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] Josée Dupuis,[2,3] Inga Prokopenko,[4,5] Adam Barker,[6] Emma Ahlqvist,[7] Denis Rybin,[8] John R. Petrie,[9] Mary E. Travers,[4] Nabila 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 Kanoni,[13,17] John F. Peden,[5,18] Fabiola Turrini,[7,19] Stefan Gustafsson,[20] Carina Zabena,[21,22] Peter Almgren,[7] David J.P. Barker,[23] Daniel Barnes,[6] Elaine M. Dennison,[24] Johan G. Eriksson,[25,26,27,28] Per Eriksson,[1] Elodie Eury,[10,11] Lasse Folkersen,[29] Caroline S. Fox,[3,30] Timothy M. Frayling,[31] Anuj Goel,[5,18] Harvest F. Gu,[32] Momoko Horikoshi,[4,5] Bo Isomaa,[27,33] Anne U. Jackson,[34] Karen A. Jameson,[24] Eero Kajantie,[25,35] Julie Kerr-Conte,[10,36] Teemu Kuulasmaa,[37] Johanna Kuusisto,[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 Osmond,[24] Laura Pascoe,[40] Felicity Payne,[14] Avan A. Sayer,[24] Bengt Sennblad,[1] Angela Silveira,[1] Alena Stančáková,[37] Kathy Stirrups,[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,[31] Weijia Xie,[31] Björn Zethelius,[44] the DIAGRAM Consortium,* the GIANT Consortium,* the MuTHER Consortium,* the CARDIoGRAM Consortium,* the C4D Consortium,* Halit Ongen,[5,18,45] Anders Mälarstig,[1] Jemma C. Hopewell,[46] Danish Saleheen,[47,48] John Chambers,[49,50] Sarah Parish,[46] John Danesh,[47] Jaspal Kooner,[50,51] Claes-Göran Östenson,[32] Lars Lind,[41] Cyrus C. Cooper,[24] Manuel Serrano-Ríos,[21,22] Ele Ferrannini,[52] Tom J. Forsen,[28,53] Robert Clarke,[46] Maria Grazia Franzosi,[54] Udo Seedorf,[55] Hugh Watkins,[5,18] Philippe Froguel,[10,11,56] Paul Johnson,[4,57] Panos Deloukas,[13] Francis S. Collins,[58] Markku Laakso,[37] Emmanouil T. Dermitzakis,[12] Michael Boehnke,[34] Mark I. McCarthy,[4,5,59] Nicholas J. Wareham,[6] Leif Groop,[7] François Pattou,[10,36] Anna L. Gloyn,[4] George V. Dedoussis,[17] Valeriya Lyssenko,[7] James B. Meigs,[60,61] Inês Barroso,[14,62] Richard M. Watanabe,[63,64] Erik Ingelsson,[20] Claudia Langenberg,[6] Anders Hamsten,[1] and Jose C. Florez[15,16,61]

From the [1]Atherosclerosis Research Unit, Department of Medicine Solna, Karolinska Institutet, Karolinska University Hospital Solna, Stockholm, Sweden; the [2]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts; the [3]National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts; the [4]Oxford Centre for Diabetes Endocrinology and Metabolism, University of Oxford, Oxford, U.K.; the [5]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, U.K.; the [6]MRC Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, U.K.; the [7]Department of Clinical Sciences, Diabetes and Endocrinology, University Hospital and Malmö, Lund University, Malmö, Sweden; the [8]Boston University Data Coordinating Center, Boston, Massachusetts; the [9]BHF Cardiovascular Research Centre, University of Glasgow, Glasgow, U.K.; the [10]Université Lille-Nord de France, Lille, France; the [11]CNRS UMR 8199, Institut Pasteur de Lille, Lille, France; the [12]Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland; the [13]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, U.K.; the [14]Metabolic Disease Group, Wellcome Trust Sanger Institute, Hinxton, Cambridge, U.K.; the [15]Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts; the [16]Center for Human Genetic Research and Diabetes Research Center (Diabetes Unit), Massachusetts General Hospital, Boston, Massachusetts; the [17]Department of Dietetics-Nutrition, Harokopio University, Athens, Greece; the [18]Department of Cardiovascular Medicine, University of Oxford, Oxford, U.K.; the [19]Department of Medicine, University of Verona, Verona, Italy; the [20]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; the [21]CIBER de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), Madrid, Spain; the [22]Fundación Investigación Biomédica del Hospital Clínico San Carlos, Madrid, Spain; the [23]Heart Research Center, Oregon Health and Science University, Portland, Oregon; the [24]MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton General Hospital, Southampton, U.K.; the [25]National Institute for Health and Welfare, Helsinki, Finland; the [26]Helsinki University Central Hospital, Unit of General Practice, Helsinki, Finland; the [27]Folkhälsan Research Centre, Helsinki, Finland; the [28]Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland; the [29]Experimental Cardiovascular Research Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden; the [30]Division of Endocrinology, Diabetes, and Hypertension, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; the [31]Institute of Biomedical and Clinical Sciences, Peninsula Medical School, University of Exeter, Exeter, U.K.; the [32]Endocrinology and Diabetes Unit, Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden; the [33]Malmska Municipal Health Care Center and Hospital, Jakobstad, Finland; the [34]Center for Statistical Genetics, Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, Michigan; the [35]Hospital for Children and Adolescents, Helsinki University Central Hospital and University of Helsinki, Helsinki, Finland; the [36]INSERM UMR 859, Lille, France; the [37]Department of Medicine, University of Kuopio and Kuopio University Hospital, Kuopio, Finland; the [38]First Department of Propaedeutic Medicine, Laiko General Hospital, Athens University Medical School, Athens, Greece; the [39]National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland; the [40]Institute of Cell and Molecular Biosciences, Newcastle University, Newcastle, U.K.; the [41]Department of Medical Sciences, Molecular Medicine, Science for Life Laboratory, Uppsala University, Uppsala, Sweden; the [42]Department of Medicine, Helsinki University Central Hospital, and Research Program of Molecular Medicine, University of Helsinki, Helsinki, Finland; the [43]Institute of Cellular Medicine, Newcastle University, Newcastle, U.K.; the [44]Department of Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden; the [45]Department of Cardiovascular Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford, U.K.; the [46]Clinical Trial Service Unit, University of Oxford, Oxford, U.K.; the [47]Department of Public Health and Primary Care, University of Cambridge, Cambridge, U.K.; the [48]Center for Non-Communicable Diseases Pakistan, Karachi,

