Journal of Genetics (2019) 98:8
https://doi.org/10.1007/s12041-018-1046-1

RESEARCH ARTICLE

Genomewide association study of C-peptide surfaces key regulatory genes in Indians

KHUSHDEEP BANDESH[1,2], GAURI PRASAD[1,2], ANIL KUMAR GIRI[1,2], V. SAROJA VORUGANTI[3], NANCY F. BUTTE[4], SHELLEY A. COLE[5], ANTHONY G. COMUZZIE[5], INDICO CONSORTIUM[†], NIKHIL TANDON[6]* and DWAIPAYAN BHARADWAJ[2,7]*

[1]*Genomics and Molecular Medicine Unit, CSIR-Institute of Genomics and Integrative Biology, New Delhi 110 020, India*
[2]*Academy of Scientific and Innovative Research, CSIR-Institute of Genomics and Integrative Biology Campus, New Delhi 110 020, India*
[3]*Department of Nutrition and UNC Nutrition Research Institute, University of North Carolina at Chapel Hill, Kannapolis, NC 28081, USA*
[4]*USDA/ARS Children's Nutrition Research Center, Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA*
[5]*Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX, USA*
[6]*Department of Endocrinology and Metabolism, All India Institute of Medical Sciences, New Delhi 110 029, India*
[7]*Systems Genomics Laboratory, School of Biotechnology, Jawaharlal Nehru University, New Delhi 110 067, India*
*For correspondence. E-mail: Nikhil Tandon, nikhil_tandon@hotmail.com; Dwaipayan Bharadwaj, db@jnu.ac.in.

Received 30 July 2018; accepted 6 September 2018; published online 13 February 2019

Abstract. Insulin is a commonly used measure of pancreatic β-cell function but exhibits a short half-life in the human body. During biosynthesis, insulin release is accompanied by C-peptide at an equimolar concentration which has a much higher plasma half-life and is therefore projected as a precise measure of β-cell activity than insulin. Despite this, genetic studies of metabolic traits have neglected the regulatory potential of C-peptide for therapeutic intervention of type-2 diabetes. The present study is aimed to search genomewide variants governing C-peptide levels in genetically diverse and high risk population for metabolic diseases—Indians. We performed whole genome genotyping in 877 healthy Indians of Indo-European origin followed by replication of variants with $P \leq 1 \times 10^{-3}$ in an independent sample-set of 1829 Indians. Lead-associated signals were also tested *in-silico* in 773 Hispanics. To secure biological rationale for observed association, we further carried out DNA methylation quantitative trait loci analysis in 233 Indians and publicly available regulatory data was mined. We discovered novel lncRNA gene *AC073333.8* with the strongest association with C-peptide levels in Indians that however missed genomewide significance. Also, noncoding genes, *RP1-209A6.1* and *RPS3AP5*; protein gene regulators, *ZNF831* and *ETS2*; and solute carrier protein gene *SLC15A5* retained robust association with C-peptide after meta-analysis. Integration of methylation data revealed *ETS2* and *ZNF831* single-nucleotide polymorphisms as significant meth-QTLs in Indians. All genes showed reasonable expression in the human lung, signifying alternate important organs for C-peptide biology. Our findings mirror polygenic nature of C-peptide where multiple small-effect size variants in the regulatory genome principally govern the trait biology.

Keywords. C-peptide; genetic variants; genomewide association study; Hispanics; Indians; meth-QTLs.

[†] Members of the INDICO consortium are listed in the supplementary data.

*KB researched the data and wrote the manuscript. KB and GP performed the analysis. AKG performed DNA methylation study. VSV, NFB, SAC and AGC provided summary statistics data for lead variants to perform meta-analysis in Hispanic population. VSV also analytically reviewed the manuscript. NT provided intellectual inputs. DB conceived, supervised the study and critically evaluated the manuscript.

Electronic supplementary material: The online version of this article (https://doi.org/10.1007/s12041-018-1046-1) contains supplementary material, which is available to authorized users.

Introduction

Decades of rigorous efforts in understanding type-2 diabetes (T2D) genetics have mirrored the disease as apparently an outcome of altered quantitative traits (Plomin *et al.* 2009). Dissection of heritability of such traits, especially in nondiseased individuals can reflect various intermediary disease mechanisms that are neglected otherwise. In this regard, C-peptide, a 31 amino-acid long cleavage product of insulin synthesis has been largely overlooked. Owing to a higher plasma half-life (~30 min) than insulin (~4 min), C-peptide levels are a precise and worthy measure of insulin secretion (Faber *et al.* 1978). Since its discovery, C-peptide has been considered as a biologically inert molecule. However today, several studies detail its independent functional activity. Comprehensive studies in animal models of diabetes and early clinical trials in type-1 diabetic patients have established that replacement of C-peptide in physiological concentrations is helpful in improving diabetes-induced functional and structural abnormalities of peripheral nerves, kidneys and brain (Wahren *et al.* 2002). C-peptide binds to specific G-protein coupled receptors on human cell membranes and enhances H_4K_{16} acetylation at promoters of ribosomal RNA genes upon cellular entry (Rigler *et al.* 1999; Lindahl *et al.* 2010). Importantly, C-peptide increases nitric oxide synthesis to relax vascular smooth muscles and restores nerve conduction velocity by enhancing Na^+/K^+ ATPase activity (Zhong *et al.* 2004). Besides, C-peptide stimulates several transcription factors that are essential for general cellular processes (Hills and Brunskill 2009). Specific cellular functions and benefits in preventing micro-vascular complications evidently signify that the C-peptide itself has a major regulatory role independent of T2D aetiology.

