Genetic variants associated with glycine metabolism and their role in insulin sensitivity and type 2 diabetes

Weijia Xie¹, Andrew R. Wood¹, Valeriya Lyssenko², Michael N. Weedon¹, Joshua W. Knowles³, Sami Alkayyali², Themistocles L. Assimes³, Thomas Quertermous³, Fahim Abbasi³, Jussi Paananen⁴, Hans Häring⁵, Torben Hansen⁶, Oluf Pedersen⁶, Ulf Smith⁷, Markku Laakso⁴, MAGIC investigators, the DIAGRAM consortium, the GENESIS consortium, the RISC consortium, Jacqueline M. Dekker⁸, John J. Nolan⁹, Leif Groop², Ele Ferrannini¹⁰, Klaus-Peter Adam¹¹, Walter E. Gall¹¹, Timothy M. Frayling¹, Mark Walker¹²

Affiliations

¹Genetics of Complex Traits, Peninsula School of Medicine, University of Exeter, UK.²Lund University Diabetes Center, Department of Clinical Sciences, Diabetes and Endocrinology, Lund University, Malmoe, Sweden.³Department of Medicine, Stanford University School of Medicine, Stanford, California, USA.⁴Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland.⁵Department of Internal Medicine, Division of Endocrinology, Diabetology, Nephrology, Vascular Medicine and Clinical Chemistry, University of Tübingen, Tübingen, Germany.⁶Novo Nordisk Foundation Center for Basic Metabolic Research Faculty of Health Science, University of Copenhagen, Copenhagen, Denmark.⁷The Lundberg Laboratory for Diabetes Research, Department of Molecular and Clinical Medicine, Sahlgrenska Academy, Gothenburg, Sweden.⁸Department of Epidemiology and Biostatistics, VrijeUniversiteit Medical Center, Amsterdam, The Netherlands.⁹Steno Diabetes Center, Gentofte, Denmark.¹⁰Department of Internal Medicine, University of Pisa, Pisa, Italy.¹¹Metabolon, Inc., Research Triangle Park, North Carolina, US.¹²Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne, UK.¹³Faculty of Health Sciences, University of Southern Denmark,
Odense, Denmark.\textsuperscript{14} Hagedorn Research Institute, Copenhagen, Denmark.\textsuperscript{15} Institute of Biomedical Science, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark.\textsuperscript{16} Faculty of Health Sciences, University of Aarhus, Aarhus, Denmark.

Corresponding author:

Timothy M. Frayling

Peninsula College of Medicine and Dentistry
University of Exeter
Magdalen Road
Exeter
EX1 2LU
tel. +44 1392 262935
fax. +44 1392 262926
email: tim.frayling@pms.ac.uk

Word count: 3,996
Number of figures: 5
Number of tables: 3

Short Title: Glycine metabolism and insulin sensitivity
Abstract

Circulating metabolites associated with insulin sensitivity may represent useful biomarkers, but their causal role in insulin sensitivity and diabetes is less certain. We previously identified novel metabolites correlated with insulin sensitivity measured by the hyperinsulinemic-euglycemic clamp. The top-ranking metabolites were in the glutathione and glycine biosynthesis pathways. We aimed to identify common genetic variants associated with metabolites in these pathways and test their role in insulin sensitivity and type 2 diabetes. Using 1,004 non-diabetic individuals from the RISC study we performed a genome wide association study of 14 insulin sensitivity-related metabolites and one metabolite ratio. We replicated our results in the Botnia study (n=342). We assessed the association of these variants with diabetes-related traits in genome wide association study meta-analyses (GENESIS (including RISC, EUGENE2 and Stanford), MAGIC, DIAGRAM). We identified four associations with three metabolites; glycine (rs715 at CPS1), serine (rs478093 at PHGDH), betaine (rs499368 at SLC6A12; rs17823642 at BHMT); and one association signal with glycine/serine ratio (rs1107366 at ALDH1L1). There was no robust evidence for association between these variants and insulin resistance or diabetes. Genetic variants associated with genes in the glycine biosynthesis pathways do not provide consistent evidence for a role of glycine in diabetes-related traits.
Using mass-spectrometry based metabolomic approaches, recent studies have identified associations between small molecules and insulin sensitivity and type 2 diabetes (1-6). Previous studies in the RISC study identified novel associations between insulin sensitivity and small molecules including amino acids glycine, cysteine, isoleucine, and creatine, and the organic acids alpha-hydroxybutyrate (α-HB) and alpha-ketobutyrate (α-KB). Glycine was the amino acid most strongly associated with increased insulin sensitivity (4), a finding consistent with other studies (7-9)

Whilst some metabolites may represent important biomarkers, the causal directions of their associations to diabetes related traits are uncertain. It is important to understand the causal role or otherwise of these molecules in order to avoid an increasingly confusing picture of diabetes disease processes.

The identification of genetic variants strongly associated with metabolites may provide useful tools to help understand causal directions of correlated phenotypes. Genetic variants are unlikely to be influenced by disease processes or environmental factors, therefore provide robust tools in Mendelian randomisation to assess causal directions of correlated phenotypes (10). Recently, the principle of Mendelian randomisation has been used to provide evidence for a causal association between reduced sex hormone binding globulin (SHBG) levels and reduced B-type natriuretic peptide (BNP) levels and type 2 diabetes (11, 12), but evidence against causality between raised triglycerides and increased insulin resistance and type 2 diabetes (13).

In this study, we focused on the associations of glycine and glutathione biosynthesis pathways to type 2 diabetes, because, apart from the strong correlations identified in the RISC study, some other recent studies have provided evidence that high glycine level is associated with increased insulin sensitivity and decreased type 2 diabetes risks (7-9). In addition, type 2 diabetes patients have unrestrained gluconeogenesis and severely deficient
glutathione synthesis (14, 15). Glycine supplementation can improve deficient glutathione synthesis in type 2 diabetic patients, and glutathione supplementation can improve insulin sensitivity in non-diabetic individuals (15, 16). We hypothesized that glycine and glutathione pathways contribute to diabetes and insulin resistance. We aimed to identify genetic variants influencing circulating levels of metabolites in the glycine and glutathione pathways. We tested these variants in Mendelian randomisation analyses to examine the potential causal role of these metabolites in insulin resistance and type 2 diabetes.
RESEARCH DESIGN AND METHODS

1. Study participants.
We analysed non-diabetic participants of European ancestry from four studies who provided DNA for genome-wide genotyping and underwent a direct measure of insulin sensitivity. The four studies include: RISC (Relationship between Insulin Sensitivity and Cardiovascular disease risk; n=957), Botnia (n=339), EUGENE2 consortium (European network on Functional Genomics of type 2 diabetes; n=577) and Stanford Insulin Suppression Test (IST) Cohort (n=263). The descriptive characteristics of the RISC participants are shown in Table 1. In brief, we excluded individuals with cryptic relatedness using PLINK pairwise IBD estimation (PI_HAT>0.2). We excluded individuals with lipid disorders or diabetes, lipid medications, pregnancy, fasting plasma glucose ≥7.0 mmol/L; 2-h plasma glucose [on a 75-g oral glucose tolerance test (OGTT)] ≥11.0 mmol/L. The individual study characteristics, genotyping and phenotyping details are provided in the Online Supplemental Materials. We performed GWAS for metabolites in the RISC study, replicated the GWAS findings in the Botnia study, and carried out Mendelian randomisation analyses in RISC, EUGENE2 and Stanford IST to test the associations of genetic variants to insulin sensitivity.

2. Selection and Measurement of metabolites in RISC and Botnia studies.
We selected 14 metabolites for GWAS. The metabolites were selected based on the study of Gall et al. (2010) (4). We selected metabolites that were both available in the RISC study, and associated with insulin sensitivity (3, 5). Details are shown in Table 2. We selected metabolite ratios from the 14 metabolites based on two criteria: 1) the two metabolites were linked by one-step enzymatic reactions; 2) the ratio was associated with insulin sensitivity measured by hyperinsulinemic-euglycemic clamp (M-value). The glycine-serine ratio was the only one that satisfied both criteria (Figure 2d).
In both RISC and Botnia studies, metabolites were measured using multiple platform mass spectrometry technology (ultra-high performance liquid chromatography (UHPLC) and gas chromatography (GC)), as previously described (17-19).

3. GWAS of metabolites and metabolite ratios in RISC

The plasma concentrations of metabolites were fitted in a linear regression model with adjustment for age, sex and centres. Then the standardized residuals were normalized by inverse-normal transformation prior to GWAS. We performed GWAS with each metabolite using MACH2QTL based on an additive genetic model (20, 21).

For the glycine-serine ratio, we log 10-transformed the ratio, then adjusted for age, sex and centre in linear regression analyses. We performed GWAS using MACH2QTL in the same way as single metabolites’ concentrations.

4. Candidate-region association study of metabolites in RISC

Some of the key enzymes and transporters involved in the metabolism and transport of metabolites are known. We selected 34 genes for the fourteen metabolites, consisting of carrier-encoding or enzyme-encoding genes involved in the rate-limiting steps of the relevant biosynthetic pathways. The genes selected are listed in Table 2. We classified SNPs within 300kb of these genes as candidate SNPs. To prioritise SNPs for follow up, we corrected for multiple testing of the total number of SNPs in each candidate region (a conservative threshold given the correlation between SNPs). However, we still used $P\text{-value} < 3 \times 10^{-9}$ ($5 \times 10^{-8}$ corrected for 15 tests (14 metabolites and one ratio)) in all available studies as the final criteria for association.

5. Selection of SNPs for genotyping in Botnia and meta-analysis in RISC


We used two statistical thresholds to select SNPs for replication. First, we used $P\text{-value} < 5 \times 10^{-8}$, as the standard for genome-wide significance in the context of common SNPs.
Second, for the 34 candidate genes (see Table 2 and Methods section 4) we divided 0.05 by the total number of SNPs in the gene ±300kb. We meta-analysed SNP-metabolite results from RISC and Botnia using an inverse-variance weighted approach as implemented in STATA command “metan”. For use in Mendelian randomisation analyses we used SNPs reaching $P$-value $< 3 \times 10^{-9}$ in the meta-analysis of the two studies.

5.2. Metabolite ratio.

To validate SNPs associated with the glycine/serine ratio, we linked results from the recently published KORA and UKtwins studies (22) to our GWAS results. We meta-analysed our results with those from the KORA and/or UKtwins studies with a significant threshold of $P$-value $< 3 \times 10^{-9}$ when including all available studies.

6. Effects of associated SNPs on other metabolites in the glycine and glutathione biosynthesis pathways.

We performed further analyses for the five SNPs associated with metabolites in the glycine or glutathione biosynthesis pathways, which include glycine, serine, betaine, α-HB, α-KB, and glycine/serine ratio. We tested the associations of each SNP against the other metabolite traits. We performed association analyses in the linear regression model described in Method section 3 in STATA (version 10.1).

