

Biomarkers for type 2 diabetes and impaired fasting glucose using a non-targeted metabolomics approach

Cristina Menni^{1*}, Eric Fauman^{2*}, Idil Erte¹, John R. B. Perry^{1,3,4,5}, Gabi Kastenmüller⁶, So-Youn Shin^{7,8}, Ann-Kristin Petersen⁹, Craig Hyde¹⁰, Maria Psatha¹, Kirsten J. Ward¹, Wei Yuan¹, Mike Milburn¹¹, Colin N.A. Palmer¹², Timothy M. Frayling⁴, Jeff Trimmer¹³, Jordana T. Bell¹, Christian Gieger⁹, Rob Mohnney¹¹, Mary Julia Brosnan^{13*}, Karsten Suhre^{6,14†}, Nicole Soranzo^{7†‡}, Tim D Spector^{1†‡}

¹ Department of Twin Research and Genetic Epidemiology, Kings College London, London SE1 7EH, UK.

² Pfizer Worldwide Research and Development, Computational Sciences Center of Emphasis, 35 Cambridgepark Drive, Cambridge MA 02140, USA.

³ Wellcome Trust Centre for Human Genetics, Roosevelt Drive, University of Oxford, Oxford, UK.

⁴ Genetics of Complex Traits, Exeter Medical School, University of Exeter, UK.

⁵ Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, USA.

⁶ Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, Ingolstädter Landstrasse 1, Neuherberg, 85764, Germany.

⁷ Human Genetics, Wellcome Trust Sanger Institute, Hinxton CB10 1HH, UK.

⁸ MRC Centre for Causal Analyses in Translational Epidemiology, School of Social and Community Medicine, University of Bristol, Bristol, UK.

⁹ Institute of Genetic Epidemiology, Helmholtz Zentrum München, Ingolstädter Landstrasse 1, Neuherberg, 85764, Germany.

¹⁰ Pfizer Worldwide Research and Development, Clinical Research Statistics, 558 Eastern Point Rd, Groton CT 06340, USA.

¹¹ Metabolon Inc., Raleigh-Durham, NC 27709, USA.

¹² Biomedical Research Institute, University of Dundee, Ninewells Hospital and Medical School, Dundee, UK.

¹³ Cardiovascular and Metabolic Diseases, Pfizer Worldwide Research and Development, 620 Memorial Drive, Cambridge, MA 02139, USA.

¹⁴ Department of Physiology and Biophysics, Weill Cornell Medical College in Qatar, Education City, Qatar Foundation, Doha, Qatar.

*These authors contributed equally

†These authors contributed equally

‡Corresponding authors

Correspondence to:

Prof Tim D Spector

Department of Twin Research and Genetic Epidemiology

Kings College London

London SE1 7EH, UK

Dr Nicole Soranzo

Human Genetics

Wellcome Trust Sanger Institute

Genome Campus

Hinxton, CB10 1HH

Tel. +44 (0)1223 492364

Fax. +44 (0)1223 491919

ns6@sanger.ac.uk

Running title: Metabolite biomarkers of diabetes

Word count: 1987

Number of tables: 2

Number of figures: 2

ABSTRACT

Using a non-targeted metabolomics approach of 447 fasting plasma metabolites, we searched for novel molecular markers that arise before and after hyperglycaemia in a large population-based cohort of 2,204 females (115 Type 2 Diabetes-T2D cases, 192 individuals with impaired fasting glucose- IFG and 1,897 controls) from TwinsUK.

Forty-two metabolites from three major fuel sources, carbohydrates, lipids and proteins, were found to significantly correlate with T2D after adjusting for multiple testing; of these, 22 were previously reported as associated with T2D or insulin resistance. Fourteen metabolites were found to be associated with IFG. Among the metabolites identified, the branched-chain-keto-acid metabolite 3-methyl-2-oxovalerate, was the strongest predictive biomarker for IFG after glucose (OR=1.65, 95%CI=1.39,1.95, $P=8.46 \times 10^{-9}$) and was moderately heritable ($h^2=0.20$). The association was replicated in an independent population (n=720, OR=1.68, 95%CI=1.34, 2.11, $P=6.52 \times 10^{-6}$) and validated in 189 Twins with urine metabolomics taken at the same time as plasma (OR=1.87, 95%CI=1.27,2.75, $P=1 \times 10^{-3}$). Results confirm an important role for catabolism of branched-chain-amino-acids in T2D and IFG.

In conclusion, this T2D-IFG biomarker study has surveyed the broadest panel of non-targeted metabolites to date, revealing both novel and known associated metabolites and providing potential novel targets for clinical prediction and a deeper understanding of causal mechanisms.

Introduction

Currently, stratification of individuals at risk for Type 2 Diabetes (T2D) within the general population is based on well-established factors such as age, BMI and fasting glucose (1). Although these factors contribute considerably to disease risk, they may not identify at risk individuals before the disease process is well underway.

Recently, a number of studies have found several metabolites to be correlated with insulin resistance and T2D (2-6), and T2D-associated metabolic profiles have been identified ten to fifteen years before the diagnosis/onset of the disease (7-9). . To help preventive strategies, and maximize the potential for existing effective interventions, it is important to characterize the molecular changes that take place in the development of T2D.

We aim to understand other biochemical changes, in addition to hyperglycemia, that take place at the onset of T2D using the largest metabolomic screening approach to date. We assess over 400 metabolites to determine which metabolomic profiles are correlated with T2D and impaired fasting glucose (IFG) in a large cohort of females from TwinsUK with independent replication.

Research Design and Methods

We analyzed data from 2,204 females from TwinsUK for whom non-targeted plasma metabolomic profiling was available along with glucose/diabetic information (10). Subjects were classified into three groups based on fasting glucose levels at time of initial sampling and at subsequent visits (on average 2.08(1.21) visits): T2D cases (fasting glucose ≥ 7 mmol/L/physician's letter confirming diagnosis), individuals with IFG

(5.6mmol/L<fasting glucose<7 mmol/L), T2D controls (3.9mmol/L<fasting glucose<5mmol/L).

Metabolomics measurements (on plasma and urine)

Non-targeted metabolite detection and quantification was conducted by the metabolomics provider *Metabolon Inc.* (Durham, USA) on TwinsUK fasting plasma samples as described previously (11) and on 187 spot urine samples taken at the same time as plasma.

Replication cohort for 3-methyl-2-oxovalerate

We included 536 individuals with IFG and 184 controls identified via fasting glucose from the follow-up study KORA F4 (Cooperative Health Research in the Region of Augsburg) (12) with fasting metabolomic profiles for 3-methyl-2-oxovalerate.

Statistical analysis

We inverse normalised the data as the metabolite concentrations were not normally distributed. To avoid spurious false-positive associations due to small sample size, we excluded metabolic traits with more than 20% missing values.

For each T2D-control and IFG-control contrast we ran random intercept logistic regressions adjusting for age and BMI at the time of sampling, metabolite batch and family relatedness. We used a conservative Bonferroni correction to account for multiple testing thus giving a significant threshold of 1×10^{-4} (0.05/447).

Taking advantage of the twin design of our study, for each metabolite significantly associated with one or more contrasts, we estimated heritability using structural equation modelling. For contrasts between each disease class and controls, we ran a stepwise linear regression

including all the significant metabolites to look for metabolites independently associated with T2D and IFG respectively.

We investigated further the role of 3-methyl-2-oxovalerate, the strongest predictive biomarker after glucose, by: (i) replicating the result in an independent population (KORA) (ii) validating the result in urine (TwinsUK); (iii) investigating the underlying genetic influences using GWAS data and (iv) assessing causality of the metabolite-IFG association by Mendelian Randomisation.

As 3-methyl-2-oxovalerate was associated with SNP rs1440581 (13) we tested this SNP for association with T2D status using cases and controls from the DIAGRAM consortium and by genotyping rs1440581 in 4961 T2D cases and 5948 controls from GoDarts (KASPar system, KBiosciences; genotyping success rate >95%, HWE $P > 0.05$).

