Molecular Genetics and Genomics (2020) 295:1013–1026
https://doi.org/10.1007/s00438-020-01678-6

ORIGINAL ARTICLE

Multifaceted genome-wide study identifies novel regulatory loci in *SLC22A11* and *ZNF45* for body mass index in Indians

Anil K. Giri[1,2] · Gauri Prasad[1,2] · Khushdeep Bandesh[1,2] · Vaisak Parekatt[1] · Anubha Mahajan[3] · Priyanka Banerjee[1] · Yasmeen Kauser[1,2] · Shraddha Chakraborty[1,2] · Donaka Rajashekar[1] · INDICO · Abhay Sharma[1,2] · Sandeep Kumar Mathur[4] · Analabha Basu[5] · Mark I. McCarthy[3] · Nikhil Tandon[6] · Dwaipayan Bharadwaj[2,7]

Received: 19 October 2019 / Accepted: 11 April 2020 / Published online: 4 May 2020

Abstract

Obesity, a risk factor for multiple diseases (e.g. diabetes, hypertension, cancers) originates through complex interactions between genes and prevailing environment (food habit and lifestyle) that varies across populations. Indians exhibit a unique obesity phenotype with high abdominal adiposity for a given body weight compared to matched white populations suggesting presence of population-specific genetic and environmental factors influencing obesity. However, Indian population-specific genetic contributors for obesity have not been explored yet. Therefore, to identify potential genetic contributors, we performed a two-staged genome-wide association study (GWAS) for body mass index (BMI), a common measure to evaluate obesity in 5973 Indian adults and the lead findings were further replicated in 1286 Indian adolescents. Our study revealed novel association of variants—rs6913677 in *BAI3* gene ($p = 1.08 \times 10^{-8}$) and rs2078267 in *SLC22A11* gene ($p = 4.62 \times 10^{-8}$) at GWAS significance, and of rs8100011 in *ZNF45* gene ($p = 1.04 \times 10^{-7}$) with near GWAS significance. As genetic loci may dictate the phenotype through modulation of epigenetic processes, we overlapped genetic data of identified signals with their DNA methylation patterns in 236 Indian individuals and performed methylation quantitative trait loci (meth-QTL) analysis. Further, functional roles of discovered variants and underlying genes were speculated using publicly available gene regulatory databases (ENCODE, JASPAR, GeneHancer, GTEx). The identified variants in *BAI3* and *SLC22A11* genes were found to dictate methylation patterns at unique CpGs harboring critical *cis*-regulatory elements. Further, *BAI3, SLC22A11* and *ZNF45* variants were located in repressive chromatin, active enhancer, and active chromatin regions, respectively, in human subcutaneous adipose tissue in ENCODE database. Additionally, these genomic regions represented potential binding sites for key transcription factors implicated in obesity and/or metabolic disorders. Interestingly, GTEx portal identify rs8100011 as a robust *cis*-expression quantitative trait locus (*cis*-eQTL) in subcutaneous adipose tissue ($p = 1.6 \times 10^{-7}$), and *ZNF45* gene expression in skeletal muscle of Indian subjects showed an inverse correlation with BMI indicating its possible role in obesity. In conclusion, our study discovered 3 novel population-specific functional genetic variants (rs6913677, rs2078267, rs8100011) in 2 novel (*SLC22A11* and *ZNF45*) and 1 earlier reported gene (*BAI3*) for BMI in Indians. Our study decodes key genomic loci underlying obesity phenotype in Indians that may serve as prospective drug targets in future.

Keywords *BAI3* · *SLC22A11* · *ZNF45* · Body mass index · GWAS · DNA methylation · Gene regulation

Communicated by Stefan Hohmann.

Anil K. Giri, Gauri Prasad and Khushdeep Bandesh contributed equally to this study.

Membership of INDICO consortium has been listed in acknowledgements section.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00438-020-01678-6) contains supplementary material, which is available to authorized users.

Extended author information available on the last page of the article

Introduction

Obesity has turned to be leading cause for more than 200 medical disorders that affects millions of people worldwide and raises huge economic burden on global health systems (Malik et al. 2013; Tremmel et al. 2017). Since 1980, the prevalence of obesity has been doubled in 73 countries including India (Collaborators GBDO et al. 2017).

Obesity represents a chronic, heterogeneous and complex disorder that precipitates in an individual via

gene–environment interactions (Willyard 2014). Genetic factors contribute nearly 40–70% of inter-individual variability in BMI, a commonly used parameter to assess obesity (Willyard 2014). So far, genome-wide association studies (GWAS) have identified 227 genetic loci implicated in diverse biological pathways (central nervous system, food sensing, digestion, adipogenesis, insulin signaling, lipid metabolism, muscle/liver biology, and gut microbiome) that may play decisive roles in the development of obese phenotype (Pigeyre et al. 2016).