7. Mendelian randomization analyses


We tested the role of metabolite-associated SNPs reaching genome wide significance with two diabetes related traits: hyperinsulinemic-euglycemic clamp ($M$-value corrected for kg body weight), which was a measure of whole body insulin sensitivity; and fasting insulin. $M$-value based measures of insulin sensitivity were corrected for age, sex and centre, and converted to SD units and inverse-normalised. Fasting insulin was natural log transformed,
corrected for age, sex and centre, and converted to SD units. Using RISC data, we calculated two estimates of the association between metabolite-SNPs and diabetes related traits for each metabolite trait. First, we calculated an estimated expected effect if there was a causal association between metabolites and diabetes-related measures, using a triangulation approach as shown in Supplementary Figure 1: we calculated the correlation between standardised metabolite levels and the two diabetes related traits. We then multiplied these standardized effects by that between metabolite SNPs and metabolites to estimate an approximate expected effect size of the association between metabolite SNPs and M-value and fasting insulin. We calculated approximate expected 95% CIs based on the observed effects and standard errors using the Taylor series expansion of the ratio of two means (23).

Second, we tested the observed effect between metabolite SNPs and the two diabetes related traits. For the clamp-based measures of insulin sensitivity, we used three studies (RISC, EUGENE2 and Stanford (IST)) and meta-analysed results using the program METAL (24). In EUGENE2, insulin sensitivity was measured using the same hyperinsulinemic-euglycemic clamp based protocol as that used by RISC (25). In the Stanford study, insulin sensitivity was measured by steady-state plasma glucose (SSPG) method. The SSPG value is highly inversely correlated to M-value ($r = -0.93$, $P$-value $<0.001$) (26), so meta-analyses were performed between the three studies by reversing the signs of the effect sizes in Stanford.

For fasting insulin measures of insulin sensitivity, we used data from the MAGIC (Meta-Analyses of Glucose and Insulin-related traits Consortium), consisting of a meta-analysis of 23 genome-wide association studies with 27,589 individuals (27).


We used data from the DIAGRAM consortium (Diabetes Genetics Replication And Meta-analysis Consortium) to assess the association of metabolite SNPs with type 2 diabetes.
This data consists of 8,130 patients with type 2 diabetes and 38,987 controls from eight GWAS (28).

7.3. Instrumental variable analysis

In this RISC study, where we had measures of SNPs, metabolites and insulin sensitivity, we performed instrumental variable analyses using the two-stage least squares regression approach (2SLS) implemented in the STATA command “ivreg2”. Using the 2SLS, the instrumental variable estimator $\beta_{IV}$ provides an estimate of the causal effects of exposure (i.e. metabolites) on outcome (i.e. insulin sensitivity) even in the presence of unmeasured confounders (10) (29).
RESULTS

1. GWAS and replication of insulin sensitivity-related metabolites

In the RISC study we identified eight signals of interest, either at genome wide significance or reaching a locus-wide nominal level of significance around one of the candidate genes (Supplementary Table 2). One of these signals represented a widely reported association between SNPs in the FADs gene cluster and fatty acids (27, 30, 31) (in our case adrenate) and we did not pursue this association further. We successfully genotyped SNPs representing six of the remaining seven signals in the Botnia study (Table 3).

After meta-analysis of RISC and Botnia data (where available), we identified four association signals with three separate single metabolites - the amino acids glycine, serine and betaine - at \( P \)-value < \( 3 \times 10^{-9} \). For ratios of metabolites, we identified one signal for glycine/serine ratio that when meta-analysed with published KORA data reached \( P \)-value < \( 3 \times 10^{-9} \). Details of the associations are given in Table 3 & Figure 1. Two SNPs associated with serine were taken forward from the RISC GWAS, but did not replicate in the Botnia study (Table 3).

1.1. Novel associations between SNPs in two loci and betaine levels

We identified an association between rs499368 in the SLC6A12 (solute carrier family 6) gene and betaine levels (\( P \)-value=\( 1.46 \times 10^{-10} \)) (Figure 1a). This signal has not previously been reported with any other trait and was not captured at \( r^2 > 0.8 \) in the published KORA or UKtwins data. The second association occurred between rs17823642 near a candidate gene BHMT (betaine-homocysteine methyltransferase) and betaine levels (\( P \)-value=\( 2.3 \times 10^{-9} \)) (Figure 1b). This signal has not previously been reported with any other trait at genome wide significance but was captured at \( r^2 = 1.0 \) by rs7732845 in the published KORA data (but not UKtwins) data and a meta-analysis of RISC, Botnia and KORA data (\( P \)-value for KORA alone = \( 1.98 \times 10^{-6} \)) confirms very robust evidence of association (meta-analysed \( P \)-value=\( 6.07 \times 10^{-14} \)).
1.2. A SNP in a known locus is associated with glycine levels.

We identified an association between rs715 in the 3’ UTR of the *CPS1* (carbamoyl-phosphate synthetase 1) gene and glycine levels at genome wide significance (*P*-value=$3.30\times10^{-50}$) (Figure 1c). This signal was not captured at $r^2$ >0.8 in the published KORA or UKtwins data but a SNP (rs4673558) with an $r^2$ = 0.21 with rs715 is associated with glycine levels with *P*-value=$4.3\times10^{-11}$ in the UKtwins data.

1.3. A SNP in a known locus is associated with serine levels.

We identified an association between rs478093 near the *PHGDH* (3-phosphoglycerate dehydrogenase) gene and serine levels at genome wide significance (*P*-value = $1.52\times10^{-9}$) (SNP not available in Botnia study) (Figure 1d). This signal was previously reported as associated with serine and ratios of metabolites involving serine in the published KORA and UKtwins data (based on rs477992 ($r^2$ = 0.93), meta-analysed *P*-value=$1.94\times10^{-14}$) (22).

1.4. A SNP in a novel locus is associated with glycine-serine ratios.

We identified a previously unreported association between rs1107366 near the *ALDH1L1* (aldehyde dehydrogenase 1 family, member L1) gene and glycine-serine ratios (*P*-value=$2.25\times10^{-56}$) (Figure 1e). This signal reached genome wide significance in combination with data from the KORA and UKtwins studies (meta-analysed *P*-value=$2.8\times10^{-12}$) (Table 3).

1.5. The association between rs715 in *CPS1* and glycine levels is highly sex specific.

The SNPs in the *CPS1* locus have been previously reported with sex-specific effect on glycine and homocysteine levels (rs7422339, $r^2$ =0.92 with rs715) (32, 33). We observed a similar sex-specific association between this signal and glycine levels (Supplementary Figure 2). The association was weak in males ($\beta$ = -0.19; 95%CI (-0.31,-0.08); *P*-value = $1.1\times10^{-3}$), but more than four times the effect size in females ($\beta$ = -0.84; 95%CI (-0.98,-0.69); *P*-value = $4.5\times10^{-28}$). The Z-test for the null hypothesis of no sex-specific effect was rejected
at $P$-value $= 2.67 \times 10^{-13}$. This result was consistent with the female-specific association reported before (32, 33). There was no evidence that the association was different between pre and post menopausal women (pre: effect $= -0.78$; 95% CI (-0.95,-0.61); $P$-value $= 1.57 \times 10^{-19}$; post: effect $= -1.04$; 95% CI (-1.35,-0.73); $P$-value $= 4.57 \times 10^{-11}$). We did not observe any evidence of sex-specific effects for the other metabolite SNPs.

2. Effects of metabolite-associated SNPs on other metabolites in the glycine and glutathione biosynthesis pathways.

The effects of the five confirmed signals on the other four single metabolites levels and glycine/serine ratio are shown in Table 4, with the last column showing the effects of metabolite-insulin sensitivity associations.

3. Associations of metabolite-associated SNPs with fasting insulin-based and clamp-based measures of insulin sensitivity

Associations between the five metabolite-associated SNPs, metabolite levels, and diabetes related traits (fasting insulin and hyperinsulinemic-euglycemic clamp (M-value)) are shown in Table 5.

There were strong correlations between fasting insulin and metabolite levels ($0.10 < r < 0.20$), as expected given that the metabolites were selected as those correlated with a measure of insulin sensitivity. These strong correlations between phenotypes meant the MAGIC data had $>90\%$ power to detect at $P$-value $<0.01$ based on the estimated expected effects between metabolite SNPs and fasting insulin. However, there were no associations between metabolite SNPs and fasting insulin. The effect sizes observed in the MAGIC data were all smaller than those expected based on the triangulation calculations.

We identified a nominal association between the glycine/serine ratio associated SNP rs1107366 near $ALDH1$ (aldehyde dehydrogenase) and clamp-based measures of insulin sensitivity ($\beta = 0.09$ SDs (95% CIs: 0.03, 0.15), where the allele that raises glycine/serine
ratios increases clamp-based insulin sensitivity $P\text{-value} = 0.005$), (Table 5, Supplementary Table 3). The observed effect size was larger than the expected effect (expected $\beta_{\text{SNP-IS}} = 0.03$ SDs; (95% CIs: 0.01, 0.04). Results of instrumental variable analyses in RISC were consistent with the main results: the glycine/serine ratio predicted from the rs1107366 genotypes is associated with clamp-based insulin sensitivity ($\beta_{\text{IV}} = 1.00; 95\% \text{ CI} (0.24, 1.76); P\text{-value}=0.01$).

No other metabolite SNPs were associated with clamp-based measures of insulin sensitivity in either the triangulation analyses or instrumental variable analyses, including the other glycine and serine signals (rs715 and rs478093),


There was no evidence of association between four of the five metabolite SNPs and type 2 diabetes, based on the meta-analysis of case control studies reported by the DIAGRAM study (Supplementary Table 4). The rs715 SNP in CPS1 associated with glycine levels was poorly captured in the type 2 diabetes GWAS meta-analysis (28).
DISCUSSION

Using a genome wide approach, we have identified five associations between genetic variants and circulating levels of three metabolites and one metabolite ratio. These metabolites occur in pathways strongly correlated with the gold standard measure of insulin sensitivity (hyperinsulinemic-euglycemic clamp) in the RISC study – primarily the glycine biosynthesis pathway. Three of these associations have been previously identified at genome wide levels of significance. We have tested the association of these variants with insulin sensitivity using the largest collective set of studies with these measures, including RISC, EUGENE2 and Stanford.

1. Amino acid associated SNPs and diabetes related traits.

Glycine and serine are glycogenic amino acids involved in hepatic gluconeogenesis and glutathione biosynthesis, which are potentially important pathways in diabetes and insulin resistance. In the RISC study, glycine, serine and betaine were positively correlated with clamp-based measures of insulin sensitivity and negatively correlated with fasting insulin. These associations are in line with previous findings (7-9, 34).