Results

Metabolites associated with T2D and IFG

Levels of 447 fasting plasma metabolites (281 known and 176 unknown) were obtained for 115 T2D cases, 192 individuals with IFG and 1,897 normo-glycaemic controls. The demographic characteristics are presented in **Table 1**. After adjusting for age, BMI, metabolite batch and family relatedness, 42 of the 447 metabolites tested showed significant differences among T2D cases and controls with a Bonferroni corrected cut-off of 1×10^{-4} ($=0.05/447$). As depicted in **Figure 1a**, the 42 metabolites fall into 3 principal classes: 12 are lipids (primarily medium and long chain free fatty acids), 7 are carbohydrates, 9 are branched-chain-amino-acids (BCAAs) or derivatives and 14 are unknown. Beside glucose, a one standard deviation change in metabolite level resulted in T2D effect sizes ranging from OR 1.05 to 3.36 for adrenate (22:4n6) and mannose respectively (**Table 2**).

We repeated the analysis for the IFG group contrasting with controls. This revealed 14 significantly associated metabolites, 8 of which were also identified for T2D (**Table 2**). Six of the 14 metabolites are related to BCAA catabolism, 3 are carbohydrates and 2 are lipids (**Figure 1b**). Two metabolites were independently associated with IFG in the stepwise regression including these 14 metabolites: glucose and 3-methyl-2-oxovalerate. Using 1297 monozygotic and 1200 dizygotic twin pairs, we estimated heritability for each metabolite identified in one or more contrasts. The calculated heritabilities ranged from 0 to 65%.

Effect sizes, association statistics, heritability estimates and literature references for both contrasts are shown in **Table 2**.

Investigating the role of 3-methyl-2-oxovalerate in IFG

3-methyl-2-oxovalerate is the branched-chain-keto-acid (BCKA) derivative of isoleucine, one of three BCAAs. We found it to be significantly associated with IFG in 536 individuals with IFG and 184 normo-glycaemic controls from the KORA population (OR=1.68, 95%CI=1.34,2.11, $P=6.52 \times 10^{-6}$) and in the inverse variance fixed-effect meta-analysis of the results (OR=1.66, 95%CI=1.45;1.90, $P=2.62 \times 10^{-13}$), thus replicating our result.

We next studied 94 individuals with IFG and 95 control subjects from TwinsUK with urine metabolomic profiles available at the same time as plasma sampling. 3-methyl-2-oxovalerate correlated significantly with IFG (OR= 1.87, 95%CI=1.27;2.75, $P=1 \times 10^{-3}$) thus suggesting that urine could also be used to test for elevated 3-methyl-2-oxovalerate.

Genetics of 3-methyl-2-oxovalerate and GWAS

3-methyl-2-oxovalerate has a heritability $h^2=0.20$ (95%CI=0.08;0.33) (**Table 2**). Our companion metabolite GWAS (13) revealed that 3-methyl-2-oxovalerate is strongly

associated with SNPs upstream of the *PPMIK* gene on chromosome 4 (top hit SNP rs1440581, Beta=-0.014(0.017), $P=1.21 \times 10^{-16}$).

We assessed whether the association between 3-methyl-2-oxovalerate and IFG is consistent with a causal hypothesis. Given the magnitude of effect between 3-methyl-2-oxovalerate and T2D and between rs1440581 and 3-methyl-2-oxovalerate, we theoretically estimated assuming causality (using Mendelian Randomization) that the biomarker raising allele 'C' would be associated with increased risk of T2D (OR=1.10,95%CI=1.03;1.18). We obtained in the actual data a meta-analysed test statistic of OR=1.03(95%CI=1.00;1.05, $P=0.08$) after analyzing rs1440581 in 17,132 T2D cases and 62,810 controls (DIAGRAM consortium (14)+replication in GoDARTs (15)).

Discussion

Using the largest biochemical screening approach to date (447 metabolites), we searched for molecular markers that arise before and after hyperglycaemia in a large cross-sectional population of women. We identified forty-two metabolites with high statistical significance associated with T2D and fourteen metabolites associated with IFG. Although diabetes is considered to be primarily a disorder of glucose, we find other dimensions, apart from carbohydrates, in the metabolic space that associate with T2D and IFG: namely lipids and amino-acids.

Though many metabolites identified have previously been associated with T2D or insulin resistance (**Table 2**), we are the first to report their associations with IFG. Moreover as IFG presents itself before T2D in prospective studies, this could improve disease prediction and early intervention. Also, this is the first study on IFG using a wide untargeted platform such as *Metabolon* (a previous IFG study (16) used a different platform with little overlap). We

also report the novel association of the BCKA 3-methyl-2-oxovalerate with IFG both in plasma and in urine.

Carbohydrates

As expected, glucose itself showed the strongest association with both T2D and IFG, followed by mannose which is consistent with previous findings (3; 5; 17-20) and emphasize the importance of other glucose and non-glucose pathways. In particular, dimethylarginine (SDMA+ADMA) have been more associated with the micro and macrovascular complications than with the pathogenesis of diabetes itself; while the association of malate and arabinose with T2D were never reported.

Lipids

T2D patients often present with elevated lipid profiles and within this study lipids (primarily the free fatty acids) make up the second largest group of T2D/IFG-associated metabolites.

Lipids with the longest chain (adrenate (22:4n6) and arachidonate (20:4n6)) are elevated in IFG patients compared to controls. Similarly, lipids with shorter chain (5-dodecenoate (12:1n7), heptanoate (7:0) and pelargonate (9:0)) are depleted in T2D patients relative to controls.

In contrast, the fatty acid chains found in triglyceride molecules in diabetes seem to act differently. Rhee *et al.* found that triglycerides containing longer chain fatty acids were associated with a decreased risk of diabetes, while triglycerides containing shorter chains were associated with an increased risk (9). This contrasting pattern of association may reflect alterations in triglyceride lipolysis, which could either contribute to or be a result of the dysregulation of glucose metabolism.

Among the lipids identified, the novel associations include the fatty acid 15-methylpalmitate and the medium fatty acid 5-dodecenoate (12:1n7).

Amino-acids

The third major group of metabolites are amino-acids. Within this group the BCAAs, valine, isoleucine and leucine and their BCKAs, 3-methyl-2-oxovalerate, 4-methyl-2-oxopentanoate and 3-methyl-2-oxobutyrate are significantly elevated in both individuals with IFG and subjects T2D compared to controls.

Elevated BCAA levels have previously been associated with increased risk of incident T2D (3; 5; 7; 21) and independently predict future T2D onset (7). Breakdown products of BCAAs (propionylcarnitine and α -methylbutyrylcarnitine and isovalerylcarnitine) were also found to be elevated (22). However, whereas previous targeted panels did not include BCKAs, the non-targeted approach used here highlighted specific effects on these important intermediates in BCAA catabolism (see **Figure 2**). These suggest that it may be the breakdown of BCAAs which is associated with diabetes, and not specifically the elevated levels of BCAAs themselves. Consistent with this idea, a knock-out of the mouse *BCAT2* gene, which blocks the first step in BCAA metabolism, results in greatly elevated plasma levels of BCAAs and yet these animals have improved glucose control, insulin sensitivity and resistance to diet induced obesity (23).

3-methyl-2-oxovalerate

Among the metabolites identified, the BCKA 3-methyl-2-oxovalerate is the strongest predictor of IFG after and independently of glucose. BCAA catabolism occurs primarily in the mitochondria, proceeding through BCAA transaminase, and then through the branched-chain alpha-keto-acid dehydrogenase, a complex of three separate gene products. In our

companion GWAS (13), SNP rs1440581 had the strongest associations with all BCAAs, BCKAs and the C3-acylcarnitine, propionylcarnitine. This SNP is upstream of *PPMIK* mitochondrial phosphatase which dephosphorylates and thereby activates the BCKD, clearly highlighting the importance of mitochondrial function for plasma levels of BCAAs and BCKAs (24). The centrality of mitochondrial function to BCAA catabolism and metabolic disease has been noted before (25). BCAA dysregulation could be a cause and/or consequence of mitochondrial dysfunction. Increased BCAA catabolism, resulting in increased BCAA catabolic intermediates, may impair mitochondrial oxidation of glucose and lipids, potentially resulting in mitochondrial stress and impaired insulin secretion and action. Reduced mitochondrial function in T2D and IFG may reduce the capacity of the mitochondria to break down BCAAs, resulting in elevated levels of BCAAs and BCKAs.

The current study has several strengths. It used a non-targeted metabolomic approach that identifies a wide-range of biochemicals beside lipids. TwinsUK has phenotypic longitudinal data available which allowed us to accurately classify subjects as cases, IFG and controls. The availability of urine metabolites, genetic data and twin design enabled us to explore the biological implication of 3-methyl-2-oxovalerate further. Finally, the robustness of our results is highlighted by the fact that we confirm many previous findings and our main association reported is clearly replicated in an independent cohort and validated in urine.