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 *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 (*TCF7L2*, *SLC30A8*, *VPS13C/C2CD4A/B*, and *ARAP1*, formerly *CENTD2*) increase T2D risk. The proinsulin-raising allele of *ARAP1* was associated with a lower fasting glucose ($P = 1.7 \times 10^{-4}$), improved β-cell function ($P = 1.1 \times 10^{-5}$), and lower risk of T2D (odds ratio 0.88; $P = 7.8 \times 10^{-6}$). 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 (*TCF7L2*, *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 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 ($n = 5,759$), 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.

Pakistan; [49]Epidemiology and Biostatistics, Imperial College London, Norfolk Place, London, U.K.; [50]Cardiology, Ealing Hospital NHS Trust, Middlesex, U.K.; the [51]National Heart and Lung Institute, Imperial College London, London, U.K.; the [52]Department of Internal Medicine and CNR Institute of Clinical Physiology, University of Pisa School of Medicine, Pisa, Italy; the [53]Vaasa Health Care Center, Vaasa, Finland; the [54]Department of Cardiovascular Research, Mario Negri Institute for Pharmacological Research, Milan, Italy; the [55]Leibniz Institute for Arteriosclerosis Research, University of Münster, Münster, Germany; the [56]Department of Genomics of Common Disease, School of Public Health, Imperial College London, Hammersmith Hospital, London, U.K.; the [57]DRWF Human Islet Isolation Facility and Oxford Islet Transplant Programme, University of Oxford, Oxford, U.K.; the [58]National Institutes of Health, Bethesda, Maryland; the [59]Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, U.K.; the [60]General Medicine Division, Massachusetts General Hospital, Boston, Massachusetts; the [61]Department of Medicine, Harvard Medical School, Boston, Massachusetts; the [62]University of Cambridge Metabolic Research Laboratories, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, U.K.; the [63]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California; and the [64]Department of Physiology and Biophysics, Keck School of Medicine, University of Southern California, Los Angeles, California.

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 = 1,649), 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 (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 ≥ 7 mmol/L (3); all participants were of European descent.

