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. DOI: 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,2,3 Inga Prokopenko,4,5 Adam Barker,6 Emma Ahlvqvist,7 Denis Rybin8 John R. Petrie,9 Mary E. Travers,4 Nabila Bouatia-Naji,10,11 Antigone S. Dimas5,12 Alexandra Nica,12,13 Eleanor Wheeler,14 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,1 Elodie Eury,10,11 Lasse Folkesen,29 Caroline S. Fox,3,30 Timothy M. Frayling,21 Anuj Goel,3,18 Harvest F. Gu,32 Momoko Horikoshi,4,5 Bo Isomaa,27,33 Anne U. Jackson,24 Karen A. Jensen,24 Eero Kajantie,25,35 Julie Kerr-Conte10,36 Teemu Kuulasmaa,37 Johanna Kuusiisto,37 Ruth J.F. Loos,6 Jian’an Luan,6 Konstantinos Makrilakis,38 Alisa K. Manning,2 María Teresa Martínez-Larrad,21,22 Narisu Narisu,39 Maria Nastase Mannila,1 John Öhrvik,1 Clive Osmond24 Laura Pascoe,40 Felicity Payne14 Avan A. Sayer,24 Bengt Sennblad,1 Angela Silvera,1 Alena Stancáková,37 Kathy Stitturp,13 Amy J. Swift,39 Ann-Christine Syvänen,41 Tiinamaija Tuomi,27,42 Ferdinand M. van ’t Hooft,1 Mark Walker,43 Michael N. Weedon,41 WeiJia Xie,31 Björn Zethelius,4,44 the DIAGRAM Consortium,* the GIANT Consortium,* the MuTHER Consortium,*, the CARDIOGRAM Consortium,* the C4D Consortium,* Halit Ongen,5,18,45 Anders Mälarstig,1

From the 1Atherosclerosis Research Unit, Department of Medicine Solna, Karolinska Institutet, Karolinska University Hospital Solna, Stockholm, Sweden; the 2Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts; the 3National Heart, Lung, and Blood Institute, Bethesda, Maryland; the 4Oxford Centre for Diabetes Endocrinology and Metabolism, University of Oxford, Oxford, U.K.; the 5Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, U.K.; the 6MRC Epidemiology Unit, Institute of Metabolic Science, Addenbrooke’s Hospital, Cambridge, U.K.; the 7Department of Clinical Sciences, Diabetes and Endocrinology, University Hospital and Malmo, Lund University, Malmo, Sweden; the 8Boston University Data Coordinating Center, Boston, Massachusetts; the 9BHF Cardiovascular Research Centre, University of Glasgow, Glasgow, U.K.; the 10Université Lille-Nord de France, Lille, France; the 11CNRS UMR 8189, Institut Pasteur de Lille, Lille, France; the 12Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland; the 13Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, U.K.; the 14Metabolic Disease Group, Wellcome Trust Sanger Institute, Hinxton, Cambridge, U.K.; the 15Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts; the 16Center for Human Genetic Research and Diabetes Research Center (Diabetes Unit), Massachusetts General Hospital, Boston, Massachusetts; the 17Department of Dietetics-Nutrition, Harokopio University, Athens, Greece; the 18Department of Cardiovascular Medicine, University of Oxford, Oxford, U.K.; the 19Department of Medicine, University of Verona, Verona, Italy; the 20Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; the 21CIBER de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), Madrid, Spain; the 22Fundación Investigación Biomédica del Hospital Clínico San Carlos, Madrid, Spain; the 23Heart Research Center, Oregon Health and Science University, Portland, Oregon; the 24MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton General Hospital, Southampton, U.K.; the 25National Institute for Health and Welfare, Helsinki, Finland; the 26Helsinki University Central Hospital, Unit of General Practice, Helsinki, Finland; the 27Folkhalsan Research Centre, Helsinki, Finland; the 28Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland; the 29Experimental Cardiovascular Research Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden; the 30Division of Endocrinology, Diabetes, and Hypertension, Brigham and Women’s Hospital, Harvard Medical School, Boston, Massachusetts; the 31Institute of Biomedical and Clinical Sciences, Peninsula Medical School, the 32Endocrinology and Diabetes Unit, Brigham and Women’s Hospital, Harvard Medical School, Boston, Massachusetts; the 33Malmska Municipal Health Care Center and Hospital, Jakobstad, Finland; the 34Center for Human Genetic Research and Diabetes Research Center, University of Oregon Health Science Center, Portland, Oregon; the 35Center for Statistical Genetics, Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, Michigan; the 36Hospital for Children and Adolescents, Helsinki University Central Hospital and University of Helsinki, Helsinki, Finland; the 37INSERM UMR 859, Lille, France; the 38Center of Medicine, University of Kuopio and Kuopio University Hospital, Kuopio, Finland; the 39National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland; the 40Institute of Cell and Molecular Biosciences, Newcastle University, Newcastle, U.K.; the 41Department of Medical Sciences, Molecular Medicine, Science for Life Laboratory, Uppsala University, Uppsala, Sweden; the 42Department of Medicine, Helsinki University Central Hospital, and Research Program of Molecular Medicine, University of Helsinki, Helsinki, Finland; the 43Institute of Cellular Medicine, Newcastle University, Newcastle, U.K.; the 44Department of Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden; the 45Department of Cardiovascular Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford, U.K.; the 46Clinical Trial Service Unit, University of Oxford, Oxford, U.K.; the 47Department of Public Health and Primary Care, University of Cambridge, Cambridge, U.K.; the 48Center for Non-Communicable Diseases Pakistan, Karachi,
The proinsulin-raising allele of 

