal differences between prediabetic subjects with or without dyslipidaemia. The Cardiovascular Risk in Young Finns Study

Short title: Pathway enrichment in prediabetes

Jaakko Laaksonen[a], Tuukka Taipale[a], Ilkka Seppälä[a], Emma Raitoharju[a], Nina Mononen[a], Leo-Pekka Lyytikäinen[a], Melanie Waldenberger[b,c], Thomas Illig[b,d,e], Nina Hutri-Kähönen[f], Tapani Rönnemaa[g,h], Markus Juonala[g,h], Jorma Viikari[g,h], Mika Kähönen[i], Olli Raitakari[j,k], Terho Lehtimäki[a]

[a]Department of Clinical Chemistry, Fimlab Laboratories and Faculty of Medicine and Life Sciences, University of Tampere

[b]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health

[c]Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health

[d]Hannover Unified Biobank, Hannover Medical School

[e]Institute for Human Genetics, Hannover Medical School

[f]Department of Paediatrics, Tampere University Hospital and Faculty of Medicine and Life Sciences, University of Tampere

[g]Department of Medicine, University of Turku

[h]Division of Medicine, Turku University Hospital

[i]Department of Clinical Physiology, Tampere University Hospital and Faculty of Medicine and Life Sciences, University of Tampere

[j]Department of Clinical Physiology and Nuclear Medicine, University of Turku and Turku University Hospital

[k]Research Centre for Applied and Preventive Cardiovascular Medicine, University of Turku

Corresponding author:

Jaakko Laaksonen

Faculty of Medicine and Life Sciences, Arvo building, room D339

PO Box 100

FI-33014 University of Tampere, Finland

Tel.: +358 407641187

Fax: +358 331174168

e-mail: laaksonen.jaakko.h@student.uta.fi

Abstract word count: 247

Word count: 4,088 (including the abbreviation list, acknowledgements, funding, disclosure statement and author contributions)

Keywords: prediabetes, dyslipidaemia, gene set enrichment analysis, gene expression

Abstract

BACKGROUND: Prediabetes (PR) often occurs together with dyslipidaemia (D), which is paradoxically treated with statins predisposing to type 2 diabetes mellitus. We examined peripheral blood pathway profiles in prediabetic subjects with (PR_D) and without dyslipidaemia (PR_0) and compared these to non-prediabetic controls (C) without dyslipidaemia (C_0).

METHODS: The participants were from the Cardiovascular Risk in Young Finns Study, including 1,240 subjects aged 34-49 years. Genome-wide expression data of peripheral blood and gene set enrichment analysis were used to investigate the differentially expressed genes and enriched pathways between different subtypes of prediabetes.

RESULTS: Pathways for cholesterol synthesis, interleukin-12 (IL12)-mediated signalling events and downstream signalling in naïve CD8+ T cells were up-regulated in the PR_0 group in comparison to controls (C_0) and the up-regulation of these pathways was independent of waist circumference, blood pressure, smoking status and insulin. Adjustment for CRP left the CD8+ T cell signalling and IL12-mediated signalling event pathway up-regulated. The cholesterol synthesis pathway was also up-regulated when all prediabetic subjects (PR_0 and PR_D) were compared to the non-prediabetic control group. No pathways were up- or down-regulated when the PR_D group was compared to the C_0 group. Five genes in the PR_0 group and one in the PR_D group was significantly differentially expressed in comparison to the C_0 group.

CONCLUSIONS: Blood cell gene expression profiles differ significantly between prediabetic subjects with and without dyslipidaemia. Whether this classification may be used in detection of prediabetic individuals at a high risk of cardiovascular complications remains to be examined.

Abbreviations: BH-FDR, Benjamini-Hochberg false discovery rate; BP, blood pressure; C_0, normoglycaemic and normolipidaemic subjects; C_D, normoglycaemic and dyslipidaemic subjects; CV, cardiovascular; FDR, false discovery rate; FWER, family-wise error rate; GSEA, gene set enrichment analysis; IFG, impaired fasting glucose; IGT, impaired glucose tolerance; KEGG, Kyoto Encyclopedia of Genes and Genomes; NCI PID, National Cancer Institute Pathway Interaction Database; NES, normalised enrichment score; PC, principal component; PR_0, prediabetic and normolipidaemic subjects; PR_D, prediabetic and dyslipidaemic subjects; TC, total cholesterol; TG, triglycerides; T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus; YFS, Young Finns Study.

Introduction

The prevalence of type 2 diabetes mellitus (T2DM) and the subsequent development of its cardiovascular complications (CV) are increasing worldwide. A potential risk factor for the development of T2DM is prediabetes (PR), which may be defined as an intermediate state between normal glucose metabolism and T2DM. The American Diabetes Association defines prediabetes as impaired glucose tolerance (IGT), impaired fasting glucose (IFG) or elevated HbA_{1c} [1]. Deficiency in the insulin secretion of pancreatic beta cells has a more pronounced role in prediabetes than insulin resistance [2], but most persons with prediabetes are also insulin-resistant [3].

Dyslipidaemia, which is defined by elevated total cholesterol (TC), LDL cholesterol (LDL-C), non-HDL cholesterol (non-HDL-C) or triglycerides (TG), or low HDL cholesterol (HDL-C), [4] commonly occurs together with T2DM. A typical pattern of lipid abnormalities in diabetic patients includes hypertriglyceridaemia, low HDL-C and a presence of small dense LDL-C particles [5]. The same pattern, known as atherogenic dyslipidaemia, has been shown to occur already during the prediabetic stage [6,7]. Metabolically, low absorption efficiency and high synthesis of cholesterol are also related to an elevated serum glucose level and insulin resistance [8,9]. However, whether the progression or development of CV complications will differ in prediabetic patients with and without dyslipidaemia is not well known.

In atherogenic dyslipidaemia, statins are widely used as first-line drugs. Since the treatment goal for serum LDL-C is more stringent in patients with T2DM than in those without T2DM, high-dose statin treatment may be needed [4]. The use of statins in prediabetic subjects may be complicated, since these drugs can increase the risk of incident T2DM [10], and by affecting insulin sensitivity, insulin secretion and glucose transport increase plasma

glucose levels [11]. Despite this controversy, statins yield an overall benefit in terms of preventing vascular events and, therefore, these drugs are widely recommended in current treatment guidelines also for the treatment of dyslipidaemia in T2DM patients [4]. With this clinical background, it is rational to seek better understanding of the metabolic differences of PR subphenotypes with and without dyslipidaemia.