To date, only two genomewide association studies (GWAS) have been conducted for C-peptide in European, Hispanic, Afro-Caribbean and Thai populations (Comuzzie *et al.* 2012; Hayes *et al.* 2013) and a meta-GWAS in Europeans (Roshandel *et al.* 2018), contributing 11 unique single-nucleotide polymorphism (SNP) associations. However, none of these studies have included the Indian population. Among various linguistic groups in India, Indo-Europeans are a genetically diverse population that comprises several endogamous groups contributing to mixed ethnicity and high risk for T2D and related traits (Tabassum *et al.* 2011; Giri *et al.* 2016). A lack of C-peptide genetic studies in this inherently distinct and highly susceptible population guided us to search genomewide for associated variants to find out novel biology if any. Over the last decade, genetic discoveries for physiological traits have progressed with an extra-ordinary pace. However, the understanding of biology beneath such genetic associations lags well-behind. Translation from genetics to physiology is challenging in securing mechanistic inferences to pinpoint potential biomarkers. In such a scenario, integration of epigenetic and gene regulatory information can be fruitful in pinpointing underlying driving mechanisms. Variation in DNA methylation pattern affects gene transcription by altering binding of specific methyl CpG binding proteins. Gene regulatory sites are often seats for GWAS variants that influence gene expression in a tissue-specific manner. Thus, to identify under-pinning biological processes, we also studied DNA methylation signatures at identified genetic loci and probed publicly available human tissue gene expression data for C-peptide-associated genes. Therefore, in the current study we aimed to identify variants for C-peptide in healthy Indians that may affect the trait leading to plausible biological insights into functional mechanism. This study is a sole attempt for conducting GWAS for C-peptide in Indians.

Materials and methods

Ethical approval

This study was conducted in accordance with the principles of the Helsinki Declaration and was approved by the Ethics Committee of All India Institute of Medical Sciences and CSIR-Institute of Genomics and Integrative Biology. Informed written consent was obtained from all study participants.

Study population

The present GWAS was executed in two-stages, discovery phase and replication phase. The study participants are part of INdian DIabetes Consortium (INdian DIabetes Consortium 2011) and were collected from Diabetes Awareness Camps piloted in and around Delhi. The samples served as control individuals (normoglycemic) in T2D GWAS previously done in our laboratory (Tabassum *et al.* 2013). Sample details and criteria for selection have been outlined in a flow diagram (figure 1). The Indo-European ethnicity was defined by an individual's birth place and place of origin of the parents if residing in northern part of Indian subcontinent. Anthropometric and clinical characteristics of study participants are provided in (table 1 in electronic supplementary material at http://www.ias.ac.in/jgenet/). Plasma C-peptide levels were measured by electrochemiluminescence immunoassay using Elecsys 2010 (Roche Diagnostics). The Hispanic study population is a part of VIVA LA FAMILIA study recruited between 2000 and 2005 in Houston, Texas, USA and comprises 815 children and adolescents from 263 Hispanic families between the ages of 4 and 19 years (Butte *et al.* 2006; Comuzzie *et al.* 2012).

Discovery phase

Illumina Human 610-Quad BeadChips (Illumina, San Diego, USA) were used for the genomewide scan. The GenCall algorithm was employed to ascertain genotype