Using Mendelian randomisation analyses, we assessed if these amino acids play a causal role in insulin sensitivity and type 2 diabetes risks. Our Mendelian randomisation analyses do not support a causal association between genetically-changed glycine, serine and betaine levels and insulin sensitivity levels as measured by fasting insulin, or type 2 diabetes. However, for the clamp-based measures of insulin sensitivity, we observed a suggestive association of a glycine/serine ratio-associated SNP rs1107366 (near the ALDH1L1 gene). The allele that raises glycine/serine ratios increases clamp-based insulin sensitivity. The rs1107366-insulin sensitivity association needs further replication in larger sample size, especially given that other glycine and/or serine-associated SNPs (e.g. rs715 and rs478093) are not associated with insulin sensitivity. The rs715 variant in CPS1 for example explains a greater proportion of the
variance in glycine levels (~13% compared to ~2% for rs1107366). It is also possible that rs1107366 could influence insulin sensitivity via non-glycine mediated (pleiotropic) effects. In an insulin resistant state the increase of hepatic gluconeogenesis would result in greater consumption of glycogenic amino acids, which may be accentuated in individuals with genetically influenced lower levels of these molecules and is consistent with the hypothesis of reverse causality (35). However causal mechanisms in both directions remain plausible because gluconeogenesis is controlled in many different ways.

2. Biology of metabolite levels

Our data highlight some candidate genes and protein products important in controlling circulating metabolite levels. At each locus there is a clear candidate gene, although we cannot be certain which gene is affected by the associated SNP.

2.1. The betaine and serine signals are in or near functionally-relevant genes

The $SLC6A12$ (solute carrier family 6) gene is a highly plausible candidate for influencing betaine levels. Betaine is an osmolyte used by cells for protection against hyperosmotic environments (36), and $SLC6A12$ encodes a highly conserved osmoregulator which controls cellular volume by extrusion of betaine (37, 38). Previous studies have shown that hyperosmolarity could induce insulin resistance by impairing IRS-1 tyrosine phosphorylation, and degradation of IRS1 and IRS2 in adipocytes (39, 40). This connection may partly explain the association of betaine with insulin sensitivity.

The $BHMT$ (betaine-homocysteine methyltransferase) gene encodes a cytosolic enzyme that catalyses the conversion of betaine and homocysteine to dimethylglycine and methionine, respectively. The rs17823642 SNP is highly correlated with three SNPs associated with $BHMT$ enzyme activity and protein level ($rs41272270$, $r^2 = 1$; $rs16876512$, $r^2 = 0.925$; $rs6875201$, $r^2 = 0.925$) (41).
The previously described serine GWAS signal rs478093 is in *PHGDH* (3-phosphoglycerate dehydrogenase; also known as *PGDH* (see *3-PGDH* in Figure 2a), the gene product of which catalyzes the first and rate-limiting step in the phosphorylated pathway of serine biosynthesis.

2.2. The effect of variation at the *CPS1* locus on glycine level

The rs715 SNP represents the same association signal as that identified between the SNP rs2216405 near *CPS1* ($r^2=0.47$ with rs715) and glycine levels (31), but the variance in glycine levels explained by rs715 (12.87%) in our study compared to the variance explained by the rs2216405 SNP (8.64%) suggests that rs715 is a better marker for the causal variant. A recent study reported an association between rs715 and glycine levels specific to females, consistent with our results (32).

The enzyme encoded by *CPS1* (carbamoyl-phosphate synthetase) catalyzes synthesis of carbamoyl phosphate from ammonia and bicarbonate (Figure 2b). Patients with defects in the function or expression of CPS1 suffer from life-threatening hyperammonemia(42). It is possible that variants at this locus perturb the conversion of ammonia and bicarbonate to carbamoyl phosphate. We hypothesise that *CPS1* variants may cause excess ammonia, which may then lead to increased production of glycine and tetrahydrofolate in the glycine cleavage system (Figure 2d & Figure 3).

2.3. *ALDH1L1* as a candidate enzyme involved in glycine metabolism

Glycine is a key component of the folate pathway (Figure 3). The protein product of *ALDH1L1* catalyzes the conversion of 10-formyltetrahydrofolate, NADP, and water to tetrahydrofolate, NADPH, and carbon dioxide (Figure 2c). Our association between a SNP near the *ALDH1L1* gene and glycine/serine ratio implicates this enzyme in glycine/serine conversion rate. This is in accordance with the knowledge that the glycine cleavage system, which accounts for ~41% of whole body glycine flux, is tightly linked with tetrahydrofolate in folate metabolism (Figure 2d & Figure 3) (43).
2.4. Sex-specific effect at CPS1 locus on glycine and homocysteine levels

We observed a sex-specific association of CPS1 variants on glycine levels. This is consistent with the findings of Mittelstrass et al. (2011) (32). The variants at the CPS1 locus also have female-specific effects on homocysteine levels (33), and the glycine-raising allele is associated with raised homocysteine. (44).

2.5. The links between glycine, serine, homocysteine and betaine in folate and homocysteine metabolism

Glycine, serine and betaine are linked to homocysteine and folate metabolism (Figure 3) (44, 45). The CPS1 variant rs715 is strongly correlated with rs7422339 ($r^2 = 0.92$) that was previously reported as associated with homocysteine and folate levels (33). House et. al. (44) demonstrated that the plasma concentrations of homocysteine, glycine and serine were all elevated in folate-deficient rats. From the link between betaine and homocysteine in the reaction catalysed by BHMT, we hypothesized that, if rs17823642 is associated with betaine level through reduced functioning of BHMT then it would result in elevation of not only betaine but also homocysteine levels (46). We assessed the effect of rs17823642 on homocysteine levels in an independent European study (InCHIANTI). We observed a nominal association in female ($n = 575; \beta = 0.26; 95\% \text{ CI} (0.07, 0.44); P\text{-value} = 5.9 \times 10^{-3}$) but not in male ($n = 458; \beta = -0.02; 95\% \text{ CI} (-0.25, 0.21); P\text{-value} = 0.87$), where the betaine-raising allele also correlated with increased homocysteine levels. However, this association requires further confirmation with larger sample size.

3. Limitations

There are a number of limitations in our study. First, the triangulation approach for estimating expected effects does not take into account the complicated feedback mechanisms and interactions involved in controlling metabolite levels. The SNPs we identified are associated
with several metabolites (**Table 4**), which means they do not provide specific instruments for one metabolite. However, the associations for any one SNP are all with metabolites closely connected in well annotated pathways and therefore provide an instrument to test the relationship between alterations in those pathways and diabetes related outcomes.

Second, our estimated effects are approximate, with metabolite levels only measured in RISC, and observed estimates coming from separate studies. Nevertheless, the size of the MAGIC study provided very good power to see the expected very small effects on fasting insulin levels.

Finally, we have only been able to assess some of the many metabolites associated with insulin sensitivity. The metabolites selected in our study were those most strongly associated with clamp-measured insulin sensitivity in the RISC study (4) and we focused efforts on these metabolite traits. Some recent studies have reported other metabolites associated with dysglycemia (e.g. branched chain and aromatic amino acids) (3, 5, 47, 48), but these metabolites have not been measured in RISC study.

In conclusion, our study provides novel insight into the genetic regulation of metabolite levels, particularly those involved in the glycine-related pathways closely correlated to insulin sensitivity. Genetic variants associated with metabolite levels provide an important approach to helping unravel the functional role of the metabolic pathways that influence diabetes related traits.
ACKNOWLEDGEMENTS

This work was supported by grants from the Wellcome Trust 083270/Z/07/Z and MRC G0601261. The RISC study is supported by European Union Grant QLG1-CT-2001-01252 and AstraZeneca. The EUGENE2 study is supported by a grant from the European Community’s FP6 EUGENE2 no. LSHM-CT-2004-512013. We thank Merck Research Labs for conducting DNA genotyping on the RISC samples and Stanford samples. J.K. is supported by an American Heart Association National Fellowship to Faculty Transition Award (10FTF3360005). W.E.G. is an employee of Metabolon, Inc., Research Triangle Park, NC, USA, as a Director in the Diagnostics division. Metabolon is a diagnostics and services company offering the industry’s leading biochemical profiling platform.

No potential conflicts of interest relevant to this article were reported.

Author Contributions: W.X. and T.M.F. designed the study, researched data, performed analyses, and co-wrote the initial draft of the article. A.R.W. and M.N.W. analysed data. F.A., T.L.A., T.Q and J.P. contributed data and reviewed the manuscript. T.H. and H.H. provided samples and data from individual studies, contributed to discussion and reviewed the manuscript. O.P. and E.F. provided samples and data from individual studies and reviewed the manuscript. J.J.N. performed phenotyping in individual studies and reviewed the manuscript. W.E.G. performed phenotyping in individual studies, contributed to discussion and reviewed the manuscript. V.L. contributed data and reviewed/edited the manuscript. U.S. provided samples and data from individual studies and reviewed the manuscript. K.A. provided samples and data from individual studies and reviewed the manuscript. J.M.D. provided samples and data from individual studies and reviewed the manuscript. M.L. and W.M. contributed data and reviewed/edited the manuscript. S.A. contributed data and reviewed/edited the manuscript. J.W.K. contributed data, contributed to
discussion and reviewed/edited the manuscript. L.G. contributed to discussion and
reviewed/edited the manuscript.

Prof. Timothy M. Frayling is the guarantor of this work, had full access to all the data, and
takes full responsibility for the integrity of data and the accuracy of data analysis.
Table 1: Summary details of RISC individuals and relevant characteristics

<table>
<thead>
<tr>
<th>Name (units)</th>
<th>Age (yrs)</th>
<th>BMI (kg/m²)</th>
<th>FI (pmol/l)</th>
<th>M-value* (micromol/kgbodywt/min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>1,004</td>
<td>1,004</td>
<td>973</td>
<td>1,004</td>
</tr>
<tr>
<td>mean</td>
<td>43.91</td>
<td>25.42</td>
<td>34.3</td>
<td>39.84</td>
</tr>
<tr>
<td>sd of mean</td>
<td>8.37</td>
<td>4.04</td>
<td>18.55</td>
<td>16.2</td>
</tr>
<tr>
<td>median</td>
<td>44</td>
<td>24.9</td>
<td>30</td>
<td>38.33</td>
</tr>
<tr>
<td>min</td>
<td>30</td>
<td>16.9</td>
<td>3</td>
<td>4.92</td>
</tr>
<tr>
<td>max</td>
<td>61</td>
<td>43.9</td>
<td>116</td>
<td>114.25</td>
</tr>
</tbody>
</table>

**correlation with age**  
**correlation with BMI**  
**correlation with FI**  
**correlation with M-value**

\[ r = 0.30; P = 8.3 \times 10^-6 \]  
\[ r = 0.01; P = 0.31 \]  
\[ r = -0.04; P = 0.01 \]  
\[ r = -0.12; P = 7.5 \times 10^-58 \]  
\[ r = -0.49; P = 6.4 \times 10^{-44} \]