RESERCH 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 nonobese diabetic 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 (PCSK1) has been implicated in obesity, and four (TCFTL2, SLCO3A8, VPS13C/C2CD4A/B, and ARAPI, formerly CDEND2) increase T2D risk. The proinsulin-raising allele of ARAPI was associated with a lower fasting glucose (P = 1.7 × 10^-8), improved β-cell function (P = 1.1 × 10^-9), and lower risk of T2D (odds ratio 0.88; P = 7.8 × 10^-9). Notably, PCSK1 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 glycemic 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, SLCO3A8, 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 nonobese diabetic 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 glycemic loci ARAPI, MADD (two independent signals), TCFTL2, VPS13C/C2CD4A/B, SLCO3A8, and PCSK1. 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 glycemic 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 Fenland 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 \((n = 1,508)\), 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 (GHRAS) \((n = 668)\), and Stockholm Diabetes Prevention Program (SDPP) \((n = 399)\), for a total of 16,378 participants (with maximal sample for any one SNP of 15,898). We excluded individuals with known diabetes, on antidiabetic treatment, or with fasting glucose \(\geq 7\) 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 immunosorbent or immunometric assays. Fasting insulin \((\text{pmol/L})\) was measured using either enzyme-linked immunosorbent, immunofluorescent, or radioimmunometric assays (Supplementary Table 1).

**Genotyping.** Genome-wide commercial arrays (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 has comparable predictive value \((18)\) and displayed better statistical performance in pilot studies and adequate heritability in the Framingham Heart Study, one of the larger cohorts examined here \(\left(h^2 = 0.36\right.)\) for the proinsulin-to-insulin ratio. In Framingham, correlation between 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 (if applicable), and age squared (Framingham only) to evaluate the association under an additive genetic model. Association analysis was performed by individual studies using SNPTST \((19)\), STATA \((20)\), PLINK \((21)\), or LMEKIN (R kinship package) software \((22)\). Genome-wide association inflation coefficients were estimated for each discovery cohort using the genomic control \((GC)\) method \((23)\) and applied subsequently to each individual SNP association test statistics to correct for cryptic relatedness. The \(\lambda\) GC value for the final meta-analysis of proinsulin adjusted for fasting insulin was 1.01. The inverse-variance fixed effects meta-analysis method was used to evaluate 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 \((\text{ASAP2})\) 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 (Supplementary Table 1; 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.

