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 Rybin,8 John R. Petrie,9 Mary E. Travers,10 Nabiha Bouatia-Naji,10,11 Antigone S. Dimas,5,12 Alexandra Nica,12,13 Eleanor Wheeler,14 Han Chen,2 Benjamin F. Voight,15,16 Jalal Taneera,7 Stavroula Kaxaki,13,17 John F. Peden,5,18 Fabiola Tururn,7,19 Stefan Gustafsson,20 Carina Zabena,21,22 Peter Almgren,23 David J.P. Barker,24 Daniel Barnes,6 Elaine M. Dennisson,24 Johan G. Eriksson,25,26,27,28 Per Eriksson,29 Elodie Eury,10,11 Lasse Folkersken,29 Caroline S. Fox,3,30 Timothy M. Frayling,31 Anuj Goel,32,33 Momoko Horikoshi,39 Maria Nastase Mannila,4 John Öhrvik,4 Clive Osmond,24 Laura Pascoe,40 Felicity Payne14 Avan A. Sayer,2,4 Bengt Sennblad,1 Angela Silveira,1 Alena Stancakova,37 Kathy Stirrups,13 Amy J. Swift,39 Ann-Christine Svingen,41 Tiinaajamaa Tuomi,27,42 Ferdinand M van ’t Hof,39 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,* the 28Department of General Practice and Primary Health Care, University of Oxford, Oxford, U.K.; the 2Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts; the 63Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California; and the 64Department of Physiology and Biophysics, Keck School of Medicine, Keck School of Medicine, Los Angeles, California.

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 3Department of Nutrition and Dietetics-Nutrition, Harokopio University, Athens, Greece; the 4Institute of Public Health, Imperial College London, London, 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 Uppsala and Malmo, L.U., Malmo, Sweden; the 8Boston University Data Coordinating Center, Boston, MA; the 9UoGH BHF Cardiovascular Research Centre, University of Glasgow, Glasgow, U.K.; the 10University Lille-Nord de France, Lille, France; the 11CNRS UMR5199, Institut Pasteur de Lille, France; the 12Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland; the 13Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, U.K.; the 14Center for Statistical Genetics, Department of Biostatistics, Broad Institute, Cambridge, MA; the 15Program in Medical Genetics and Development, University of Geneva Medical School, Geneva, Switzerland; the 16Wellcome Trust Sanger Institute, Hinxton, Cambridge, U.K.; the 17Department of Clinical Sciences, Diabetes and Endocrinology, University Hospital Uppsala and Malmo, L.U., Malmo, Sweden; the 18Department of Medicine, Karolinska Institutet, Stockholm, Sweden; the 19Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; the 20Department of Dietetics-Nutrition, Harokopio University, Athens, Greece; the 21Department of Cardiovacular Medicine, University of Oxford, Oxford, U.K.; the 22Department of Medicine, University of Verona, Verona, Italy; the 23Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; the 24CIBER de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), Madrid, Spain; the 25Fundacion Investigacion Biomédica del Hospital Clinico San Carlos, Madrid, Spain; the 26Heart Research Center, Oregon Health and Science University, Portland, Oregon; the 27MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton General Hospital, Southampton, U.K.; the 28National Institute for Health and Welfare, Finland; the 29Experimental Cardiovascular Research Unit, Imperial College London, London, U.K.; the 30Department of Biomedical and Clinical Sciences, Peninsula Medical School, University of Exeter, Exeter, U.K.; the 31Institute of Non-Infectious and Diabetes Unit, Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden; the 32Department of Preventive Medicine, Keck School of Medicine, Los Angeles, California; and the 33Endocrinology and Diabetes Unit, Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden; the 34Center for Statistical Genetics, Department of Bioinformatics, School of Public Health, University of Exeter, Exeter, U.K.; the 35Hospital for Children and Adolescents, Helsinki University Central Hospital and University of Helsinki, HUS, Helsinki, Finland; the 36INSERM UMR859, Lille, France; the 37Department of Medicine, University of Kuopio and Kuopio University Hospital, Kuopio, Finland; the 38First Department of Propaedeutic Medicine, Laiko General Hospital, Athens, Greece; the 39National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland; the 40Institute of Cell and Metabolic Medicine, University of Cambridge, Cambridge, U.K.; the 41National Institute for Health and Care Research, Cambridge, U.K.; the 42Department of Medical Sciences, Molecular Medicine, Science for Life Laboratory, Uppsala University, Uppsala, Sweden; the 43Department of Medicine, Helsinki University Central Hospital, and Research Program of Molecular Medicine, University of Helsinki, Helsinki, Finland; the 44Institute of Cellular Medicine, Newcastle University, Newcastle, Newcastle, U.K.; the 45Department of Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden; the 46Department of Cardiovascular Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford, U.K.; the 47Clinical Trial Service Unit, University of Oxford, Oxford, U.K.; the 48Center for Non-Communicable Diseases Pakistan, Karachi, Pakistan; the 49Epidemiology and Biostatistics, Imperial College London, London, U.K.; the 50Cardiovascular Imaging Research Centre, Imperial College London, London, U.K.; the 51National Heart and Lung Institute, Imperial College London, London, U.K.; the 52Department of Internal Medicine and CINR Institute of Clinical Physiology, University of Pisa School of Medicine, Pisa, Italy; the 53Center for Genomics, Biomedical Sciences, Addenbrooke’s Hospital, Cambridge, U.K.; the 54Department of Cardiovascular Research, Mario Negri Institute for Pharmacological Research, Milan, Italy; the 55Leibniz Institute for Arteriosclerosis Research, University of Munich, Germany; the 56Department of Genomics of Common Disease, School of Public Health, Imperial College London, Hammersmith Hospital, London, U.K.; the 57DRWF Human Islet Isolation Facility and Oxford Islet Transplant Programme, University of Oxford, Oxford, U.K.; the 58National Institutes of Health, Bethesda, Maryland; the 59Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, U.K.; the 60Clinical Medicine Division, Massachusetts General Hospital, Boston, Massachusetts; the 61Department of Medicine, Harvard Medical School, Boston, Massachusetts; the 62University of Cambridge Metabolic Research Laboratories, Institute of Metabolic Science, Addenbrooke’s Hospital, Cambridge, U.K.; the 63Department of Preventive Medicine, Keck School of Medicine, Los Angeles, California; and the 64Department of Physiology and Biophysics, Keck School of Medicine, Los Angeles, California.