Proinsulin and insulin measurements. Proinsulin (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 (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 (h^2 = 0.36 vs. 0.34 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 SNPTEST (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 λ 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 (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 and expression quantitative trait loci (eQTL) studies are described in the Supplementary Data.
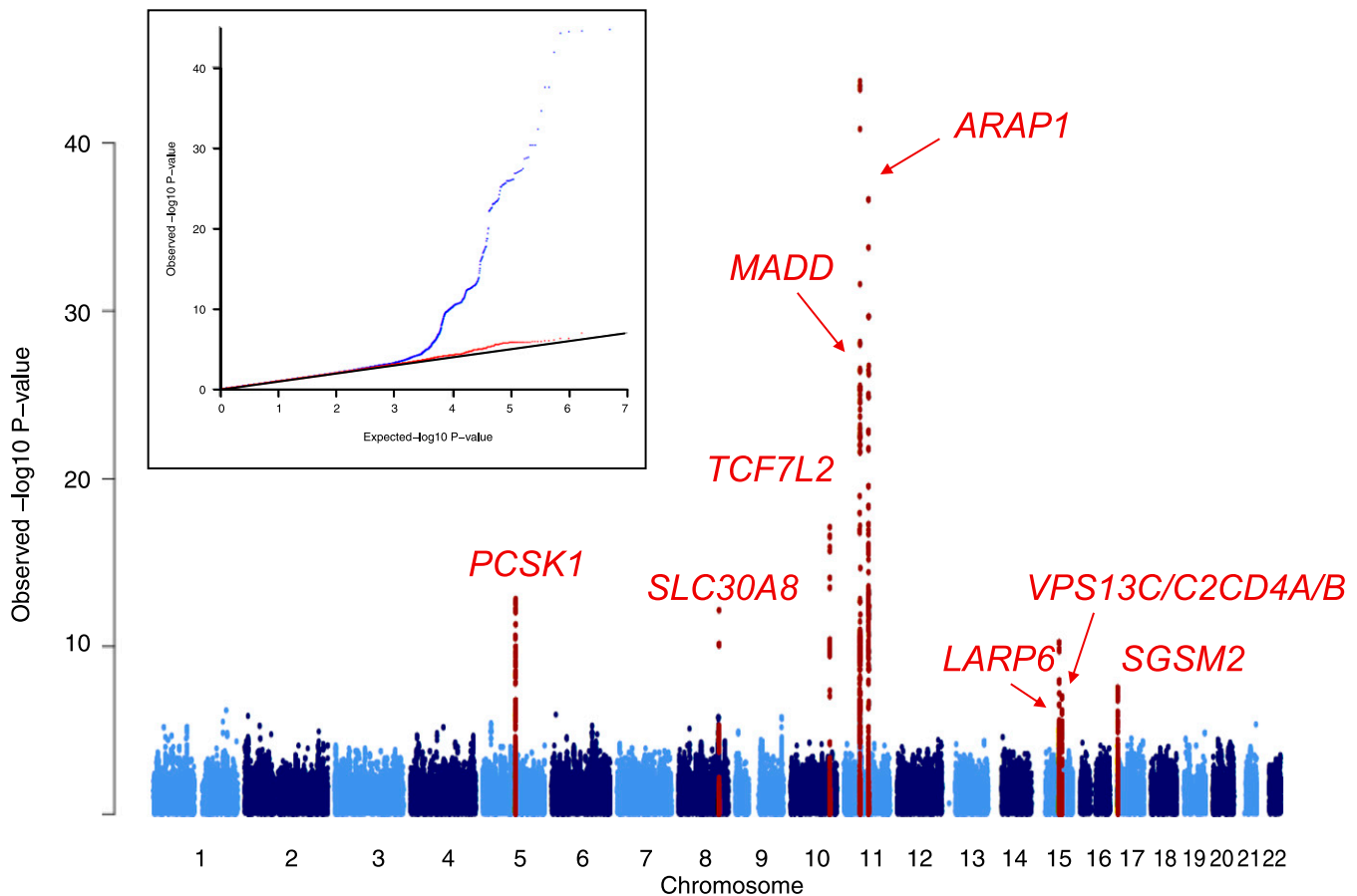

FIG. 1. Manhattan plot of the association *P* values for fasting proinsulin adjusted for fasting insulin. Directly genotyped and imputed SNPs are plotted with their meta-analysis *P* values (as −log10 values) as a function of genomic position (NCBI Build 36). The SNPs that achieved genome-wide significance ($P < 5 \times 10^{-8}$) 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.

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 ($P < 1 \times 10^{-5}$; Fig. 1). The clean dataset showed no 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 ($P < 5 \times 10^{-8}$), 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 1.5×10^{-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 sex 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 associated with fasting proinsulin levels at genome-wide levels of statistical significance

SNP	Nearest gene	CHR	Position	Allele (effect/ other)	Freq.	Discovery P value ($n = 10,538$–10,701)	Replication β (SE) ($n = 5,894$–15,898)	Replication β (SE), BMI adjusted ($n = 5,893$–15,890)	Replication β (SE), BMI + FG adjusted ($n = 5,744$–15,741)	Combined P value ($n = 19,168$–26,599)	Heterogeneity I^2 in % (Q-test P value)
rs11603334	*ARAP1*	11	72110633	A/G	0.15	3.3×10^{-38}	0.0928 (0.0053)	0.0943 (0.0056)	0.0938 (0.0052)	7.7×10^{-102}	77.2 (1.9×10^{-7})
rs10501320	*MADD*	11	47250375	G/C	0.72	2.1×10^{-45}	0.0805 (0.0064)	0.0748 (0.0069)	0.0775 (0.0064)	1.1×10^{-88}	68.1 (8.7×10^{-4})
rs6235	*PCSK1*	5	95754654	G/C	0.28	7.7×10^{-14}	0.0394 (0.005)	0.0407 (0.0052)	0.0438 (0.0049)	9.8×10^{-27}	30.2 (0.14)
rs7903146	*TCF7L2*	10	114748339	T/C	0.30	3.5×10^{-18}	0.0321 (0.007)	0.0330 (0.0076)	0.0303 (0.0063)	2.3×10^{-20}	66.2 (1.6×10^{-3})
rs4502156	*VPS13C/C2CD4A/B*	15	60170447	T/C	0.58	3.8×10^{-11}	0.0294 (0.0044)	0.0278 (0.0046)	0.0260 (0.0043)	3.5×10^{-20}	45.1 (0.03)
rs11558471	*SLC30A8*	8	118254914	A/G	0.69	4.2×10^{-13}	0.0280 (0.0049)	0.0280 (0.0053)	0.0273 (0.0048)	5.8×10^{-18}	58.5 (0.01)
rs10838687	*MADD* (2nd signal)*	11	47269468	T/G	0.80	8.6×10^{-7}	0.0253 (0.0050)	0.0240 (0.0049)	0.0246 (0.0049)	6.9×10^{-12}	18.4 (0.26)
rs1549318	*LARP6*	15	68896201	T/C	0.61	8.0×10^{-8}	0.0192 (0.0051)	0.0209 (0.0056)	0.0180 (0.005)	2.4×10^{-10}	52.7 (0.02)
rs4790333	*SGSM2*	17	2209453	T/C	0.45	2.2×10^{-8}	0.0154 (0.0043)	0.0179 (0.0042)	0.0179 (0.0042)	3.0×10^{-9}	56.8 (3.5×10^{-3})
rs9727115	*SNX7***	1	98949841	G/A	0.64	5.3×10^{-6}	0.0133 (0.0051)	0.0130 (0.005)	0.0134 (0.0045)	2.4×10^{-7}	0.00 (0.65)

