Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes

Citation for published version:
DIAGRAM Consortium, GIANT Consortium, Morris, A, Rudan, I, Campbell, H & Wilson, J 2011, 'Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes', Diabetes, vol. 60, no. 10, pp. 2624-34. https://doi.org/10.2337/db11-0415

Digital Object Identifier (DOI):
10.2337/db11-0415

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Diabetes

Publisher Rights Statement:
© 2011 by the American Diabetes Association.

Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. See http://creativecommons.org/licenses/by-nc-nd/3.0/ for details.

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Genome-Wide Association Identifies Nine Common Variants Associated With Fasting Proinsulin Levels and Provides New Insights Into the Pathophysiology of Type 2 Diabetes

Rona J. Strawbridge,¹ Joséé Dupuis,² Inga Prokopenko,⁴,⁵ Adam Barker,⁶ Emma Ahlvqvist,⁷ Denis Rybin,⁸ John R. Petrie,⁹ Mary E. Travers,⁴ Nabila Bouatia-Naji,¹⁰,¹¹ Antigone S. Dimas,¹²,¹³ Alexandra Nica,¹⁴,¹⁵ Eleanor Wheeler,¹⁶ Han Chen,² Benjamin F. Voight,¹⁵,¹⁶ Jalal Taneera,⁷ Stavroula Kanoni,¹³,¹⁷ John F. Pedersen,⁵,¹⁸ Fabiola Turrini,⁷,¹⁹ Stefan Gustafsson,²⁰ Carina Zabena,²¹,²² Peter Almgren,⁷ David J.P. Barker,²¹ Daniel Barnes,⁶ Elaine M. Dennison,²⁴ Johan G. Eriksson,²⁵,²⁶,²⁷,²⁸ Per Eriksson,⁷ Elodie Eury,¹⁰,¹¹ Lasse Folkersten,²⁹ Caroline S. Fox,³,³⁰ Timothy M. Frayling,² Anuj Goel,³,¹⁸ Harvest F. Gu,³² Momoko Horikoshi,²⁵ Bo Isomaa,²⁷,³³ Jaspal Kooner,⁵⁰,⁵¹ Claes-Göran Östenson,²³ Lars Lind,⁴¹ Cyrus C. Cooper,²⁴ Manuel Serrano-Jemma C. Hopewell,⁴⁶ Danish Saleheen,⁴⁷,⁴⁸ John Danesh,⁴⁷ Teemu Kuulasmaa,³⁷ Johanna Kuusisto,³⁷ Ruth J.F. Loos,⁶ Jian’an Lu,⁶ Konstantinos Makrilakis,³⁸ Alisa K. Manning,² María Teresa Martínez-Larrad,²¹,²² Nariesi Narisu,³⁹ Maria Nastase Mannila,¹ John Öhrvik,¹ Clive Osmond,²⁴ Laura Pascoe,⁴⁰ Felicity Payne,¹⁴ Avan A. Sayer,²⁴ Bengt Sennblad,¹ Angela Silveira,¹ Alena Stančáková,³⁷ Kathy Stirtups,¹³ Amy J. Swift,³⁹ Ann-Christine Syvänni,²¹ Tiinanajia Tuomi,²⁷,⁴² Ferdinand M. van ’t Hooft,²¹ Mark Walker,⁴³ Michael N. Weedon,³¹ Weijia Xie,³¹ Björn Zethelius,⁴⁴ the DIAGRAM Consortium,* the GIANT Consortium,* the MuTHER Consortium,* the CARDIOGRAM Consortium,* the C4D Consortium,* Halit Ongen,⁵ National Institute for Health and Welfare, Helsinki, Finland; the ²Atherosclerosis Research Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden; the ³Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts; the ⁴MRC Epidemiology Unit, Institute of Metabolic Science, Addenbrooke’s Hospital, Cambridge, U.K.; the ⁵Department of Clinical Sciences, Diabetes and Endocrinology, University Hospital and Malmö University, Malmö, Sweden; the ⁶Oxford Centre for Diabetes Endocrinology and Metabolism, University of Oxford, Oxford, U.K.; the ⁷Wellcome Trust Genome Campus, Hinxton, U.K.; the ⁸Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; the ⁹BHF Cardiovascular Research Centre, University of Oxford, Oxford, U.K.; the ¹⁰CNRS UMR 8199, Institut Pasteur de Lille, Lille, France; the ¹¹CNRS UMR 859, INSERM UMR 859, Unité de Recherche, Institut National de Recherche en Santé et en Médecine, Paris, France; the ¹²MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton General Hospital, Southampton, U.K.; the ¹³National Institute for Health and Welfare, Helsinki, Finland; the ¹⁴Lund University, Malmö, Sweden; the ¹⁵Center for Human Genetic Research and Diabetes Research Center (Diabetes Unit), Massachusetts General Hospital, Boston, Massachusetts; the ¹⁶Division of Endocrinology, Diabetes, and Hypertension, Brigham and Women’s Hospital, Harvard Medical School, Boston, Massachusetts; the ¹⁷Center for Biomedical and Clinical Sciences, Peninsula Medical School, Torbay Hospital and University of Plymouth, Plymouth, U.K.; the ¹⁸Framingham Heart Study, Framingham, Massachusetts; the ¹⁹Department of Cardiovascular Medicine, University of Geneva, Geneva, Switzerland; the ²⁰National Institute for Health and Welfare, Helsinki, Finland; the ²¹Department of Cardiovascular Medicine, University of Oxford, Oxford, U.K.; the ²²Humana Research Institute, Inc., Seattle, WA, USA; the ²³Institut National de la Santé et de la Recherche Médicale, UMR 1188, Hôpital Saint-Louis, Paris, France; the ²⁴Center for Human Genetic Research and Diabetes Research Center (Diabetes Unit), Massachusetts General Hospital, Boston, Massachusetts; the ²⁵National Institute for Health and Welfare, Helsinki, Finland; the ²⁶Center for Statistics and Evaluation, Institute of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, Michigan; the ²⁷Folkhälsan Research Centre, Helsinki University Central Hospital and Research Program of Molecular Medicine, University of Helsinki, Helsinki, Finland; the ²⁸Center for Human Genetic Research and Diabetes Research Center (Diabetes Unit), Massachusetts General Hospital, Boston, Massachusetts; the ²⁹Helsinki University Central Hospital, Unit of General Practice, Helsinki, Finland; the ³⁰Folkhalsan Research Centre, Helsinki, Finland; the ³¹Center for Medical Research andassociate Health Studies, University of Helsinki, Helsinki, Finland; the ³²Department of Public Health and Caring Sciences, Karolinska Institutet, Stockholm, Sweden; the ³³DIABETES, VOL. 60, OCTOBER 2011 diabetes.diabetesjournals.org
OBJECTIVE—Proinsulin is a precursor of mature insulin and C-peptide. Higher circulating proinsulin levels are associated with impaired β-cell function, raised glucose levels, insulin resistance, and type 2 diabetes (T2D). Studies of the insulin processing pathway could provide new insights about T2D pathophysiology.