Gene expression and pathway profiling of blood and tissue samples may provide better understanding of the pathogenesis of prediabetes and its association to dyslipidaemia. An earlier metabolomics investigation of plasma from non-diabetic subjects with reduced insulin sensitivity showed alterations in lipid metabolic pathways, steroid hormone biosynthesis and bile acid metabolism [12]. The majority of these pathways and certain amino acid metabolism pathways have also been found to be differentially regulated in the liver of pigs with impaired incretin (a group of metabolic hormones that stimulate a decrease in blood glucose levels) function [13]. Altered pathways in lipid metabolism, insulin action, inflammatory response and complex oxidative processes have also been revealed from subcutaneous adipose and muscle tissue from non-diabetic, insulin-resistant subjects [14,15]. However, these earlier metabolic studies did not take into account whether plasma lipid abnormalities were present in the subjects.

It has been suggested that insulin resistance elicits dyslipidaemia either mechanically or by means of genetic linkage, but further validation is still needed [2,3]. We aimed to identify the metabolic pathways and gene expression associated with the prediabetic state, with special respect to a division based on the subjects' dyslipidaemia status. In order to pinpoint dysregulated metabolic pathways associated with these PR subphenotypes, we performed a gene set enrichment pathway analysis (GSEA) and also otherwise analysed gene-wise differences between PR subphenotypes and controls.

Materials and Methods

Population

The Cardiovascular Risk in Young Finns Study (YFS) is a Finnish longitudinal population study on the evolution of cardiovascular risk factors from childhood to adulthood, the sample and methods have been described in detail elsewhere [16,17]. The present study included 1,240 subjects who were not diagnosed with T1DM or T2DM (in 2012), were not on medication for hypertension or hypercholesterolaemia and for whom complete gene expression data as well as data on lipids, glucose and clinical characteristics were available.

The study plan was approved by the ethics committees of all participating universities, and the study protocol of each study phase corresponded with the proposal by the World Health Organization. All subjects gave written informed consent, and the study was conducted in accordance with the Declaration of Helsinki.

Clinical and biochemical measurements

Height and weight were measured and body mass index (BMI) was calculated as weight in kilograms divided by height in metres squared. Waist circumference was measured using an anthropometric tape at the midpoint between the iliac crest and the lowest rib to the nearest 0.1 cm. Blood pressure was measured three times after a 5-min rest with a random zero sphygmomanometer and was estimated as the average of the three measurements.

Venous blood samples were drawn after an overnight fast for the determination of serum lipid levels, glucose, insulin, glycated haemoglobin A1c (HbA1c) and high-sensitive C-reactive protein (hs-CRP). Standard enzymatic methods were used for serum TC, TG and HDL-C determinations. LDL-C was calculated by the Friedewald formula in participants with TG levels < 4.0 mmol/l [18]. Non-HDL-C was calculated as TC – HDL-C. Glucose

concentrations were determined by the enzymatic hexokinase method. Serum insulin was measured with immunoassay and HbA1c with an immunoturbidimetric method. Hs-CRP was determined immunoturbidimetrically. Details of all of the methods have been previously described elsewhere [19].

RNA isolation, microarrays and data processing

RNA was isolated and the gene expression levels were analysed using commercially available kits. Expression data was analysed in R (http://www.r-project.org/) using the Bioconductor packages (http://www.bioconductor.org/). Details of the process have been previously described elsewhere [20].

Definition of prediabetes and dyslipidaemia

The classification of prediabetes was based on fasting plasma glucose and HbA1c according to the criteria of the American Diabetes Association [1]. People with impaired IFG, i.e. prediabetes, were defined as having a fasting plasma glucose level of 5.6–6.9 mmol/l or HbA1c of 5.7–6.4% (38–46 mmol/l) and not diagnosed with T2DM. The diagnosis of T2DM included subjects with a fasting plasma glucose level of over 7.0 mmol/l or HbA1c of over 6.5% (48 mmol/l), or those with reported use of oral glucose-lowering medication or insulin (but had not reported having T1DM) or who had a reported diagnosis of T2DM by a physician.

Dyslipidaemia was defined according to the European guidelines [21]. The criteria for dyslipidaemia were TC > 5.0 mmol/l, LDL-C > 3.0 mmol/l, HDL-C < 1.0 mmol/l in men and < 1.2 mmol/l in women, non-HDL-C > 3.8 mmol/l or TG > 1.7 mmol/l.

Statistical analysis

Gene sets were collected from five publicly available collections: BioCarta (http://cgap.nci.nih.gov/Pathways/BioCarta_Pathways), KEGG (Kyoto Encyclopedia of Genes and Genomes, http://www.genome.jp/kegg/), Reactome (http://www.reactome.org/), NCI PID (National Cancer Institute Pathway Interaction Database, http://pid.nci.nih.gov/) and HumanCyc (http://humancyc.org/). Enrichment analysis was performed by using each gene set separately. In order to avoid too narrowly or too broadly defined functional gene sets, pathways containing less than 10 or more than 200 genes were excluded. As a result, 1,078 pathways were included in the study. The reduced number of pathways potentially increases the power of the analysis by decreasing the multiple testing correction burden.

The study population was divided into four subphenotypes as follows: prediabetic individuals with (PR_D) or without dyslipidaemia (PR_0), and normoglycaemic (non-prediabetic) control (C) subjects with (C_D) or without (C_0) dyslipidaemia. All prediabetic subjects (PR), regardless of dyslipidaemia status, were compared to the non-prediabetic control group (C_0 and C_D together). The PR_0 and PR_D groups were individually compared to the C_0 group and also to each other (PR_0 vs. PR_D). We also did similar analyses and examined whether the results differ when the dyslipidaemia status definition was based on the high LDL-C level (LDL > 3.0 mmol/l) only. The baseline characteristics of the groups were compared using the t-test for continuous variables and a χ^2 test for proportions.

Potential population stratification was taken into account by using principal components (PCs) computed from all genotypes as covariates [22]. Based on a scree plot, the seven first PCs were used. In addition to the PCs, the analyses were adjusted by age, sex, BMI or waist circumference, smoking, insulin, systolic and diastolic blood pressure, and hs-CRP. R language was used for adjusting the gene expression data. After the adjustment, GSEA software (http://www.broad.mit.edu/gsea) [23,24] was used to analyse the association of gene pathways with the phenotype. The pathways were considered to be significantly up- or

down-regulated when the false discovery rate (FDR) was smaller than 0.10 and the family-wise error rate (FWER) was smaller than 0.05 after 1,000 permutation cycles. FDR < 0.25 can be considered significant according to the criteria recommended by Subramanian et al. [23].

The expression of individual genes in the same setting was analysed using the phenoTest R package with a Benjamini-Hochberg-FDR-corrected p-value of ≤ 0.05 and log2 fold change of ≥ 1.5 as the significance level. The analysis was adjusted with age, sex, BMI and the first seven PCs. For boxplots, the statistical significance of the difference in gene expression was assessed using the nonparametric Wilcoxon signed-rank test.