β-Coefficients for the effect allele are shown after adjustments for sex, age, geographic covariates (if applicable), and age squared (Framingham only), besides the additional adjustments denoted in the column headings. The combined P values are shown for the joint analysis with the basic adjustments listed above. CHR, chromosome; Freq, frequency; SNP, single-nucleotide polymorphism. *This second SNP in *MADD* was selected for further testing based on a suggestive P value (8.6×10^{-7}) when analyses for chromosome 11 were conditioned for the top two signals (*MADD* and *ARAP1*). The conditional analysis results are shown. **SNX7 only reached genome-wide significance after adjusting for fasting glucose ($P = 5.4 \times 10^{-9}$). Another signal (rs306549 in *DDX31*) reached genome-wide significance in women ($P = 2.0 \times 10^{-8}$) but not in men ($P = 0.17$; see text for details).

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 $P < 1.5 \times 10^{-3}$, 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 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 *SGSM2*) were also examined in relation to other glucometabolic 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 *SGSM2* 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 *SGSM2* 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 *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 *SGSM2*) 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 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 nonsynonymous variant (S690T), which is in perfect LD with rs6234, another missense variant (Q665E); both were
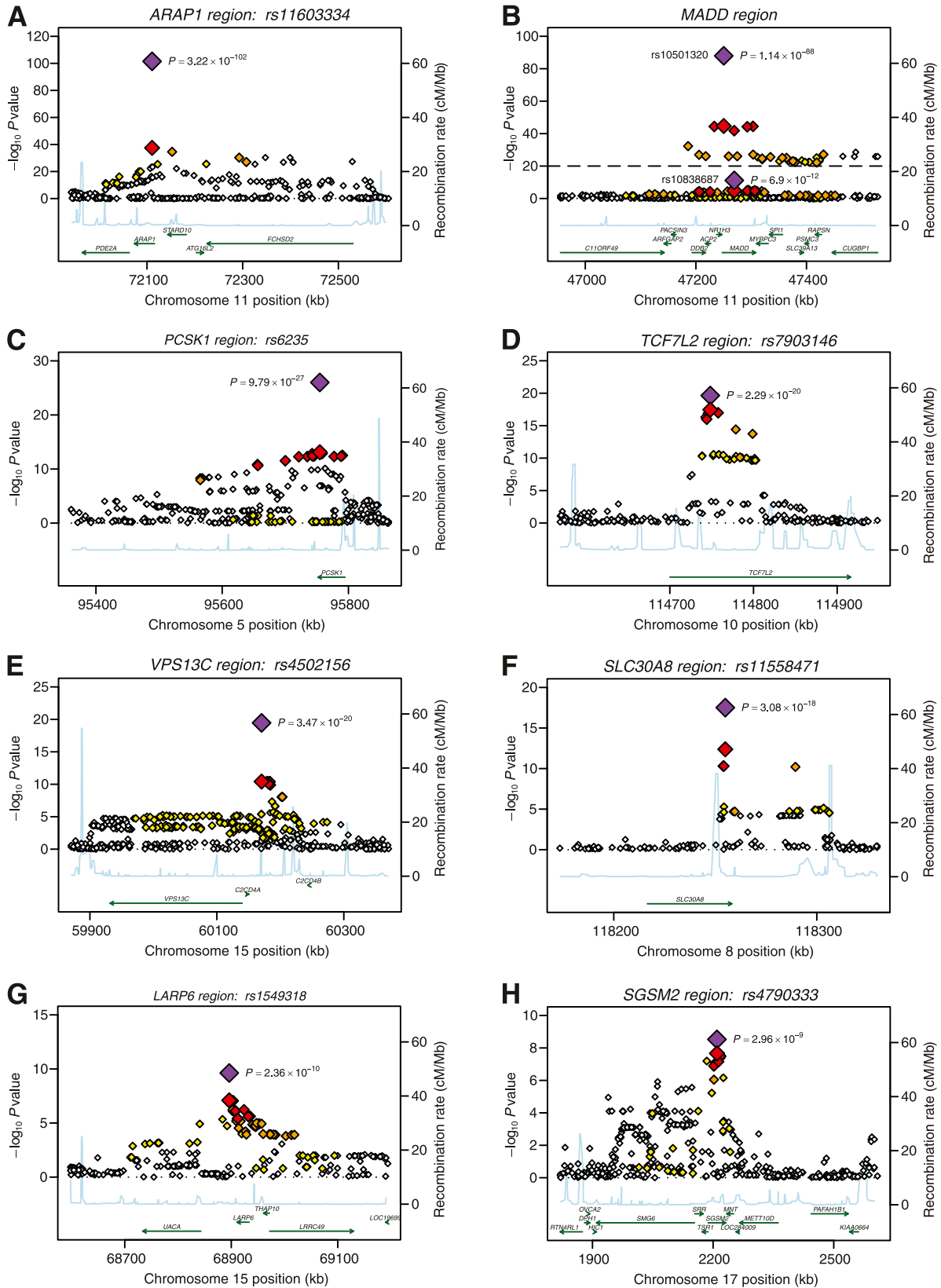
FIG. 2. Regional plots of eight genomic regions containing novel genome-wide significant associations. For each region, directly genotyped and imputed SNPs are plotted with their meta-analysis *P* values (as −log10 values) as a function of genomic position (NCBI Build 36). In each panel, the stage 1 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