RESEARCH DESIGN AND METHODS—We have conducted a meta-analysis of genome-wide association tests of ~2.5 million genotyped or imputed single nucleotide polymorphisms (SNPs) and fasting proinsulin levels in 10,701 nondiabetic adults of European ancestry, with follow-up of 23 loci in up to 16,378 individuals, using additive genetic models adjusted for age, sex, fasting insulin, and study-specific covariates.

RESULTS—Nine SNPs at eight loci were associated with proinsulin levels (P < 5 × 10⁻⁸). Two loci (LARP6 and SGSM2) have not been previously related to metabolic traits, one (MADD) has been associated with fasting glucose, one (PCSKit) has been implicated in obesity, and four (TCFTL2, SLC30A8, VPS13C/C2CD4A/B, and ARAPI, formerly CENTD2) increase T2D risk. The proinsulin-raising allele of ARAPI was associated with a lower fasting glucose (P = 1.7 × 10⁻⁶), improved β-cell function (P = 1.1 × 10⁻⁷), and lower risk of T2D (odds ratio 0.88; P = 7.8 × 10⁻⁸). Notably, PCSKit encodes the protein prohormone convertase 1/3, the first enzyme in the insulin processing pathway. A genotype score composed of the nine proinsulin-raising alleles was not associated with coronary disease in two large case-control datasets.

CONCLUSIONS—We have identified nine genetic variants associated with fasting proinsulin. Our findings illuminate the biology underlying glucose homeostasis and T2D development in humans and argue against a direct role of proinsulin in coronary artery disease pathogenesis. Diabetes 60:2624–2634, 2011

Genome-wide association studies (GWAS) have uncovered dozens of common genetic variants associated with risk for type 2 diabetes (T2D; reviewed in [1]). Known associated variants in these loci account for only a small proportion of the heritable component of T2D (1), suggesting that additional loci await discovery. The Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) was created under the premise that genome-wide analysis of continuous diabetes-related traits could not only identify loci regulating variation in these glycomic traits, but also yield additional T2D susceptibility loci and insights into the underlying physiology of these loci (2–5). In addition, the genetic study of T2D endophenotypes may help clarify the pathophysiologic heterogeneity of this disease by elucidating the respective roles of β-cell function, insulin secretion, processing and sensitivity, and glucose metabolism (6).

Discovery of novel genetic determinants of insulin secretion and action has primarily focused on insulin levels (3,4,7,8). Proinsulin is the molecular precursor for insulin and has relatively low insulin-like activity, and its enzymatic conversion into mature insulin and C-peptide is a critical step in insulin production and secretion (Supplementary Fig. 1). Although hyperinsulinemia typically denotes insulin resistance, high proinsulin in relation to circulating levels of mature insulin can indicate β-cell stress as a result of insulin resistance, impaired β-cell function, and/or insulin processing and secretion abnormalities (9) (Supplementary Fig. 2). There is good evidence that higher proinsulin predicts future T2D (10) and coronary artery disease (CAD) (11–13), even after taking fasting glucose levels into account. Interestingly, some loci previously associated with fasting glucose levels (MADD) or risk of T2D (TCFTL2, SLC30A8, CDKAL1) are also associated with higher circulating proinsulin (6,14–17). Therefore, genome-wide analysis of proinsulin levels could reveal additional novel loci increasing susceptibility for T2D and perhaps CAD.

Thus, to identify novel loci influencing proinsulin processing and secretion and potentially increasing susceptibility for T2D, we performed a meta-analysis of ~2.5 million directly genotyped or imputed autosomal single nucleotide polymorphisms (SNPs) from four GWAS of fasting proinsulin levels (adjusted for concomitant fasting insulin) including 10,701 nondiabetic adult men and women of European descent. Follow-up of 23 lead SNPs from the most significant association signals in up to 16,378 additional individuals of European ancestry detected nine genome-wide significant associations with proinsulin levels, including two novel signals in or near LARP6 and SGSM2, and the known glycomic loci ARAPI, MADD (two independent signals), TCFTL2, VPS13C/C2CD4A/B, SLC30A8, and PCSKit. Here we describe these genetic associations, perform fine-mapping to identify potential causal variants, assess gene expression in human tissues, and define their impact on other glycomic quantitative traits and risk of both T2D and CAD.