Corresponding author: Jose C. Florez, jcflorez@partners.org.

Received 23 March 2011 and accepted 29 June 2011.

DOI: 10.2337/db11-0415


* A full list of the DIAGRAM Consortium, the GIANT Consortium, the MuTHER Consortium, the CARDioGRAM Consortium, and the C4D Consortium is given in the Supplementary Data. © 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.
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 ($P < 5 \times 10^{-8}$). Two loci (LARP6 and SGLT2) 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 (TCF7L2, SLC30A8, VPS13C/C2CD4A/B, and ARAP1, formerly CENT2) increase T2D risk. The proinsulin-raising allele of ARAP1 was associated with a lower fasting glucose ($P = 1.7 \times 10^{-8}$), improved β-cell function ($P = 1.1 \times 10^{-10}$), and lower risk of T2D (odds ratio 0.88; $P = 7.8 \times 10^{-5}$). 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.

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 (TCF7L2, SLC30A8, CDA1L) 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 SGLT2, and the known glycemic loci ARAP1, MADD (two independent signals), TCF7L2, VPS13C/C2CD4A/B, SLC30A8, 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 ($n = 5,759$), 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 METSIM ($n = 5,122$), Botnia-PP (2,280), HBCS ($n = 1,544$), the Ely study ($n = 1,568$), the Hertfordshire study ($n = 1,016$), ULSAM ($n = 930$), RCC ($n = 914$), PIVUS ($n = 912$), Segovia ($n = 911$), GHRS ($n = 668$), and 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
Genetic variants associated with proinsulin were identified through a two-stage association study. In the discovery stage, we used genome-wide commercial arrays (Affymetrix 500K, MIPS 50K, and Illumina HumanM610K) to genotype participants 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 × 10⁻⁵; Fig. 1). The clean dataset showed no significant deviation from Hardy-Weinberg equilibrium, and the minor allele frequencies were in the typical range for European populations. We meta-analyzed stage 1 and stage 2 results using inverse-variance weighted fixed effects meta-analysis methods including up to 27,079 participants.