![Manhattan plot of the association \(P\) values for fasting proinsulin adjusted for fasting insulin. 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). The SNPs that achieved genome-wide significance \((P < 5 \times 10^{-8})\) on follow-up are shown in red. Insert: Quantile-quantile (Q-Q) plot for fasting proinsulin adjusted for fasting insulin. The expected null distribution is plotted along the diagonal, the entire distribution of observed \(P\) values is plotted in blue, and a distribution that excludes the nine novel findings is plotted in red.](image)
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). Cohort 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, SLC30A8, MADD, VPS13C/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, all 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 5.1% 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>
<thead>
<tr>
<th>SNP</th>
<th>Nearest gene (CHR)</th>
<th>Pos (other) Freq.</th>
<th>Discovery P value</th>
<th>Replication β (SE)</th>
<th>Replication (a)</th>
<th>BMI adjusted (a)</th>
<th>BMI + FG adjusted (a)</th>
<th>BMI + FG adjusted (a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs295167</td>
<td>TCF7L2</td>
<td>1 90943321 G/A</td>
<td>0.64</td>
<td>5.3</td>
<td>5.8</td>
<td>5.8</td>
<td>5.8</td>
<td>5.8</td>
</tr>
<tr>
<td>rs4790333</td>
<td>SLC30A8</td>
<td>10 16793869 T/C</td>
<td>0.45</td>
<td>2.2</td>
<td>2.2</td>
<td>2.2</td>
<td>2.2</td>
<td>2.2</td>
</tr>
<tr>
<td>rs9727115</td>
<td>SGSM2</td>
<td>21 32855427 A/G</td>
<td>0.69</td>
<td>4.2</td>
<td>5.9</td>
<td>5.9</td>
<td>5.9</td>
<td>5.9</td>
</tr>
<tr>
<td>rs11603334</td>
<td>VPS13C/C2CD4A/B</td>
<td>10 1052.7 (0.02)</td>
<td>0.0253 (0.0050)</td>
<td>0.0240 (0.0049)</td>
<td>0.0246 (0.0049)</td>
<td>6.9</td>
<td>6.9</td>
<td>6.9</td>
</tr>
<tr>
<td>rs10501320</td>
<td>MADD</td>
<td>8 118254914 A/G</td>
<td>0.69</td>
<td>4.2</td>
<td>5.9</td>
<td>5.9</td>
<td>5.9</td>
<td>5.9</td>
</tr>
<tr>
<td>rs10838687</td>
<td>VPS13C/C2CD4A/B</td>
<td>8 11286084 T/C</td>
<td>0.45</td>
<td>2.2</td>
<td>2.2</td>
<td>2.2</td>
<td>2.2</td>
<td>2.2</td>
</tr>
<tr>
<td>rs306549</td>
<td>SGSM2</td>
<td>15 60170447 T/C</td>
<td>0.58</td>
<td>3.8</td>
<td>5.9</td>
<td>5.9</td>
<td>5.9</td>
<td>5.9</td>
</tr>
<tr>
<td>rs6235</td>
<td>LARP6</td>
<td>15 68896201 T/C</td>
<td>0.61</td>
<td>8.0</td>
<td>5.9</td>
<td>5.9</td>
<td>5.9</td>
<td>5.9</td>
</tr>
</tbody>
</table>
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 ($\beta$-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 $\beta$-coefficients) and meta-analyzed the resulting $z$ scores. Our findings were essentially unchanged, suggesting that heterogeneity in the $\beta$-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 $P < 1.5 \times 10^{-3}$, 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 $\beta$-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 SGS3M2) were also examined in relation to other glucometabolic traits (fasting and 2-h postload glucose and insulin, homeostasis model assessment estimates of $\beta$-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 