* M-value for the clamp expressed per kg body weight

FI = fasting insulin, M-value = hyperinsulinemic-euglycemic clamp based measure of insulin sensitivity (the higher the M-value, the higher the insulin sensitivity).
<table>
<thead>
<tr>
<th>Metabolites (mmol/l)</th>
<th>Candidate genes</th>
<th>Mean (min, max)</th>
<th>SD</th>
<th>Median</th>
<th>Correlation with M-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-hydroxybutyrate</td>
<td>LDHA, LDHB, LDHC, LDHD, α-HBDH</td>
<td>4.50 (1.09, 13.35)</td>
<td>1.75</td>
<td>4.23</td>
<td>r = -0.35; $P = 1.5 \times 10^{-30}$</td>
</tr>
<tr>
<td>adrenate</td>
<td></td>
<td>0.20 (0.06, 1.19)</td>
<td>0.09</td>
<td>0.18</td>
<td>r = -0.19; $P = 9.0 \times 10^{-10}$</td>
</tr>
<tr>
<td>α-ketoglutaric acid</td>
<td></td>
<td>1.09 (0.00, 2.92)</td>
<td>0.39</td>
<td>1.09</td>
<td>r = -0.21; $P = 1.4 \times 10^{-11}$</td>
</tr>
<tr>
<td>α-ketobutyrate</td>
<td>CBS</td>
<td>0.38 (0.00, 1.16)</td>
<td>0.21</td>
<td>0.36</td>
<td>r = -0.28; $P = 4.4 \times 10^{-20}$</td>
</tr>
<tr>
<td>betaine</td>
<td>BHMT, CHDH,</td>
<td>4.26 (1.16, 11.92)</td>
<td>1.38</td>
<td>4.13</td>
<td>r = 0.06; $P = 6.6 \times 10^{-11}$</td>
</tr>
<tr>
<td>creatine</td>
<td>CKMT1A/1B, CKMT2</td>
<td>4.55 (1.05, 14.3)</td>
<td>2.19</td>
<td>4.13</td>
<td>r = -0.18; $P = 1.4 \times 10^{-8}$</td>
</tr>
<tr>
<td>decanoylcarnitine</td>
<td>CPTIC, SLC25A20</td>
<td>0.04 (0, 0.35)</td>
<td>0.04</td>
<td>0.03</td>
<td>r = 0.14; $P = 1.3 \times 10^{-5}$</td>
</tr>
<tr>
<td>glutamate</td>
<td>NAGS, SIRT4, GLUD1</td>
<td>18.19 (4.99, 100.26)</td>
<td>12.21</td>
<td>14.38</td>
<td>r = 0.06; $P = 0.06$</td>
</tr>
<tr>
<td>glutamate</td>
<td>NAGS, SIRT4, GLUD1</td>
<td>17.35 (7.05, 10.53)</td>
<td>5.26</td>
<td>16.12</td>
<td>r = 0.24; $P = 7.2 \times 10^{-15}$</td>
</tr>
<tr>
<td>ketovaleine</td>
<td>BCAT2, BCKDHA, BCKDHB</td>
<td>1.59 (0.12, 2.99)</td>
<td>0.40</td>
<td>1.62</td>
<td>r = -0.23; $P = 7.1 \times 10^{-14}$</td>
</tr>
<tr>
<td>linoleoylgpc</td>
<td>PLA2G5, PLA2G12A, PLA2G2D</td>
<td>15.65 (5.24, 39.89)</td>
<td>5.18</td>
<td>15.15</td>
<td>r = -0.29; $P = 2.4 \times 10^{-21}$</td>
</tr>
<tr>
<td>oleate</td>
<td>OLAH, ACSL1</td>
<td>85.14 (11.91, 569.89)</td>
<td>36.67</td>
<td>81.57</td>
<td>r = -0.17; $P = 3.2 \times 10^{-8}$</td>
</tr>
<tr>
<td>oleoylgpc</td>
<td>OLAH, LCLAT, PLD1</td>
<td>9.81 (3.14, 22.64)</td>
<td>2.93</td>
<td>9.49</td>
<td>r = 0.27; $P = 1.2 \times 10^{-18}$</td>
</tr>
<tr>
<td>serine</td>
<td>PSPH, PHGDH, CBS, SDS, SHMT2</td>
<td>10.90 (4.61, 21.75)</td>
<td>2.23</td>
<td>10.70</td>
<td>r = 0.14; $P = 4.6 \times 10^{-6}$</td>
</tr>
</tbody>
</table>
### Table 3: Associations between SNPs and insulin sensitivity-associated metabolites in the RISC (n=1,004) and Botnia (n=339) studies

<table>
<thead>
<tr>
<th>SNP</th>
<th>Trait</th>
<th>Genes</th>
<th>Effect allele/other allele</th>
<th>RISC effect allele freq</th>
<th>RISC effect (95% CIs)</th>
<th>RISC P</th>
<th>Botnia effect (95% CIs)</th>
<th>Botnia P</th>
<th>Meta effect (95% CIs)</th>
<th>Meta P</th>
<th>Meta N</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs478093</td>
<td>serine</td>
<td>PHGDH</td>
<td>G/A</td>
<td>0.71</td>
<td>0.30 (0.20, 0.39)</td>
<td>1.5×10⁻⁹</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1.9×10⁻¹⁴</td>
<td>2,772</td>
</tr>
<tr>
<td>rs715</td>
<td>glycine</td>
<td>CPS1</td>
<td>T/C</td>
<td>0.68</td>
<td>-0.58 (-0.68, -0.48)</td>
<td>5.3×10⁻³⁰</td>
<td>-0.70 (-0.54, -0.86)</td>
<td>2.0×10⁻¹⁶</td>
<td>-0.61 (-0.53, -0.69)</td>
<td>3.3×10⁻⁵⁰</td>
<td>1,343</td>
</tr>
<tr>
<td>rs1107366</td>
<td>glycine/serine</td>
<td>ALDH1L1</td>
<td>G/A</td>
<td>0.51</td>
<td>0.02 (0.01, 0.03)</td>
<td>5.0×10⁻⁶</td>
<td>0.016 (0.002, 0.03)</td>
<td>0.024</td>
<td>0.019 (0.011, 0.027)</td>
<td>2.3×10⁻⁶</td>
<td>1,343</td>
</tr>
<tr>
<td>rs17823642</td>
<td>betaine</td>
<td>BHMT, BHMT2</td>
<td>C/T</td>
<td>0.89</td>
<td>0.35 (0.21, 0.49)</td>
<td>1.4×10⁻⁶</td>
<td>0.52 (0.24, 0.80)</td>
<td>3.5×10⁻⁴</td>
<td>0.38 (0.51, 0.26)</td>
<td>2.4×10⁻⁹</td>
<td>1,345</td>
</tr>
<tr>
<td>rs13233754</td>
<td>serine</td>
<td>PSPH</td>
<td>G/A</td>
<td>0.97</td>
<td>-0.61 (-0.88, -0.33)</td>
<td>2.0×10⁻⁵</td>
<td>0.02 (-0.59, 0.62)</td>
<td>0.96</td>
<td>-0.50 (-0.25, -0.75)</td>
<td>1.0×10⁻⁴</td>
<td>1,343</td>
</tr>
<tr>
<td>rs4275190</td>
<td>serine</td>
<td>PSPH</td>
<td>T/C</td>
<td>0.68</td>
<td>0.17 (0.08, 0.27)</td>
<td>4.0×10⁻⁴</td>
<td>0.08 (-0.26, 0.43)</td>
<td>0.63</td>
<td>0.16 (0.07, 0.26)</td>
<td>3.9×10⁻⁴</td>
<td>1,344</td>
</tr>
<tr>
<td>rs499368</td>
<td>betaine</td>
<td>SLC6A12, SLC6A13</td>
<td>A/T</td>
<td>0.51</td>
<td>-0.46 (-0.61, -0.3)</td>
<td>8.1×10⁻⁹</td>
<td>-0.25 (-0.41, -0.1)</td>
<td>1.5×10⁻³</td>
<td>-0.36 (-0.46, -0.25)</td>
<td>1.5×10⁻¹⁰</td>
<td>1,342</td>
</tr>
</tbody>
</table>

NA = genotype data not available in Botnia (This signal was previously reported as associated with serine and ratios of metabolites involving serine in the published KORA and UKtwins data (based on rs477992 (\(r^2 = 0.93\)), meta-analysed \(P\)-value=1.94×10⁻¹⁴). *The glycine/serine ratio was normalised by log 10 transformation, and the residuals after adjusting for age, sex and centre were used as the trait in GWAS. The other traits were normalised by inverse normal transformation and standardised (i.e. in SD unit). †This association was genome wide significant when meta-analysed with published KORA data (\(P\)-value =3.71×10⁻¹¹; n=2,687), and therefore followed up in the Botnia study: the \(P\)-value when combining RISC, KORA and Botnia results is 2.8×10⁻¹² (n =3,026). Gene names: PHGDH: phosphoglycerate dehydrogenase; CPS1:
carbamoyl-phosphate synthase 1; *ALDH1L*: aldehyde dehydrogenase 1 family, member L1; *BHMT*: betaine–homocysteine S-methyltransferase/2; *PSPH*: phosphoserine phosphatase; *SLC6A12/13*: solute carrier family 6 (neurotransmitter transporter, betaine/GABA), member 12/13;
Table 4: Associations between SNPs and five metabolites in the glycine and glutathione biosynthesis pathway in RISC study

<table>
<thead>
<tr>
<th>Metabolites</th>
<th>Glycine</th>
<th>Glycine/serine</th>
<th>Serine</th>
<th>Betaine</th>
<th>Betaine</th>
<th>Insulin Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rs715</td>
<td>rs1107366</td>
<td>rs478093</td>
<td>rs17823642</td>
<td>rs499368</td>
<td>(M-value)</td>
</tr>
<tr>
<td>glycine</td>
<td>-0.58</td>
<td>5.3×10^{-30}</td>
<td>-0.17</td>
<td>7.5×10^{-5}</td>
<td>-0.10</td>
<td>0.05</td>
</tr>
<tr>
<td>serine</td>
<td>-0.24</td>
<td>1.9×10^{-6}</td>
<td>0.03</td>
<td>0.57</td>
<td>-0.30</td>
<td>1.5×10^{-9}</td>
</tr>
<tr>
<td>betaine</td>
<td>0.24</td>
<td>1.3×10^{-6}</td>
<td>-0.08</td>
<td>0.07</td>
<td>0.09</td>
<td>0.06</td>
</tr>
<tr>
<td>αHB</td>
<td>-0.07</td>
<td>0.20</td>
<td>0.09</td>
<td>0.03</td>
<td>0.09</td>
<td>0.07</td>
</tr>
<tr>
<td>αKB</td>
<td>0.00</td>
<td>0.97</td>
<td>0.05</td>
<td>0.22</td>
<td>0.06</td>
<td>0.23</td>
</tr>
<tr>
<td>glycine/serine</td>
<td>-0.05</td>
<td>9.8×10^{-20}</td>
<td>-0.02</td>
<td>3.45×10^{-6}</td>
<td>0.01</td>
<td>4.7×10^{-3}</td>
</tr>
</tbody>
</table>

The single metabolites and glycine/serine ratio were log10 transformed and all associations were fitted in linear regression model with adjustment for age, sex and centres. The cells of the table are coloured on a gradient greyscale, where the grey colour represents positive effect and white colour represents negative effect of SNPs on metabolite traits in the association analyses. The effect sizes of SNPs on single metabolites are not comparable to that in Table 3 and Table 5, where the metabolite traits are in SD units and under inverse-normal transformation.
Table 5: Summary of associations between SNPs, insulin sensitivity-associated metabolite, and diabetes-related traits (fasting insulin and insulin sensitivity)