Results

The demographics of the study population, when the division is based on any type of dyslipidaemia, or hypercholesterolaemia defined as a high LDL cholesterol (LDL > 3.0 mmol/l) only, are presented in Table 1. Of the non-medicated subjects with prediabetes, 79.5% had dyslipidaemia and 66.0% hypercholesterolaemia defined as LDL > 3.0 mmol/l. When comparing all PR individuals to the non-prediabetic control group (C), GSEA identified up-regulation of cholesterol biosynthesis pathways in all of the used but differently adjusted models 1–3 (FDR < 0.014 for all) (Table 2). A positive normalized enrichment score (NES) indicated that all the pathways were up-regulated. The leading-edge subsets containing the most up-regulated genes are almost identical in the KEGG steroid and HumanCyc cholesterol biosynthesis pathways (**Supplementary table 1**). In the PR vs. C group comparison, superpathway of methionine degradation remained significantly enriched in models 1 and 2 (FDR < 0.006 and < 0.015, respectively). However, in model 3, the additional adjustment with hs-CRP abolished the association.

Two additional pathways, cholesterol biosynthesis II (via 24,25-dihydrolanosterol) and cholesterol biosynthesis III (via desmosterol) from HumanCyc, were

closely and significantly co-enriched with the cholesterol biosynthesis pathway because they all share the same genes. Hence, they are not shown in the tables.

These KEGG and HumanCyc pathways were also up-regulated when the PR_0 group was compared to the corresponding control group without prediabetes and dyslipidaemia (C_0) after adjustment for age, sex, BMI and the first seven PCs (Table 3). In addition, in this setting, the pathways for interleukin (IL)-12-mediated signalling events and downstream signalling in naive CD8+ T cells were also significantly up-regulated. When further adjusted for waist circumference (instead of BMI), blood pressure, smoking and insulin, all other pathways except the HumanCyc superpathway of cholesterol synthesis remained significant. In the fully adjusted model 3, the association of steroid and cholesterol biosynthesis pathways was abolished after additional adjustment for hs-CRP, leaving only the downstream signalling in naive CD8+ T cells and IL12-mediated signalling events significantly up-regulated.

When the PR_D group was compared to the C_0 group, no pathways were significantly enriched in any of the models. In the PR_0 vs. PR_D group comparison one Reactome pathway, Cytochrome P450 – arranged by substrate type was up-regulated in the PR_D group (NES 2.09, $p < 0.001$, FDR 0.041, FWER 0.026) after adjustment for age, sex, BMI and the seven first PCs. Further adjustment in models 2 and 3, similar as in other analyses, abolished the association.

When comparing prediabetic subjects with normal LDL-C (< 3.0 mmol/l) (PR_{NC}) to normoglycaemic subjects with normal LDL-C (C_{NC}), one pathway for cholesterol biosynthesis remained significantly up-regulated in all models 1–3 (Table 4). In prediabetic subjects with high LDL-C (≥ 3.0 mmol/l) (PR_{HC}), no pathways were significantly enriched as compared to the C_{NC} group. When comparing the two prediabetic groups to each other (PR_{NC}

vs. PR_{HC}), one pathway from NCI PID was significantly up-regulated. Regulation of cytoplasmic and nuclear SMAD2/3 signalling was enriched in the PR_{HC} group in models 1 (NES 2.03, p 0.002, FDR 0.029, FWER 0.022) and 2 (NES 2.00, $p<0.001$, FDR 0.042, FWER 0.033) but not in model 3.

The expression of individual pathway genes was not statistically significant in any PR phenotype when compared to the C_0 group. This is explained by the fact that GSEA considers all expressed genes by rank without a fold-change threshold. Therefore, we also tested gene-wise differences between PR subphenotype groups with a less stringent cut-off value for log2 fold change (FC >1.2) and a Benjamini-Hochberg-FDR-corrected p-value of ≤ 0.05 (**Supplementary table 2**). In gene-wise analysis, we identified five genes up-regulated in the PR_0 group as compared to the C_0 group, including type 1 neurotrophic tyrosine kinase receptor (*NTRK1*); granzyme B (*GZMB*); perforin 1 (*PRF1*); killer cell immunoglobulin-like receptor, two domains, long cytoplasmic tail, 4 (*KIR2DL4*); and family with sequence similarity 179 member A (*FAM179A*). One gene, secretory leukocyte peptidase inhibitor (*SLPI*), was up-regulated in the PR_D subjects as compared to the C_0 group. The trend analyses for these six genes are shown in **Figure 1.** In PR vs. C, PR_0 vs. PR_D and PR_{NC} vs. PR_{HC} group comparisons no genes were differentially expressed.

Discussion

Our analysis of peripheral blood cell mRNA expression shows, for the first time, that the pathway profiles differ significantly between prediabetic subphenotypes with and without dyslipidaemia. We observed that, compared to normoglycaemic and normolipidaemic controls, the cholesterol biosynthesis pathway was up-regulated in normolipidaemic prediabetic individuals but not in those with both prediabetes and dyslipidaemia. Also, pathways related to the immune response were up-regulated only in the

PR_0 group. It is not surprising that pathway analysis identified differences between prediabetic subphenotypes. However, using the most recent pathway databases our analysis pinpointed the specific pathways which were up-regulated.

The enrichment of the cholesterol biosynthesis pathway was independent of both BMI and waist circumference. Parallel results have been reported by Gylling et al. [9] who assayed cholesterol precursors and markers of cholesterol synthesis and absorption from plasma. In their study, markers of cholesterol synthesis were already increased in subjects with IFG, and cholesterol metabolism was regulated more by peripheral insulin sensitivity than obesity. In the present study, the enrichment remained significant until the analysis was adjusted for hs-CRP. This suggests that an up-regulated cholesterol biosynthesis pathway is related to the increased overall inflammation as measured by hs-CRP. Interestingly, serum hs-CRP concentration has been previously associated with dietary cholesterol absorption but not synthesis of cholesterol in subjects with IFG or IGT and features of the metabolic syndrome cholesterol metabolism [25].

The mechanism through which prediabetes in the absence of dyslipidaemia regulates cholesterol metabolism gene pathways in blood cells remains open. A possible reason for the up-regulation of cholesterol synthesis in the PR_0 group is the cholesterol deprivation inside the blood cells due to the lack of extra cholesterol available in the plasma. In leukocytes, the expression of certain genes which are included in the KEGG Steroid biosynthesis pathway has been found to be associated with plasma lipid levels. The expression is hypothesized to be activated by peroxisome proliferative activated receptors (PPARs). [26] However, we did not observe changes in the expression of PPARs. Also, we did not observe gene-wise changes in the expression of HMG-CoA reductase (*HMGCR*), the rate-limiting step in cholesterol metabolism or sterol-regulatory element-binding protein (*SREBP-2*), which regulates the transcription of *HMGCR*. The upregulated pathways consist

of the latter half of the cholesterol biosynthesis pathway, with farnesyl pyrophosphate being the first intermediate. The first sterol intermediate is lanosterol, and the subsequent reactions define the post-squalene part of the pathway. In this portion of the pathway, the demethylation of lanosterol has been suggested to act as the rate-limiting step [27].