SGS3M2 and decreased levels at TCF7L2), HOMA-B (decreased at MADD, SLC30A8, VPS13C/C2CD4A/B, and TCF7L2 and decreased at PCSK1, ARAP1, and LARP6), insulin resistance as measured by HOMA-IR (increased at LARP6 and SGS3M2 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; $P = 1.7 \times 10^{-6}$), lower A1C (0.023%; $P = 0.02$), and a lower risk of T2D (odds ratio [OR] 0.88; $P = 7.8 \times 10^{-6}$, Table 3). The two novel loci (LARP6 and SGS3M2) 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 $\beta$-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 $\beta$-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. We evaluated whether any of the associated SNPs was in strong LD with a potentially causal variant. We used SNPer (34) to classify all SNPs in strong LD with the lead SNP ($r^2 \geq 0.8$) with a 1-Mb region. We found that PCSK1 rs6235 codes for a nonsynonymous variant (S690T), which is in perfect LD with rs6234, another missense variant (Q665E); both were
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>Alles proinsulin-raising/other</th>
<th>32,33-split proinsulin (n = 4,103–6,433)</th>
<th>Fasting insulin (n = 36,775–38,335)</th>
<th>HOMA-B (n = 35,046–36,606)</th>
<th>Insulinogenic index (n = 10,992–14,956)</th>
<th>C-peptide (n = 1,537–2,956)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>β (P value)</td>
<td>β (P value)</td>
<td>β (P value)</td>
<td>β (P value)</td>
<td>β (P value)</td>
</tr>
<tr>
<td>rs11603334</td>
<td>ARAP1</td>
<td>A/G</td>
<td>0.146 (1.2 × 10^{-25})</td>
<td>0.012 (0.02)</td>
<td>0.016 (1.4 × 10^{-4})</td>
<td>0.049 (1.1 × 10^{-5})</td>
<td>0.017 (0.22)</td>
</tr>
<tr>
<td>rs10503120</td>
<td>MADD</td>
<td>G/C</td>
<td>0.124 (1.4 × 10^{-20})</td>
<td>0.000 (0.45)</td>
<td>0.005 (1.4 × 10^{-5})</td>
<td>0.005 (0.67)</td>
<td>0.03 (0.46)</td>
</tr>
<tr>
<td>rs6235</td>
<td>PCSK1</td>
<td>G/C</td>
<td>0.057 (4.5 × 10^{-6})</td>
<td>0.001 (0.78)</td>
<td>0.008 (0.02)</td>
<td>0.004 (0.70)</td>
<td>0.004 (0.8)</td>
</tr>
<tr>
<td>rs7903146</td>
<td>TCFL2</td>
<td>T/C</td>
<td>0.042 (1.5 × 10^{-3})</td>
<td>0.005 (0.005)</td>
<td>0.000 (0.02)</td>
<td>0.003 (0.001)</td>
<td>0.003 (0.8)</td>
</tr>
<tr>
<td>rs4502156</td>
<td>VPS13C/C2CD4A/B</td>
<td>T/C</td>
<td>0.039 (4.5 × 10^{-4})</td>
<td>0.000 (0.67)</td>
<td>0.000 (0.04)</td>
<td>0.000 (0.001)</td>
<td>0.000 (0.8)</td>
</tr>
<tr>
<td>rs1158471</td>
<td>SLC30A8</td>
<td>A/G</td>
<td>0.059 (1.1 × 10^{-5})</td>
<td>0.000 (0.21)</td>
<td>0.000 (0.06)</td>
<td>0.000 (0.001)</td>
<td>0.000 (0.8)</td>
</tr>
<tr>
<td>rs10838887</td>
<td>MADD (2nd signal)</td>
<td>T/G</td>
<td>0.006 (0.71)</td>
<td>0.000 (0.22)</td>
<td>0.000 (0.06)</td>
<td>0.000 (0.001)</td>
<td>0.000 (0.8)</td>
</tr>
<tr>
<td>rs1548318</td>
<td>LARP6</td>
<td>T/C</td>
<td>0.043 (1.2 × 10^{-4})</td>
<td>0.000 (0.02)</td>
<td>0.000 (0.07)</td>
<td>0.000 (0.88)</td>
<td>0.000 (0.8)</td>
</tr>
<tr>
<td>rs4790333</td>
<td>SGSM2</td>
<td>T/C</td>
<td>0.040 (3.2 × 10^{-4})</td>
<td>0.008 (0.04)</td>
<td>0.000 (0.55)</td>
<td>0.000 (0.33)</td>
<td>0.000 (0.15)</td>
</tr>
</tbody>
</table>