Our study has some limitations. Our discovery sample consisted of women only and some metabolites might be influenced by gender-specific hormones. Unknown metabolites might not really be new, but merely not yet identified. Finally, our Mendelian Randomization analysis, was unable to firmly support or reject causality for the association of 3-methyl-2-oxovalerate and IFG. To explore this further additional 3-methyl-2-oxovalerate associated variants need to be identified and tested, boosting power and reducing the impact of potential unwanted pleiotropic confounding.

Conclusion

Here we find evidence that multiple metabolites from three major fuel sources, carbohydrates, lipids and proteins are robust risk factors for the development of both IFG and T2D. Further work is encouraged by these data, including understanding the role of diet and microbiota on the free fatty acids relationships with T2D.

Acknowledgments

Conflict of interest: Michael V. Milburn and Robert P. Mohny are employees of Metabolon Inc. Eric Fauman, Craig Hyde, Jeff Trimmer and Julia Brosnan are full time employees and share holders of Pfizer.

We gratefully acknowledge the contributions of Sally John, Carol Hicks, Li Xi and Vinicius Bonato from Pfizer Worldwide Research and Development, USA.

We thank Gabriela Surdulescu and Dylan Hodgkiss from the Department of Twin Research & Genetic epidemiology, King's College London, UK, for sample selection, sample handling and shipment.

We are grateful to the the DIAGRAM consortium for sharing data <http://diagram-consortium.org/>.

Finally, we wish to express our appreciation to all study participants of the TwinsUK and KORA studies for donating their blood and time.

TDS is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Author Contributions: *Conceived and designed the experiments:* JT, CG, MJB, KS, NS, TDS. *Performed the experiments:* CNAP, MM, RM. *Analyzed the data:* CM, IE, CH. *Analyzed the replication dataset:* GK. *Contributed reagents/materials/analysis tools:* JRBP, SYS, AKP, CNAP, TMF, WY. *Wrote the paper:* CM, EF, JRBP, MP, MJB, TS, *Reviewed/edited manuscript:* KJW, TMF, JTB, RM, KS, NS.

Funding

Part of this work was funded by Pfizer Worldwide Research and Development. The study was funded by the Wellcome Trust; European Community's Seventh Framework Programme

(FP7/2007-2013). The study also receives support from the National Institute for Health Research (NIHR) Clinical Research Facility at Guy's & St Thomas' NHS Foundation Trust and NIHR Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. Tim Spector is an NIHR senior Investigator and is holder of an ERC Advanced Principal Investigator award. TDS is an NIHR senior Investigator and is holder of an ERC Advanced Principal Investigator award.

JRBP is supported by the Wellcome Trust as a Sir Henry Wellcome Postdoctoral Research Fellow (092447/Z/10/Z).

SYS is supported by a Post-Doctoral Research Fellowship from the Oak Foundation.

NS's team is supported by the Wellcome Trust (Grant Codes WT098051 and WT091310), the EUFP7 (EPIGENESYS Grant Code 257082 and BLUEPRINT Grant Code HEALTH-F5-2011-282510)"

References

1. Noble D, Mathur R, Dent T, Meads C, Greenhalgh T: Risk models and scores for type 2 diabetes: systematic review. *BMJ* 2011;343:d7163
2. Galazis N, Iacovou C, Haoula Z, Atiomo W: Metabolomic biomarkers of impaired glucose tolerance and type 2 diabetes mellitus with a potential for risk stratification in women with polycystic ovary syndrome. *Eur J Obstet Gynecol Reprod Biol* 2012;160:121-130
3. Gall WE, Beebe K, Lawton KA, Adam KP, Mitchell MW, Nakhle PJ, Ryals JA, Milburn MV, Nannipieri M, Camastra S, Natali A, Ferrannini E: alpha-hydroxybutyrate is an early biomarker of insulin resistance and glucose intolerance in a nondiabetic population. *PLoS One* 2010;5:e10883
4. Stancakova A, Civelek M, Saleem NK, Soininen P, Kangas AJ, Cederberg H, Paananen J, Pihlajamaki J, Bonnycastle LL, Morken MA, Boehnke M, Pajukanta P, Lusi AJ, Collins FS, Kuusisto J, Ala-Korpela M, Laakso M: Hyperglycemia and a Common Variant of GCKR Are Associated With the Levels of Eight Amino Acids in 9,369 Finnish Men. *Diabetes* 2012;61:1895-1902
5. Suhre K, Meisinger C, Doring A, Altmaier E, Belcredi P, Gieger C, Chang D, Milburn MV, Gall WE, Weinberger KM, Mewes HW, Hrabce de Angelis M, Wichmann HE, Kronenberg F, Adamski J, Illig T: Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. *PLoS One* 2010;5:e13953
6. Floegel A, Stefan N, Yu Z, Muhlenbruch K, Drogan D, Joost HG, Fritsche A, Haring HU, Hrabce de Angelis M, Peters A, Roden M, Prehn C, Wang-Sattler R, Illig T, Schulze MB, Adamski J, Boeing H, Pischon T: Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. *Diabetes* 2013;62:639-648
7. Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, Lewis GD, Fox CS, Jacques PF, Fernandez C, O'Donnell CJ, Carr SA, Mootha VK, Florez JC, Souza A,

Melander O, Clish CB, Gerszten RE: Metabolite profiles and the risk of developing diabetes. *Nat Med* 2011;17:448-453

8. Tabak AG, Jokela M, Akbaraly TN, Brunner EJ, Kivimaki M, Witte DR: Trajectories of glycaemia, insulin sensitivity, and insulin secretion before diagnosis of type 2 diabetes: an analysis from the Whitehall II study. *Lancet* 2009;373:2215-2221

9. Rhee EP, Cheng S, Larson MG, Walford GA, Lewis GD, McCabe E, Yang E, Farrell L, Fox CS, O'Donnell CJ, Carr SA, Vasan RS, Florez JC, Clish CB, Wang TJ, Gerszten RE: Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. *J Clin Invest* 2011;121:1402-1411

10. Moayyeri A, Hammond CJ, Valdes AM, Spector TD: Cohort Profile: TwinsUK and Healthy Ageing Twin Study. *Int J Epidemiol* 2012;

11. Menni C, Kastenmuller G, Petersen AK, Bell JT, Psatha M, Tsai PC, Gieger C, Schulz H, Erte I, John S, Brosnan MJ, Wilson SG, Tsaprouni L, Lim EM, Stuckey BGA, Deloukas P, Mohny RP, Suhre K, Spector TD, Valdes AM: Metabolomic markers reveal novel pathways of aging and early development in human populations. *Int J Epidemiol* 2013;In press

12. Wichmann HE, Gieger C, Illig T: KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* 2005;67 Suppl 1:S26-30

13. Shin SY, Fauman E, Petersen A, Krumsiek J, Huang J, Erte I, Yang TP, Walter K, Menni C, Chen L, Valdes A, L. HC, Wang V, Ziemek D, Roberts P, Xi L, Mohny RP, Milburn MV, John SL, Trimmer J, Theis FJ, Overington JP, Kastenmuller G, K S, Brosnan MJ, Gieger C, Spector T, Soranzo N: Hundreds of genetic loci underpin metabolic diversity in humans. Submitted 2013;

14. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segre AV, Steinthorsdottir V, Strawbridge RJ, Khan H, Grallert H, Mahajan A, Prokopenko I, Kang HM, Dina C, Esko T, Fraser RM, Kanoni S, Kumar A, Lagou V, Langenberg C, Luan J, Lindgren CM, Muller-