RESEARCH DESIGN AND METHODS

Cohort/study description. Four cohorts contributed to the discovery meta-analysis through the contribution of phenotypic and GWAS data. These included the Framingham Heart Study (n = 5,759), Precocious Coronary Artery Disease (PROCARDIS) (n = 3,259), the Finnland study (n = 1,372), and the Diabetes Genetics Initiative (DGI) (n = 311), for a total of 10,701 participants.
Eleven cohorts contributed to the follow-up efforts; these included Metabolic Syndrome in Men (METSIM) \( (n = 5,122) \), Botnia Prevalence, Prediction and Prevention of diabetes (Botnia-PPP) \( (n = 2,280) \), Helsinki Birth Cohort Study (HBCS) \( (n = 1,640) \), the Ely study \( (n = 1,568) \), the Hertfordshire study \( (n = 1,016) \), Uppsala Longitudinal Study of Adult Men (ULSAM) \( (n = 939) \), Relationship between Insulin Sensitivity and Cardiovascular disease (RISC) \( (n = 914) \), Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS) \( (n = 912) \), Segovia \( (n = 911) \), the Greek Health Randomized Aging Study (GHRA5) \( (n = 668) \), and Stockholm Diabetes Prevention Program (SDPP) \( (n = 399) \), for a total of 16,378 participants \( (\text{with maximal sample for any one SNP of 15,898}) \). We excluded individuals with known diabetes, on antidiabetic treatment, or with fasting glucose \( \geq 7 \text{ mmol/L} \) \( (3) \); all participants were of European descent.

**Proinsulin and insulin measurements.** Proinsulin \( (\text{pmol/L}) \) was measured from fasting whole blood, plasma, or serum or a combination of these using enzyme-linked immunosorbsorbent or immunometric assays. Fasting insulin \( (\text{pmol/L}) \) was measured using either enzyme-linked immunosorbsorbent, immunofluorescent, or radioimmunometric assays \( (\text{Supplementary Table 1}) \).

**Genotyping.** Genome-wide commercial arrays \( (\text{Affymetrix 500K, MIPS 50K, and Illumina Human1M/610K}) \) were used by the four discovery cohorts as described in Supplementary Table 1. Imputation and quality control methods are described in the Supplementary Data.

**Statistical analyses.** We aimed to identify genetic variants associated with high proinsulin levels relative to an individual’s fasting insulin levels. This can be done by examining proinsulin-to-insulin ratios or by statistically adjusting proinsulin for fasting insulin. We chose the latter because the adjusted trait and the ratio was 0.95, and the quantile-quantile GWAS plots were comparable.

We used a linear regression model with natural log transformed fasting proinsulin as the dependent variable and genotypes as predictors, with adjustment for natural-log transformed fasting insulin values, sex, age, geographical covariates \( (\text{if applicable}) \), and age squared \( (\text{Framingham only}) \) to estimate the pooled regression estimates for additively coded SNPs using METAL \( (24) \). Sex interaction effects were evaluated with a function in the GWAMA software \( (25) \).

**Follow-up SNP selection and analysis.** We carried forward to stage 2 the most significant SNP from each of 21 independent loci that showed association with proinsulin in stage 1 analyses at \( P < 1 \times 10^{-5} \). Additionally, two SNPs near the \( P < 1 \times 10^{-5} \) threshold \( (\text{ASAP2 and a gene desert region}) \) were carried forward as a result of biological plausibility \( (\text{ASAP2 is involved in vesicular transport}) \) and/or consistency of direction of effect in all discovery stage 1 studies \( (\text{both loci}) \). We genotyped these 23 variants in 11 additional stage 2 studies totaling 16,378 nondiabetic participants of European ancestry \( (\text{Supplementary Table 1}; \text{genotyping assays and conditions are available upon request}) \). We meta-analyzed stage 1 and stage 2 results using inverse-variance weighted fixed effects meta-analysis methods, including up to 27,070 participants.

Additional analyses and expression and expression quantitative trait loci \( (\text{eQTL}) \) studies are described in the Supplementary Data.
RESULTS

Genome-wide association meta-analysis (stage 1). We conducted a two-stage association study in individuals of European descent (total $n = 27,079$, with $n = 10,701$ in the discovery stage). **Haplotype and phenotype information can be found in Supplementary Table 1, and the study design is outlined in Supplementary Fig. 3.** A total of 21 independent variants (including two SNPs identified during conditional analyses, see below) met our statistical threshold for follow-up ($P < 1 \times 10^{-8}$; Fig. 1). The clean dataset showed no systematic deviation from the null expectation, with the exception of the tail of the distribution (Fig. 1, inset).

Follow-up studies (stage 2) and global (stage 1 + stage 2) meta-analysis for 23 loci. We followed up 23 SNPs (the 21 mentioned above plus 2 others that approached our significance threshold and were selected as a result of biological plausibility; see RESEARCH DESIGN AND METHODS) in 11 cohorts totaling up to 16,378 nondiabetic individuals of European descent (Table 1 and Supplementary Table 2). Joint meta-analysis of discovery and follow-up cohorts ($n = 27,079$) revealed nine signals at eight loci reaching genome-wide significance ($P < 5 \times 10^{-8}$), of which two are novel (SGSM2, LARP6), five have previously been associated with glucose metabolism and/or T2D (TCF7L2, SLC3A2, MADD, VPSJ3/C2CD4A/B, and ARAP1), and one (PCSK1) has been previously implicated in obesity and associated with proinsulin levels, although not at genome-wide significance (Table 1 and Fig. 2). Adjusting for BMI, fasting glucose, or both did not attenuate these signals. Of note, when adjusting for fasting glucose or both fasting glucose and BMI (but not BMI alone), one other locus, SNX7, reached genome-wide significance ($P = 5.4 \times 10^{-9}$ and $1.5 \times 10^{-8}$, respectively).