SNP	Nearest gene	Alleles proinsulin-raising/other	32,33-split proinsulin (n = 4,103–6,343)		Fasting insulin (n = 36,775–38,335)		HOMA-B (n = 35,046–36,606)		Insulinogenic index (n = 10,902–14,956)		C-peptide (n = 1,537–2,956)	
			β	P value	β	P value	β	P value	β	P value	β	P value
rs11603334	ARAP1	A/G	0.146	1.2×10^{-25}	0.012	0.02	0.016	1.4×10^{-4}	0.049	1.1×10^{-5}	0.017	0.22
rs10501320	MADD	G/C	0.124	1.4×10^{-20}	−0.003	0.45	−0.015	8.1×10^{-5}	−0.005	0.67	-3.0×10^{-4}	0.98
rs6235	PCSK1	G/C	0.057	4.5×10^{-6}	0.001	0.78	0.008	0.02	−0.004	0.70	−0.008	0.46
rs7903146	TCF7L2	T/C	0.042	1.5×10^{-3}	−0.012	0.005	−0.020	1.4×10^{-7}	−0.063	9.6×10^{-7}	0.000	0.98
rs452156	VPS13C/C2CD4A/B	T/C	0.039	4.5×10^{-4}	−0.002	0.67	−0.010	0.004	−0.053	3.5×10^{-9}	−0.017	0.11
rs11558471	SLC30A8	A/G	0.059	1.1×10^{-5}	−0.005	0.21	−0.017	1.4×10^{-6}	−0.032	0.001	0.003	0.81
rs10838687	MADD (2nd signal)	T/G	0.006	0.71	0.006	0.22	0.007	0.06	0.0029	0.79	−0.0084	0.47
rs1549318	LARP6	T/C	0.043	1.2×10^{-4}	0.009	0.02	0.009	0.007	−0.001	0.88	−0.014	0.18
rs4790333	SGSM2	T/C	0.040	3.2×10^{-4}	0.008	0.04	0.002	0.55	0.009	0.33	−0.015	0.15

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

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 transcripts, in tissues available to us that had been paired to genetic data. We found a significant association (*P* = 0.01 permutation threshold) of rs1549318 with expression levels of *LARP6* in adipose tissue. SNP rs1549318 is located ~37 kb from *LARP6*, and the proinsulin-raising T allele is associated with lower levels of expression. Analysis of an eQTL database from human liver indicated that the proinsulin-raising A allele of the lead SNP at the *SGSM2* locus (rs179456) was associated with increased liver expression of *TRPS1* (*P* = 0.004).

CAD. We constructed unweighted and weighted genotype scores composed of the nine genome-wide significant proinsulin-raising alleles, with weights defined by the β-coefficients from our replication meta-analysis, and tested the association of these scores with CAD in the Coronary Artery Disease Genome-wide Replication And Meta-Analysis (CARDIoGRAM) (37) (*n* = ~22,000 CAD case subjects and 60,000 control subjects) and C4D (38) (*n* = 15,420 CAD case subjects and 15,062 control subjects) datasets. Neither weighted nor unweighted genotype scores reached nominal significance in either dataset (*P* = 0.47 and 0.81 for unweighted and weighted scores in CARDIoGRAM, respectively; *P* = 0.60 and 0.43 for unweighted and weighted scores in C4D, respectively).

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*, *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.