<table>
<thead>
<tr>
<th>SNP</th>
<th>Effect allele/other allele</th>
<th>Metabolite</th>
<th>SNP-met effect (SD) (95% CIs)</th>
<th>SNP-met effect (SD) (95% CIs)</th>
<th>Met-trait effect (SD) (95% CI)</th>
<th>Met-trait effect (SD) (95% CI)</th>
<th>Expected SNP-trait effect (SD) (95% CIs)</th>
<th>Observed SNP-trait effect (SD) (95% CIs)</th>
<th>Observed SNP-trait effect (SD) (95% CIs)</th>
<th>Observed SNP-trait effect (SD) (95% CIs)</th>
<th>Observed SNP-trait effect (SD) (95% CIs)</th>
<th>P</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs478093 G/A</td>
<td>serine</td>
<td>0.31 (0.22, 0.40)</td>
<td>1.9×10^{-10}</td>
<td>-0.19 (-0.25, -0.13)</td>
<td>2.0×10^{-9}</td>
<td>-0.06 (-0.08, -0.03)</td>
<td>1.1×10^{-3}</td>
<td>(-6.93×10^{-3}, 9.13×10^{-3})</td>
<td>0.79 (0.17, 0.35)</td>
<td>37795 (7779, 11509)</td>
<td>0.01</td>
<td>14381</td>
<td></td>
</tr>
<tr>
<td>rs715 T/C</td>
<td>glycine</td>
<td>-0.58 (-0.67, -0.48)</td>
<td>5.0×10^{-12}</td>
<td>-0.17 (-0.23, -0.11)</td>
<td>1.0×10^{-7}</td>
<td>0.10 (0.06, 0.14)</td>
<td>1.9×10^{-4}</td>
<td>(-4.83×10^{-3}, 0.02)</td>
<td>0.21 (0.03, 0.39)</td>
<td>14381 (1248, 2649)</td>
<td>0.07</td>
<td>38227</td>
<td></td>
</tr>
<tr>
<td>rs1107366 G/A</td>
<td>glycine/serine</td>
<td>0.21 (0.12, 0.29)</td>
<td>4.1×10^{-4}</td>
<td>-0.05 (-0.11, 0.01)</td>
<td>0.12</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.05</td>
<td>27174</td>
<td></td>
</tr>
<tr>
<td>rs17823642 C/T</td>
<td>betaine</td>
<td>0.35 (0.21, 0.49)</td>
<td>1.3×10^{-4}</td>
<td>-0.13 (-0.2, -0.07)</td>
<td>2.4×10^{-5}</td>
<td>-0.05 (-0.08, -0.02)</td>
<td>1.9×10^{-4}</td>
<td>(-0.01, 0.01)</td>
<td>0.76 (0.04, 0.16)</td>
<td>38227 (2623, 5024)</td>
<td>0.45</td>
<td>1884</td>
<td></td>
</tr>
<tr>
<td>rs499368 A/T</td>
<td>betaine</td>
<td>-0.46 (-0.61, -0.30)</td>
<td>6.4×10^{-9}</td>
<td>-0.13 (-0.2, -0.07)</td>
<td>2.4×10^{-5}</td>
<td>0.06</td>
<td>4.0×10^{-4}</td>
<td>(-8.62×10^{-3}, 9.42×10^{-3})</td>
<td>0.93 (0.01, 0.19)</td>
<td>27174 (2518, 4935)</td>
<td>0.64</td>
<td>1884</td>
<td></td>
</tr>
</tbody>
</table>

**Fasting Insulin**

**Hyperinsulinemic-euglycemic Clamp**
SNP-met: the associations between the five metabolite-associated SNPs and metabolites.

Met-trait: the associations between metabolites and diabetes related continuous traits.

Expected SNP-trait effect: the approximate expected effect sizes of SNPs on fasting insulin and clamp-based measures of insulin sensitivity, based on the product of SNP-met effect and met-trait effect.

Observed SNP-trait effect: the observed effects of SNPs on fasting insulin and clamp-based measures of insulin sensitivity.

SNP-metabolite, metabolite-trait and expected SNP-trait effect sizes are in SD units and based on RISC study. Observed SNP-trait values are in SD units and are taken from a meta-analysis of the RISC, EUGENE2 and Stanford studies.

NA: The expected and observed correlations between rs1107366 and fasting insulin are not shown, because glycine/serine ratios are not associated with fasting insulin ($P$-value = 0.12). *$P$-values for SNP-met associations differ very slightly from those in Table 3 due to different statistical methods used in MACH. †Observed SNP-fasting insulin values (pmol/L) are normalised by natural log transformation in the MAGIC group and so are not directly comparable to the expected trait values – an approximate conversion factor from these units to SDs is $1/0.5 = 2.0$ for natural log insulin (given the median SD for natural log insulin across the MAGIC studies is 0.5).

‡The effect direction is different from Table 4, because the effects were reported on allele A in Table 4; and in Table 5 the glycine/serine ratio was inverse-normal transformed for the consistency with the other two associations in the triangulation analysis.
Figure legends

**Figure 1.** Regional association plots of the five SNP-metabolite associations in RISC cohort

In each plot, the top panel shows the name and location of genes in the UCSC Genome Browser. The $-\log_{10}$ of $P$-values of the imputed SNPs are plotted on the y-axis against genomic position (NCBI Build 36) on the x-axis. The top signal is represented by a diamond. Estimated recombination rates (taken from HapMap) are plotted to reflect the local linkage disequilibrium structure around the associated SNPs and their correlated proxies (according to a greyscale from $r^2 = 0$ to 1, based on pairwise $r^2$ values from HapMap Phase II CEU).

**Figure 2.** Schematics of metabolic pathways relevant to SNP-metabolite associations

**Figure 3.** Links between glycine, serine, folate, homocysteine and betaine in folate metabolism and homocysteine metabolism

Modified from (44) and (45)

Enzymes: (1) dihydrofolate reductase; (2) serine hydroxymethyltransferase (SHMT); (3) glycine synthase (also called glycine cleavage enzyme); (4) methylenetetrahydrofolate reductase (MTHFR); (5) methionine synthase (the other name of 5-methyltetrahydrofolate-homocysteine methyltransferase (MHMT)) (6) betaine-homocysteinemethyltransferase (BHMT). (Modified from (44) and (45)).
REFERENCES

Figure 1. Regional association plots of the five SNP-metabolite associations in RISC cohort

In each plot, the top panel shows the name and location of genes in the UCSC Genome Browser. The $-\log_{10}$ of $P$-values of the imputed SNPs are plotted on the $y$-axis against genomic position (NCBI Build 36) on the $x$-axis. The top signal is represented by a diamond.

Estimated recombination rates (taken from HapMap) are plotted to reflect the local linkage disequilibrium structure around the associated SNPs and their correlated proxies (according to a greyscale from $r^2 = 0$ to 1, based on pairwise $r^2$ values from HapMap Phase II CEU).
Figure 2. Schematics of metabolic pathways relevant to SNP-metabolite associations
177x145mm (96 x 96 DPI)
Figure 3. Links between glycine, serine, folic acid, homocysteine and betaine in folate metabolism and homocysteine metabolism

Enzymes: (1) dihydrofolate reductase; (2) serine hydroxymethyltransferase (SHMT); (3) glycine synthase (also called glycine cleavage enzyme); (4) methylenetetrahydrofolate reductase (MTHFR); (5) methionine synthase (the other name of 5-methyltetrahydrofolate-homocysteine methyltransferase (MHMT)) (6) betaine-homocysteine methyltransferase (BHMT). (Modified from (40) and (41)).
Online Supplemental Materials for
Genetic variants associated with glycine metabolism and their role in insulin sensitivity and type 2 diabetes

Weijia Xie\textsuperscript{1}, Andrew R. Wood\textsuperscript{1}, Valeriya Lyssenko\textsuperscript{2}, Michael N. Weedon\textsuperscript{1}, Joshua W. Knowles\textsuperscript{3}, Sami Alkayyali\textsuperscript{2}, Themistocles L. Assimes\textsuperscript{3}, Thomas Quertermous\textsuperscript{3}, Fahim Abbasi\textsuperscript{3}, Jussi Paananen\textsuperscript{4}, Hans Häring\textsuperscript{5}, Torben Hansen\textsuperscript{6,13}, Oluf Pedersen\textsuperscript{6,14,15,16}, Ulf Smith\textsuperscript{7}, Markku Laakso\textsuperscript{4}, MAGIC investigators, the DIAGRAM consortium, the GENESIS consortium, the RISC consortium, Jacqueline M. Dekker\textsuperscript{8}, John J. Nolan\textsuperscript{9}, Leif Groop\textsuperscript{2}, Ele Ferrannini\textsuperscript{10}, Klaus-Peter Adam\textsuperscript{11}, Walter E. Gall\textsuperscript{11}, Timothy M. Frayling\textsuperscript{1}, Mark Walker\textsuperscript{12}

\textbf{Affiliations}

\textsuperscript{1}Genetics of Complex Traits, Peninsula School of Medicine, University of Exeter, UK. \textsuperscript{2}Lund University Diabetes Center, Department of Clinical Sciences, Diabetes and Endocrinology, Lund University, Malmoe, Sweden. \textsuperscript{3}Department of Medicine, Stanford University School of Medicine, Stanford, California, USA. \textsuperscript{4}Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland. \textsuperscript{5}Department of Internal Medicine, Division of Endocrinology, Diabetology, Nephrology, Vascular Medicine and Clinical Chemistry, University of Tübingen, Tübingen, Germany. \textsuperscript{6}Novo Nordisk Foundation Center for Basic Metabolic Research Faculty of Health Science, University of Copenhagen, Copenhagen, Denmark. \textsuperscript{7}The Lundberg Laboratory for Diabetes Research, Department of Molecular and Clinical Medicine, Sahlgrenska Academy, Gothenburg, Sweden. \textsuperscript{8}Department of Epidemiology and Biostatistics, Vrije Universiteit Medical Center, Amsterdam, The Netherlands. \textsuperscript{9}Steno Diabetes Center, Gentofte, Denmark. \textsuperscript{10}Department of Internal Medicine, University of Pisa, Pisa, Italy. \textsuperscript{11}Metabolon, Inc., Research Triangle Park, North Carolina, US. \textsuperscript{12}Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne, UK. \textsuperscript{13}Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark. \textsuperscript{14}Hagedorn Research Institute, Copenhagen, Denmark. \textsuperscript{15}Institute of Biomedical Science, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark. \textsuperscript{16}Faculty of Health Sciences, University of Aarhus, Aarhus, Denmark.
Corresponding author:

Timothy M. Frayling
Peninsula College of Medicine and Dentistry
University of Exeter
Magdalen Road
Exeter
EX1 2LU
tel. +44 1392 262935
fax. +44 1392 262926
email: tim.frayling@pms.ac.uk

This file includes:

1, Characteristics of individual studies
2, Genotyping information of individual studies
3, Phenotyping information of individual studies
Tables S1 to S4
Figures S1 to S2
MAGIC investigators
DIAGRAM consortium
GENESIS consortium
References
Supporting online material

1. Characteristics of individual studies

1.1. RISC (Relationship between Insulin Sensitivity and Cardiovascular disease)

The RISC study is based on unrelated individuals of European descent, aged 30–60 years with no clinical signs of disease, recruited from 19 centres in 14 countries. A summary of individuals and relevant characteristics used in this study is given in Table 1. The RISC methodologies have been described in detail previously (1). In brief, the initial exclusion criteria were: treatment for obesity, hypertension, lipid disorders or diabetes, pregnancy, cardiovascular or chronic lung disease, weight change of 5 kg or more in last 6 months, cancer (in last 5 yr), and renal failure. Exclusion criteria after screening were: arterial blood pressure 140/90 mm Hg or higher; fasting plasma glucose 7.0 mmol/liter or greater; 2-h plasma glucose [on a 75-g oral glucose tolerance test (OGTT)] 11.0 mmol/L or greater; total serum cholesterol 7.8 mmol/L or greater; serum triglycerides 4.6 mmol/L or greater; and electrocardiogram abnormalities.