Conditional analyses on the two strongest signals revealed that the MADD locus harbors two independent signals 19 kb apart (rs10501320 and rs10838687; $r^2 = 0.068$ in HapMap CEU), whereas a second independent signal near ARAP1 did not replicate (Fig. 2B, Table 1, and Supplementary Table 2). Among the nine replicated SNPs, individual loci explained between 0.2 and 1.4% of the variance in proinsulin in the discovery samples and up to 2.3% of the variance in the follow-up samples. Together, the nine genome-wide significant SNPs explained between 5.4 and 7.7% of the proinsulin variance in the discovery samples and 8.1% of the variance in the RISC cohort, one of the few follow-up cohorts with genotypes available for all nine SNPs.

Heterogeneity and sex-stratified analyses. We noted some degree of heterogeneity in our joint meta-analyses (Table 1). Part of the heterogeneity arose from the METSIM sample, which enrolled only men; exclusion of this cohort from our meta-analysis reduced the heterogeneity. We also stratified our analyses by sex and tested for a SNP × sex interaction (26). Our overall findings remained essentially unchanged after stratification, and heterogeneity was attenuated (e.g., $I^2 = 77.2%$, heterogeneity $P = 1.9 \times 10^{-7}$ for combined men and women, whereas $I^2 = 64.6%$, heterogeneity $P = 4.5 \times 10^{-4}$ [men] and $I^2 = 55.6%$, heterogeneity $P = 0.01$ [women] in stratified analyses). Furthermore, tests for interaction with sex among SNPs that reached our follow-up significance threshold revealed a locus (rs306549 in DDX31) where a genome-wide significant association was seen in women ($P = 2.0 \times 10^{-8}$; Supplementary Fig. 4A) but not men ($P = 0.17$; Supplementary Fig. 4B; sex interaction $P = 8.9 \times 10^{-5}$). Although removal

### Table 1: Loci with fasting proinsulin levels at statistical significance

<table>
<thead>
<tr>
<th>SNP</th>
<th>Nearest gene</th>
<th>CHR</th>
<th>Position</th>
<th>Allele</th>
<th>Discovery Chr</th>
<th>Freq.</th>
<th>Follow-up Chr</th>
<th>P-value</th>
<th>Combined Chr</th>
<th>Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs6235</td>
<td>MADD</td>
<td>11</td>
<td>47250375</td>
<td>A</td>
<td>0.69</td>
<td>4.2</td>
<td>0.0321</td>
<td>10.701</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs10838687</td>
<td>SLC3A2</td>
<td>11</td>
<td>47250375</td>
<td>G</td>
<td>0.65</td>
<td>5.3</td>
<td>0.0280</td>
<td>10.701</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs10501320</td>
<td>VPSJ3</td>
<td>11</td>
<td>47250375</td>
<td>A</td>
<td>0.64</td>
<td>5.3</td>
<td>0.0280</td>
<td>10.701</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs10838687</td>
<td>MADD</td>
<td>11</td>
<td>47250375</td>
<td>G</td>
<td>0.65</td>
<td>5.3</td>
<td>0.0280</td>
<td>10.701</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs6235</td>
<td>MADD</td>
<td>11</td>
<td>47250375</td>
<td>A</td>
<td>0.69</td>
<td>4.2</td>
<td>0.0321</td>
<td>10.701</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Abbreviations:** Freq., frequency; Combined P value, P-value for the combined analysis; Heterogeneity, heterogeneity in % (Q-test).
of the METSIM cohort improved the heterogeneity score and produced nominal significance for the association in men (P = 0.02), the effect size remained threefold stronger in women than in men (β-coefficient 0.0427 vs. 0.0165, respectively).

To provide further reassurance regarding any residual heterogeneity, we repeated our meta-analyses based on P values (rather than β-coefficients) and meta-analyzed the resulting z scores. Our findings were essentially unchanged, suggesting that heterogeneity in the β-estimates across cohorts has not produced spurious results.

**Exploration of proinsulin processing mechanisms.**

Proinsulin is initially cleaved to 32,33-split proinsulin and further to insulin and C-peptide before secretion (Supplementary Fig. 1); we were therefore interested in the effects of the nine top SNPs on these traits. The proinsulin-raising alleles of each SNP were consistently associated with higher 32,33-split proinsulin levels, with effect sizes following the rank order of proinsulin effect sizes. Nearly all associations reached nominal conventional levels of statistical significance in this smaller dataset of 4,103–6,343 individuals with measures of 32,33-split proinsulin levels (all P < 1.5 × 10^{-3}, with the exception of the conditional signal at MADD). The insulinogenic index (27), which measures dynamic insulin secretion during the first 30 min after an oral glucose load and was available in 14,956 subjects, showed nominal associations for four loci. Of these, the proinsulin-raising alleles were associated with a lower insulinogenic index at VPS13C/C2CD4A/B, TCF7L2, and SLC30A8 and higher at ARAP1 (Table 2).

We detected no nominal associations with fasting C-peptide (P > 0.05). Given the differences in hepatic clearance of insulin and C-peptide, we also performed sensitivity analyses to account for any possible impact this may have had on our results. We adjusted proinsulin levels for fasting C-peptide rather than fasting insulin in two cohorts (Ely and Botnia-PPP); comparison of β-estimates showed that the majority of loci had very similar effect sizes and the same rank order was preserved, arguing against noticeable discrepancies between the two adjustment schemes.