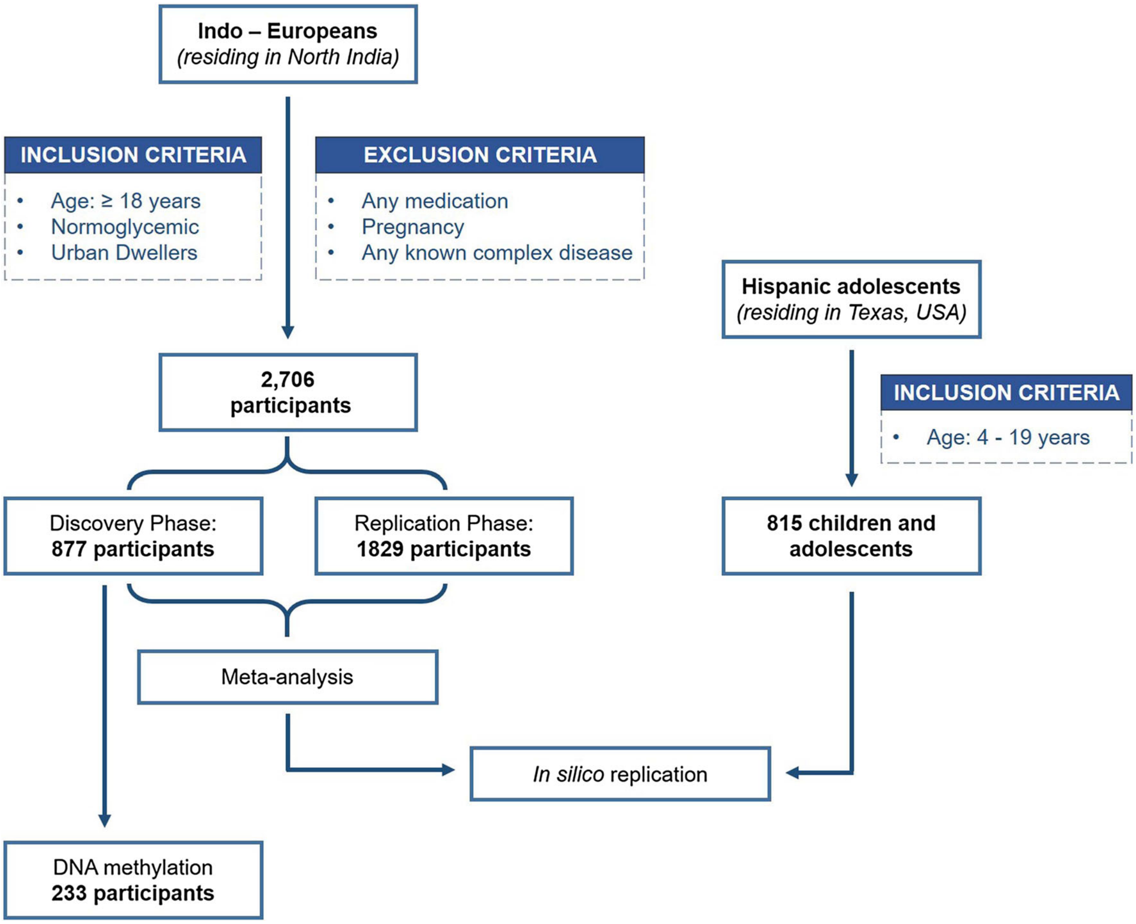

Figure 1. Flow diagram of study participants. Sample numbers for various phases of the study and their criteria for selection are outlined.

calls (GenomeStudio, Illumina). A stringent quality control (QC) check for samples and SNPs was executed. Samples with low genotype call rate ($<95\%$), heterozygosity and discordant sex were rejected. Identity-by-descent check was implemented to remove related or duplicate samples (pi_hat > 0.1875 for relatedness and pi_hat > 0.98 for duplication). SNP calls with missing rate $> 5\%$, MAF < 0.01 were excluded. SNPs with MAF 0.01–0.05 and Hardy Weinberg equilibrium (HWE) $P < 10^{-4}$ were removed. Also SNPs with MAF > 0.5 and HWE $P < 10^{-7}$ were discarded. Linkage disequilibrium (LD) pruning of SNPs was done with genotyped autosomal SNPs applying the –indep-pairwise option of PLINK with a window size of 50 SNPs, step size of 5 and $r^2 < 0.2$. Population stratification was checked through principal component analysis using HapMap Phase III population, CEU, CHB and YRI and the outliers were removed. Finally, a total of 552,011 autosomal SNPs were retained after stringent QC for association testing in 877 healthy normoglycemic individuals. C-peptide values were inverse normalized exercising an inbuilt R command (www.r-project.org). Association of QC-passed SNPs with C-peptide was tested by linear regression analysis under additive model adjusting for age, sex, body-mass index (BMI) and first three principal components using PLINK. Any deviation of observed P values from the P values expected under a null hypothesis was checked by constructing a quantile–quantile plot (QQ plot). The observed $-\log_{10}$ adjusted P values were plotted against expected $-\log_{10}$ P values. Median χ^2 statistics was used to compute genomic inflation factor λ using PLINK. QQ and Manhattan plots were constructed using R (www.r-project.org).

Validation phase and meta-analysis

Associated SNPs with discovery phase P value $<10^{-3}$ were selected for replication in an independent sample set using Illumina Golden Gate assay. SNPs with genotype call $<90\%$, genotype confidence score <0.25, GenTrans score <0.60, cluster separation score <0.40, MAF <0.01 and HWE $P < 1 \times 10^{-5}$ were removed. Also, samples with call rate $<90\%$ were excluded. Finally, a set of 1350 SNPs were tested for association in an independent set of 1829 healthy normoglycemic individuals. QC-passed SNPs were assessed for association with inverse normalized values of C-peptide through linear regression model adjusting for age, sex, BMI and first three principal components. An inverse variance method was used to meta-analyse discovery and replication phase results under a fixed effect model by METAL (www.sph.umich.edu/csg/abecasis/Metal).

In-silico replication in Hispanic population-VIVA LA FAMILIA study

For *in-silico* replication of identified signals, association data on C-peptide was requested from the VIVA LA FAMILIA Study which identifies genetic variants influencing paediatric obesity and its comorbidities in Hispanic population (Butte *et al.* 2006; Comuzzie *et al.* 2012). The data were obtained for 815 children and adolescents of 263 Hispanic families (Comuzzie *et al.* 2012). The C-peptide association in Hispanics was adjusted for age, sex and BMI. We meta-analysed by combining summary statistics of our lead signals of C-peptide GWAS following a sample size based analysis by METAL. An overall z-statistic and P value were calculated from a weighted sum of the individual statistics. Weights were taken proportional to the square-root of the number of individuals examined in each sample and selected such that the squared weights sum to 1.0. The z-statistic summarized the magnitude and the direction of effect relative to the reference allele and all studies are aligned to the same reference allele.

DNA methylation analysis

We studied DNA methylation signatures in peripheral blood using Infinium HumanMethylation450 K BeadChips of 233 normoglycemic Indians considered in discovery phase of GWAS. Data generation has been described previously (Giri *et al.* 2017). Sample QC involved extreme missingness, sex disparity checks and samples with failed bisulphite conversion (samples having intensity three SD away from mean intensity for C1, C2, C3 and C4 probes). CpGs with bead count <3 in 5% of samples and detection P value >0.01 for $<1\%$ of samples were excluded. CpGs in sex chromosomes (X and Y), established cross-hybridization probes and polymorphic CpGs were also removed. CpGs with 100% call rate in all the samples have only been considered for analysis. Methylation outlier value for CpGs was fixed by fixMeth-Outliers command in minifi. Methylation data were regressed for confounders such as cell composition, age, sex, BMI, bisulphite conversion efficiency and plate number. Methylation data were extracted for identified genetic variants and tested for SNP-CpG association using linear regression model in PLINK.