1.2. Botnia

To replicate association signals we used metabolite data and bespoke genotyping in the Botnia study. The Botnia study started in 1990 at the West coast of Finland with an aim to identify genes associated with susceptibility to type 2 diabetes (2). All subjects were given information about exercise and healthy diet and exposed at 2-3 years intervals to a new OGTT. 342 non-diabetic individuals with available metabolite measurements were included in the current analyses. Characteristics of the Botnia individuals are listed in Supplementary Table 1.

1.3. EUGENE2 (European network on Functional Genomics of Type 2 Diabetes)

The participants included in this study were healthy, non-diabetic offspring of patients with type 2 diabetes. For inclusion, one of the parents had to have type 2 diabetes and the other
parent normal glucose tolerance evaluated by an OGTT or a lack of history of type 2 diabetes. The probands were randomly selected from regions of four centres in Europe: Copenhagen, Denmark (n=278), Gothenburg, Sweden (n=100), Kuopio, Finland (n=217) and Tubingen, Germany (n=149). All study participants underwent a standard medical history, routine laboratory testing, assessment of lifestyle factors (alcohol consumption, activity, smoking status), and an OGTT. The protocol was approved by the Ethics Committee of the corresponding study centres, and informed written consent was obtained from all participants. Further details about the EUGENE2 consortium are provided in (3).

1.4. Stanford Insulin Suppression Test (IST) Cohort

Stanford (IST) Cohort (n=381) includes a subset of all subjects participating in various clinical research studies at Stanford University Medical that called for at least one insulin suppression test (IST) between 2002 and 2007. Volunteers for these studies came from the surrounding Stanford communities and were generally free of major chronic medical conditions at the time of their IST. Subjects were not eligible to participate in any protocol if they reported being on medications known to influence insulin sensitivity including corticosteroids, metformin, sulfonylureas or thiazolidinediones. Steady state plasma glucose (SSPG) derived from an IST was used as a direct measure of insulin sensitivity. In this study, we include only those genotyped individuals reporting white-non Hispanic ancestry (n=263).

2. Genotyping information of individual studies

2.1. Genotyping in RISC

Samples were genotyped on the Affymetrix 6.0 microarray platform, and 909,508 single nucleotide polymorphisms (SNPs) were called in total. We then used MACH to impute the genotypes of all HapMap version 22 SNPs on Chr 1-22 in all individuals (1, 4, 5). The exclusion criteria for quality control were: SNP-wise call rate <98%, imputation quality $r^2$ hat <0.3, minor allele frequency (MAF) <5%, non-European descent and sex-mismatches.
2.2. Genotyping in Botnia

The allelic discrimination method was used with a TaqMan assay on the ABI 7900 platform (Applied Biosystems, Foster City, CA, USA) to genotype replication SNPs in 2,770 non-diabetic individuals from the Botnia study.

2.3. Genotyping in EUGENE2

Samples were genotyped on the Illumina 550K platform in the Kuopio centre. In total, 561,301 SNPs were called. The autosomal SNPs were then imputed in HapMap Phase II CEU panel using MACH (6). SNPs of low imputation quality (SNP-wise call rate <95%, imputation quality $r^2_{\text{hat}} <0.3$, MAF < 5%) were dropped from analysis. Individuals showing non-European ancestry were removed.

2.4. Genotyping in Stanford IST

Samples from the Stanford IST cohort were genotyped on the Affymetrix 6.0 microarray platform. In total, 909,508 SNPs were called. Genotype imputation was conducted using the MACH program with HapMap Phase II reference (6). Quality control steps removed SNPs of SNP-wise call rate <90%, MAF<2%, imputation quality $r^2_{\text{hat}} <0.3$, and Hardy-Weinberg $P$-value <1x10^{-6}. Individuals showing non-European ancestry were removed.

3. Phenotyping information of individual studies

3.1. RISC

Insulin sensitivity (M-value) was measured by hyperinsulinemic euglycemic clamp as previously described (1). Exogenous insulin was administered as a primed-continuous intravenous infusion at a rate of 240 pmol min$^{-1}$m$^{-2}$ for 120min, simultaneously with a variable 20% (wt/vol) glucose infusion. This was adjusted every 5 to 10 min to maintain plasma glucose levels within 0.8 mmol/l of the target glucose level (4.5–5.5 mmol/l). Insulin sensitivity was assessed as the mean glucose infusion rate over the last 40 min of the clamp, corrected for the body weight (M-value) (micromol/kg bodywt/min). To ensure consistency
across study centres, the clamp procedure was standardized. Fasting insulin was measured by a two-sited, time-resolved fluoroimmunoassay (AutoDELFIA Insulin kit, Wallac Oy, Turku, Finland) using monoclonal antibodies.

### 3.2. EUGENE2

Participants underwent 75-g oral (OGTT) and intravenous glucose tolerance tests (IVGTT). A bolus of glucose (300 mg/kg in a 50% solution) was given into the antecubital vein within 30 s in the IVGTT. At 60 min after the glucose bolus a euglyceamic-hyperinsulineamic clamp was initiated (insulin infusion: 240 pmol m−2 min−1 for 120 min) to evaluate insulin sensitivity (7). Glucose was clamped at 5.0 mmol/l for the next 120 min by infusion of 20% glucose at various rates according to glucose measurements performed at 5 min intervals. The glucose disposal during the clamp was expressed as the amount of glucose infused per kilogram body weight per minute during the last 60 min of the clamp examination (micromol/kgbodywt/min).

### 3.3. Stanford IST

Insulin sensitivity was measured by steady-state plasma glucose (SSPG) method. The SSPG value is highly inversely correlated to M-value ($r = -0.93, P<0.001$) (8).
**Supplementary Table 1**: Summary details of Botnia individuals and relevant characteristics

<table>
<thead>
<tr>
<th>Phenotypes</th>
<th>name (unit)</th>
<th>age (yrs)</th>
<th>BMI (kg/m²)</th>
<th>FI (pmol/l)</th>
<th>FG (mmol/l)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>342 342</td>
<td>342 342</td>
<td>338 338</td>
<td>342 342</td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>46.98</td>
<td>25.63</td>
<td>31.9</td>
<td>5.63</td>
<td></td>
</tr>
<tr>
<td>sd</td>
<td>14.07</td>
<td>3.63</td>
<td>19.18</td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td>median</td>
<td>46.58</td>
<td>25.39</td>
<td>26.46</td>
<td>5.65</td>
<td></td>
</tr>
<tr>
<td>min</td>
<td>5.31</td>
<td>15.58</td>
<td>6.54</td>
<td>4.18</td>
<td></td>
</tr>
<tr>
<td>max</td>
<td>78.7</td>
<td>41.23</td>
<td>134.04</td>
<td>6.89</td>
<td></td>
</tr>
</tbody>
</table>

FI: fasting insulin; FG: fasting glucose
**Supplementary Table 2:** Eight GWAS and candidate-region signals of insulin-sensitivity associated metabolites in the RISC (n=1,004) and Botnia (n=339) studies

<table>
<thead>
<tr>
<th>Chr</th>
<th>Pos</th>
<th>SNP</th>
<th>‘Analysis</th>
<th>Trait</th>
<th>Genes</th>
<th>Effect allele/other allele</th>
<th>RISC effect allele freq</th>
<th>RISC effect (95% CIs)</th>
<th>RISC P</th>
<th>Botnia effect (95% CIs)</th>
<th>Botnia P</th>
<th>Meta effect (95% CIs)</th>
<th>Meta P</th>
<th>Meta N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>120056649</td>
<td>rs478093</td>
<td>GWAS</td>
<td>serine</td>
<td><em>PHGDH</em></td>
<td>G/A</td>
<td>0.71</td>
<td>(0.20, 0.39)</td>
<td>1.5×10⁻⁹</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>2</td>
<td>211251300</td>
<td>rs715</td>
<td>GWAS</td>
<td>glycine</td>
<td><em>CPS1</em></td>
<td>T/C</td>
<td>0.68</td>
<td>(-0.68, -0.48)</td>
<td>5.3×10⁻⁻³⁰</td>
<td>-0.70</td>
<td>(-0.54, -0.86)</td>
<td>2.0×10⁻¹⁶</td>
<td>-0.61</td>
<td>(0.33, 0.39)</td>
</tr>
<tr>
<td>3</td>
<td>127386855</td>
<td>rs1107366</td>
<td>GWAS</td>
<td>glycine/serine</td>
<td><em>ALDH11L1, KLF15</em></td>
<td>G/A</td>
<td>0.51</td>
<td>(0.01, 0.03)</td>
<td>5×10⁻⁶</td>
<td>0.02</td>
<td>(0.002, 0.03)</td>
<td>0.02</td>
<td>(0.011, 0.027)</td>
<td>2.3×10⁻⁶</td>
</tr>
<tr>
<td>5</td>
<td>78377053</td>
<td>rs17823642</td>
<td>candidate</td>
<td>betaine</td>
<td><em>BHMT, BHMT2</em></td>
<td>C/T</td>
<td>0.89</td>
<td>(0.21, 0.49)</td>
<td>1.4×10⁻⁶</td>
<td>0.52</td>
<td>(0.8, 0.24)</td>
<td>3.5×10⁻⁴</td>
<td>0.38</td>
<td>2.4×10⁻⁹</td>
</tr>
<tr>
<td>7</td>
<td>55908571</td>
<td>rs13233754</td>
<td>candidate</td>
<td>serine</td>
<td><em>PSPH</em></td>
<td>G/A</td>
<td>0.97</td>
<td>(-0.38, -0.33)</td>
<td>2.0×10⁻⁵</td>
<td>0.02</td>
<td>(-0.59, 0.62)</td>
<td>0.96</td>
<td>(-0.59, 0.62)</td>
<td>1.0×10⁻⁴</td>
</tr>
<tr>
<td>7</td>
<td>56071882</td>
<td>rs4275190</td>
<td>candidate</td>
<td>serine</td>
<td><em>PSPH</em></td>
<td>T/C</td>
<td>0.68</td>
<td>(0.08, 0.27)</td>
<td>4.0×10⁻⁴</td>
<td>0.08</td>
<td>(-0.26, 0.43)</td>
<td>0.63</td>
<td>(0.07, 0.26)</td>
<td>3.9×10⁻⁴</td>
</tr>
<tr>
<td>11</td>
<td>61322484</td>
<td>rs174541</td>
<td>GWAS</td>
<td>adrenate</td>
<td><em>FADS1-3, FEN1</em></td>
<td>T/C</td>
<td>0.65</td>
<td>(0.19, 0.37)</td>
<td>2.9×10⁻⁹</td>
<td>widely reported signal, not followed up</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>191181</td>
<td>rs499368</td>
<td>GWAS</td>
<td>betaine</td>
<td><em>SLC6A12, SLC6A13</em></td>
<td>A/T</td>
<td>0.51</td>
<td>(-0.61, -0.3)</td>
<td>8.1×10⁻⁹</td>
<td>-0.25</td>
<td>(-0.41, -0.1)</td>
<td>1.5×10⁻³</td>
<td>-0.36</td>
<td>(0.46, 0.25)</td>
</tr>
</tbody>
</table>