Nurasyid M, Pechlivanis S, Rayner NW, Scott LJ, Wiltshire S, Yengo L, Kinnunen L, Rossin EJ, Raychaudhuri S, Johnson AD, Dimas AS, Loos RJ, Vedantam S, Chen H, Florez JC, Fox C, Liu CT, Rybin D, Couper DJ, Kao WH, Li M, Cornelis MC, Kraft P, Sun Q, van Dam RM, Stringham HM, Chines PS, Fischer K, Fontanillas P, Holmen OL, Hunt SE, Jackson AU, Kong A, Lawrence R, Meyer J, Perry JR, Platou CG, Potter S, Rehnberg E, Robertson N, Sivapalaratnam S, Stancakova A, Stirrups K, Thorleifsson G, Tikkanen E, Wood AR, Almgren P, Atalay M, Benediktsson R, Bonnycastle LL, Burt N, Carey J, Charpentier G, Crenshaw AT, Doney AS, Dorkhan M, Edkins S, Emilsson V, Eury E, Forsen T, Gertow K, Gigante B, Grant GB, Groves CJ, Guiducci C, Herder C, Hreidarsson AB, Hui J, James A, Jonsson A, Rathmann W, Klopp N, Kravic J, Krjutskov K, Langford C, Leander K, Lindholm E, Lobbens S, Mannisto S, Mirza G, Muhleisen TW, Musk B, Parkin M, Rallidis L, Saramies J, Sennblad B, Shah S, Sigurethsson G, Silveira A, Steinbach G, Thorand B, Trakalo J, Veglia F, Wennauer R, Winckler W, Zabaneh D, Campbell H, van Duijn C, Uitterlinden AG, Hofman A, Sijbrands E, Abecasis GR, Owen KR, Zeggini E, Trip MD, Forouhi NG, Syvanen AC, Eriksson JG, Peltonen L, Nothen MM, Balkau B, Palmer CN, Lyssenko V, Tuomi T, Isomaa B, Hunter DJ, Qi L, Shuldiner AR, Roden M, Barroso I, Wilsgaard T, Beilby J, Hovingh K, Price JF, Wilson JF, Rauramaa R, Lakka TA, Lind L, Dedoussis G, Njolstad I, Pedersen NL, Khaw KT, Wareham NJ, Keinanen-Kiukaanniemi SM, Saaristo TE, Korpi-Hyovalti E, Saltevo J, Laakso M, Kuusisto J, Metspalu A, Collins FS, Mohlke KL, Bergman RN, Tuomilehto J, Boehm BO, Gieger C, Hveem K, Cauchi S, Froguel P, Baldassarre D, Tremoli E, Humphries SE, Saleheen D, Danesh J, Ingelsson E, Ripatti S, Salomaa V, Erbel R, Jockel KH, Moebus S, Peters A, Illig T, de Faire U, Hamsten A, Morris AD, Donnelly PJ, Frayling TM, Hattersley AT, Boerwinkle E, Melander O, Kathiresan S, Nilsson PM, Deloukas P, Thorsteinsdottir U, Groop LC, Stefansson K, Hu F, Pankow JS, Dupuis J, Meigs JB, Altshuler D, Boehnke M, McCarthy MI: Large-scale association analysis provides

insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* 2012;44:981-990

15. Zhou K, Bellenguez C, Spencer CC, Bennett AJ, Coleman RL, Tavendale R, Hawley SA, Donnelly LA, Schofield C, Groves CJ, Burch L, Carr F, Strange A, Freeman C, Blackwell JM, Bramon E, Brown MA, Casas JP, Corvin A, Craddock N, Deloukas P, Dronov S, Duncanson A, Edkins S, Gray E, Hunt S, Jankowski J, Langford C, Markus HS, Mathew CG, Plomin R, Rautanen A, Sawcer SJ, Samani NJ, Trembath R, Viswanathan AC, Wood NW, Harries LW, Hattersley AT, Doney AS, Colhoun H, Morris AD, Sutherland C, Hardie DG, Peltonen L, McCarthy MI, Holman RR, Palmer CN, Donnelly P, Pearson ER: Common variants near ATM are associated with glycemic response to metformin in type 2 diabetes. *Nat Genet* 2011;43:117-120

16. Wang-Sattler R, Yu Z, Herder C, Messias AC, Floegel A, He Y, Heim K, Campillos M, Holzapfel C, Thorand B, Grallert H, Xu T, Bader E, Huth C, Mittelstrass K, Doring A, Meisinger C, Gieger C, Prehn C, Roemisch-Margl W, Carstensen M, Xie L, Yamanaka-Okumura H, Xing G, Ceglarek U, Thiery J, Giani G, Lickert H, Lin X, Li Y, Boeing H, Joost HG, de Angelis MH, Rathmann W, Suhre K, Prokisch H, Peters A, Meitinger T, Roden M, Wichmann HE, Pischon T, Adamski J, Illig T: Novel biomarkers for pre-diabetes identified by metabolomics. *Mol Syst Biol* 2012;8:615

17. Crawford SO, Hoogeveen RC, Brancati FL, Astor BC, Ballantyne CM, Schmidt MI, Young JH: Association of blood lactate with type 2 diabetes: the Atherosclerosis Risk in Communities Carotid MRI Study. *Int J Epidemiol* 2010;39:1647-1655

18. van Doorn M, Vogels J, Tas A, van Hoogdalem EJ, Burggraaf J, Cohen A, van der Greef J: Evaluation of metabolite profiles as biomarkers for the pharmacological effects of thiazolidinediones in Type 2 diabetes mellitus patients and healthy volunteers. *Br J Clin Pharmacol* 2007;63:562-574

19. Zhang X, Wang Y, Hao F, Zhou X, Han X, Tang H, Ji L: Human serum metabonomic analysis reveals progression axes for glucose intolerance and insulin resistance statuses. *J Proteome Res* 2009;8:5188-5195
20. Zhao X, Fritsche J, Wang J, Chen J, Rittig K, Schmitt-Kopplin P, Fritsche A, Haring HU, Schleicher ED, Xu G, Lehmann R: Metabonomic fingerprints of fasting plasma and spot urine reveal human pre-diabetic metabolic traits. *Metabolomics* 2010;6:362-374
21. Fiehn O, Garvey WT, Newman JW, Lok KH, Hoppel CL, Adams SH: Plasma metabolomic profiles reflective of glucose homeostasis in non-diabetic and type 2 diabetic obese African-American women. *PLoS One* 2010;5:e15234
22. Newgard CB, An J, Bain JR, Muehlbauer MJ, Stevens RD, Lien LF, Haqq AM, Shah SH, Arlotto M, Slentz CA, Rochon J, Gallup D, Ilkayeva O, Wenner BR, Yancy WS, Jr., Eisenson H, Musante G, Surwit RS, Millington DS, Butler MD, Svetkey LP: A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab* 2009;9:311-326
23. She P, Reid TM, Bronson SK, Vary TC, Hajnal A, Lynch CJ, Hutson SM: Disruption of BCATm in mice leads to increased energy expenditure associated with the activation of a futile protein turnover cycle. *Cell Metab* 2007;6:181-194
24. Lu G, Sun H, She P, Youn JY, Warburton S, Ping P, Vondriska TM, Cai H, Lynch CJ, Wang Y: Protein phosphatase 2Cm is a critical regulator of branched-chain amino acid catabolism in mice and cultured cells. *J Clin Invest* 2009;119:1678-1687
25. Newgard CB: Interplay between lipids and branched-chain amino acids in development of insulin resistance. *Cell Metab* 2012;15:606-614
26. Wu JH, Lemaitre RN, Manichaikul A, Guan W, Tanaka T, Foy M, Kabagambe EK, Djousse L, Siscovick D, Fretts AM, Johnson C, King IB, Psaty BM, McKnight B, Rich SS, Chen YD, Nettleton JA, Tang W, Bandinelli S, Jacobs DR, Jr., Browning BL, Laurie CC, Gu

X, Tsai MY, Steffen LM, Ferrucci L, Fornage M, Mozaffarian D: Genome-Wide Association Study Identifies Novel Loci Associated With Concentrations of Four Plasma Phospholipid Fatty Acids in the De Novo Lipogenesis Pathway: Results from the CHARGE Consortium. *Circ Cardiovasc Genet* 2013;