Since isolated IFG and IGT are characterised by different patterns of lipid changes [28], they presumably affect the lipid metabolism by distinct mechanisms. The down-regulation of the cholesterol biosynthesis pathway has been previously associated in insulin resistance in adipose tissue [14], but in our study, the same pathway was up-regulated in peripheral blood. Analogous results have been published related to the mitogen-activated protein kinase (MAPK) signalling pathway in insulin resistance in metabolic syndrome – the pathway is up-regulated in muscle tissue but down-regulated in blood [29,30]. On the other hand, one study has demonstrated that the mechanisms which regulate gene expression in liver and mononuclear leukocytes are similar and that these leukocytes can be used to predict the level of expression of *HMGCR* and LDL receptor (*LDLR*) genes [31]. This could indicate that also the hepatic cholesterol production is increased in the PR_0 group, although we did not observe increased expression of these two genes.

Recent data suggests that non-alcoholic fatty liver disease (NAFLD) results mainly from disturbed hepatic cholesterol homeostasis and the hepatic accumulation of free cholesterol [32]. If the cholesterol synthesis pathway is up-regulated in the liver, the newly synthesized cholesterol may promote the pathogenesis of NAFLD, since cholesterol export and bile acid synthesis pathways were not up-regulated. This hypothesis is supported by a Japanese study that showed a positive association between NAFLD and IFG, independently of T2DM risk factors [33]. NAFLD is also considered to be a consequence of insulin resistance [34], but it is also an independent risk factor of T2DM, particularly in individuals with IFG [35].

Based on above reasons, potentially increased hepatic cholesterol production would imply that the onset of prediabetes launches a cascade leading to hypercholesterolaemia and/or NAFLD, which highlights the importance of early detection of prediabetes and prevention of T2DM through lifestyle intervention. Guidelines in condensed form have been provided for building up an effective intervention program, the IMAGE toolkit [36] gives also instructions for evaluation and quality assurance. In addition to working at the patient level, actions at policy and environmental levels are needed for sustainable diabetes prevention [37].

The enrichment of the IL12 signalling pathway remained significant in all models when the PR_0 group was compared to the C_0 group. Elevated IL12 levels have been previously shown to be dependent on hs-CRP [38] and peripheral insulin resistance [39] in T2DM. Since the IL12 pathway remained up-regulated when the analysis was adjusted with both serum insulin level and hs-CRP, it may be suggested that, in the PR_0 group, the activation of the IL12 pathway is mediated by another mechanism. When only the LDL-C levels were taken into account, the pathway profiles were similar to the ones of all prediabetic subjects, i.e. when the dyslipidaemia status was not considered. This suggests that the up-regulation of IL12-mediated and CD8+ T cell pathways could be partly related to hypertriglyceridaemia or low HDL-C.

The enrichment of cholesterol biosynthesis and inflammation related pathways was seen in the PR_0 group when compared to the C_0 group but not when compared to the PR_D group, which implies that the metabolic differences between the two PR subphenotypes are small. However, whether this difference will evolve over time requires longitudinal studies.

The analysis of individual genes revealed only a moderate increase in gene expression. Some of the genes have been previously associated with metabolic dysfunction.

Only one gene, secretory leukocyte protease inhibitor (*SLPI*) which is a potent inhibitor of the inflammatory cascade [40], was up-regulated in the PR_D group. The up-regulation of *SLPI* has also been previously shown to correlate negatively with HDL cholesterol and positively with HbA1c. This may be due to an attempt to counterbalance the low-grade inflammation associated with prediabetes and dyslipidaemia. [41] The expression of *LDLR* or scavenger receptor genes [42] in the PR_0 or PR_D groups was not different compared to the C_0 group. Since scavenger receptors are key molecules in the formation of atherosclerotic plaques [43], our results imply that prediabetes combined with dyslipidaemia does not directly cause atherosclerosis, which is also stated by Grundy [2].

Three out of five genes that were slightly up-regulated in the PR_0 group – granzyme B (*GZMB*); perforin 1 (*PRF1*); and killer cell immunoglobulin-like receptor, two domains, long cytoplasmic tail, 4 (*KIR2DL4*) – have been found to be down-regulated after exposure to high blood glucose in normoglycaemic controls. In T2DM patients, the expression levels of these genes has been reported to be low already and hardly affected by hyperglycaemia. [44] These genes are typically expressed in cells with cytotoxic functions, such as CD8+ T cells [45]. *GZMB* and *PRF1* are also included in the downstream signalling in the naive CD8+ T cells pathway, which was up-regulated in the PR_0 group but not in the PR_D group in comparison to the C_0 group. The reason why *GZMB*, *PRF1* and *KIR2DL4* were up-regulated in the PR_0 group but not in the PR_D group and expressed in lower levels in T2DM patients in the study by van der Pouw Kraan et al. [44] might be that one or more of the components of dyslipidaemia co-regulate the expression of these genes; the T2DM patients in their study [44] met the elevated triglycerides criterion of dyslipidaemia used in our study. However, in another study, the plasma level of granzyme B correlated positively with fasting glucose and HbA1c, as well as with triglycerides, total cholesterol and LDL cholesterol [46].

The present study has some limitations. A major one is that profiling gene expression from peripheral blood cells makes it challenging to speculate how the expression levels represent the gene expression in other tissues. Another limitation is that no glucose tolerance tests were performed on the study population and the definition of prediabetes was based only on fasting plasma glucose and HbA1c levels. Some studies [47,48], but not all [49], have shown that IGT is a better predictor of cardiovascular complications than IFG. This raises the question whether there are differences in gene expression and pathway profiles when prediabetes is diagnosed by either IFG or IGT. However, the HbA1c cut-off point for prediabetes has a high specificity to identify cases of IGT [1] and also subsequent 6-year diabetes incidence [50]. In addition, the Finnish gene pool has been shown to be distinctive and the results may not be directly generalizable to populations with a different ethnic background. We also recognize that microarray studies are limited by multiple testing problems and false positives.

In summary, our data indicates that blood cell gene expression pathway profiles differ significantly between prediabetic subphenotypes with and without dyslipidaemia. The pathway analysis identified up-regulated pathways, including cholesterol biosynthesis, IL12-mediated signalling and signalling in naïve CD8+ T cells in prediabetic individuals only in the absence of dyslipidaemia. However, whether this classification may be used in e.g. early-phase detection of individuals at a high risk of cardiovascular complications should be further examined in longitudinal studies. The clinical implication is that physicians should actively screen patients for prediabetes and dyslipidaemia and encourage especially those with prediabetes to permanent lifestyle changes with active follow-ups.