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

**DISCUSSION**

We report the first meta-analysis of genome-wide association studies of proinsulin levels, aiming to capture an increase in proinsulin relative to the nonspecific increase in insulin processing (Supplementary Fig. 2). We have found that higher proinsulin levels are associated with increased insulin resistance (34). Thus, we hypothesize that the failure of the β-cell to process proinsulin adequately in response to metabolic demands may be a key determinant of proinsulin levels. Interestingly, the proinsulin-raising allele of the lead SNP at the TRPS1 locus (rs179456) was associated with increased proinsulin levels, suggesting an important role for the proinsulin-raising allele in the regulation of proinsulin production.

**LIVER GENOME-WIDE ASSOCIATION STUDY OF PROINSULIN**

We constructed the unweighted and weighted genome-wide association analysis datasets for circulating fasting proinsulin. We adjusted for fasting insulin or more sophisticated measures of insulin resistance (Supplementary Fig. 2). Loci that simply increased proinsulin levels were not associated with proinsulin-raising alleles. The T allele (encoding trypsin) at the TRPS1 locus (rs179456) was associated with increased proinsulin levels, suggesting an important role for the proinsulin-raising allele in the regulation of proinsulin production.

**EVOLUTIONARY DETERMINANTS OF PROINSULIN**

The knowledge that TCF7L2 (35, 36) is a perfect proxy for the known T2D-associated SNP rs7903146 has been leveraged to identify disease-associated signals. Interestingly, the proinsulin-raising determinants of proinsulin can serve to identify disease-associated signals. We found a significant association at the proinsulin-raising A allele of the lead SNP at the VPS13C/C2CD4A/B locus (rs1548318) with increased proinsulin levels. The T allele (encoding trypsin) at the TRPS1 locus (rs179456) was associated with increased proinsulin levels, suggesting an important role for the proinsulin-raising allele in the regulation of proinsulin production.

**SLC30A8]**

SLC30A8 is a perfect proxy for the known T2D-associated SNP rs7903146. At 30 kb from the lead signal (rs179456), the proinsulin-associated SNP rs11558471 (37) is predicted to be within high LD (r^2 > 0.8) correlations in the HapMap CEU with potentially functional SNPs within 1 Mb of the lead signals. We also tested whether any of the proinsulin-associated SNPs might influence proximal (cis) expression of human transcripts in tissues available to us that had been paired for fasting insulin or more sophisticated measures of insulin resistance and were found to be strongly influenced by the proinsulin-raising allele. A meta-analysis of an eQTL database from human liver indicated that the T allele is associated with lower levels of expression. Analyses of the lead signals.

**TCF7L2**

The knowledge that TCF7L2 (35, 36) is a perfect proxy for the known T2D-associated SNP rs7903146 has been leveraged to identify disease-associated signals. Interestingly, the proinsulin-raising determinants of proinsulin can serve to identify disease-associated signals. We found a significant association at the proinsulin-raising A allele of the lead SNP at the VPS13C/C2CD4A/B locus (rs1548318) with increased proinsulin levels. The T allele (encoding trypsin) at the TRPS1 locus (rs179456) was associated with increased proinsulin levels, suggesting an important role for the proinsulin-raising allele in the regulation of proinsulin production.

**DISCUSSION**

We report the first meta-analysis of genome-wide association studies of proinsulin levels, aiming to capture an increase in proinsulin relative to the nonspecific increase in insulin processing (Supplementary Fig. 2). We have found that higher proinsulin levels are associated with increased insulin resistance (34). Thus, we hypothesize that the failure of the β-cell to process proinsulin adequately in response to metabolic demands may be a key determinant of proinsulin levels. Interestingly, the proinsulin-raising allele of the lead SNP at the TRPS1 locus (rs179456) was associated with increased proinsulin levels, suggesting an important role for the proinsulin-raising allele in the regulation of proinsulin production.