Power calculation

We calculated the power of study using Quanto software (http://biostats.usc.edu/Quanto.html) under an additive genetic model for a range of allele frequencies (0.01–0.5). Two-tailed test at significance level of 0.05 with effect size ranging from 0.0001 to 1.514 was used for calculating power. For final power analysis, MAF for combined samples and calculated effect size from meta-analysis was used (figure 1 in electronic supplementary material).

Results

Under the null distribution, QQ plots indicated a good agreement (genomic inflation factor $\lambda = 1$, mean $\chi^2 = 0.97$) (figure 2 in electronic supplementary material). In discovery phase, the strongest association was observed for *HDAC4* variant, rs1796447 ($P = 2.85 \times 10^{-6}$) (figure 2). This association receded in meta-analysis ($P = 0.002$). Two other variants residing on chromosome 20 presented zinc-finger protein 831 (*ZNF831*) as a second lead signal in discovery phase (rs3026620, $P = 4.17 \times 10^{-6}$; rs6128532, $P = 3.49 \times 10^{-5}$). The association with *ZNF831* was sustained by two variants, rs6128525 and rs2064865 in meta-analysis ($P = 7.51 \times 10^{-5}$ and $P = 6.11 \times 10^{-4}$ respectively). Besides, a region *FHIT* marked by three variants on chromosome 3, featured strong association in discovery phase but lost when meta-analysed (rs17684830, $P = 7.32 \times 10^{-6}$; rs6414610, $P = 1.54 \times 10^{-5}$; rs7373003, $P = 5.52 \times 10^{-5}$ in discovery phase). Also, *GABRA6* locus on chromosome 5 followed a similar trend for association to C-peptide levels with significant discovery phase P values of 8.25×10^{-6} and 2.31×10^{-5} for SNPs rs4454083 and rs1444740, respectively. In meta-analysis, *GABRA6* displayed nominal significance (rs4454083, $P = 0.007$). A region on chromosome 10 mapping to nonprotein coding *RPS3AP5* locus was observed as another robustly associated signal in discovery phase (rs11201130, $P = 7.6 \times 10^{-6}$; rs12783251, $P = 2.21 \times 10^{-5}$). This association remained significant in meta-analysis with rs11201130 ($P = 5.38 \times 10^{-4}$).

In meta-analysis combining discovery and replication phases, the strongest association was observed at a novel locus, long noncoding RNA (lncRNA) gene, *AC073333.8* on chromosome 7 (two variants rs1674809 and rs818488) (table 1). Another novel lncRNA gene, *RP1-209A6.1* surfaced as lead signal after meta-analysing discovery and replication phases. Besides, two other variants, an intergenic variant rs10138141 in *ETS2* and an exonic variant in *SLC15A5* featured suggestive association (table 1).

An *in silico* replication and meta-analysis of summary statistics of three lead signals (rs1013841, rs9460790 and rs1527014) in Hispanic children and adolescents did not improve the association status (table 2). All three variants retained nominal association with C-peptide after meta-analysis ($P < 0.01$), however the directionality of association of variants rs9460790 and rs1527014 was observed to be reversed in Hispanics. Heterogeneity was observed to be more than 70% across populations for all three variants