Acknowledgements

None.

Funding

The Young Finns Study has been financially supported by the Academy of Finland [grant numbers 286284 (T.L.), 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), 41071 (Skidi), and 285902 (E.R.)]; the Social Insurance Institution of Finland; the Kuopio, Tampere and Turku University Hospital Medical Funds [grant number X51001 (T.L.)]; Laboratoriolääketieteen Edistämissäätiö (J.L.); the Juho Vainio Foundation; the Paavo Nurmi Foundation; the Finnish Foundation of Cardiovascular Research; the Finnish Cultural Foundation; the Tampere Tuberculosis Foundation; the Emil Aaltonen Foundation; and the Yrjö Jahnsson Foundation.

Disclosure statement

The authors declare that there is no duality of interest associated with this manuscript.

Contribution statement

JL contributed to the study design, statistical analyses, data interpretation and drafting of the manuscript. TT and LPL contributed to the statistical analyses and critical revision of the manuscript. IS contributed to the study design, statistical analyses and critical revision of the manuscript. ER, NM, MW, TI, NHK, TR and MJ contributed to the data collection and critical revision of the manuscript. JV contributed to the initial design of YFS, cohort collection and critical revision of the manuscript. MK contributed to obtaining funding, cohort collection and critical revision of the manuscript. OR leads YFS and contributed to obtaining funding, as well as cohort collection and critical revision of the manuscript. TL supervised the research and contributed to the study design, obtaining funding and cohort collection, in addition to reviewing and editing the manuscript. All authors have read and approved the final manuscript.

1. American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care*. 2010;33 Suppl 1:S62-9.

2. Grundy SM. Pre-diabetes, metabolic syndrome, and cardiovascular risk. *J Am Coll Cardiol*. 2012;59(7):635-643.

3. Ferrannini E, Gastaldelli A, Iozzo P. Pathophysiology of prediabetes. *Med Clin North Am*. 2011;95(2):327-339.

4. Perk J, De Backer G, Gohlke H, et al. European guidelines on cardiovascular disease prevention in clinical practice (version 2012). the fifth joint task force of the european society of cardiology and other societies on cardiovascular disease prevention in clinical practice (constituted by representatives of nine societies and by invited experts). *Eur Heart J*. 2012;33(13):1635-1701.

5. Jacobs MJ, Kleisli T, Pio JR, et al. Prevalence and control of dyslipidemia among persons with diabetes in the united states. *Diabetes Res Clin Pract*. 2005;70(3):263-269.

6. Haffner SM, Stern MP, Hazuda HP, Mitchell BD, Patterson JK. Cardiovascular risk factors in confirmed prediabetic individuals: Does the clock for coronary heart disease start ticking before the onset of clinical diabetes? *JAMA*. 1990;263(21):2893-2898.

7. Festa A, Williams K, Hanley AJ, et al. Nuclear magnetic resonance lipoprotein abnormalities in prediabetic subjects in the insulin resistance atherosclerosis study. *Circulation*. 2005;111(25):3465-3472.

8. Pihlajamaki J, Gylling H, Miettinen TA, Laakso M. Insulin resistance is associated with increased cholesterol synthesis and decreased cholesterol absorption in normoglycemic men. *J Lipid Res*. 2004;45(3):507-512.

9. Gylling H, Hallikainen M, Pihlajamaki J, et al. Insulin sensitivity regulates cholesterol metabolism to a greater extent than obesity: Lessons from the METSIM study. *J Lipid Res*. 2010;51(8):2422-2427.

10. Carter AA, Gomes T, Camacho X, Juurlink DN, Shah BR, Mamdani MM. Risk of incident diabetes among patients treated with statins: Population based study. *BMJ*. 2013;346:f2610.

11. Cederberg H, Stančáková A, Yaluri N, Modi S, Kuusisto J, Laakso M. Increased risk of diabetes with statin treatment is associated with impaired insulin sensitivity and insulin secretion: A 6 year follow-up study of the METSIM cohort. *Diabetologia*. 2015;58(5):1109-1117.

12. Lucio M, Fekete A, Weigert C, et al. Insulin sensitivity is reflected by characteristic metabolic fingerprints--a fourier transform mass spectrometric non-targeted metabolomics approach. *PLoS One*. 2010;5(10):e13317.

13. Renner S, Romisch-Margl W, Prehn C, et al. Changing metabolic signatures of amino acids and lipids during the prediabetic period in a pig model with impaired incretin function and reduced beta-cell mass. *Diabetes*. 2012;61(8):2166-2175.

14. Elbein SC, Kern PA, Rasouli N, Yao-Borengasser A, Sharma NK, Das SK. Global gene expression profiles of subcutaneous adipose and muscle from glucose-tolerant, insulin-sensitive, and insulin-resistant individuals matched for BMI. *Diabetes*. 2011;60(3):1019-1029.

15. Patti ME, Butte AJ, Crunkhorn S, et al. Coordinated reduction of genes of oxidative metabolism in humans with insulin resistance and diabetes: Potential role of PGC1 and NRF1. *Proc Natl Acad Sci U S A*. 2003;100(14):8466-8471.

16. Raitakari OT, Juonala M, Ronnemaa T, et al. Cohort profile: The cardiovascular risk in young finns study. *Int J Epidemiol*. 2008;37(6):1220-1226.

17. Juonala M, Viikari JS, Raitakari OT. Main findings from the prospective cardiovascular risk in young finns study. *Curr Opin Lipidol*. 2013;24(1):57-64.

18. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem*. 1972;18(6):499-502.

19. Suomela E, Oikonen M, Virtanen J, et al. Prevalence and determinants of fatty liver in normal-weight and overweight young adults. the cardiovascular risk in young finns study. *Ann Med*. 2015;47(1):40-46.

20. Turpeinen H, Seppälä I, Lyytikäinen L, et al. A genome-wide expression quantitative trait loci analysis of proprotein convertase subtilisin/kexin enzymes identifies a novel regulatory gene variant for FURIN expression and blood pressure. *Hum Genet*. 2015;134(6):627-636.

21. Perk J, De Backer G, Gohlke H, et al. European guidelines on cardiovascular disease prevention in clinical practice (version 2012). the fifth joint task force of the european society of cardiology and other societies on cardiovascular disease prevention in clinical practice (constituted by representatives of nine societies and by invited experts). *Eur Heart J*. 2012;33(13):1635-1701.

22. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38(8):904-909.

23. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545-15550.

24. Mootha VK, Lindgren CM, Eriksson K, et al. PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet*. 2003;34(3):267-273.