**LIVER GENOME-WIDE ASSOCIATION STUDY OF PROINSULIN**

We constructed the unweighted and weighted genome-wide association analysis datasets for circulating fasting proinsulin. We adjusted for fasting insulin or more sophisticated measures of insulin resistance (Supplementary Fig. 2). Loci that simply increased proinsulin levels were not associated with proinsulin-raising alleles. The T allele (encoding trypsin) at the TRPS1 locus (rs179456) was associated with increased proinsulin levels, suggesting an important role for the proinsulin-raising allele in the regulation of proinsulin production.

**EVOLUTIONARY DETERMINANTS OF PROINSULIN**

The knowledge that TCF7L2 (35, 36) is a perfect proxy for the known T2D-associated SNP rs7903146 has been leveraged to identify disease-associated signals. Interestingly, the proinsulin-raising determinants of proinsulin can serve to identify disease-associated signals. We found a significant association at the proinsulin-raising A allele of the lead SNP at the VPS13C/C2CD4A/B locus (rs1548318) with increased proinsulin levels. The T allele (encoding trypsin) at the TRPS1 locus (rs179456) was associated with increased proinsulin levels, suggesting an important role for the proinsulin-raising allele in the regulation of proinsulin production.

**SLC30A8]**

SLC30A8 is a perfect proxy for the known T2D-associated SNP rs7903146. At 30 kb from the lead signal (rs179456), the proinsulin-associated SNP rs11558471 (37) is predicted to be within high LD (r^2 > 0.8) correlations in the HapMap CEU with potentially functional SNPs within 1 Mb of the lead signals. We also tested whether any of the proinsulin-associated SNPs might influence proximal (cis) expression of human transcripts in tissues available to us that had been paired for fasting insulin or more sophisticated measures of insulin resistance and were found to be strongly influenced by the proinsulin-raising allele. A meta-analysis of an eQTL database from human liver indicated that the T allele is associated with lower levels of expression. Analyses of the lead signals.**
TABLE 3

<table>
<thead>
<tr>
<th>SNP</th>
<th>remembered resistant alleles</th>
<th>T2D</th>
<th>2-h glucose</th>
<th>HOMA-IR</th>
<th>Matsuda index</th>
<th>2-h insulin</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs4790333</td>
<td>T/C</td>
<td>1.01</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td>rs1549318</td>
<td>T/C</td>
<td>1.02</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>rs11071657</td>
<td>T/C</td>
<td>1.02</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>rs17271305</td>
<td>T/C</td>
<td>1.02</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>rs11603334</td>
<td>A/G</td>
<td>1.14</td>
<td>0.014</td>
<td>0.014</td>
<td>0.014</td>
<td>0.014</td>
</tr>
<tr>
<td>rs4502156</td>
<td>T/C</td>
<td>2.6</td>
<td>0.013</td>
<td>0.013</td>
<td>0.013</td>
<td>0.013</td>
</tr>
<tr>
<td>rs7903146</td>
<td>T/G</td>
<td>4.8</td>
<td>0.012</td>
<td>0.012</td>
<td>0.012</td>
<td>0.012</td>
</tr>
</tbody>
</table>

**Note:** Although four biological candidates (TCF7L2, SLC30A8, and RORA) are known to associate with the same genomic region, we show that the T2D-associated SNP is in strong LD with those in Japanese (44), and that the same LD is associated with T2D in Europeans (44), with supportive evidence found in Europeans (3,44). Here we show that the strongest association with proinsulin levels, the associated SNP rs1549318, is located within the regulatory region of TCF7L2, which is involved in the regulation of transcription. The expression of TCF7L2 is increased in insulin-resistant cells, indicating that the role of its protein product in the divergent effects of these loci on T2D is that both diseases are mediated by a ribonucleoprotein that affects the transfer of phospholipids to membranes.
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 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 an 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
32. Li Y, Abecasis GR. Mach 1.0: rapid haplotype reconstruction and missing genotype inference (Abstract). Am J Hum Genet 2006;S70:2290
34. Riva A, Kohane IS. A SNP-centric database for the investigation of the human genome. BMC Bioinformatics 2004;5:33