**Association with other glycemic traits.**

To clarify potential mechanisms, the top nine signals (ARAP1, two at MADD, PCSK1, TCF7L2, VPS13C/C2CD4A/B, SLC30A8, LARP6, and S6GSM2) were also examined in relation to other glucoregulatory traits (fasting and 2-h postload glucose and insulin, homeostasis model assessment estimates of β-cell function [HOMA-B] and insulin resistance [HOMA-IR] (28), glycated hemoglobin [A1C], T2D, and BMI [Table 3]). We investigated results available from MAGIC meta-analyses of GWAS of glycemic traits (3–5) and obtained T2D and BMI results in collaboration with the Diabetes Genetics Replication And Meta-analysis (DIAGRAM) (29) and Genomewide Investigation of Anthropometric measures (GIANT) (30) consortia, respectively. Nominal associations (P < 0.05) were found for fasting glucose (with the proinsulin-raising allele increasing fasting glucose levels at MADD, SLC30A8, TCF7L2, and VPS13C/C2CD4A/B and decreasing fasting glucose levels at ARAP1 and PCSK1), fasting insulin (increased levels at ARAP1, LARP6, and S6GSM2 and decreased levels at TCF7L2), HOMA-B (decreased at MADD, SLC30A8, VPS13C/C2CD4A/B, and TCF7L2 and increased at PCSK1, ARAP1, and LARP6), insulin resistance as measured by HOMA-IR (increased at LARP6 and S6GSM2 and decreased at TCF7L2), and 2-h postload glucose (decreased at SLC30A8 and VPS13C/C2CD4A/B and increased at ARAP1 and TCF7L2).

We detected no significant associations for 2-h postload insulin or insulin sensitivity as estimated by the Matsuda index (31) (Table 3).

Associations with T2D were confirmed for four known T2D loci (SLC30A8, ARAP1, VPS13C/C2CD4A/B, and TCF7L2; Table 3). Counterintuitively, the proinsulin-raising allele of ARAP1 (formerly known as CENDT2 and reported as such in DIAGRAM+) (29) was associated with a lower fasting glucose (0.019 mg/dL per A allele; P = 1.7 × 10^{-5}), lower A1C (0.023%; P = 0.02), and a lower risk of T2D (odds ratio [OR] 0.88; P = 7.8 × 10^{-6}; Table 3). The two novel loci (LARP6 and S6GSM2) did not show significant associations with T2D (OR [95% CI]: 1.01 [0.95–1.07] and 1.01 [0.96–1.08], respectively), indicating that if they increase T2D risk they do so to an extent confined within the bounds of narrow 95% CI.

**Fine-mapping, copy number variants, and tissue expression.**

We used MACH (32) or IMPUTE (19) applied to the 1000 Genomes CEU reference panel (www.1000genomes.org) to carry out imputation of ~8 million autosomal SNPs with minor allele frequency >1%. Analysis of 1000 Genomes-imputed data in the four discovery cohorts indicates that although there are low-frequency (1–5%) genetic variants that influence levels of circulating proinsulin, these are found in the same loci that contain common proinsulin-influencing variants, and none of them yield substantially stronger signals than the index SNP at each locus (Supplementary Fig. 5).

Using current databases of copy number variants (33) and the SNAP software (http://www.broadinstitute.org/mpg/snap/index.php; CEU, HapMap release 22), we checked whether any of the proinsulin-associated SNPs were within 500 kb and in linkage disequilibrium (LD) with any of the SNPs known to tag copy number variants in the human genome. No copy number variant tag SNPs with r^2 > 0.3 were found within 500 kb of our lead SNPs.

To guide identification of the gene responsible for each association signal, we also examined the gene expression profile of selected genes in each associated region across a range of human tissues, including islets and fluorescence-activated cell (FAC)-sorted β-cells (Fig. 3A–F and Supplementary Fig. 6). We defined 1-Mb intervals around the lead SNP at each locus and prioritized biologically plausible genes as gleaned from the literature (see Box in Supplementary Data). We were able to demonstrate β-cell expression of most genes examined (Fig. 3F). However, at the LARP6 locus, CT62 is expressed exclusively in testis, likely excluding it as a relevant gene in this context. At the ARAP1 locus, STARD10 is expressed more strongly in pancreatic and islet tissue than any other tissue type; similarly, at the VPS13C locus both C2CD4A and C2CD4B demonstrate higher expression in pancreas and islets than all other tissue types.

We also studied the expression of the transcript for the gene closest to the index SNP at each of the nine replicated loci in human islets isolated from 55 nondiabetic and 9 diabetic individuals. Of the nine loci, PCSK1 (P = 0.02) and MADD (P = 0.07) demonstrated 35–45% lower expression in subjects with T2D compared with control subjects.

**Functional explanation.**

We evaluated whether any of the associated SNPs was in strong LD with a potentially causal variant. We used SNPper (34) to classify all SNPs within 500 kb and in linkage disequilibrium (LD) with any of the proinsulin-associated SNPs with minor allele frequency <5% (DIAGRAM+). However, at the lead SNP of each locus, using current databases of copy number variants (33) and the SNAP software (http://www.broadinstitute.org/mpg/snap/index.php; CEU, HapMap release 22), we checked whether any of the proinsulin-associated SNPs were within 500 kb and in linkage disequilibrium (LD) with any of the SNPs known to tag copy number variants in the human genome. No copy number variant tag SNPs with r^2 > 0.3 were found within 500 kb of our lead SNPs.