25. Hallikainen M, Kolehmainen M, Schwab U, et al. Serum adipokines are associated with cholesterol metabolism in the metabolic syndrome. *Clinica chimica acta*. 2007;383(1):126-132.

26. Ma J, Dempsey AA, Stamatiou D, Marshall KW, Liew C. Identifying leukocyte gene expression patterns associated with plasma lipid levels in human subjects. *Atherosclerosis*. 2007;191(1):63-72.

27. Gaylor JL. Membrane-bound enzymes of cholesterol synthesis from lanosterol. *Biochem Biophys Res Commun*. 2002;292(5):1139-1146.

28. Lorenzo C, Hartnett S, Hanley AJ, et al. Impaired fasting glucose and impaired glucose tolerance have distinct lipoprotein and apolipoprotein changes: The insulin resistance atherosclerosis study. *The Journal of Clinical Endocrinology & Metabolism*. 2013;98(4):1622-1630.

29. Pasini E, Flati V, Paiardi S, et al. Intracellular molecular effects of insulin resistance in patients with metabolic syndrome. *Cardiovascular diabetology*. 2010;9:46-46.

30. Tangen SE, Tsinajinnie D, Nuñez M, Shaibi GQ, Mandarino LJ, Coletta DK. Whole blood gene expression profiles in insulin resistant latinos with the metabolic syndrome. *PloS one*. 2013;8(12):e84002.

31. Powell EE, Kroon PA. Low density lipoprotein receptor and 3-hydroxy-3-methylglutaryl coenzyme A reductase gene expression in human mononuclear leukocytes is regulated coordinately and parallels gene expression in human liver. *J Clin Invest*. 1994;93(5):2168-2174.

32. Arguello G, Balboa E, Arrese M, Zanlungo S. Recent insights on the role of cholesterol in non-alcoholic fatty liver disease. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*. 2015;1852(9):1765-1778.

33. Jimba S, Nakagami T, Takahashi M, et al. Prevalence of non-alcoholic fatty liver disease and its association with impaired glucose metabolism in japanese adults. *Diabetic Med*. 2005;22(9):1141-1145.

34. Gaggini M, Morelli M, Buzzigoli E, DeFronzo RA, Bugianesi E, Gastaldelli A. Non-alcoholic fatty liver disease (NAFLD) and its connection with insulin resistance, dyslipidemia, atherosclerosis and coronary heart disease. *Nutrients*. 2013;5(5):1544-1560.

35. Bae JC, Rhee EJ, Lee WY, et al. Combined effect of nonalcoholic fatty liver disease and impaired fasting glucose on the development of type 2 diabetes: A 4-year retrospective longitudinal study. *Diabetes Care*. 2011;34(3):727-729.

36. Lindström J, Neumann A, Sheppard K, et al. Take action to prevent diabetes–the IMAGE toolkit for the prevention of type 2 diabetes in europe. *Hormone and Metabolic research*. 2010;42(S 01):S37-S55.

37. Schwarz PE, Riemenschneider H. Slowing down the progression of type 2 diabetes: We need fair, innovative, and disruptive action on environmental and policy levels! *Diabetes Care*. 2016;39 Suppl 2:S121-6.

38. Mishra M, Kumar H, Bajpai S, Singh RK, Tripathi K. Level of serum IL-12 and its correlation with endothelial dysfunction, insulin resistance, proinflammatory cytokines and lipid profile in newly diagnosed type 2 diabetes. *Diabetes Res Clin Pract*. 2011;94(2):255-261.

39. Wegner M, Winiarska H, Bobkiewicz-Kozłowska T, Dworacka M. IL-12 serum levels in patients with type 2 diabetes treated with sulphonylureas. *Cytokine*. 2008;42(3):312-316.

40. Jin F, Nathan C, Radzioch D, Ding A. Secretory leukocyte protease inhibitor: A macrophage product induced by and antagonistic to bacterial lipopolysaccharide. *Cell*. 1997;88(3):417-426.

41. López-Bermejo A, Ortega FJ, Castro A, Ricart W, Fernández-Real J. The alarm secretory leukocyte protease inhibitor increases with progressive metabolic dysfunction. *Clinica Chimica Acta*. 2011;412(11):1122-1126.

42. Krieger M. Structures and functions of multiligand lipoprotein receptors: Macrophage scavenger receptors and LDL receptor-related protein (LRP). *Annu Rev Biochem*. 1994;63(1):601-637.

43. Kzhyshkowska J, Neyen C, Gordon S. Role of macrophage scavenger receptors in atherosclerosis. *Immunobiology*. 2012;217(5):492-502.

44. van der Pouw Kraan, Tineke CTM, Chen WJ, Bunck MC, et al. Metabolic changes in type 2 diabetes are reflected in peripheral blood cells, revealing aberrant cytotoxicity, a viral signature, and hypoxia inducible factor activity. *BMC medical genomics*. 2015;8(1):20.

45. Obata-Onai A, Hashimoto S, Onai N, et al. Comprehensive gene expression analysis of human NK cells and CD8(+) T lymphocytes. *Int Immunol*. 2002;14(10):1085-1098.

46. El Mesallamy HO, Hamdy NM, Mostafa DM, Amin AI. The serine protease granzyme b as an inflammatory marker, in relation to the insulin receptor cleavage in human obesity and type 2 diabetes mellitus. *Journal of Interferon & Cytokine Research*. 2014;34(3):179-186.

47. DECODE Study Group, the European Diabetes Epidemiology Group. Glucose tolerance and cardiovascular mortality: Comparison of fasting and 2-hour diagnostic criteria. *Arch Intern Med*. 2001;161(3):397-405.

48. Qiao Q, Pyorala K, Pyorala M, et al. Two-hour glucose is a better risk predictor for incident coronary heart disease and cardiovascular mortality than fasting glucose. *Eur Heart J*. 2002;23(16):1267-1275.

49. Barr EL, Zimmet PZ, Welborn TA, et al. Risk of cardiovascular and all-cause mortality in individuals with diabetes mellitus, impaired fasting glucose, and impaired glucose tolerance: The australian diabetes, obesity, and lifestyle study (AusDiab). *Circulation*. 2007;116(2):151-157.

50. Droumaguet C, Balkau B, Simon D, et al. Use of HbA1c in predicting progression to diabetes in french men and women: Data from an epidemiological study on the insulin resistance syndrome (DESIR). *Diabetes Care*. 2006;29(7):1619-1625.

Table 1. Demographics of the study population according to prediabetes/control (PR/C), dyslipidaemia (D/0) and hypercholesterolaemia subtype status (HC/NC).