Table 1. Demographic characteristics of the study populations. *Mean (SD)*

		Cases	Impaired fasting glucose	Controls
TUK plasma	N	115	192	1897
	Females, N(%)	115(100%)	192(100%)	1897(100%)
	Age, years	63.00(9.61)	60.01(12.40)	50.02(14.43)
	BMI, kg/m ²	30.58(6.32)	27.89(5.66)	25.42(4.55)
	Fasting glucose, mmol/L	7.58(2.90)	6.02(0.36)	4.46(0.29)
	Fasting insulin, mg/dl	19.28(22.77)	11.66(10.81)	6.90(4.43)
	Creatinine	74.80(15.98)	76.67(13.09)	74.98(17.76)
TUK urine	N	-	94	95
	Females, N(%)	-	94(100%)	95(100%)
	Age, years	-	63.74(8.03)	63.17(9.32)
	BMI, kg/m ²	-	29.22(5.17)	26.59(4.65)
	Fasting glucose, mmol/L	-	5.94(0.27)	4.53(0.29)
	Fasting insulin, mg/dl	-	13.50(11.74)	6.02(2.76)
	Creatinine	-	76.50(15.60)	78.00(14.7)
KORA	N	-	536	184
	Females, N(%)	-	197(37%)	124(67%)
	Age, years	-	64.2(5.3)	63.2(5.7)
	BMI, kg/m ²	-	29.1(4.1)	26.4(3.8)
	Fasting glucose, mmol/L	-	6.04(0.35)	4.75(0.18)

Table 2. List of significant metabolites in one or more comparison. Metabolite concentrations provided by *Metabolon* are relative quantification that were median normalized using the run day median and then inverse normalised. All analyses are adjusted for age, BMI, batch effect and family relatedness. Heritability estimates are calculated on 1200 DZ and 1297 MZ pairs. Upward arrow (↑) indicates metabolite concentration increases in subjects with T2D or IFG. Downward arrow (↓) indicates metabolite concentration decreased in subjects with T2D or IFG. Note that when the metabolite is significantly associated to both contrasts, the direction is the same in both. However, if the metabolite is not significantly associated in both contrast, the arrow represents the concentration of the significant contrast. The ids of the unknown metabolites are consistently used by *Metabolon*. The fact that they are unknown means that their chemical identity has not yet been identified.

Metabolite	Super-p	Sub-p	T2D-controls		IFG -controls		h^2 [95%CI]	Previously reported as associated with T2D or IR*
			OR [95%CI]	P	OR[95%CI]	P		
2-hydroxybutyrate (AHB)	a-a	Cysteine, methionine, SAM, taurine metabolism	↑ 1.62[1.26,2.08]	1.63x10 ⁻⁴	1.67[1.39,2.02]	7.09X10 ⁻⁸	0.32[0.18,0.45]	Gall 2010 (3)†
N-acetylglucine	a-a	Glycine, serine and threonine metabolism	↓ 0.59[0.46,0.76]	6.26x10 ⁻⁵	0.89[0.76,1.05]	1.69X10 ⁻¹	0.45 [0.32,0.55]	Floegel 2013 (6) Fiehn 2010 (21)
citrulline	a-a	Urea cycle; arginine-, proline-, metabolism	↓ 0.54[0.42,0.71]	6.14x10 ⁻⁶	0.83[0.79,0.99]	3.92X10 ⁻²	0.58[0.49,0.61]	Suhre 2010 (5)

dimethylarginine (SDMA + ADMA)	a-a	Urea cycle; arginine-, proline-, metabolism	↓	0.55[0.41,0.74]	5.70×10^{-5}	0.94[0.79,1.12]	4.84×10^{-1}	0.10[0,0.26]	
proline	a-a	Urea cycle; arginine-, proline-, metabolism	↑	1.86[1.42,2.44]	5.85×10^{-6}	1.14[0.95,1.37]	1.45×10^{-1}	0.54[0.43,0.65]	Suhre 2010 (5)
3-methyl-2-oxobutyrate	a-a	Valine, leucine and isoleucine metabolism	↑	1.45[1.15,1.83]	1.96×10^{-3}	1.54[1.30,1.82]	7.74×10^{-7}	0.01[0,0.15]	Gall 2010 (3) †
3-methyl-2-oxovalerate	a-a	Valine, leucine and isoleucine metabolism	↑	2.18[1.68,2.84]	7.17×10^{-9}	1.65[1.39,1.95]	8.46×10^{-9}	0.20[0.08,0.33]	Gall 2010 (3) †
4-methyl-2-oxopentanoate	a-a	Valine, leucine and isoleucine metabolism	↑	2.00[1.54,2.59]	2.04×10^{-7}	1.58[1.33,1.87]	1.30×10^{-7}	0.17[0.04,0.29]	Fiehn 2010(21)
isoleucine	a-a	Valine, leucine and isoleucine metabolism	↑	2.35[1.79,3.08]	6.31×10^{-10}	1.55[1.29,1.86]	2.69×10^{-6}	0.50[0.37,0.58]	Suhre 2010 (5) Gall 2010 (3) † Wang 2011 (7)
leucine	a-a	Valine, leucine and isoleucine metabolism	↑	2.20[1.70,2.85]	1.99×10^{-9}	1.46[1.21,1.76]	7.82×10^{-5}	0.46[0.33,0.55]	Suhre 2010 (5) Wang 2011(7) Fiehn 2010 (21)
valine	a-a	Valine, leucine and isoleucine metabolism	↑	2.19[1.67,2.89]	3.42×10^{-8}	1.36[1.14,1.62]	5.50×10^{-4}	0.42[0.29,0.53]	Suhre 2010 (5) Wang 2011(7)

fructose	ch	Fructose, mannose, galactose, starch, and sucrose metabolism	↑	2.21[1.68,2.92]	2.14X10 ⁻⁸	1.55[1.27,1.89]	1.33X10 ⁻⁵	0.21[0.05,0.29]	Suhre 2010 (5) Fiehn 2010 (21)
mannose	ch	Fructose, mannose, galactose, starch, and sucrose metabolism	↑	3.36[2.44,4.62]	9.83x10 ⁻¹⁴	1.83[1.49,2.24]	4.93X10 ⁻⁹	0.36[0.22,0.44]	Suhre 2010 (5) Gall 2010 (3) †
1,5-anhydroglucitol (1,5-AG)	ch	Glycolysis, gluconeogenesis, pyruvate metabolism	↓	0.38[0.28,0.50]	2.42x10 ⁻¹¹	0.77[0.65,0.92]	4.48X10 ⁻³	0.65[0.58,0.68]	Suhre 2010 (5)
glucose	ch	Glycolysis, gluconeogenesis, pyruvate metabolism	↑	5.5[3.57,8.48]	1.12x10 ⁻¹⁴	3.18[2.37,4.26]	1.32x10 ⁻¹⁴	0.18[0.03,0.33]	Suhre 2010 (5) Fiehn 2010(21)
lactate	ch	Glycolysis, gluconeogenesis, pyruvate metabolism	↑	1.84[1.43,2.37]	2.11x10 ⁻⁶	1.16[0.97,1.39]	9.85X10 ⁻²	0.12[0.01,0.24]	Crawford 2010 (17)
arabinose	ch	Nucleotide sugars, pentose metabolism	↑	2.07[1.50,2.85]	8.31X10 ⁻⁶	1.36[1.13,1.64]	1.17X10 ⁻³	0.22[0.06,0.34]	
malate	ch	Krebs cycle	↑	1.63[1.27,2.07]	9.12X10 ⁻⁵	1.26[1.05,1.51]	1.42X10 ⁻²	0.29[0.13,0.34]	
octanoylcarnitine	l	Carnitine metabolism	↓	0.62[0.49,0.79]	9.88X10 ⁻⁵	0.77[0.65,0.91]	2.99X10 ⁻³	0.52[0.41,0.64]	Gall 2010 (3) †
15-methylpalmitate (isobar with 2-methylpalmitate)	l	Fatty acid, branched	↓	0.61[0.48,0.77]	4.77x10 ⁻⁵	1.09[0.90,1.32]	4.04x10 ⁻¹	0.08[0,0.24]	
10-heptadecenoate (17:1n7)	l	Long chain fatty acid	↓	0.56[0.45,0.72]	2.31x10 ⁻⁶	0.98[0.80,1.19]	8.18x10 ⁻¹	0.27[0.15,0.40]	Suhre 2010 (5)