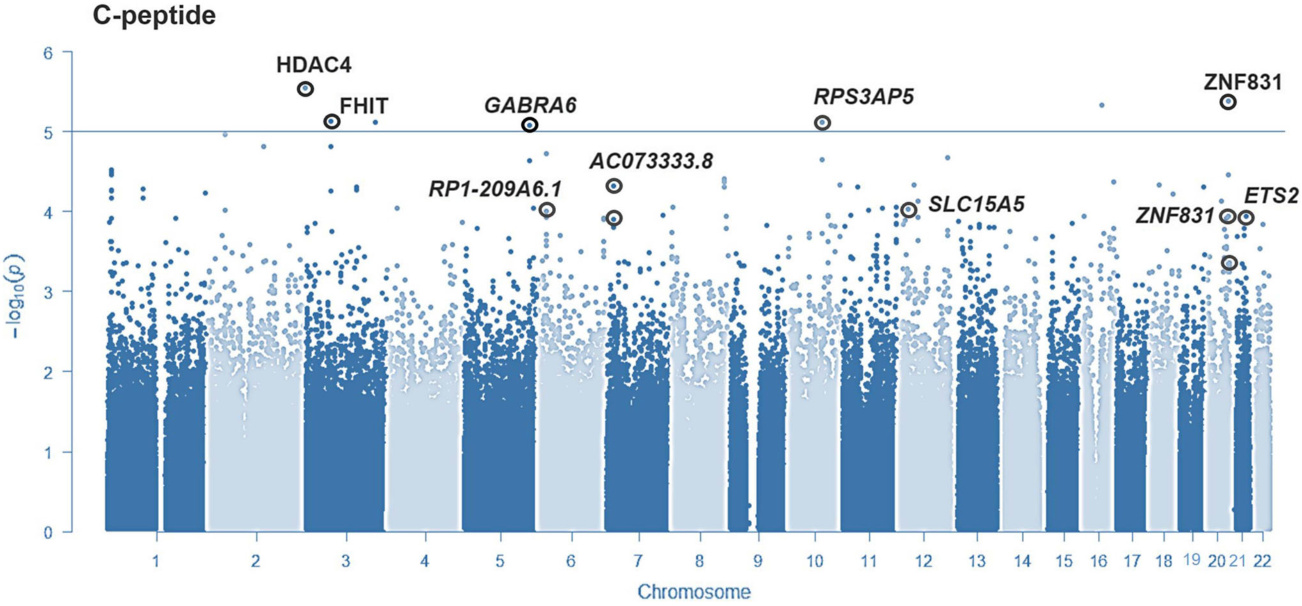

Figure 2. Manhattan plot of associated P values for C-peptide in Indians. Lead variants showing the strongest association after meta-analysis are shown in italics. The $-\log_{10}$ P values for association of directly genotyped SNPs are plotted as a function of genomic position (National Center for Biotechnology Information Build 37). P values were determined using linear regression adjusted for age, sex, BMI, PC1, PC2 and PC3 in discovery phase analysis.

(table 2). Also, we evaluated the status of previously known GWAS variants in our study. For C-peptide, we did not observe any association for reported GWAS locus, *GCKR* (table 2 in electronic supplementary material). To attest functional relevance of identified variants, we studied and integrated whole genome DNA methylation data from peripheral blood in Indians. We found that certain genetic variants within *ETS2* and *ZNF831* loci were strongly associated with variable DNA methylation pattern at multiple genic CpG sites in Indians (table 3). Further, while investigating the publicly available human tissue gene expression data (The GTEx Consortium 2015), we noticed that all associated genes (except for *RP1-209A6.1*) were expressed in human adipose tissue (figure 3 in electronic supplementary material). Moreover, we observed that all the six genes showed considerable expression in lungs, in particular.

Discussion

The present study attempted to identify genomewide common variants that govern plasma C-peptide in Indians. Our results highlight polygenic nature of C-peptide inheritance. It is well-established that the genetic variance in complex diseases/traits is mainly contributed by multiple SNPs with smaller effects that may often be missed out due to stringent GWAS P value thresholds and multiple testing corrections (Yang *et al.* 2010; Fransen *et al.* 2015). In view of a limited sample size ($n = 2706$), we identified multiple such variants that fail to attain GWAS significance but are associated nominally with C-peptide in Indians. Trait-associated GWAS SNPs are highly enriched in gene regulatory regions (Nica *et al.* 2010). To understand biological rationale for our observed associations, we studied DNA methylation marks and explored publicly available gene regulatory data (gene expression, histone active marks, repressive marks, DNase I HS sites, transcription factor binding sites etc.). Associated loci generally harbour several noncoding variants whose signal spans several genes. Multiple lines of evidence indicate noncoding RNA transcription as a proxy for functional activity of such putative genomic regions. Seemingly, in the present study, we obtained lead-associated variants in noncoding RNA genes, *AC073333.8, RP1-209A6.1* and *RPS3AP5*. *AC073333.8*, represented by two variants, surfaced as the strongest signal for association to C-peptide in Indians. Both variants reside in a highly active chromatin region (DNase I hypersensitive) of a novel long noncoding RNA, *AC073333.8* that is exonic antisense to a basic leucine zipper and W2 domain 2 protein, BZW2. *AC073333.8* lncRNA is plausibly a regulator of *BZW2* gene that needs to be validated experimentally. Previously, *BZW2* has been documented for association to Alzheimer's disease (Sherva *et al.* 2013) but never with plasma C-peptide levels. *BZW2* has been earlier reported to get significantly upregulated upon administration of insulin analogue in murine colon cancer model (Hvid *et al.* 2012). C-peptide is well-known to regulate the availability of monomeric insulin in physiological processes (Jörnvall

Table 1. SNPs showing association with C-peptide levels at meta-analysis P value $<10^{-3}$.

						Discovery phase			Replication phase			Meta-analysis				
SNP	CHR	Alleles (effect/other)	MAF	Nearby gene	SNP location	n	P value	Effect size	n	P value	Effect size	P value	Effect size	Dir	Het I Sq	Het-P value
C-peptide																
rs1674809	7	A/G	0.441	*AC073333.8*	Intronic	877	1.25×10^{-4}	0.189	1825	0.009	0.112	6.44×10^{-6}	0.145	++	29.1	0.235
rs818488	7	A/G	0.418	*AC073333.8*	Intronic	877	4.81×10^{-5}	0.202	1829	0.018	0.102	7.93×10^{-6}	0.145	++	57.4	0.126
rs9460790	6	A/C	0.419	*RP1-209A6.1*	Intronic	877	9.95×10^{-5}	−0.188	1828	0.059	−0.081	6.07×10^{-5}	−0.128	−−	63.9	0.096
rs6128525	20	A/G	0.428	*ZNF831*	Intergenic	876	1.16×10^{-4}	0.184	1740	0.07	0.08	7.51×10^{-5}	−0.128	−−	61.4	0.107
rs11201130	10	A/G	0.04	*RPS3AP5*	Intergenic	876	7.60×10^{-6}	0.562	1829	0.656	0.054	5.38×10^{-4}	−0.301	−−	88.2	0.004
rs1013841	21	A/G	0.129	*ETS2*	Intergenic	877	1.15×10^{-4}	0.268	1829	0.281	0.071	5.56×10^{-4}	−0.165	−−	76.5	0.039
rs2064865	20	A/G	0.459	*ZNF831*	3′-UTR	877	5.65×10^{-4}	0.163	1826	0.14	0.064	6.11×10^{-4}	0.109	++	58.6	0.12
rs1527014	12	A/G	0.018	*SLC15A5*	Exonic	877	1.45×10^{-4}	−0.628	1825	0.375	−0.156	6.84×10^{-4}	0.408	++	73.8	0.051