Several single nucleotide polymorphisms (SNPs) in/near genes like *FTO, MC4R, NEGR1, SH2B1, TMEM18, BDNF, FAIM2* and *SEC16B* have been replicated for BMI in genome-wide studies across multiple populations (Pigeyre et al. 2016). These signals, however, merely elucidate less than 10% of heterogeneity in BMI presentation in a population, suggesting that a large fraction of genetic determinants remains unknown along with epigenetic and environmental factors (Pigeyre et al. 2016).

One of the reasons for missing heritability can be attributed to population bias in GWAS that have mainly focused on European population, leaving genetic architecture of other populations largely unexplored (Popejoy and Fullerton 2016). Thus, delineating the unknown genetic architecture of remaining global populations may reveal novel population-specific genetic variants, explaining the remaining heritability.

In this context, genetic studies exploring Indian populations can add novel biological insight into BMI because of the following: (1) Indian population is comprised of 4693 diverse communities and various endogamous groups (Bhasin et al. 1994). This unique genetic diversity was mirrored in our previous genetic study that attributed novel population-specific genetic variants within 2q21 region of the human genome for type 2 diabetes etiology in Indians (Tabassum et al. 2013). (2) Indians present distinct features of obese phenotype that includes greater total, truncal, intra-abdominal, and subcutaneous adipose tissues, as compared to White Caucasians (Misra and Shrivastava 2013), indicating presence of population-specific genetic component for obesity. (3) Indians traditionally consume foods enriched in starch and raw sugars, a major contributor to obesity. Presence of unique environmental factors also suggests for presence of unique variants able to act in that environment as gene–environment interaction contributes significantly to obesity (Andersen et al. 2017; Qi and Cho 2008).

With an intention to identify population-specific genetic variants for BMI that can explain population-specific features of obesity in Indians, we executed a two-phase GWAS in 5973 Indian adults and tested the lead findings in a separate cohort of 1286 Indian adolescents of Indo-European ethnicity. Our study uncovered novel functional genetic candidates within *BAI3, SLC22A11* and *ZNF45* that may regulate obesity biology in Indians and warrants functional investigation for mechanistic insight.

Materials and methods

Study population

The adult participants of Indo-European descent with more than 18 years of age were enrolled through health awareness camps executed in/around Delhi. These participants were also members of Indian Diabetes Consortium (INDICO) (INDICO Consortium 2011) and served as normoglycemic controls for type 2 diabetes (T2D) GWAS executed earlier in our laboratory (Tabassum et al. 2013).

Further, adolescent participants of Indo-European origin between the age group of 10–17 years were sampled as part of a GWAS for childhood obesity and related traits in Indians. Participants were enrolled through school health surveys piloted in different zones of Delhi NCR (north, south, east, west, and central regions) as described previously (Giri et al. 2018; Tabassum et al. 2012; Prakash Dwivedi et al. 2013; Bandesh et al. 2019a). These subjects are well characterized for anthropometric as well as biochemical measurements (Giri et al. 2018; Tabassum et al. 2012; Prakash Dwivedi et al. 2013; Bandesh et al. 2019a). Overall, 21.6% (1289/5973) and 18.9% (243/1286) of the individuals had BMI > 30 in the adult and adolescent cohort, respectively.

Blood samples were drawn from participants after overnight fast, and genomic DNA was extracted from peripheral blood using salt precipitation method. Height, weight, waist and hip circumferences (WC and HC) were calculated using standard methods as described previously (INDICO Consortium 2011). Detailed phenotypic characteristics of adult study subjects are presented in Table S1.

Genome-wide association study

Discovery phase

Discovery phase samples were scanned genome-wide using Illumina Human 610-Quad BeadChips (Illumina Inc., San Diego, CA) as part of GWAS studies performed previously for T2D and related metabolic traits in our laboratory (Tabassum et al. 2013; Giri et al. 2016; Bandesh et al. 2019c; Prasad et al. 2019a). Genotype data was processed through Genome Studio software and further analyzed using PLINK (Purcell et al. 2007). Complete analysis pipeline employed in the study is shown in Fig. S1.

We followed stringent sample and SNP quality control (QC) prior to association analysis. 24 Samples with missing data for > 5% of SNPs, and 27 samples with sex-discrepancy between calculated sex and reported sex were removed. We

also discarded 11 samples with extremely low or high heterozygosity (mean ± 3 SD). Further, 87 related samples with Pi-hat score > 0.1875 and 27 potential population outliers (mean ± 6 SD) were detected using analysis of first five principal components and were excluded. Principal components were calculated using GCTA tool (http://www.complextraitgenomics.com/software/gcta/) (Yang et al. 2011). We did not find any samples with miss-matched ethnic background (Fig. S2) based on principal component analysis with 3 major global populations sequenced in 1000 genome project. Further, 29,633 SNPs with minor allele frequency (MAF) < 0.01 were excluded from the analysis. From SNPs with MAF > 0.01, 8191 SNPs with less than 97% call rate or with Hardy–Weinberg Equilibrium (HWE) $p < 10^{-5}$ were also removed.

BMI values of QC passed individuals were inverse normalized using an inbuilt command in R (http://www.r-project.org/). Association of QC passed 537,246 SNPs was tested with inverse-normalized BMI levels of 1142 adult individuals using linear regression model in PLINK (Purcell et al. 2007). Age, sex and first two principal components were used as covariates in the model.