adrenate (22:4n6)	l	Long chain fatty acid	↑	0.95[0.72,1.25]	7.16×10^{-1}	1.54[1.29,1.85]	2.02×10^{-6}	0.25[0.10,0.34]	Suhre 2010 (5) Gall 2010 (3) †
arachidonate (20:4n6)	l	Long chain fatty acid	↑	0.91[0.71,1.16]	4.56×10^{-1}	1.5[1.26,1.78]	3.66×10^{-6}	0.27[0.15,0.40]	Suhre 2010 (5) Fiehn 2010 (21)
myristate (14:0)	l	Long chain fatty acid	↓	0.64[0.52,0.80]	6.30×10^{-5}	1.02[0.84,1.24]	8.45×10^{-1}	0.22[0.08,0.34]	Suhre 2010 (5)
myristoleate (14:1n5)	l	Long chain fatty acid	↓	0.44[0.34,0.57]	1.22×10^{-10}	0.87[0.72,1.05]	1.55×10^{-1}	0.33[0.22,0.45]	Suhre 2010 (5)
palmitoleate (16:1n7)	l	Long chain fatty acid	↓	0.50[0.39,0.64]	5.10×10^{-8}	0.95[0.78,1.16]	6.01×10^{-1}	0[0,0.1263]	
pentadecanoate (15:0)	l	Long chain fatty acid	↓	0.65[0.52,0.80]	7.02×10^{-5}	1.14[0.96,1.35]	1.28×10^{-1}	0.01[0,0.17]	Suhre 2010 (5)
5-dodecenoate (12:1n7)	l	Medium chain fatty acid	↓	0.43[0.33,0.55]	1.02×10^{-10}	0.89[0.73,1.08]	2.27×10^{-1}	0.40[0.28,0.52]	
heptanoate (7:0)	l	Medium chain fatty acid	↓	0.53[0.41,0.69]	2.14×10^{-6}	0.95[0.81,1.12]	5.61×10^{-1}	0.24[0.10,0.38]	Suhre 2010 (5)
pelargonate (9:0)	l	Medium chain fatty acid	↓	0.57[0.44,0.75]	5.54×10^{-5}	1.01[0.84,1.20]	9.55×10^{-1}	0.37[0.23,0.41]	Suhre 2010 (5)
palmitoyl sphingomyelin	l	Sphingolipid	↓	0.44[0.34,0.57]	5.30×10^{-10}	0.74[0.62,0.88]	5.17×10^{-4}	0.32[0.20,0.45]	Wu 2012 (26)
cholesterol	l	Sterol/Steroid	↓	0.52[0.40,0.68]	1.25×10^{-6}	0.88[0.74,1.05]	1.64×10^{-1}	0.26[0.15,0.39]	Suhre 2010 (5)
urate	n	Purine metabolism, urate metabolism	↑	0.92[0.70,1.22]	5.79×10^{-1}	1.45[1.20,1.74]	8.98×10^{-5}	0.54[0.53,0.65]	Suhre 2010 (5)
erythritol	x	Sugar, sugar substitute, starch	↑	1.31[1.02,1.69]	3.79×10^{-2}	1.49[1.22,1.81]	6.93×10^{-5}	0.28[0.15,0.41]	
X - 06246	u	Unknown	↑	1.71[1.36,2.15]	5.00×10^{-6}	1.43[1.21,1.69]	3.37×10^{-5}	0.10[0,0.25]	

X - 08402	u	Unknown	↓	0.50[0.40,0.64]	1.70×10^{-8}	0.89[0.75,1.06]	1.91×10^{-1}	0.27[0.15,0.39]
X - 10500	u	Unknown	↓	0.49[0.37,0.63]	1.24×10^{-7}	0.80[0.68,0.94]	7.18×10^{-3}	0.07[0,0.21]
X - 10506	u	Unknown	↑	1.82[1.35,2.45]	7.21×10^{-5}	1.36[1.16,1.60]	1.71×10^{-4}	0[0,0.16]
X - 10510	u	Unknown	↓	0.37[0.29,0.48]	2.93×10^{-13}	0.74[0.63,0.88]	5.09×10^{-4}	0.26[0.14,0.38]
X - 11315	u	Unknown	↓	0.51[0.39,0.66]	7.26×10^{-7}	0.96[0.81,1.15]	6.88×10^{-1}	0.63[0.53,0.66]
X - 11423	u	Unknown	↓	0.56[0.43,0.72]	7.60×10^{-6}	0.91[0.77,1.07]	2.53×10^{-1}	0.017[0,0.07]
X - 11497	u	Unknown	↓	0.58[0.45,0.76]	4.97×10^{-5}	0.83[0.70,0.98]	2.72×10^{-2}	0.20[0.037,0.32]
X - 11550	u	Unknown	↓	0.5[0.37,0.68]	5.50×10^{-6}	0.89[0.74,1.07]	2.28×10^{-1}	0.25[0.09,0.34]
X - 12442	u	Unknown	↓	0.57[0.45,0.73]	9.77×10^{-6}	0.87[0.73,1.05]	1.49×10^{-1}	0.39[0.26,0.52]
X - 12450	u	Unknown	↓	0.6[0.49,0.75]	6.35×10^{-6}	1.08[0.91,1.29]	3.77×10^{-1}	0.18[0.03,0.331]
X - 12696	u	Unknown	↓	0.47[0.35,0.63]	2.92×10^{-7}	0.78[0.66,0.93]	4.94×10^{-3}	0.47[0.33,0.52]
X - 13215	u	Unknown	↓	0.62[0.50,0.79]	6.38×10^{-5}	0.99[0.82,1.19]	9.29×10^{-1}	0.26[0.11,0.34]
X - 13496	u	Unknown	↓	0.60[0.48,0.75]	7.70×10^{-6}	0.93[0.79,1.10]	3.91×10^{-1}	0.04[0,0.20]

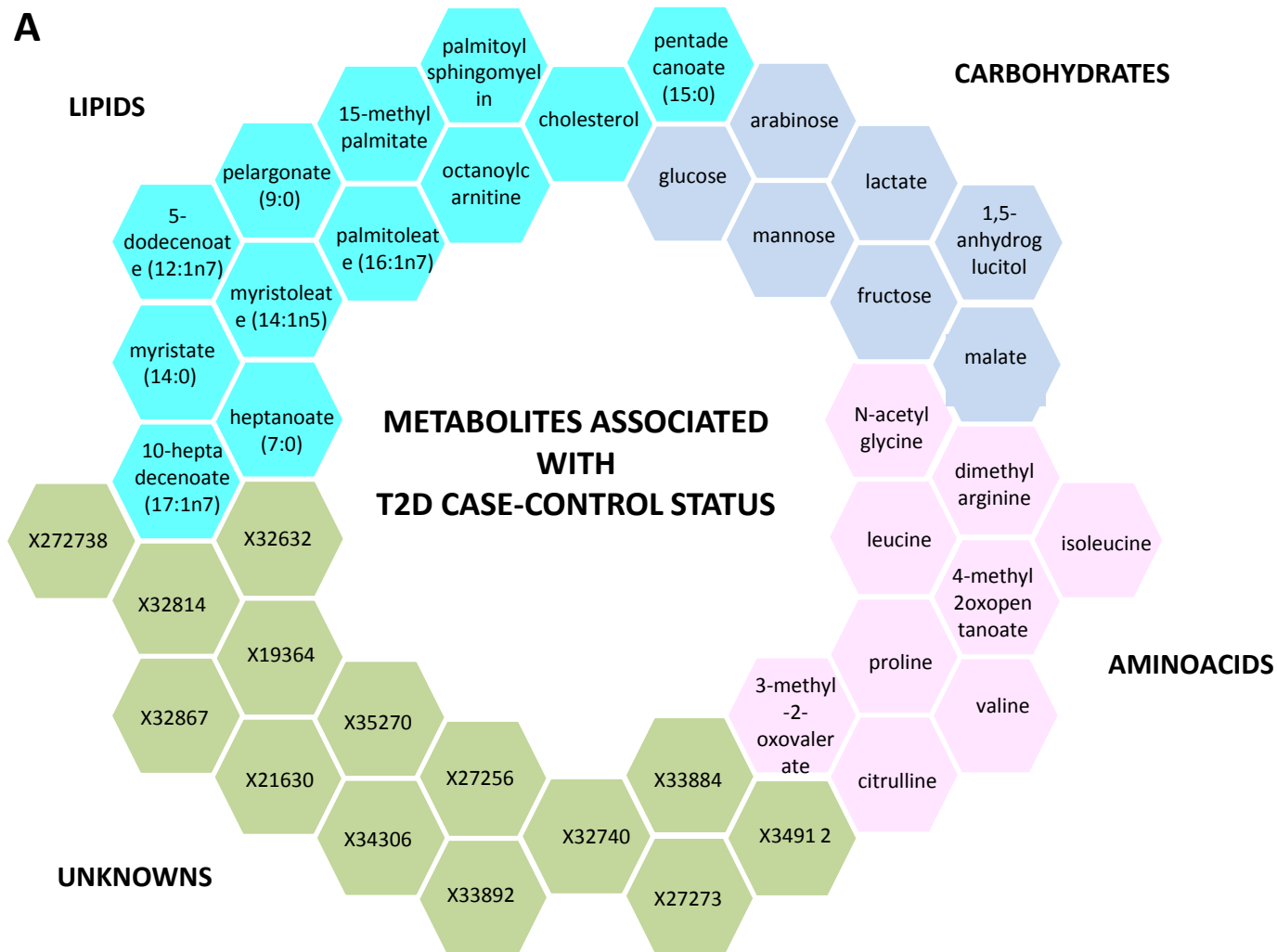
a-a = amino acid, ch = carbohydrate, l= lipid, n= nucleotide, x=xenobiotic, u=unknown.