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
Association of proinsulin loci with other glycemic traits

SNP	Nearest gene	Alleles proinsulin-raising/other		Fasting glucose (n = 44,601–46,186)	HOMA-IR (n = 35,512–37,072)	Matsuda index (n = 7,055–9,561)	2-h glucose (n = 15,088–15,252)	2-h insulin (n = 6,923–7,083)	HbA1c (n = 33,736–44,731)	T2D (n = 8,130 case subjects/38,987 control subjects)	BMI (n = 123,260–123,865)
rs11603334	ARAP1	A/G	β/OR	−0.019	0.016	−0.014	0.009	−0.050	−0.023	0.88	0.012
			P value	1.7×10^{-4}	0.105	0.226	0.043	0.294	0.016	7.8×10^{-6}	0.065
rs10501320	MADD	G/C	β/OR	0.025	−0.015	−0.001	0.034	0.015	0.004	1.01	0.014
			P value	3.1×10^{-9}	0.009	0.605	0.109	0.459	0.595	0.001	0.013
rs6235	PCSK1	G/C	β/OR	−0.014	0.001	−0.008	0.015	0.015	0.013	1.02	0.016
			P value	4.8×10^{-4}	0.775	0.435	0.273	0.273	0.401	0.629	0.002
rs7903146	TCF7L2	T/C	β/OR	0.023	−0.010	0.014	0.108	0.004	0.015	1.40	−0.019
			P value	2.8×10^{-8}	0.837	0.244	4.3×10^{-7}	0.772	0.053	2.2×10^{-51}	2.4×10^{-4}
rs4502156	VPS13C/C2CD4A/B	T/C	β/OR	0.020	0.002	−0.004	−0.071	0.012	0.029	1.07	−0.003
			P value	8.4×10^{-8}	0.034	0.658	2.1×10^{-4}	0.321	0.002	2.4×10^{-4}	0.528
rs11558471	SLC30A8	A/G	β/OR	0.027	−0.017	0.014	−0.002	0.080	0.013	1.14	−0.012
			P value	2.6×10^{-11}	0.648	0.197	1.4×10^{-4}	0.275	0.779	3.3×10^{-8}	0.027
rs10838687	MADD (2nd signal)	T/G	β/OR	−0.011	0.004	−0.006	0.014	0.003	−0.001	0.98	−0.013
			P value	0.014	0.714	0.578	0.512	0.830	0.876	0.383	0.020
rs1549318	LARP6	T/C	β/OR	−0.002	0.009	0.001	0.008	−0.010	−0.005	1.01	0.003
			P value	0.698	0.044	0.906	0.621	0.556	0.444	0.667	0.583
rs4790333	SGSM2	T/C	β/OR	0.004	0.009	0.009	−0.0002	0.008	−0.005	1.01	−0.005
			P value	0.328	0.023	0.338	0.992	0.519	0.425	0.626	0.261

β-Coefficients are adjusted for age, sex, and study-specific covariates (if applicable). OR is for T2D only as a dichotomous trait.
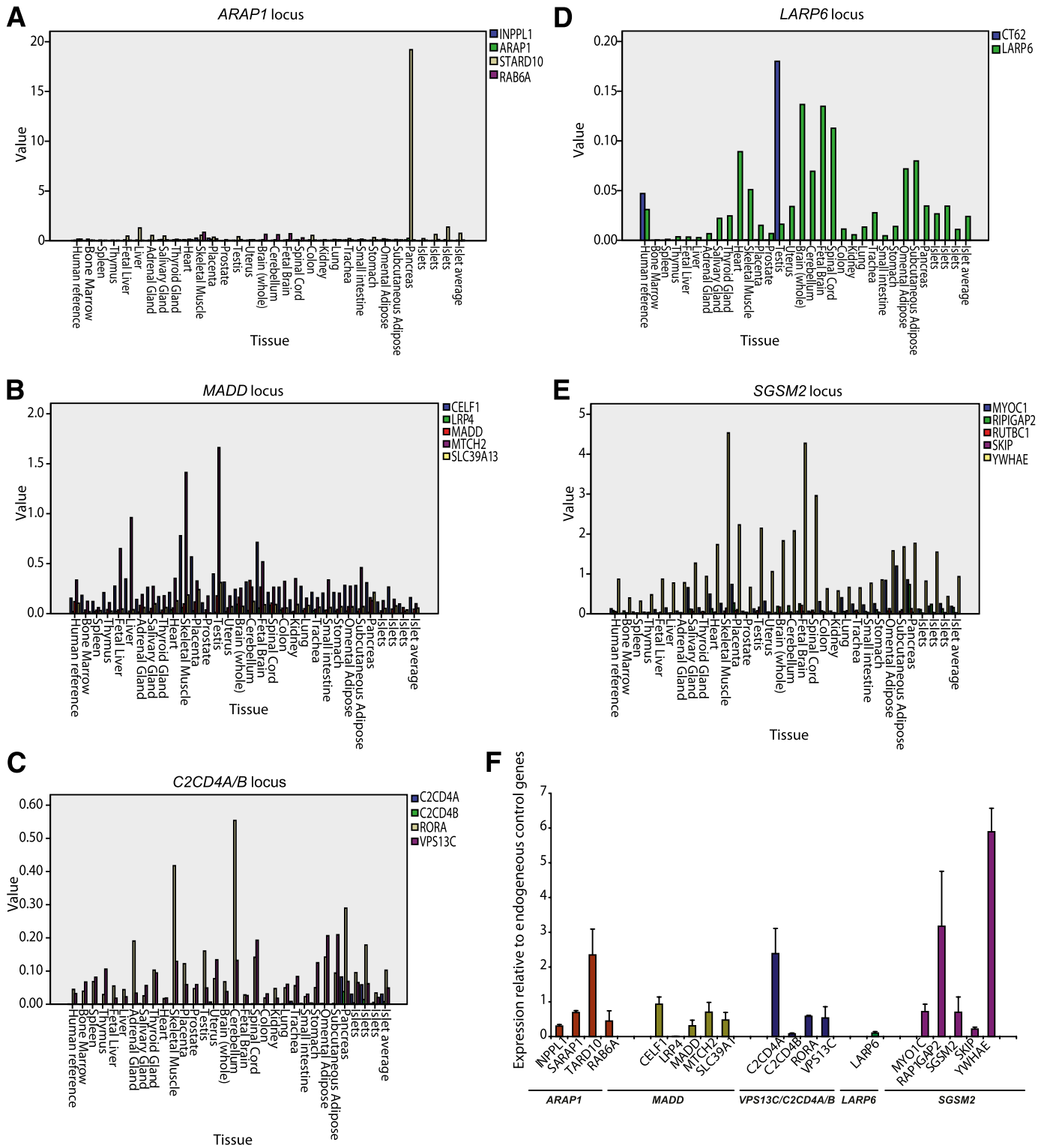
FIG. 3. Expression profiles of biologically plausible genes within each associated locus across a range of human tissue types, including islet preparations from three donors. Expression levels determined with respect to the geometric mean of three endogenous control assays. *A*: *ARAP1* region; *B*: *MADD* region; *C*: *VPS13C/C2CD4A/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.