Association analysis with plasma C-peptide levels, adjusted for age, sex, BMI and three principal components. SNP location is according to the position with respect to the gene. Effect size has been calculated with respect to the minor allele. Meta-analysis was done using METAL using fixed effect inverse variance method. CHR, chromosome; n, sample number; Dir, direction; Het-P, P value for heterogeneity in effect sizes in meta-analysis; Het-I Sq, χ^2 value for heterogeneity test. Direction ++/−− features a concordance between the discovery and replication phase.

Table 2. *In silico* replication and meta-analysis of novel SNPs rs1013841, rs9460790 and rs1527014 in Hispanic population.

				Meta-analysis Indo-Europeans					Indo-Europeans and Hispanics				
SNP	Effect allele	Other allele	Nearby gene	P value	Effect size	Dir	Het I Sq	Het-P value	Z score	P value	Dir	Het I Sq	Het-P value
rs1013841	A	G	*ETS2*	5.56×10^{-4}	−0.165	−−	76.5	0.039	−3.077	0.0021	−−−	72	0.028
rs9460790	A	C	*RP1-209A6.1*	6.07×10^{-5}	−0.128	−−	63.9	0.096	−3.109	0.0019	−−+	78.4	0.009
rs1527014	A	G	*SLC15A5*	6.84×10^{-4}	0.408	++	73.8	0.051	2.546	0.011	++−	77.1	0.013

Association data in Hispanics obtained from previously conducted C-peptide GWAS in Hispanic children and adolescents. Meta-analysis was done using METAL using sample size based analysis. Dir, direction; Het-P, P value for heterogeneity in effect sizes in meta-analysis; Het-I Sq, χ^2 value for heterogeneity test; Z score, summarizes the magnitude and the direction of effect relative to the reference allele and all studies are aligned to the same reference allele. Negative Z scores indicate allele associated with lower C-peptide levels, whereas positive Z score indicates allele associated with higher C-peptide levels. Direction: −−−features a concordance between association in Indo-Europeans and Hispanics whereas −−+/++− indicates a discordance.

Table 3. Meth-QTL analysis for novel variants in 233 Indians who have been genotyped in discovery phase.

SNP					CpG							
Name	CHR	BP	Gene	A1	Name	CHR	BP	Gene	CpG location	β	SE	*P* value
rs1013841	21	40173345	*ETS2*	G	cg02119577	21	40195074	*ETS2*	3′-UTR	−0.082	0.007	6.9×10^{-27}
				G	cg01359822		40176598		5′-Upstream	0.068	0.015	8.38×10^{-6}
				G	cg21880624		40194679		Exonic	0.001	0.001	0.047
rs2064865	20	57833740	*ZNF831*	A	cg17286326	20	57796492	*ZNF831*	Intronic	−0.015	0.003	2.07×10^{-7}
				A	cg16473762		57797320			0.017	0.004	1.58×10^{-5}
				A	cg04068564		57795246			−0.005	0.001	3.23×10^{-5}
				A	cg09642670		57797378			0.006	0.002	0.012
rs6128525	20	57846922	*ZNF831*	G	cg17286326	20	57796492	*ZNF831*	Intronic	−0.014	0.003	4.63×10^{-6}
				G	cg04068564		57795246			−0.005	0.001	2.38×10^{-5}
				G	cg16473762		57797320			0.017	0.004	2.67×10^{-5}
				G	cg09642670		57797378			0.006	0.002	0.011
				G	cg22823230		57798303			0.008	0.004	0.026

Association results showing SNPs affecting CpGs of associated gene. *P* value has been obtained from association of SNPs with methylation level as corresponding CpGs (β value) using PLINK. CpG IDs have been given based on annotation file of Illumina 450 K BeadChip. CHR, chromosome; BP, base position; SE, standard error; A1, minor allele.

et al. 2010). Thus, during hyperinsulinaemia, variation in C-peptide levels may tend to alter *AC073333.8* activity and hence *BZW2* function.

Additionally, two other noncoding genes, *RP1-209A6.1* and *RPS3AP5* cropped up for the first time in association with C-peptide. Earlier, no association of any disease/trait has been observed for lncRNA *RP1-209A6.1*. However, *RPS3AP5*, a ribosomal protein pseudogene has been originally reported to be a GWAS locus for childhood obesity in Hispanic population (Comuzzie *et al.* 2012).