*For each metabolite, the list of publications relating previous findings does not mean to be exhaustive.

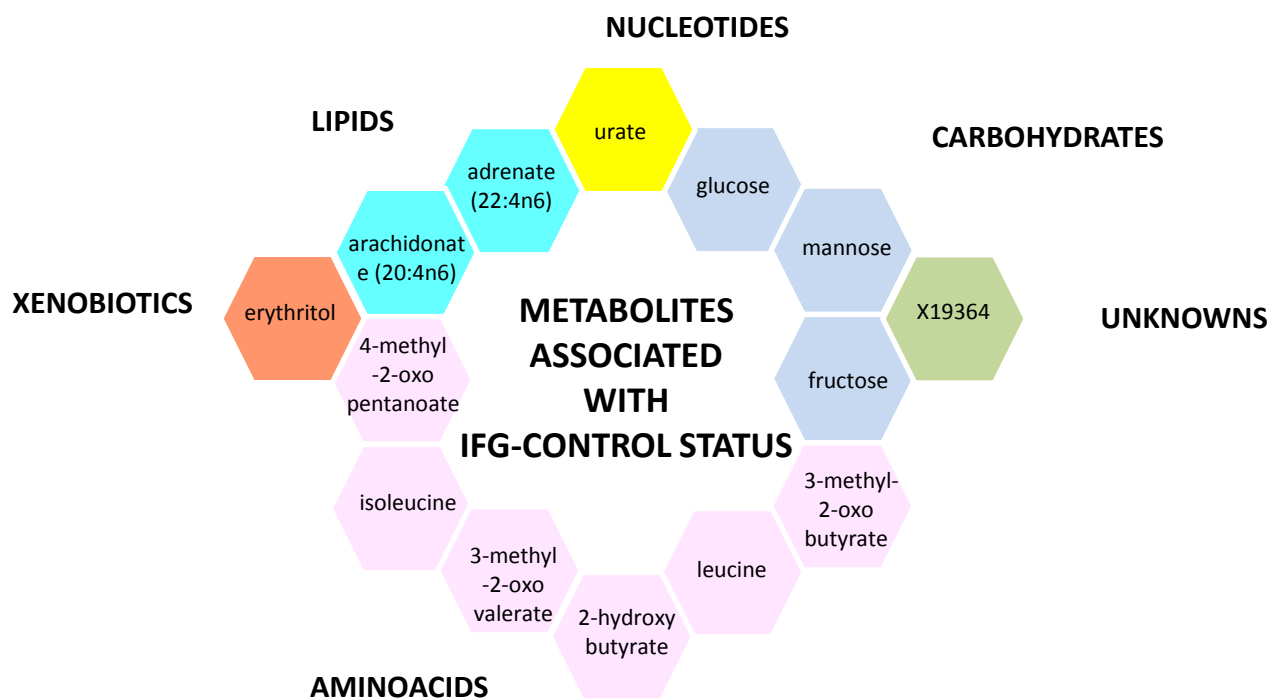
†Insulin resistance

Diabetes

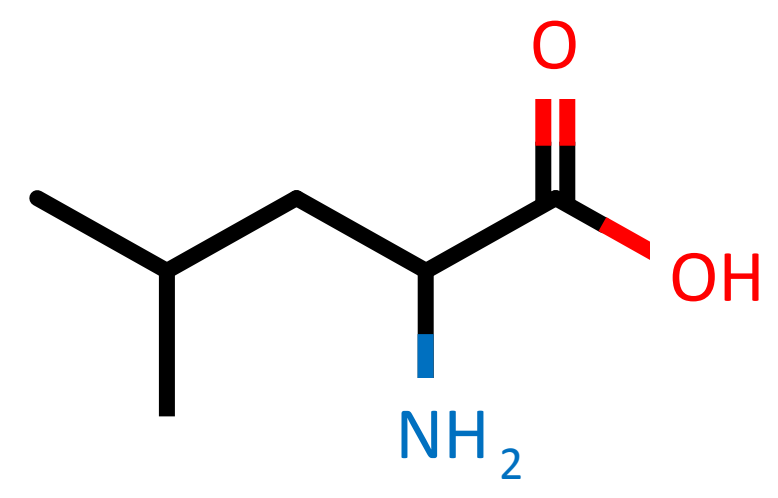
A



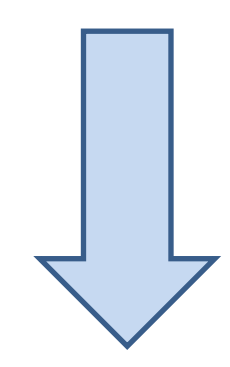
B



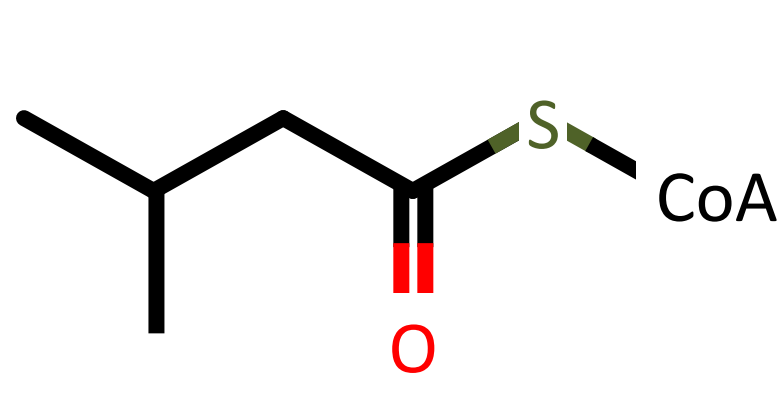
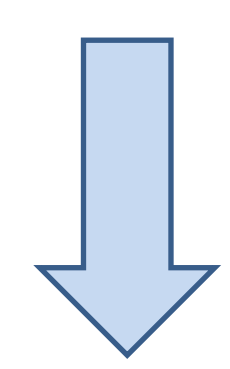
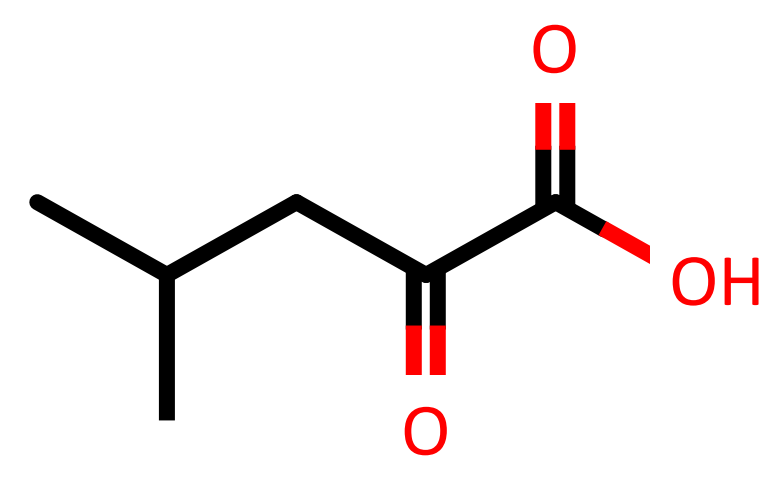
Leucine