To guide identification of the gene responsible for each association signal, we also examined the gene expression profile of selected genes in each associated region across a range of human tissues, including islets and fluorescence-activated cell (FAC)-sorted β-cells (Fig. 3A–F and Supplementary Fig. 6). We defined 1-Mb intervals around the lead SNP at each locus and prioritized biologically plausible genes as gleaned from the literature (see Box in Supplementary Data). We were able to demonstrate β-cell expression of most genes examined (Fig. 3F). However, at the LARP6 locus, CT62 is expressed exclusively in testis, likely excluding it as a relevant gene in this context. At the ARAP1 locus, STARD10 is expressed more strongly in pancreatic and islet tissue than any other tissue type; similarly, at the VPS13C locus both C2CD4A and C2CD4B demonstrate higher expression in pancreas and islets than all other tissue types.

We also studied the expression of the transcript for the gene closest to the index SNP at each of the nine replicated loci in human islets isolated from 55 nondiabetic and 9 diabetic individuals. Of the nine loci, PCSK1 (P = 0.02) and MADD (P = 0.07) demonstrated 35–45% lower expression in subjects with T2D compared with control subjects.

**Functional explanation.**

We evaluated whether any of the associated SNPs was in strong LD with a potentially causal variant. We used SNPper (34) to classify all SNPs within 500 kb and in linkage disequilibrium (LD) with any of the proinsulin-associated SNPs with minor allele frequency <5% (DIAGRAM+). However, at the lead SNP of each locus, using current databases of copy number variants (33) and the SNAP software (http://www.broadinstitute.org/mpg/snap/index.php; CEU, HapMap release 22), we checked whether any of the proinsulin-associated SNPs were within 500 kb and in linkage disequilibrium (LD) with any of the SNPs known to tag copy number variants in the human genome. No copy number variant tag SNPs with r^2 > 0.3 were found within 500 kb of our lead SNPs.
FIG. 2. Regional plots of eight genomic regions containing novel genome-wide significant associations. For each region, directly genotyped and imputed SNPs are plotted with their meta-analysis $P$ values (as $-\log_{10}$ values) as a function of genomic position (NCBI Build 36). In each panel, the stage 1 discovery SNP taken forward to stage 2 follow-up is represented by a purple diamond (with global meta-analysis $P$ value), with its stage 1 discovery $P$ value denoted by a red diamond with bolded borders. Estimated recombination rates (taken from HapMap) are plotted to reflect the local LD structure around the associated SNPs and their correlated proxies (according to a white to red scale from $r^2 = 0$ to 1, based on pairwise $r^2$ values from HapMap CEU). Gene annotations were taken from the University of California Santa Cruz genome browser. A: ARAP1 region; B: MADD region; C: PCSK1 region; D: TCF7L2 region; E: VPS13C/C2CD4A/B region; F: SLC30A8 region; G: LARP6 region; H: SGSM2 region.
### Table 2

**Association of proinsulin loci with insulin-processing traits**

<table>
<thead>
<tr>
<th>SNP</th>
<th>Nearest gene</th>
<th>Alleles</th>
<th>proinsulin-raising/other</th>
<th>C-allele (β, P)</th>
<th>Insulinogenic index (β, P)</th>
<th>Fasting insulin (β, P)</th>
<th>HOMA-B (β, P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs11603334</td>
<td>ARAP1</td>
<td>A/G</td>
<td>0.146</td>
<td>1.2 × 10⁻⁵</td>
<td>0.012</td>
<td>0.62</td>
<td>0.02</td>
</tr>
<tr>
<td>rs10503290</td>
<td>PFKL2</td>
<td>C/T</td>
<td>0.124</td>
<td>1.4 × 10⁻⁵</td>
<td>0.001</td>
<td>0.78</td>
<td>0.009</td>
</tr>
<tr>
<td>rs7003316</td>
<td>VPS13C/C2CD4A/B</td>
<td>T/C</td>
<td>0.042</td>
<td>4.1 × 10⁻⁶</td>
<td>0.005</td>
<td>0.67</td>
<td>0.002</td>
</tr>
<tr>
<td>rs110858687</td>
<td>VPS13C/C2CD4A/B</td>
<td>T/C</td>
<td>0.006</td>
<td>1.4 × 10⁻⁵</td>
<td>0.006</td>
<td>0.22</td>
<td>0.04</td>
</tr>
<tr>
<td>rs1549018</td>
<td>LARP6</td>
<td>T/C</td>
<td>0.005</td>
<td>1.4 × 10⁻⁵</td>
<td>0.007</td>
<td>0.09</td>
<td>0.002</td>
</tr>
<tr>
<td>rs179456</td>
<td>SGSM2</td>
<td>T/C</td>
<td>0.001</td>
<td>0.74</td>
<td>0.003</td>
<td>0.13</td>
<td>0.008</td>
</tr>
</tbody>
</table>

**β-Coefficients are adjusted for age, sex, and study-specific covariates (if applicable).**

- Coefficients are adjusted for age, sex, and study-specific covariates (if applicable).

---

**DISCUSSION**

We report the first meta-analysis of genome-wide association datasets for circulating fasting proinsulin. We adjusted proinsulin for fasting insulin levels, aiming to capture an increase in proinsulin relative to the nonspecific activation of the insulin processing pathway induced by generalized insulin resistance (Supplementary Fig. 2). Loci that simply influence insulin resistance are typically sought by a GWAS for fasting insulin or more sophisticated measures of insulin sensitivity (3,4,6). Thus, we hoped to identify loci that indicate the inability of the β-cell to process proinsulin adequately in response to metabolic demands.