Besides noncoding genes, other protein gene regulators, *ZNF831* and *ETS2*, also appeared as strong genetic variants and simultaneously as robust meth-QTLs in our study. The two variants that mark *ZNF831* reside in a polycomb-repressed chromatin region. *ZNF831* has previously been associated with hypertension and blood pressure at genomewide significance (International Consortium for Blood Pressure Genome-Wide Association Studies *et al.* 2001; Levy *et al.* 2009). Additionally, C-peptide has been well-elucidated to be significantly associated with hypertension in various populations (Chen *et al.* 1995; Purohit and Mathur 2013). Moreover, *ETS2*, a transcription factor that regulates developmental and apoptotic genes, has been formerly associated to genomewide significance with BMI (Locke *et al.* 2015). Higher BMI results in drastically higher plasma C-peptide levels (Chan *et al.* 2004). Also, reports have demonstrated that C-peptide upon cellular entry stimulates the activation of key transcription factors (Wahren *et al.* 2002). Thus, these studies reasonably justify the observed associations in our study. Further, we obtained an exonic nonsynonymous variant in a key solute carrier family protein, *SLC15A5* associated with C-peptide in Indians. *SLC15* is a family of membrane transporters known for key role in cellular uptake of peptides from dietary protein digestion, ultrafiltration or brain–blood efflux (Alexander *et al.* 2015). Prior studies have documented genomewide association of *SLC15A5* with visceral fat (Fox *et al.* 2012). Incidentally, C-peptide concentrations are seen to predict and increase visceral fat adiposity in nondiabetic individuals (Seidell *et al.* 1990; Tong *et al.* 2005). Hence, the presence of alternate alleles of *SLC15A5*-associated variant and modulation of C-peptide levels may be accountable for visceral adiposity in Indians to some extent.

Interestingly, we observed that all the six genes are moderately expressed in human lung tissue. C-peptide has essentially been demonstrated to alter the balance between pro-inflammatory and anti-inflammatory signalling in lungs to ameliorate inflammatory response after a haemorrhagic shock in male Wistar rats (Chima *et al.* 2011). The findings are indicative of lungs being alternatively important organs for C-peptide physiology other than pancreas and adipose tissue.

To summarize, the present study endeavoured to explore the heritability of C-peptide in Indians, a highly susceptible population for metabolic disorders. The GWAS figured several variants robustly associated with C-peptide that fail to be genomewide significant. The observations fortify the nature of polygenic traits. The study is unique to investigate and secure biological basis to observed genetic associations for varied C-peptide levels in Indians.

Acknowledgements

The authors are grateful to all the study participants. We acknowledge the support and participation of members of the INDICO consortium in data generation. We also thank VIVA LA FAMILIA study for providing the summary statistics for meta-analysis. KB acknowledges CSIR for Senior Research Fellowship (CSIR-SRF). GP acknowledges UGC for Senior Research Fellowship. This work was supported by the Council of Scientific and Industrial Research (CSIR), Government of India through Centre for Cardiovascular and Metabolic Disease Research (CARDIOMED) project (grant no. BSC0122-(14)). This work was also funded by the Department of Science and Technology-PURSE- II (DST/SR/PURSE PhaseII/11).

References

Alexander S. P. H., Kelly E., Marrion N., Peters J. A., Benson H. E., Faccenda E. *et al*. 2015 The concise guide to pharmacology 2015/16: transporters. *Br. J. Pharmacol*. **172**, 6110–6202.

Butte N. F., Cai G., Cole S. A. and Comuzzie A. G. 2006 VIVA LA FAMILIA Study: genetic and environmental contributions to childhood obesity and its comorbidities in the Hispanic population. *Am. J. Clin. Nutr*. **84**, 646–654.

Chan W. B., Tong P. C., Chow C. C., So W. Y., Ng M. C., Ma R. C. *et al*. 2004 The associations of body mass index, C-peptide and metabolic status in Chinese type 2 diabetic patients. *Diabet. Med*. **21**, 349–353.

Chen C. H., Tsai S. T., Chuang J. H., Chang M. S., Wang S. P. and Chou P. 1995 Population-based study of insulin, C-peptide, and blood pressure in Chinese with normal glucose tolerance. *Am. J. Cardiol*. **76**, 585–588.

Chima R. S., LaMontagne T., Piraino G., Hake P. W., Denenberg A. and Zingarelli B. 2011 C-peptide, a novel inhibitor of lung inflammation following hemorrhagic shock. *Am. J. Physiol.: Lung Cell. Mol. Physiol*. **300**, L730–L739.

Comuzzie A. G., Cole S. A., Laston S. L., Voruganti V. S., Haack K., Gibbs R. A. *et al*. 2012 Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. *PLoS One* **7**, e51954.

Faber O. K., Hagen C., Binder C., Markussen J., Naithani V. K., Blix P. M. *et al*. 1978 Kinetics of human connecting peptide in normal and diabetic subjects. *J. Clin. Investig*. **62**, 197–203.

Fox C. S., Liu Y., White C. C., Feitosa M., Smith A. V., Heard-Costa N. *et al*. 2012 Genome-wide association for abdominal subcutaneous and visceral adipose reveals a novel locus for visceral fat in women. *PLoS Genet*. **8**, e1002695.

Fransen E., Bonneux S., Corneveaux J. J., Schrauwen I., Di-Berardino F., White C. H. *et al*. 2015 Genome-wide association analysis demonstrates the highly polygenic character of age-related hearing impairment. *Eur. J. Hum. Genet*. **23**, 110–115.

Giri A. K., Banerjee P., Chakraborty S., Kauser Y., Undru A., Roy S. *et al*. 2016 Genome-wide association study of uric acid in Indian population and interaction of identified variants with type 2 diabetes. *Sci. Rep*. **6**, 21440.

Giri A. K., Bharadwaj S., Banerjee P., Chakraborty S., Parekatt V., Rajashekar D. *et al*. 2017 DNA methylation profiling reveals the presence of population-specific signatures correlating with phenotypic characteristics. *Mol. Genet. Genomics* **292**, 655–662.