**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 × 10⁻⁵; Fig. 1). The clean dataset showed no significant deviation from Hardy-Weinberg equilibrium, and the minor allele frequencies were in the typical range for European populations. We meta-analyzed stage 1 and stage 2 results using inverse-variance weighted fixed effects meta-analysis methods including up to 27,079 participants.

**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 × 10⁻⁵. Additionally, two SNPs near the P < 1 × 10⁻⁵ threshold (in ASAP2 and a gene desert region) were carried forward as a result of biological plausibility (ASAP2 is involved in vesicular transport) and/or consistency of direction of effect in all discovery stage 1 studies (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 27,079 participants.

**Additional analyses and expression/eQTL studies** are described in the Supplementary Material.
systematic deviation from the null expectation, with the exception of the tail of the distribution (Fig. 1, insert).

**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 × 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 × 10−9 and 1.5 × 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; r² = 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 sex stratification, and heterogeneity was attenuated (e.g., F² = 77.2%, heterogeneity P = 1.9 × 10−7 for combined men and women, whereas F² = 64.6%, heterogeneity P = 4.5 × 10−7 [men] and F² = 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 DDX3J) where a genome-wide significant association was seen in women (P = 2.0 × 10−5; Supplementary Fig. 4A) but not men (P = 0.17; Supplementary Fig. 4B; sex interaction P = 8.9 × 10−5). Although removal of the METSIM cohort improved the heterogeneity score and produced nominal significance for the association in men (P = 0.09), 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.

### Table 1

<table>
<thead>
<tr>
<th>SNP</th>
<th>Nearest gene</th>
<th>CHR</th>
<th>Position</th>
<th>Allele (effect/other)</th>
<th>Freq.</th>
<th>Discovery P value</th>
<th>Replication P value</th>
<th>Combined P value</th>
<th>Replication b (SE)</th>
<th>Allele b (SE)</th>
<th>Combined b (SE)</th>
<th>Combined P value</th>
<th>Heterogeneity F² in % (Q-test)</th>
<th>n discovery</th>
<th>n replication</th>
<th>n Combined</th>
<th>n replication (BMI adjusted)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs10501320</td>
<td>MADD</td>
<td>11</td>
<td>72110633</td>
<td>A/G</td>
<td>0.15</td>
<td>3.3 × 10−5</td>
<td>0.0825 (0.0056)</td>
<td>0.0825 (0.0056)</td>
<td>0.0825 (0.0056)</td>
<td>0.0427 (0.0056)</td>
<td>0.0427 (0.0056)</td>
<td>0.0427 (0.0056)</td>
<td>10,806 (0.03)</td>
<td>0.123 (0.05)</td>
<td>0.0307 (0.0045)</td>
<td>2.44 (10−4)</td>
<td></td>
</tr>
<tr>
<td>rs10838687</td>
<td>MADD</td>
<td>11</td>
<td>47250375</td>
<td>G/C</td>
<td>0.72</td>
<td>2.1 × 10−3</td>
<td>0.0775 (0.0066)</td>
<td>0.0775 (0.0066)</td>
<td>0.0775 (0.0066)</td>
<td>0.123 (0.05)</td>
<td>0.123 (0.05)</td>
<td>0.123 (0.05)</td>
<td>0.0307 (0.0045)</td>
<td>2.44 (10−4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs9727115</td>
<td>SNX7</td>
<td>17</td>
<td>38948914</td>
<td>G/A</td>
<td>0.64</td>
<td>5.3 × 10−6</td>
<td>0.0133 (0.0051)</td>
<td>0.0133 (0.0051)</td>
<td>0.0133 (0.0051)</td>
<td>0.0427 (0.0056)</td>
<td>0.0427 (0.0056)</td>
<td>0.0427 (0.0056)</td>
<td>0.0307 (0.0045)</td>
<td>2.44 (10−4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs306549</td>
<td>DDX3J</td>
<td>15</td>
<td>68896201</td>
<td>T/C</td>
<td>0.61</td>
<td>8.0 × 10−3</td>
<td>0.0208 (0.0056)</td>
<td>0.0208 (0.0056)</td>
<td>0.0208 (0.0056)</td>
<td>0.0427 (0.0056)</td>
<td>0.0427 (0.0056)</td>
<td>0.0427 (0.0056)</td>
<td>0.0307 (0.0045)</td>
<td>2.44 (10−4)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Exploration of proinsulin processing mechanisms. Proinsulin is initially cleaved to split 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 SGSMS2) 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 DIAGRAM (29) and 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 SGSMS2 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 SGSMS2 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). Association signals 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 CENTD2 and reported as such in DIAGRAM+) (29) was associated with a lower fasting glucose (0.019 mg/dL per A allele; \( P = 1.7 \times 10^{-4} \)), 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 SGSMS2) 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 \( \sim 8 \) 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, C62 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 exploration. 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 in strong LD with the lead SNP (\( r^2 \geq 0.8 \)) within a 1-MB region. We found that PCSK1 rs6235 codes for a non-synonymous variant (S990T), which is in perfect LD with rs6234, another missense variant (Q665E); both were predicted to be nondamaging by Polyphen (35) and SIFT (36). At SLC30A8, the proinsulin-associated SNP rs11558471 is a perfect proxy for the known T2D-associated SNP rs13266634, encoding R325W. The T allele (encoding tryptophan) is predicted to be benign by Polyphen, but damaging by SIFT. We found no other strong (\( r^2 > 0.8 \)) correlations in 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
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.
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, SLC30A8, 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 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.