We have identified nine signals at eight loci associated with higher proinsulin levels (see Box in Supplementary Data). Two of these loci (LARP6 and SGSM2) have not been previously related to metabolic traits. A 10th signal emerged after sex-stratified analyses; an explanation for the female-specific genome-wide significant association at DDX31 requires fine-mapping to identify the causal gene. Although the function of the DDX31 gene product is unknown, other members of the DEAD-box protein family have been implicated in sex-specific processes such as spermatogenesis (39). We have also replicated at the genome-wide level previously reported nominal associations of MADD, TCF7L2, VPS13C/C2CD4A/B, SLC30A8, and PCSK1 with proinsulin (6,14-17,40). The knowledge that TCF7L2, VPS13C/C2CD4A/B, and ARAP1 are established T2D loci provides reassurance that a quest for genetic determinants of proinsulin can serve to identify disease-associated signals. Interestingly, the proinsulin-raising

---

**PROCEDURE**

1. **Genome-Wide Association Study**: Performed a GWAS on circulating fasting proinsulin levels in 20,000 case subjects and 60,000 control subjects, and in 38,335 additional subjects.
2. **Replication Meta-Analysis**: Replicated the association of the top signals in the primary analysis in an independent dataset of 15,420 CAD case subjects and 15,062 control subjects.
3. **Functional Analysis**: Conducted cis-eQTL analysis to identify potential regulatory loci.
4. **Pathway Analysis**: Investigated the biological pathways associated with proinsulin levels.

---

**RESULTS**

- **Top Signals**: Identified 10 signals associated with proinsulin levels.
- **Replication**: Successfully replicated 9 of the top signals at genome-wide significance.
- **Functional Insights**: Observed significant cis-eQTL associations for the lead SNPs at 8 loci.

---

**DISCUSSION**

1. **Implications for Metabolic Disease**: The findings suggest novel pathways in the regulation of proinsulin levels that may be relevant to metabolic diseases such as T2D.
2. **Implications for GWAS**: The results highlight the importance of considering fasting proinsulin levels as a genetic trait in GWAS.

---

**CONCLUSION**

The study identifies novel genetic determinants of proinsulin levels, which may provide insights into the molecular mechanisms underlying metabolic disease.
alleles at TCF7L2, SLC30A8, and VPS13C/C2CD4A/B cause impairment of β-cell function, as estimated by HOMA-B and the insulinogenic index. By raising proinsulin but lowering insulin secretion, these loci point to defects in the insulin processing and secretion pathway, distal to the first enzymatic step. Such a hypothesis is consistent with postulated modes of action for TCF7L2 (41) and SLC30A8 (42); VPS13C, by influencing protein trafficking across membrane compartments, could also affect the same process. Further fine-mapping and functional experiments will be required to establish the precise mechanism at this locus.

**ARAP1**, which harbors the strongest proinsulin association, provides an intriguing counterpoint. Under its previous designation of **CENTD2** it was recently associated with T2D (29); however, the T2D-associated allele is associated with lower proinsulin levels, as well as lower β-cell function (HOMA-B and insulinogenic index). This suggests that the genetic defect that gives rise to T2D at this locus causes a generalized downregulation of insulin secretion (e.g., through a reduction in β-cell mass/function or very early defects in insulin processing) and stands in contrast with TCF7L2, SLC30A8, and VPS13C/C2CD4A/B. A corollary of the divergent effect of these loci on T2D is that both disproportionate elevations and reductions in proinsulin can indicate β-cell dysfunction. Of the genes that lie within 1 Mb of the **ARAP1** association signal, we have demonstrated islet expression in the four strong biological candidates we examined (**ARAP1, INPPL1, STARD10**, and **RAB6A**); however, expression of **STARD10** was much higher in pancreas than in any other human tissue, and of all genes tested at the **ARAP1** locus **STARD10** was expressed most strongly in islets, indicating that the role of its protein product in the transfer of phospholipids to membranes may be particularly relevant to this cell type.

**LARP6** is a ribonucleoprotein identified in the current study as a novel locus associated with increased fasting proinsulin levels. It is involved in the regulation of translation and subcellular localization of collagen I, in a manner dependent upon both the RNA-binding and La domains (43). The associated SNP rs1549318 is located within a region of high LD, which spans the gene and includes a number of SNPs within the RNA-binding domain. Although the link between **LARP6** and proinsulin levels is not clear, it is nominally associated with fasting insulin and HOMA-IR, but not T2D. It may therefore represent a marker of insulin resistance and perhaps other related common dysmetabolic conditions.

In previous publications we have reported the association of C2CD4B with fasting glucose (3) and that of the nearby locus VPS13C with 2-h glucose (4); C2CD4B is also associated with T2D in Japanese (44), with supportive evidence found in Europeans (3,44). Here we show that the same genomic region is associated with fasting proinsulin. The strongest association with proinsulin reported here (rs4502156) and those associated with fasting glucose and 2-h glucose may represent independent signals, since they are all in relatively weak LD in HapMap CEU Europeans: rs4502156 versus rs11071657 (best fasting glucose signal), \( r^2 = 0.306 \); rs4502156 vs. rs17271305 (best 2-h glucose signal), \( r^2 = 0.450 \); and rs11071657 versus rs17271305, \( r^2 = 0.287 \). On the other hand, in Europeans our proinsulin-associated SNP is in strong LD (\( r^2 = 0.967 \)) with the T2D-associated SNP reported by Yamauchi et al. (44). Although four strong biological candidates (C2CD4A, C2CD4B, VPS13C, and RORA, a gene that encodes a