*GWAS: signals that reach conventional threshold for genome wide significance (P-value<5x10⁻⁸). Candidate: signals within ±300kb of a candidate gene that reach threshold corrected for multiple testing of the total number of SNPs in the candidate regions. NA: genotype data not available in Botnia. Meta: the meta-analysis of RISC and Botnia studies.
The glycine/serine ratio was normalised by log 10 transformation, and the residuals after adjusting for age, sex and centre were used as the trait in GWA analysis. The other traits were normalised by inverse normal transformation and standardised (i.e. in SD unit). This P-value of rs1107366 on glycine/serine ratio was genome wide significant when meta-analysing RISC with published KORA data ($P$-value = $3.71 \times 10^{-11}$; $n$ = 2,687), therefore followed up in Botnia study, $P$-value when combining RISC, Botnia, KORA and UKtwins data is $2.8 \times 10^{-12}$ ($n$ = 3,026).

Gene names: $PHGDH$: phosphoglycerate dehydrogenase; $CPS1$: carbamoyl-phosphate synthase 1; $ALDH1L$: aldehyde dehydrogenase 1 family, member L1; $BHMT$: betaine-homocysteine S-methyltransferase/2; $PSPH$: phosphoserine phosphatase; $FADS1$-$3$: fatty acid desaturase 1-3; $FEN1$: flap structure-specific endonuclease 1; $SLC6A12/13$: solute carrier family 6 (neurotransmitter transporter, betaine/GABA), member 12/13;
Supplementary Table 3: The effect of rs1107366 on insulin resistance in RISC, EUGENE2, Stanford and meta-analysis

<table>
<thead>
<tr>
<th>Study</th>
<th>Freq (G)</th>
<th>Observed effect on insulin resistance (95% CIs)</th>
<th>Insulin sensitivity measures</th>
<th>P</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>RISC</td>
<td>0.51</td>
<td>0.10, (0.02,0.18)</td>
<td>M value</td>
<td>0.02</td>
<td>1,007</td>
</tr>
<tr>
<td>EUGENE2</td>
<td>0.46</td>
<td>-0.04, (-0.16,0.08)</td>
<td>M value</td>
<td>0.51</td>
<td>614</td>
</tr>
<tr>
<td>Stanford IST</td>
<td>0.43</td>
<td>0.15, (-0.92×10⁻³, 0.30)</td>
<td>inverse SSPG</td>
<td>0.06</td>
<td>327</td>
</tr>
<tr>
<td>Meta-analysis</td>
<td>0.48</td>
<td>0.09, (0.03,0.15)</td>
<td>/</td>
<td>5.5×10⁻³</td>
<td>1,884</td>
</tr>
</tbody>
</table>

Heterogeneity across RISC, EUGENE2 and STANFORD $P$-value = 0.07.

Supplementary Table 4: Type 2 diabetes odds ratios of metabolite-associated SNPs from the DIAGRAM consortium.

<table>
<thead>
<tr>
<th>SNP*</th>
<th>Chr</th>
<th>Position</th>
<th>Effect allele</th>
<th>Other allele</th>
<th>OR (95% CIs)</th>
<th>P</th>
<th>N (cases)</th>
<th>N (controls)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs478093</td>
<td>1</td>
<td>120,056,649</td>
<td>G</td>
<td>A</td>
<td>1.01, (0.96, 1.05)</td>
<td>0.8</td>
<td>8,130</td>
<td>38,987</td>
</tr>
<tr>
<td>rs1107366</td>
<td>3</td>
<td>127,386,855</td>
<td>G</td>
<td>A</td>
<td>0.99, (0.95, 1.03)</td>
<td>0.54</td>
<td>8,130</td>
<td>38,987</td>
</tr>
<tr>
<td>rs17823642</td>
<td>5</td>
<td>78,377,053</td>
<td>C</td>
<td>T</td>
<td>0.98, (0.92, 1.05)</td>
<td>0.63</td>
<td>8,130</td>
<td>38,987</td>
</tr>
<tr>
<td>rs499368</td>
<td>12</td>
<td>191,181</td>
<td>A</td>
<td>T</td>
<td>0.99, (0.94, 1.05)</td>
<td>0.79</td>
<td>4,751</td>
<td>33,536</td>
</tr>
</tbody>
</table>

The glycine associated SNP in CPS1 (rs715) was not captured well in the DIAGRAM studies, so details of four of the five metabolite SNPs are presented.
Supplementary Figure 1: The triangulation approach used in the Mendelian randomisation analyses

- Insulin sensitivity was measured by fasting insulin and hyperinsulinemic-euglycemic clamp (M-value)
**Supplementary Figure 2:** Regional association plots of sex-specific GWAS top signal of glycine in RISC group

In each plot, the top panel shows the name and location of genes in the UCSC Genome Browser. The $-\log_{10}$ values of the imputed SNPs are plotted on the y-axis against genomic position (NCBI Build 36) on the x-axis. The top signal is represented by a purple diamond. Estimated recombination rates (taken from HapMap) are plotted to reflect the local linkage disequilibrium structure around the associated SNPs and their correlated proxies (according to a blue to red scale from $r^2 = 0$ to 1, based on pairwise $r^2$ values from HapMap Phase II CEU).
Borecki\textsuperscript{44}, Ruth J F Loos\textsuperscript{3}, Pierre Meneton\textsuperscript{80}, Patrik KEMagnusson\textsuperscript{42}, David MNathan\textsuperscript{104,105}, Gordon H Williams\textsuperscript{69,105}, Andrew THattersley\textsuperscript{98}, Kaisa Siilander\textsuperscript{96,111}, Veikko Salomaa\textsuperscript{146}, George Davey Smith\textsuperscript{38}, Stefan R Bornstein\textsuperscript{73}, Peter Schwarz\textsuperscript{27}, Joachim Spranger\textsuperscript{67,68}, Fredrik Karpe\textsuperscript{4,107}, Alan R Shuldiner\textsuperscript{45}, Cyrus Cooper\textsuperscript{125}, George V Dedoussis\textsuperscript{84}, Manuel Serrano-Rios\textsuperscript{39}, Andrew DMorris\textsuperscript{109}, Lars Lind\textsuperscript{132}, Lyle J Palmer\textsuperscript{64,66,84}, Frank B Hu\textsuperscript{147,148}, Paul WFranks\textsuperscript{149}, Shah Ebrahim\textsuperscript{150}, Michael Marmot\textsuperscript{36}, WH Linda Kao\textsuperscript{33,151,152}, James SPankow\textsuperscript{153}, Michael J Sampson\textsuperscript{154}, Johanna Kuusisto\textsuperscript{155}, Markku Laakso\textsuperscript{155}, Torben Hansen\textsuperscript{11,136}, Oluf Pedersen\textsuperscript{31,59,157}, Peter Paul Pramstaller\textsuperscript{82,158,159}, H Erich Wichmann\textsuperscript{21,160,161}, Thomas Illig\textsuperscript{21}, Igor Rudan\textsuperscript{24,162,163}, Alan F Wright\textsuperscript{25}, Michael Stunov\textsuperscript{60}, Harry Campbell\textsuperscript{24}, James F Wilson\textsuperscript{24}, Anders Hamsten on behalf of Procardis Consortium\textsuperscript{128}, Richard N Bergman\textsuperscript{164}, Thomas A Buchanan\textsuperscript{164,165}, Francis SCollins\textsuperscript{47}, Karen LMohlke\textsuperscript{166}, Jaakko Tuomilehto\textsuperscript{94,167,168}, Timo TValle\textsuperscript{167}, David Altmüller\textsuperscript{6,7,104,105}, Jerome I Rotter\textsuperscript{62}, David SSiscovick\textsuperscript{169}, Brenda WJ H Penninx\textsuperscript{140}, Dorret I Boomsma\textsuperscript{23}, Panos Deloukas\textsuperscript{8}, Timothy DSpector\textsuperscript{8}, Timothy M Frayling\textsuperscript{28}, Luigi Ferrucci\textsuperscript{170}, Augustine Kong\textsuperscript{19}, Unnur Thorsteinsdottir\textsuperscript{19,177}, Kari Stefansson\textsuperscript{19,171}, Cornelia Mvan Duijn\textsuperscript{22}, Yurii SAulchenko\textsuperscript{22}, Antonio Cao\textsuperscript{65}, Angelo Scuteri\textsuperscript{172,177}, David Schlessinger\textsuperscript{47}, Manuela Uda\textsuperscript{65}, Aimo Ruokonen\textsuperscript{73}, Marjo-Riitta Jarvelin\textsuperscript{17,93,174}, Dawn MWaterval\textsuperscript{26}, Peter Vollenweider\textsuperscript{141}, Leena Peltonen\textsuperscript{94,96,111,112}, Vincent Mooser\textsuperscript{34}, Goncalo R Abecasis\textsuperscript{10}, Nicholas J Wareham\textsuperscript{3}, Robert Sladek\textsuperscript{40,41}, Philippe Froguel\textsuperscript{13,142}, James BMcKenzie\textsuperscript{45}, Leif Groop\textsuperscript{102}, Michael Boehnke\textsuperscript{10}, Mark I McCarthy\textsuperscript{4,5,107}, Jose CFlorez\textsuperscript{6,7,104,105} & Inès Barroso\textsuperscript{11} for the MAGIC investigators