**DISCUSSION**

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene/Region</th>
<th>Unweighted Coef</th>
<th>Weighted Coef</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs11093334</td>
<td>MADD</td>
<td>0.012</td>
<td>0.025</td>
<td>0.012</td>
</tr>
<tr>
<td>rs15932251</td>
<td>VPS13C/C2CD4A/B</td>
<td>0.009</td>
<td>0.015</td>
<td>0.004</td>
</tr>
<tr>
<td>rs16588967</td>
<td>SLC30A8</td>
<td>0.009</td>
<td>0.013</td>
<td>0.003</td>
</tr>
<tr>
<td>rs15932250</td>
<td>LARP6</td>
<td>0.007</td>
<td>0.012</td>
<td>0.003</td>
</tr>
<tr>
<td>rs16588968</td>
<td>SGSM2</td>
<td>0.006</td>
<td>0.010</td>
<td>0.002</td>
</tr>
</tbody>
</table>

β-Coefficients are adjusted for age, sex, and study-specific covariates (if applicable).
<table>
<thead>
<tr>
<th>SNP</th>
<th>Nearest gene</th>
<th>Alleles</th>
<th>Proinsulin-raising/other</th>
<th>Fasting glucose ((n = 44,601-46,186))</th>
<th>HOMA-IR ((n = 35,512-37,072))</th>
<th>Matsuda index ((n = 7,055-9,561))</th>
<th>2-h glucose ((n = 15,088-15,252))</th>
<th>2-h insulin ((n = 6,923-7,083))</th>
<th>HbA1c ((n = 33,736-44,731))</th>
<th>T2D ((n = 8,130) case subjects/ 38,987 control subjects)</th>
<th>BMI ((n = 123,260-123,865))</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs11603334</td>
<td>ARAP1</td>
<td>A/G</td>
<td>β/0R</td>
<td>0.019</td>
<td>0.105</td>
<td>0.383</td>
<td>0.002</td>
<td>0.043</td>
<td>0.000</td>
<td>0.018</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.016</td>
<td>-0.014</td>
<td>0.009</td>
<td>0.043</td>
<td>-0.050</td>
<td>0.016</td>
<td>0.004</td>
<td>0.008</td>
<td>0.012</td>
</tr>
<tr>
<td>rs10501320</td>
<td>MADD</td>
<td>G/C</td>
<td>β/0R</td>
<td>0.025</td>
<td>0.775</td>
<td>0.435</td>
<td>0.014</td>
<td>0.009</td>
<td>0.001</td>
<td>0.012</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.015</td>
<td>-0.009</td>
<td>-0.001</td>
<td>-0.012</td>
<td>0.000</td>
<td>0.016</td>
<td>0.004</td>
<td>0.008</td>
<td>0.014</td>
</tr>
<tr>
<td>rs6235</td>
<td>PCSK1</td>
<td>G/C</td>
<td>β/0R</td>
<td>-0.014</td>
<td>0.001</td>
<td>0.435</td>
<td>0.014</td>
<td>0.009</td>
<td>0.001</td>
<td>0.012</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.001</td>
<td>-0.008</td>
<td>0.034</td>
<td>0.015</td>
<td>0.013</td>
<td>0.013</td>
<td>0.053</td>
<td>0.002</td>
<td>0.012</td>
</tr>
<tr>
<td>rs7903146</td>
<td>TCF7L2</td>
<td>T/C</td>
<td>β/0R</td>
<td>0.023</td>
<td>0.034</td>
<td>0.435</td>
<td>0.014</td>
<td>0.010</td>
<td>0.004</td>
<td>0.012</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.010</td>
<td>0.014</td>
<td>0.108</td>
<td>0.004</td>
<td>0.015</td>
<td>0.140</td>
<td>0.004</td>
<td>0.019</td>
<td>0.016</td>
</tr>
<tr>
<td>rs4502156</td>
<td>VPS13C/C2CD4A/B</td>
<td>T/C</td>
<td>β/0R</td>
<td>0.020</td>