### Table 3

<table>
<thead>
<tr>
<th>Nearest gene</th>
<th>SNP</th>
<th>Fasting glucose (n = 44,480)</th>
<th>HOMA-IR</th>
<th>Matsuda index</th>
<th>2-h glucose (n = 15,252)</th>
<th>BMI (n = 123,260)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TCF7L2</strong></td>
<td>rs4502156</td>
<td>0.17 ( \times 10^{-3} )</td>
<td>0.168</td>
<td>0.142</td>
<td>0.114</td>
<td>0.026</td>
</tr>
<tr>
<td><strong>SLC30A8</strong></td>
<td>rs11071657</td>
<td>0.15 ( \times 10^{-3} )</td>
<td>0.150</td>
<td>0.132</td>
<td>0.104</td>
<td>0.024</td>
</tr>
<tr>
<td><strong>VPS13C</strong></td>
<td>rs11071657</td>
<td>0.13 ( \times 10^{-3} )</td>
<td>0.143</td>
<td>0.124</td>
<td>0.106</td>
<td>0.023</td>
</tr>
<tr>
<td><strong>MADD</strong></td>
<td>rs10501320</td>
<td>0.12 ( \times 10^{-3} )</td>
<td>0.134</td>
<td>0.116</td>
<td>0.100</td>
<td>0.022</td>
</tr>
</tbody>
</table>

## Rationale for the Study

The study aimed to identify genetic variants associated with T2D, focusing on proinsulin levels. The researchers examined a variety of genes and their associations with T2D and related metabolic markers. They observed that TCF7L2, SLC30A8, and VPS13C/C2CD4A/B are associated with T2D, while LARP6 and ARAP1 show distinct effects on proinsulin and insulin secretion, respectively. The study also highlighted the importance of considering genetic and environmental factors in understanding the complex interplay between insulin secretion and T2D.
member of the NR1 subfamily of nuclear hormone receptors) are expressed in FAC-sorted β-cells, the relative expression of the first two is much higher in islets than in other human tissues, again suggesting that these two genes, encoding nuclear factors that are upregulated in response to inflammation, may be particularly relevant to endocrine pancreatic function.

The genome-wide association of a missense variant in PCSK1 with fasting proinsulin also serves as a positive control. PCSK1 encodes the protein prohormone convertase...
L3 (PC1), which is the first enzyme in the proinsulin processing pathway, where it cleaves proinsulin to 32,33-split proinsulin (Supplementary Fig. 1). A related enzyme, PC2, acts on 32,33-split proinsulin in the second processing step. People deficient in PC1 become obese at an early age and exhibit pituitary hypofunction because of the lack of several mature peptide hormones (45), whereas PC2-null mice demonstrate increased levels of 32,33-split proinsulin (46). The rs6235 SNP reported here results in the substitution of a serine residue for threonine at position 690 of the molecule; the minor allele (Thr) is associated with higher proinsulin levels. A nominal association of the same allele with higher proinsulin levels has recently been reported (40); its association with higher BMI is only nominal here, but confirms a previous report (47). This specific amino acid change has been shown not to affect enzyme catalysis or maturation of the protein in vitro (47), but the COOH terminus of the protein (where S690T is located, adjacent to a conserved proline residue) is known to direct the correct subcellular targeting of the protein as well as stabilizing and partially inhibiting PC1. Although one might expect lower levels of the reaction product (32,33-split proinsulin) in carriers of the risk allele, the potential diversion of the substrate down its alternate path (giving rise to 65,66-split proinsulin, whose assay typically has 60% cross-reactivity with 32,33-split proinsulin) requires further study. Alternatively, if changes in the activity of PC1 also affect that of PC2 (for instance, by competing for inhibitory peptides) one might see reductions in the catalytic function of both enzymes and accumulation of both proinsulin and 32,33-split proinsulin.

Because of the reported relationship between proinsulin levels and coronary events (11–13), the identification of genetic determinants of proinsulin levels might help shed light on whether hyperproinsulinemia is a mediator of CAD or a byproduct of a shared etiologic mechanism. If hyperproinsulinemia is causally associated with an increased risk of CAD, one might expect that SNPs that specifically and selectively raise proinsulin levels should increase the risk of CAD given an adequately powered study. We have not observed such an effect for a genotype score constructed with the genome-wide significant proinsulin association signals. Assuming conservative approximations of the reported effect sizes of proinsulin on CAD (OR ~1.5 per 1-SD increase in proinsulin) (12,13), and of the nine SNPs reported here on circulating proinsulin (5%), a CAD cohort like CARDioGRAM has 99% power to detect a significant effect of proinsulin SNPs on CAD. The absence of statistical significance argues against a direct etiologic role of proinsulin on CAD.

In summary, we have identified nine loci that associate with fasting proinsulin levels. Several of these loci increase risk of T2D; interestingly, both proinsulin-raising and lowering alleles can lead to T2D through decreases in insulin secretion, indicating defects distal or proximal to the first enzymatic step in proinsulin conversion, respectively. Other genetic determinants of proinsulin levels do not necessarily lead to higher T2D risk, suggesting that it is not a mere elevation in proinsulin, but rather the specific impairment in proinsulin processing and the reaction of the β-cell to this defect that determine whether ultimately β-cell insufficiency will cause pathological hyperglycemia. The direct elevation of fasting proinsulin out of proportion to fasting insulin does not seem to increase risk of CAD.

ACKNOWLEDGMENTS

Please see the Supplementary Data.

REFERENCES


17. Stolerman ES, Manning AK, McAteer JB, et al. TCF7L2 variants are associated with increased proinsulin/insulin ratios but not obesity traits in the Framingham Heart Study. Diabetologia 2009;52:614–620


20. StataCorp. Stata Statistical Software: Release 10. College Station, TX, StataCorp LP, 2007


GENOME-WIDE ASSOCIATION STUDY OF PROINSULIN

32. Li Y, Abecasis GR. Mach 1.0: rapid haplotype reconstruction and missing genotype inference (Abstract). Am J Hum Genet 2006;S79:2290
34. Riva A, Kohane IS. A SNP-centric database for the investigation of the human genome. BMC Bioinformatics 2004;5:33