Hayes M. G., Urbanek M., Hivert M. F., Armstrong L. L., Morrison J., Guo C. *et al*. 2013 Identification of HKDC1 and BACE2 as genes influencing glycemic traits during pregnancy through genome-wide association studies. *Diabetes* **62**, 3282–3291.

Hills C. E. and Brunskill N. J. 2009 C-peptide and its intracellular signaling. *Rev. Diabet. Stud*. **6**, 138–147.

Hvid H., Fendt S. M., Blouin M. J., Birman E., Voisin G., Svendsen A. M. *et al*. 2012 Stimulation of MC38 tumor growth by insulin analog X10 involves the serine synthesis pathway. *Endocr.-Relat. Cancer* **19**, 557–574.

INdian DIabetes COnsortium 2011 INDICO: the development of a resource for epigenomic study of Indians undergoing socioeconomic transition. *HUGO J*. **5**, 65–69.

International Consortium for Blood Pressure Genome-Wide Association Studies, Ehret G. B., Munroe P. B., Rice K. M., Bochud M., Johnson A. D. *et al*. 2001 Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103–109.

Jörnvall H., Lindahl E., Astorga-Wells J., Lind J., Holmlund A., Melles E. *et al*. 2010 Oligomerization and insulin interactions of proinsulin C-peptide: threefold relationships to properties of insulin. *Biochem. Biophys. Res. Commun*. **391**, 1561–1566.

Levy D., Ehret G. B., Rice K., Verwoert G. C., Launer L. J., Dehghan A. *et al*. 2009 Genome-wide association study of blood pressure and hypertension. *Nat. Genet*. **41**, 677–687.

Lindahl E., Nyman U., Zaman F., Palmberg C., Cascante A., Shafqat J. *et al*. 2010 Proinsulin C-peptide regulates ribosomal RNA expression. *J. Biol. Chem*. **285**, 3462–3469.

Locke A. E., Kahali B., Berndt S. I., Justice A. E., Pers T. H., Day F. R. *et al*. 2015 Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206.

Nica A. C., Montgomery S. B., Dimas A. S., Stranger B. E., Beazley C., Barroso I. *et al*. 2010 Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet*. **6**, e1000895.

Plomin R., Haworth C. M. A. and Davis O. S. P. 2009 Common disorders are quantitative traits. *Nat. Rev. Genet*. **10**, 872–878.

Purohit P. and Mathur R. 2013 Hypertension association with serum lipoproteins, insulin, insulin resistance and C-peptide: unexplored forte of cardiovascular risk in hypothyroidism. *N. Am. J. Med. Sci*. **5**, 195–201.

Rigler R., Pramanik A., Jonasson P., Kratz G., Jansson O. T., Nygren P.-Å. *et al*. 1999 Specific binding of proinsulin C-peptide to human cell membranes. *Proc. Natl. Acad. Sci*. **96**, 13318–13323.

Roshandel D., Gubitosi-Klug R., Bull S. B., Canty A. J., Pezzolesi M. G., King G. L. *et al*. 2018 Meta-genome-wide association studies identify a locus on chromosome 1 and multiple variants in the MHC region for serum C-peptide in type 1 diabetes. *Diabetologia* **61**, 1098–1111.

Seidell J. C., Björntorp P., Sjöström L., Kvist H. and Sannerstedt R. 1990 Visceral fat accumulation in men is positively associated with insulin, glucose, and C-peptide levels, but negatively with testosterone levels. *Metabolism* **39**, 897–901.

Sherva R., Tripodis Y., Bennett D. A., Chibnik L. B., Crane P. K., de-Jager P. L. *et al*. 2013 Genome-wide association study of the rate of cognitive decline in Alzheimer's disease. *Alzheimer's Dementia* **10**, 45–52.

Tabassum R., Mahajan A., Chauhan G., Dwivedi O. P., Dubey H., Sharma V. *et al*. 2011 No association of TNFRSF1B variants with type 2 diabetes in Indians of Indo-European origin. *BMC Med. Genet*. **12**, 110.

Tabassum R., Chauhan G., Dwivedi O. P., Mahajan A., Jaiswal A., Kaur I. *et al*. 2013 Genome-wide association study for type 2 diabetes in Indians identifies a new susceptibility locus at 2q21. *Diabetes* **62**, 977–986.

The GTEx Consortium 2015 The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660.

Tong J., Fujimoto W. Y., Kahn S. E., Weigle D. S., McNeely M. J., Leonetti D. L. *et al*. 2005 Insulin, C-peptide, and leptin concentrations predict increased visceral adiposity at 5- and 10-year follow-ups in nondiabetic Japanese Americans. *Diabetes* **54**, 985–990.

Wahren J., Kallas A. and Sima A. A. F. 2002 The clinical potential of C-peptide replacement in type 1 diabetes. *Diabetes* **61**, 761–772.

Yang J., Benyamin B., McEvoy B. P., Gordon S., Henders A. K., Nyholt D. R. *et al*. 2010 Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet*. **42**, 565–569.

Zhong Z., Kotova O., Davidescu A., Ehren I., Ekberg K., Jornvall H. *et al*. 2004 C-peptide stimulates Na$^+$, K$^+$-ATPase via activation of ERK1/2 MAP kinases in human renal tubular cells. *Cell. Mol. Life Sci*. **61**, 2782–2790.

Corresponding editor: Indrajit Nanda