<td>0.002</td>
<td>0.435</td>
<td>0.014</td>
<td>0.009</td>
<td>0.001</td>
<td>0.012</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.002</td>
<td>-0.004</td>
<td>0.012</td>
<td>0.002</td>
<td>0.002</td>
<td>1.07</td>
<td>0.004</td>
<td>0.008</td>
<td>0.003</td>
</tr>
<tr>
<td>rs11558471</td>
<td>SLC30A8</td>
<td>A/G</td>
<td>β/0R</td>
<td>0.027</td>
<td>0.017</td>
<td>0.435</td>
<td>0.014</td>
<td>0.017</td>
<td>0.002</td>
<td>0.012</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.017</td>
<td>-0.002</td>
<td>0.000</td>
<td>0.002</td>
<td>0.000</td>
<td>-0.001</td>
<td>0.014</td>
<td>0.008</td>
<td>0.003</td>
</tr>
<tr>
<td>rs10838687</td>
<td>MADD (2nd signal)</td>
<td>T/G</td>
<td>β/0R</td>
<td>-0.011</td>
<td>0.004</td>
<td>0.435</td>
<td>0.014</td>
<td>0.017</td>
<td>0.002</td>
<td>0.012</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.004</td>
<td>-0.006</td>
<td>0.014</td>
<td>-0.003</td>
<td>0.000</td>
<td>-0.001</td>
<td>0.014</td>
<td>0.008</td>
<td>0.003</td>
</tr>
<tr>
<td>rs1549318</td>
<td>LARP6</td>
<td>T/C</td>
<td>β/0R</td>
<td>-0.002</td>
<td>0.009</td>
<td>0.435</td>
<td>0.014</td>
<td>0.017</td>
<td>0.002</td>
<td>0.012</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.009</td>
<td>0.000</td>
<td>0.008</td>
<td>-0.010</td>
<td>0.000</td>
<td>0.005</td>
<td>0.000</td>
<td>0.008</td>
<td>0.003</td>
</tr>
<tr>
<td>rs4790333</td>
<td>SGSM2</td>
<td>T/C</td>
<td>β/0R</td>
<td>0.004</td>
<td>0.009</td>
<td>0.435</td>
<td>0.014</td>
<td>0.017</td>
<td>0.002</td>
<td>0.012</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.009</td>
<td>-0.002</td>
<td>0.008</td>
<td>-0.005</td>
<td>0.000</td>
<td>1.01</td>
<td>0.004</td>
<td>0.008</td>
<td>0.003</td>
</tr>
</tbody>
</table>

β-Coeficients are adjusted for age, sex, and study-specific covariates (if applicable). OR is for T2D only as a dichotomous trait.
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

**FIG. 3.** Expression profiles of biologically plausible genes within each associated locus across a range of human tissue types, including islet preparations from three donors. Expression levels determined with respect to the geometric mean of three endogenous control assays. A: ARAP1 region; B: MADD region; C: VPS13C/C2CD4/B region; D: LARP6 region; E: SGSM2 region. F: Expression levels of genes near the proinsulin associated variants in human FAC-sorted β-cells. Data are expression means ± SD of the relative expression measured by quantitative PCR obtained from three human nondiabetic donors.
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 C-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 etiological 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 etiological 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