	C_0	PR_0	C_D	PR_D
Number of subjects	256	79	598	307
Age (years)	40.1 (4.87)	41.2 (5.00)	41.4 (5.07) [a]	42.4 (4.97) [a]
Males (%)	26.2	41.8 [a]	42.0 [a]	63.8 [a]
Total cholesterol (mmol/l)	4.36 (0.40)	4.33 (0.51)	5.42 (0.89) [a]	5.59 (0.89) [a]
LDL cholesterol (mmol/l)	2.51 (0.35)	2.53 (0.43)	3.53 (0.76) [a]	3.63 (0.79) [a]
HDL cholesterol (mmol/l)	1.49 (0.25)	1.40 (0.27) [a]	1.34 (0.35) [a]	1.25 (0.32) [a]
Non-HDL cholesterol (mmol/l)	2.88 (0.36)	2.92 (0.49)	4.09 (0.83) [a]	4.33 (0.86) [a]
Triglycerides (mmol/l)	0.82 (0.28)	0.88 (0.29)	1.25 (0.73) [a]	1.68 (1.55) [a]
Systolic BP (mmHg)	113 (12.3)	117 (12.0) [a]	118 (13.6) [a]	123 (14.0) [a]
Diastolic BP (mmHg)	70.3 (8.90)	73.9 (9.78) [a]	73.7 (10.3) [a]	78.5 (10.4) [a]
Hs-C-reactive protein (mg/l)	1.09 (1.78)	2.23 (4.47) [a]	1.37 (2.03) [a]	1.77 (2.26) [a]
Glucose (mmol/l)	4.99 (0.34)	5.64 (0.47) [a]	5.08 (0.33) [a]	5.71 (0.41) [a]
HbA1c (%)	5.29 (0.18)	5.58 (0.28) [a]	5.35 (0.17) [a]	5.63 (0.26) [a]
HbA1c (mmol/l)	34.4 (2.00)	37.4 (2.95) [a]	35.0 (1.92) [a]	38.0 (2.76) [a]
Insulin (mU/l)	5.73 (3.40)	8.68 (6.04) [a]	7.59 (5.05) [a]	11.0 (7.43) [a]
Body mass index (kg/m^2)	23.6 (3.39)	26.2 (4.90) [a]	25.6 (3.99) [a]	28.3 (4.75) [a]
Waist circumference (cm)	82.5 (10.2)	90.4 (14.3) [a]	88.9 (11.9) [a]	97.7 (12.8) [a]
Daily smokers (%)	7.81	17.7 [a]	13.2 [a]	16.2 [a]

	C_{NC}	PR_{NC}	C_{HC}	PR_{HC}
Number of subjects	362	127	480	247
Age (years)	39.9 (4.79)	40.9 (4.87)	41.8 (5.07)[b]	42.8 (4.99)[b]
Males (%)	28.7	48.0[b]	42.5[b]	63.6[b]
Total cholesterol (mmol/l)	4.36 (0.51)	4.39 (0.53)	5.64 (0.73)[b]	5.75 (0.78)[b]
LDL cholesterol (mmol/l)	2.52 (0.35)	2.53 (0.38)	3.75 (0.64)[b]	3.84 (0.67)[b]
HDL cholesterol (mmol/l)	1.41 (0.34)	1.29 (0.34)[b]	1.38 (0.32)	1.29 (0.29)[b]
Non-HDL cholesterol (mmol/l)	2.95 (0.42)	3.10 (0.52)[b]	4.27 (0.72)[b]	4.47 (0.78)[b]
Triglycerides (mmol/l)	0.97 (0.50)	1.27 (0.77)[b]	1.15 (0.54)[b]	1.39 (0.62)[b]
Systolic BP (mmHg)	114 (12.2)	119 (12.3)[b]	118 (13.9)[b]	123 (14.4)[b]
Diastolic BP (mmHg)	71.3 (9.61)	75.3 (10.0)[b]	73.6 (10.2)[b]	78.5 (10.6)[b]
Hs-C-reactive protein (mg/l)	1.39 (2.27)	2.06 (3.74)	1.19 (1.72)	1.80 (2.34)[b]
Glucose (mmol/l)	5.01 (0.35)	5.70 (0.48)[b]	5.08 (0.32)[b]	5.69 (0.40)[b]
HbA1c (%)	5.31 (0.18)	5.58 (0.26)[b]	5.36 (0.18)[b]	5.64 (0.25)[b]
HbA1c (mmol/l)	34.5 (1.95)	37.4 (2.86)[b]	35.1 (1.95)[b]	38.2 (2.69)[b]
Insulin (mU/l)	6.50 (4.42)	10.3 (7.48)[b]	7.21 (4.49)[b]	9.99 (6.13)[b]
Body mass index (kg/m^2)	24.1 (3.87)	27.3 (5.23)[b]	25.6 (3.85)[b]	28.1 (4.69)[b]
Waist circumference (cm)	84.1 (10.9)	94.1 (14.8)[b]	88.9 (11.9)[b]	96.8 (12.6)[b]
Daily smokers (%)	11.3	20.5[b]	11.4	14.6

Definitions/Abbreviations: PR_0, prediabetes without dyslipidaemia; PR_D, prediabetes with dyslipidaemia; C_D, non-prediabetic subjects with dyslipidaemia; C_0, healthy subjects without prediabetes or dyslipidaemia; **:** PR_{NC}, prediabetes without hyper-LDL cholesterolaemia; PR_{HC}, prediabetes with hyper-LDL cholesterolaemia; C_{HC}, non-prediabetic subjects with hyper-LDL cholesterolaemia; C_{NC}, healthy subjects without prediabetes or hyper-LDL cholesterolaemia; Hs, High sensitive.

Statistics: t-test or χ^2 test when appropriate. Values are mean (±SD) or proportions. [a] Difference as compared to C_0, $p<0.05$. [b] Difference as compared to C_{NC}, $p<0.05$

Table 2. Pathways enriched in all prediabetic subjects (PR) in comparison to all control subjects without prediabetes (C). All pathways were up-regulated as indicated by a positive NES.

	NES	Enrichment p-value	FDR	FWER
Model 1				
Steroid biosynthesis[a]	2.13	<0.001	0.008	0.006
Cholesterol biosynthesis[b]	2.04	<0.001	0.007	0.009
Superpathway of cholesterol biosynthesis[b]	1.91	<0.001	0.009	0.034
Superpathway of methionine degradation[b]	1.97	0.002	0.006	0.019
Model 2				
Steroid biosynthesis[a]	2.06	0.002	0.010	0.010
Cholesterol biosynthesis[b]	1.99	<0.001	0.005	0.007
Superpathway of methionine degradation[b]	1.97	0.002	0.015	0.045
Model 3				
Steroid biosynthesis[a]	2.06	<0.001	0.014	0.012
Cholesterol biosynthesis[b]	2.00	<0.001	0.003	0.006

Statistics: **Model 1**: Gene set enrichment analysis adjusted for age, sex, BMI and the first 7 PCs; **Model 2:** Model 1 + additionally adjusted for waist circumference (instead of BMI), systolic and diastolic BP, smoking and insulin; **Model 3**: Model 2 + additionally adjusted for Hs-CRP. [a] KEGG pathways, [b] HumanCyc pathways. **Abbreviations:** FDR, false discovery rate; FWER, family-wise error rate; NES, normalized enrichment score; PC, principal component; Hs, High sensitive.