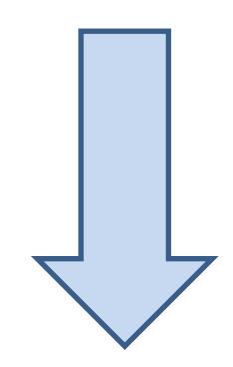
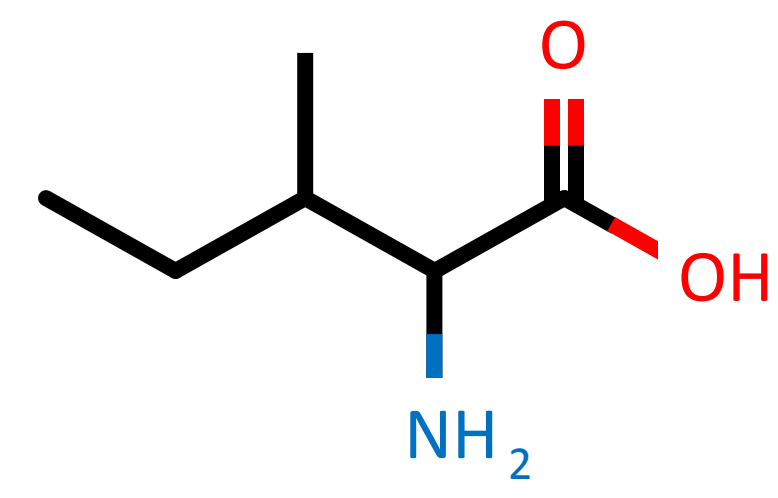
BCAT1
BCAT2



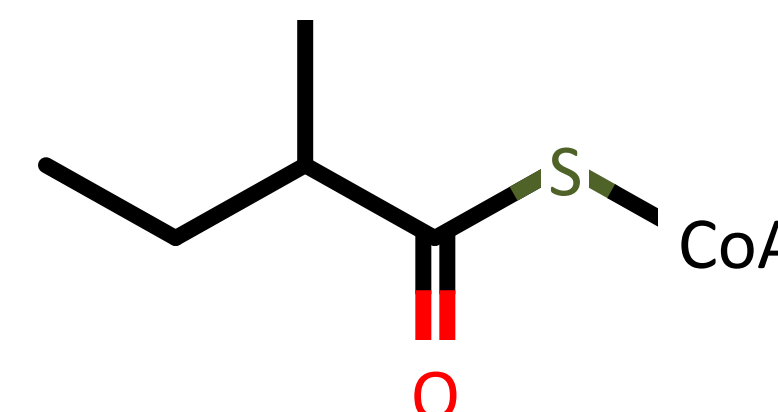
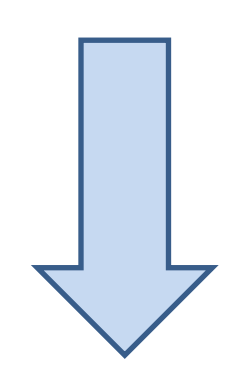
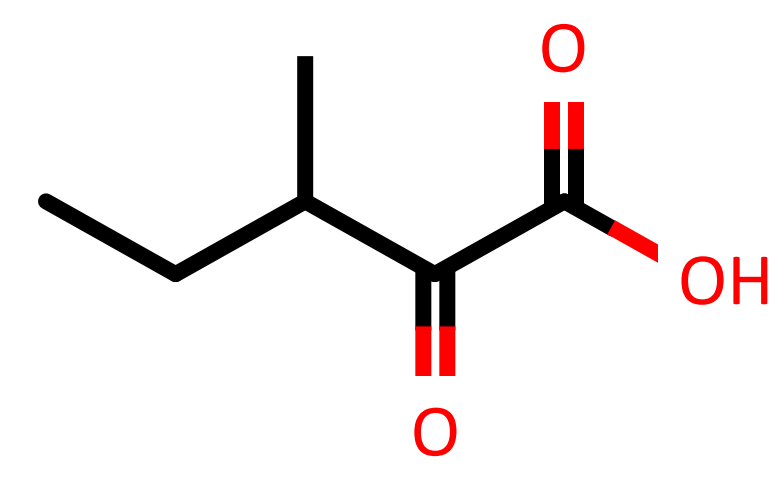
4-methyl-2-oxopentanoate



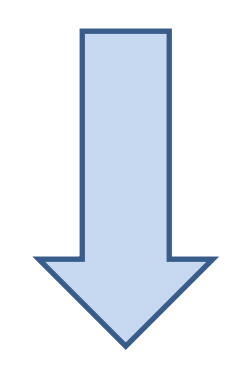
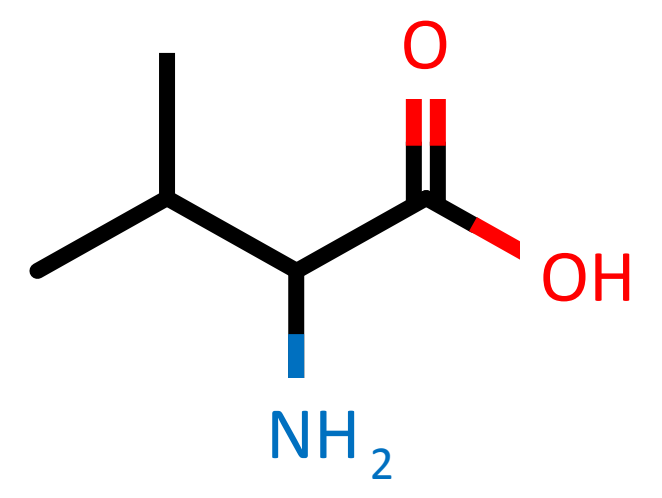
Isoleucine



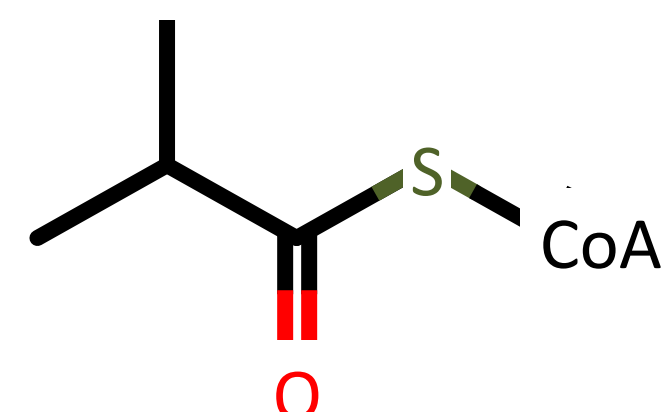
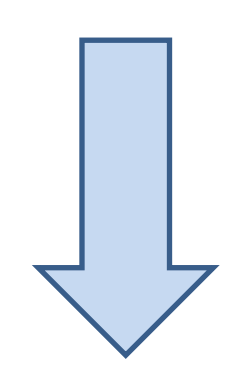
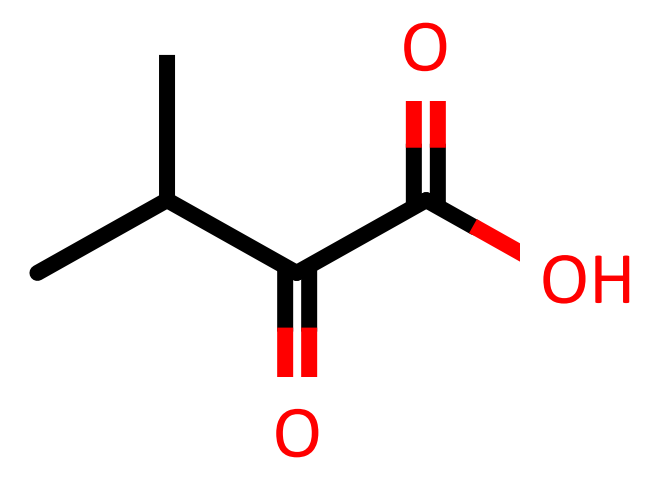
3-methyl-2-oxovalerate



Valine



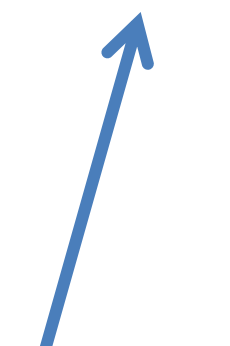
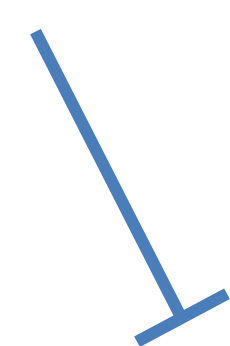
3-methyl-2-oxobutyrate



BCKDK

BCKD

PPM1K



C5 acylcarnitines

C3 acylcarnitine