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

1/3 (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 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.

REFERENCES

1. Billings LK, Florez JC. The genetics of type 2 diabetes: what have we learned from GWAS? Ann N Y Acad Sci 2010;1212:59–77
2. Prokopenko I, Langenberg C, Florez JC, et al. Variants in *MTNR1B* influence fasting glucose levels. Nat Genet 2009;41:77–81
3. Dupuis J, Langenberg C, Prokopenko I, et al.; DIAGRAM Consortium; GIANT Consortium; Global BPgen Consortium; A. Hamsten on behalf of Procardis Consortium; MAGIC investigators. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet 2010;42:105–116
4. Saxena R, Hivert MF, Langenberg C, et al.; GIANT consortium; MAGIC investigators. Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. Nat Genet 2010;42:142–148
5. Soranzo N, Sanna S, Wheeler E, et al.; WTCCC. Common variants at 10 genomic loci influence hemoglobin A₁ (C) levels via glycemic and non-glycemic pathways. Diabetes 2010;59:3229–3239
6. Ingelsson E, Langenberg C, Hivert MF, et al.; MAGIC investigators. Detailed physiologic characterization reveals diverse mechanisms for novel genetic loci regulating glucose and insulin metabolism in humans. Diabetes 2010;59:1266–1275
7. Rich SS, Goodarzi MO, Palmer ND, et al. A genome-wide association scan for acute insulin response to glucose in Hispanic-Americans: the Insulin Resistance Atherosclerosis Family Study (IRAS FS). Diabetologia 2009;52:1326–1333
8. Palmer ND, Langefeld CD, Ziegler JT, et al. Candidate loci for insulin sensitivity and disposition index from a genome-wide association analysis of Hispanic participants in the Insulin Resistance Atherosclerosis (IRAS) Family Study. Diabetologia 2010;53:281–289
9. Røder ME, Porte D Jr, Schwartz RS, Kahn SE. Disproportionately elevated proinsulin levels reflect the degree of impaired B cell secretory capacity in patients with noninsulin-dependent diabetes mellitus. J Clin Endocrinol Metab 1998;83:604–608
10. Wareham NJ, Byrne CD, Williams R, Day NE, Hales CN. Fasting proinsulin concentrations predict the development of type 2 diabetes. Diabetes Care 1999;22:262–270
11. Lindahl B, Dinesen B, Eliasson M, et al. High proinsulin concentration precedes acute myocardial infarction in a nondiabetic population. Metabolism 1999;48:1197–1202
12. Yudkin JS, May M, Elwood P, Yarnell JW, Greenwood R, Davey Smith G; aaerphilly Study. Concentrations of proinsulin like molecules predict coronary heart disease risk independently of insulin: prospective data from the Caerphilly Study. Diabetologia 2002;45:327–336
13. Zethelius B, Byberg L, Hales CN, Lithell H, Berne C. Proinsulin is an independent predictor of coronary heart disease: report from a 27-year follow-up study. Circulation 2002;105:2153–2158
14. Loos RJF, Franks PW, Francis RW, et al. TCF7L2 polymorphisms modulate proinsulin levels and beta-cell function in a British Europid population. Diabetes 2007;56:1943–1947
15. Kirchhoff K, Machicao F, Haupt A, et al. Polymorphisms in the *TCF7L2*, *CDKAL1* and *SLC30A8* genes are associated with impaired proinsulin conversion. Diabetologia 2008;51:597–601
16. González-Sánchez JL, Martínez-Larrad MT, Zabena C, Pérez-Barba M, Serrano-Ríos M. Association of variants of the *TCF7L2* gene with increases in the risk of type 2 diabetes and the proinsulin:insulin ratio in the Spanish population. Diabetologia 2008;51:1993–1997
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
18. Hanley AJ, D'Agostino R Jr, Wagenknecht LE, et al.; Insulin Resistance Atrherosclerosis Study. Increased proinsulin levels and decreased acute insulin response independently predict the incidence of type 2 diabetes in the insulin resistance atherosclerosis study. Diabetes 2002;51:1263–1270
19. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet 2007;39:906–913
20. StataCorp. *Stata Statistical Software: Release 10*. College Station, TX, StataCorp LP, 2007
21. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–575
22. R Development Core Team. *R: A language and environment for statistical computing*. Vienna, Austria, R Foundation for Statistical Computing, 2007