Table 3. Pathways enriched in prediabetic subjects without dyslipidaemia (PR_0) in comparison to control subjects without prediabetes and dyslipidaemia (C_0). All pathways were up-regulated as indicated by a positive NES.

	NES	Enrichment p-value	FDR	FWER
Model 1				
Steroid biosynthesis[a]	2.04	<0.001	0.020	0.022
Cholesterol biosynthesis[b]	2.10	<0.001	0.001	0.001
Superpathway of cholesterol biosynthesis[b]	1.94	0.002	0.007	0.015
Downstream signalling in naïve CD8+ T cells[c]	1.99	<0.001	0.020	0.038
IL12-mediated signalling events[c]	2.00	0.002	0.035	0.036
Model 2				
Steroid biosynthesis[a]	1.96	<0.001	0.048	0.044
Cholesterol biosynthesis[b]	1.97	<0.001	0.009	0.015
Downstream signalling in naïve CD8+ T cells[c]	2.09	0.002	0.006	0.007
IL12-mediated signalling events[c]	1.99	0.002	0.014	0.030
Model 3				
Downstream signalling in naïve CD8+ T cells[c]	2.08	<0.001	0.016	0.012
IL12-mediated signalling events[c]	2.06	<0.001	0.011	0.022

Statistics: Model 1: Gene set enrichment analysis adjusted for age, sex, BMI and the first 7 PCs; **Model 2:** Model 1 + additionally adjusted for waist circumference (instead of BMI), systolic and diastolic BP, smoking and insulin; **Model 3:** Model 2 + additionally adjusted for Hs-CRP. [a] KEGG pathways, [b] HumanCyc pathways, and [c] NCI PID pathways. **Abbreviations:** FDR, false discovery rate; FWER, family-wise error rate; NES, normalized enrichment score; PC, principal component; Hs, high sensitive.

Table 4. Pathways enriched in subjects who had prediabetes but no hyper-LDL cholesterolaemia (PR_{NC}) in comparison to those without prediabetes and hyper-LDL cholesterolaemia (C_{NC}) (LDL ≤ 3.0 mmol/l). All pathways were up-regulated as indicated by a positive NES.

	NES	Enrichment p-value	FDR	FWER
Model 1				
Superpathway of cholesterol biosynthesis[a]	1.92	0.002	0.052	0.030
Model 2				
Superpathway of cholesterol biosynthesis[a]	1.87	0.002	0.054	0.042
Model 3				
Superpathway of cholesterol biosynthesis[a]	1.96	<0.001	0.015	0.012

Statistics: **Model 1**: Gene set enrichment analysis adjusted for age, sex, BMI and the first 7 PCs; **Model 2**: Model 1 + additionally adjusted for waist circumference (instead of BMI), systolic and diastolic BP, smoking and insulin; **Model 3**: Model 2 + additionally adjusted for Hs-CRP. [a] HumanCyc pathways. **Abbreviations:** FDR, false discovery rate; FWER, family-wise error rate; NES, normalized enrichment score; PC, principal component; Hs, high sensitive.

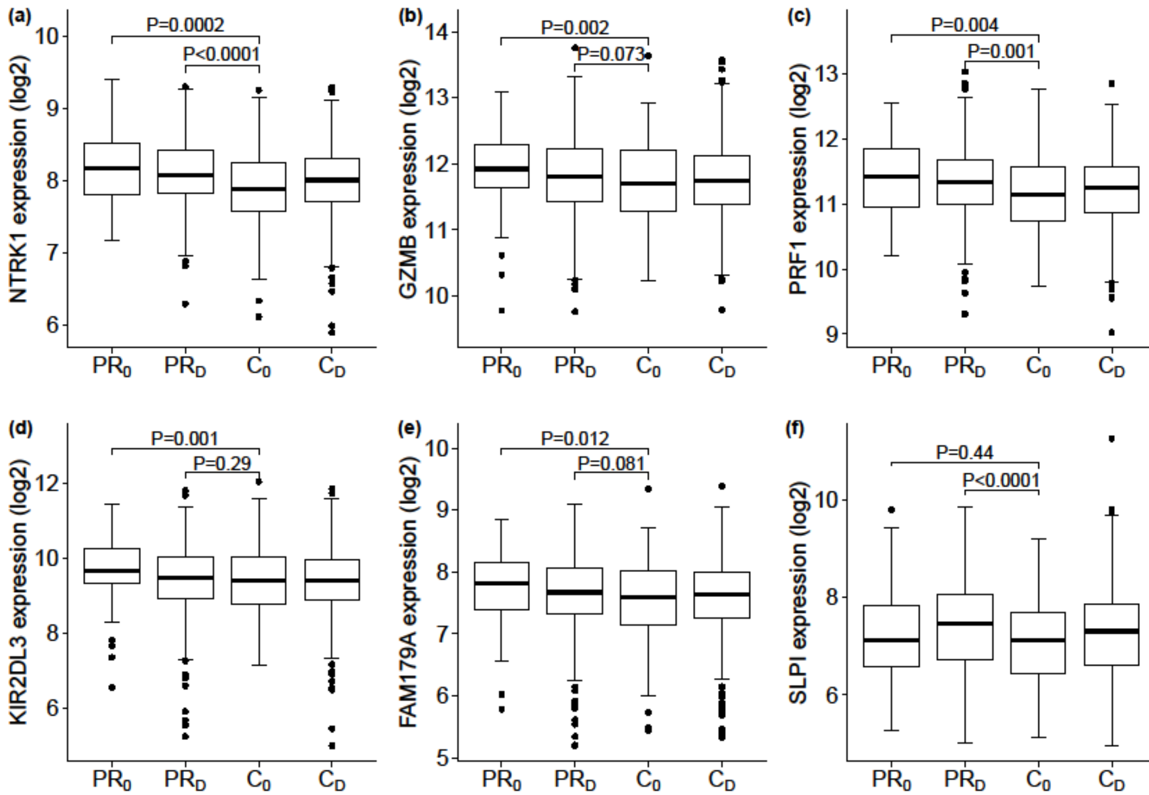

Figure 1. Gene expression changes of (a) NTRK1, (b) GZMB, (c) PRF1, (d) KIR2DL3, (e) FAM179A and (f) SLPI genes over prediabetes and control phenotypes.