\textsuperscript{1}Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA. \textsuperscript{2}National Heart, Lung, and Blood Institute’s Framingham Heart Study, Framingham, Massachusetts, USA. \textsuperscript{3}Medical Research Council (MRC), Epidemiology Unit, Institute of Metabolic Science, Addenbrooke’s Hospital, Cambridge, UK. \textsuperscript{4}Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK. \textsuperscript{5}Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. \textsuperscript{6}Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA. \textsuperscript{7}Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA. \textsuperscript{8}Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK. \textsuperscript{9}Twin Research and Genetic Epidemiology Department, King’s College London, St Thomas’ Hospital Campus, London, UK. \textsuperscript{10}Center for Statistical Genetics, Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, Michigan, USA. \textsuperscript{11}Metabolic Disease Group, Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK. \textsuperscript{12}Cardiovascular Health Research Unit and Department of Medicine, University of Washington, Seattle, Washington, USA. \textsuperscript{13}Centre National de la Recherche Scientifique–Unité Mixte de Recherche 8090, Pasteur Institute, Lille 2–Droit et Santé University, Lille, France. \textsuperscript{14}Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland. \textsuperscript{15}University Institute of Social and Preventative Medicine, Centre Hospitalier Universitaire Vaudois (CHUV) and University of Lausanne, Lausanne, Switzerland. \textsuperscript{16}Swiss Institute of Bioinformatics, Lausanne, Switzerland. \textsuperscript{17}Department of Epidemiology and Public Health, Imperial College London, Faculty of Medicine, Norfolk Place, London, UK. \textsuperscript{18}Boston University Data Coordinating Center, Boston, Massachusetts, USA. \textsuperscript{19}deCODE Genetics, Reykjavik, Iceland. \textsuperscript{20}Department of Human Genetics, Leiden University Medical Centre, Leiden, The Netherlands. \textsuperscript{21}Institute of Epidemiology, Helmholtz Zentrum Muenchen, German Research Center for Environmental Health, Neuherberg, Germany. \textsuperscript{22}Department of Epidemiology, Erasmus Medical College, Rotterdam, The Netherlands. \textsuperscript{23}Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands. \textsuperscript{24}Centre for Population Health
Sciences, University of Edinburgh, Edinburgh, UK.  
25MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, Edinburgh, UK.  
26Division of Genetics, Research and Development, GlaxoSmithKline, King of Prussia, Pennsylvania, USA.  
27Department of Cardiovascular Medicine, University of Oxford, Oxford, UK.  
28Genetics of Complex Traits, Institute of Biomedical and Clinical Sciences, Peninsula College of Medicine and Dentistry, University of Exeter, Exeter, UK.  
29National Institute of Aging, Baltimore, Maryland, USA.  
30Unit for Child and Adolescent Health and Welfare, National Institute for Health and Welfare, Biocenter Oulu, University of Oulu, Oulu, Finland.  
31Hagedorn Research Institute, Gentofte, Denmark.  
32Department of Medicine and Therapeutics, Level 7, Ninewells Hospital and Medical School, Dundee, UK.  
33Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA.  
34Department of Nutrition–Dietetics, Harokopio University, Athens, Greece.  
35General Medicine Division, Massachusetts General Hospital, Boston, Massachusetts, USA.  
36Department of Epidemiology and Public Health, University College London, London, UK.  
37Departments of Nutrition and Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA.  
38MRC Centre for Causal Analyses in Translational Epidemiology, University of Bristol, Bristol, UK.  
39Fundación para la Investigación Biomédica del Hospital Clínico San Carlos, Madrid, Spain.  
40Departments of Medicine and Human Genetics, McGill University, Montreal, Canada.  
41Genome Quebec Innovation Centre, Montreal, Canada.  
42Department of Medical Epimediology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.  
43Department of Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden.  
44Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, Missouri, USA.  
45Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, Maryland, USA.  
46INSERM U859, Université de Lille-Nord de France, Lille, France.  
47Genome Technology Branch, National Human Genome Research Institute, Bethesda, Maryland, USA.  
48The Broad Institute, Cambridge, Massachusetts, USA.  
49Leiden Genome Technology Center, Leiden University Medical Center, Leiden, The Netherlands.  
50INSERM U780, Paris Sud University, Villejuif, France.  
51The Heart Research Institute, Sydney, New South Wales, Australia.  
52PathWest Laboratory of Western Australia, Department of Molecular Genetics, J Block, QEII Medical Centre, Nedlands West Australia, Australia.  
53School of Surgery and Pathology, University of Western Australia, Nedlands West Australia, Australia.  
54Department of Social Medicine, University of Bristol, Bristol, UK.  
55Landspitali University Hospital, Reykjavik, Iceland.  
56Icelandic Heart Association, Kopavogur, Iceland.  
57The Human Genetics Center and Institute of Molecular Medicine, University of Texas Health Science Center, Houston, Texas, USA.  
58Steno Diabetes Center, Gentofte, Denmark.  
59Faculty of Health Science, University of Aarhus, Aarhus, Denmark.  
60Department of Medicine, University of Leipzig, Leipzig, Germany.  
61Endocrinology–Diabetology Unit, Corbeil-Essonnes Hospital, Essones, France.  
62Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA.  
63Clinical Trial Service Unit and Epidemiological Studies Unit, University of Oxford, Oxford, UK.  
64Centre for Genetic Epidemiology and Biostatistics, University of Western Australia, Perth, Australia.  
65Istituto di Neurogenetica e Neurofarmacologia (INN), Consiglio Nazionale delle Ricerche, c/o Cittadella Universitaria di Monserrato, Monserrato, Cagliari, Italy.  
66Western Australian Sleep Disorders Research Institute, Queen Elizabeth Medical Centre II, Perth, Australia.  
67Department of Endocrinology, Diabetes and Nutrition, Charite-Universitaetsmedizin Berlin, Berlin, Germany.  
68Department of Clinical Nutrition, German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany.  
69Division of Endocrinology, Diabetes, and Hypertension, Brigham and Women’s Hospital, Harvard Medical School, Boston, Massachusetts, USA.  
70Department of Human Genetics, Leiden University Medical Centre, Leiden, The Netherlands.  
71Department of Cardiovascular
Research, Istituto di Ricerche Farmacologiche ‘Mario Negri’, Milan, Italy. 72 Institut National de la Santé et de la Recherche Médicale, Institut National de la Recherche Agronomique, Université Paris 13, Bobigny Cedex, France. 73 Department of Medicine III, Division Prevention and Care of Diabetes, University of Dresden, Dresden, Germany. 74 Center for Human Nutrition, University of Texas Southwestern Medical Center, Dallas, Texas, USA. 75 Department of Genetics and Pathology, Rudbeck Laboratory, Uppsala University, Uppsala, Sweden. 76 Centre Hospitalier Universitaire, de Poitiers, Endocrinologie Diabetologie, CIC INSERM 0802, INSERM U927, Université de Poitiers, Unité de Formation et de Recherche, Médecine Pharmacie, Poitiers, France. 77 Department of Public Health and Clinical Medicine, Section for Nutritional Research, Umeå University, Umeå, Sweden. 78 Department of Clinical Sciences, Obstetrics and Gynecology, University of Oulu, University of Oulu, Finland. 79 Centre National de Génotypage/Institut de génomique/Commissariat à l’énergie atomique, Evry Cedex, France. 80 INSERM U872, Faculté de Médecine Paris Descartes, Paris Cedex, France. 81 Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Düsseldorf, Germany. 82 Institute of Genetic Medicine, European Academy Bozen/Bolzano (EURAC), Viale Druso, Bolzano, Italy, Affiliated Institute of the University Lübeck, Lübeck, Germany. 83 Department of Pulmonary Physiology, Sir Charles Gairdner Hospital, Perth, Australia. 84 Busselton Population Medical Research Foundation, Sir Charles Gairdner Hospital, Perth, Australia. 85 Heart Institute of Western Australia, Sir Charles Gairdner Hospital, Nedlands West Australia, Australia. 86 School of Medicine and Pharmacology, University of Western Australia, Nedlands West Australia, Australia. 87 Folkhalsan Research Centre, Helsinki, Finland. 88 Malmö Municipal Health Care Center and Hospital, Jakobstad, Finland. 89 Nuffield Department of Surgery, University of Oxford, Oxford, UK. 90 Research Centre for Prevention and Health, Glostrup University Hospital, Glostrup, Denmark. 91 Faculty of Health Science, University of Copenhagen, Copenhagen, Denmark. 92 National Institute for Health and Welfare, Unit of Population Studies, Turku, Finland. 93 Institute of Health Sciences and Biocenter Oulu, University of Oulu, Oulu, Finland. 94 Department of Public Health, Faculty of Medicine, University of Helsinki, Helsinki, Finland. 95 National Institute for Health and Welfare, Unit for Child and Adolescent Mental Health, Helsinki, Finland. 96 Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland. 97 Department of Internal Medicine and Biocenter Oulu, Oulu, Finland. 98 Diabetes Genetics, Institute of Biomedical and Clinical Science, Peninsula College of Medicine and Dentistry, University of Exeter, Exeter, UK. 99 National Institute for Health and Welfare, Unit of Living Conditions, Health and Wellbeing, Helsinki, Finland. 100 Interdisciplinary Centre for Clinical Research, University of Leipzig, Leipzig, Germany. 101 The Danish Twin Registry, Epidemiology, Institute of Public Health, University of Southern Denmark, Odense, Denmark. 102 Department of Clinical Sciences, Diabetes and Endocrinology, Lund University, University Hospital Malmö, Malmö, Sweden. 103 Gladstone Institute of Cardiovascular Disease, University of California, San Francisco, California, USA. 104 Diabetes Research Center, Diabetes Unit, Massachusetts General Hospital, Boston, Massachusetts, USA. 105 Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA. 106 Division of Cardiology, University of Ottawa Heart Institute, Ottawa, Ontario, Canada. 107 Oxford National Institute for Health Research, Biomedical Research Centre, Churchill Hospital, Oxford, UK. 108 Department of Clinical Genetics, Erasmus Medical College, Rotterdam, The Netherlands. 109 Biomedical Research Institute, University of Dundee, Ninewells Hospital and Medical School, Dundee, UK. 110 Department of Geriatric Medicine and Metabolic Disease, Second University of Naples, Naples, Italy. 111 National Institute for Health and Welfare, Unit of Public Health Genomics, Helsinki, Finland. 112 Department of Medical Genetics, University of Helsinki, Helsinki, Finland. 113 Department of Medical
These authors contributed equally to this work. Correspondence should be addressed to M.B. (boehnke@umich.edu), M.I.M. (mark.mccarthy@drl.ox.ac.uk), J.C.F. (jcflorez@partners.org) or I.B. (ib1@sanger.ac.uk).
References
Supplementary Figure 1. The triangulation approach used in the Mendelian randomisation analyses.

Insulin sensitivity was measured by fasting insulin and hyperinsulinaemic-euglycemic clamp (M-value).
Supplementary Figure 2. Regional association plots of sex-specific GWAS top signal of glycine in RISC group

In each plot, the top panel shows the name and location of genes in the UCSC Genome Browser. The $-\log_{10}$ of $P$-values of the imputed SNPs are plotted on the y-axis against genomic position (NCBI Build 36) on the x-axis. The top signal is represented by a purple diamond. Estimated recombination rates (taken from HapMap) are plotted to reflect the local linkage disequilibrium structure around the associated SNPs and their correlated proxies (according to a blue to red scale from $r^2 = 0$ to $1$, based on pairwise $r^2$ values from HapMap Phase II CEU).

283x115mm (96 x 96 DPI)