23. Devlin B, Roeder K. Genomic control for association studies. Biometrics 1999;55:997–1004

24. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190–2191

25. Mägi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. BMC Bioinformatics 2010;11:288

26. Magi R, Lindgren CM, Morris AP. Meta-analysis of sex-specific genome-wide association studies. Genet Epidemiol 2010;34:846–853

27. Seltzer HS, Allen EW, Herron AL Jr, Brennan MT. Insulin secretion in response to glycemic stimulus: relation of delayed initial release to carbohydrate intolerance in mild diabetes mellitus. J Clin Invest 1967;46: 323–335

28. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia 1985;28:412–419

29. Voight BF, Scott LJ, Steinthorsdottir V, et al.; MAGIC investigators; GIANT Consortium. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet 2010;42:579–589

30. Speliotes EK, Willer CJ, Berndt SI, et al.; MAGIC; Procardis Consortium. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet 2010;42:937–948

31. Matsuda M, DeFronzo RA. Insulin sensitivity indices obtained from oral glucose tolerance testing: comparison with the euglycemic insulin clamp. Diabetes Care 1999;22:1462–1470

32. Li Y, Abecasis GR. Mach 1.0: rapid haplotype reconstruction and missing genotype inference (Abstract). Am J Hum Genet 2006;S79:2290

33. Craddock N, Hurles ME, Cardin N, et al.; Wellcome Trust Case Control Consortium. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. Nature 2010;464: 713–720

34. Riva A, Kohane IS. A SNP-centric database for the investigation of the human genome. BMC Bioinformatics 2004;5:33

35. Sunyaev S, Ramensky V, Koch I, Lathe W 3rd, Kondrashov AS, Bork P. Prediction of deleterious human alleles. Hum Mol Genet 2001;10:591–597

36. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. Genome Res 2001;11:863–874

37. Schunkert H, König IR, Kathiresan S, et al.; Cardiogenics; CARDIoGRAM Consortium. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nat Genet 2011;43:333–338

38. Peden JF, Hopewell JC, Saleheen D, et al.; Coronary Artery Disease (C4D) Genetics Consortium. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. Nat Genet 2011;43:339–344

39. Dufau ML, Tsai-Morris C, Tang P, Khanum A. Regulation of steroidogenic enzymes and a novel testicular RNA helicase. J Steroid Biochem Mol Biol 2001;76:187–197

40. Heni M, Haupt A, Schäfer SA, et al. Association of obesity risk SNPs in PCSK1 with insulin sensitivity and proinsulin conversion. BMC Med Genet 2010;11:86

41. da Silva Xavier G, Loder MK, McDonald A, et al. TCF7L2 regulates late events in insulin secretion from pancreatic islet beta-cells. Diabetes 2009; 58:894–905

42. Nicolson TJ, Bellomo EA, Wijesekara N, et al. Insulin storage and glucose homeostasis in mice null for the granule zinc transporter ZnT8 and studies of the type 2 diabetes-associated variants. Diabetes 2009;58:2070–2083

43. Cai L, Fritz D, Stefanovic L, Stefanovic B. Binding of LARP6 to the conserved 5′ stem-loop regulates translation of mRNAs encoding type I collagen. J Mol Biol 2010;395:309–326

44. Yamauchi T, Hara K, Maeda S, et al. A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at UBE2E2 and C2CD4A-C2CD4B. Nat Genet 2010;42:864–868

45. Farooqi IS, Volders K, Stanhope R, et al. Hyperphagia and early-onset obesity due to a novel homozygous missense mutation in prohormone convertase 1/3. J Clin Endocrinol Metab 2007;92:3369–3373

46. Furuta M, Carroll R, Martin S, et al. Incomplete processing of proinsulin to insulin accompanied by elevation of Des-31,32 proinsulin intermediates in islets of mice lacking active PC2. J Biol Chem 1998;273:3431–3437

47. Benzinou M, Creemers JW, Choquet H, et al. Common nonsynonymous variants in PCSK1 confer risk of obesity. Nat Genet 2008;40:943–945