To find the deviation of p values, a quantile–quantile (QQ) plot was created between observed and theoretical distribution of p values using qqman package in R (Turner 2014).

Replication phase

The present study is part of a large genetic study to identify genetic variations regulating various metabolic traits (glycemic, lipids, nitrogen metabolism and inflammatory parameters) in apparently healthy Indian adults (Giri et al. 2016; Bandesh et al. 2019b, c; Prasad et al. 2019b). SNPs with discovery phase $p < 10^{-4}$ for BMI and other metabolic traits besides previously established signals for BMI and other traits were genotyped in validation phase individuals using GoldenGate technology (Illumina, San Diago, USA). A total of 204 samples (4.22%) were genotyped in duplicates and an error rate of less than 0.01% was detected between technical replicates.

Samples with less than 90% call rate were excluded from the analysis. SNPs with genotype confidence score < 0.25, GenTran score < 0.60, cluster separation score < 0.4 and call rate < 90% were expelled. SNPs with minor allele frequency < 0.01 and Hardy–Weinberg equilibrium $p < 10^{-5}$ were also removed. After stringent QC, a total of 4831 adult individuals were analyzed in validation phase of the study. BMI values were inverse normalized. Association analysis was carried using linear regression adjusted for age and sex.

We further tested the association of top 3 novel signals ($p \leq 10^{-7}$) observed in Indian adults during meta-analysis in 1286 Indian adolescents genotyped using Axiom™ Genome-Wide EUR 1 Array. Both the adult and adolescent cohort has similar BMI range and genetic composition (Fig. S2). Data for adolescent cohort were also analyzed using standard QC procedures (samples with < 90% call rate, SNPs with < 95% call rate and HWE $p < 10^{-5}$ were excluded) before association. Association analysis was carried out using linear regression adjusted for age, sex and first two principal components. METAL was used for meta-analysis of summary statistics of discovery and validation phases of study using fixed-effect inverse variance method (Willer et al. 2010). Conditional analysis for signals in *BAI3* gene (previously identified—rs513357 and presently identified—rs6913677) was carried in discovery phase samples assuming additive linear model. Age, sex, BMI and respective SNP genotypes were used as covariates in the model using PLINK.

Moreover, we performed an in silico replication of identified SNPs in 2078 South Asian subjects from United Kingdom Biobank (UKBB) (Sudlow et al. 2015). Previous association status of discovered variants and genes were accessed from Type 2 Diabetes Knowledge Portal (2019).

Statistical power of study

Power of study was calculated using Quanto software (http://biostats.usc.edu/Quanto.html) assuming an additive genetic model for allele frequencies ranging from 0.001 to 0.5. For power calculations, two-tailed test at significance level of 0.05 with effect size ranging from 0.1 to 0.5, obtained from literature, was used. The average BMI was taken as 25.72 kg/m^2 with a standard deviation of 5.15 kg/m^2 for power calculation.

Combined risk score analysis

To identify the cumulative effect of 14 established and 3 novel SNPs on BMI levels in Indians, we performed allele dosage analysis by classifying the subjects on the basis of number of "effective" risk alleles as described earlier (Chauhan et al. 2010). The analysis involved samples in which genotypes at all 17 SNPs were available. We calculated effective unweighted as well as weighted allele dosage score (ADS) for this purpose.

An unweighted ADS was computed as sum of number of risk alleles for all 17 SNPs per individual. However, weighted ADS was calculated as the weighted mean of the proportion of risk alleles at 17 SNPs (i.e. 1 for two risk alleles, 0.5 for one risk allele, and 0 for no risk allele) with weights as the relative effect sizes of different SNPs. The "effective" number of risk alleles was derived by multiplying weighted ADS by 34 (maximum number of risk alleles corresponding to 17 SNPs).

BMI values were inverse normalized. Effect sizes (beta) and *p* values for overall trend in total subjects were calculated using linear regression analysis in SPSS version 25.0 (https://www.ibm.com/in-en/analytics/spss-statistics-software) to identify change in BMI levels with every unit increase in number of "effective" risk alleles.

For unweighted and weighted risk score analysis, subjects with < 10 and < 80 number of "effective" risk alleles, respectively, were taken as the reference group to calculate risk of obesity for different risk groups. For this, subjects were classified into two groups—subjects with BMI < 25 kg/m^2 (normal weight) and subjects with BMI ≥ 25 kg/m^2 (overweight/obese). The *p* values and odds ratios while comparing the different groups for risk of obesity were calculated using Chi squared test statistic.

Imputation analysis

Imputation analysis of *BAI3* and *ZNF45* loci (signals with discovery $p < 0.05$) was performed in GWAS dataset as described previously (Tabassum et al. 2013). For reference panel, 1000 Genomes Phase 3 population was used. In brief, pre-phasing for respective chromosomes was carried using SHAPEIT (Delaneau et al. 2013). A region of 1 Mb on either side encompassing the LD block of the variant was imputed using IMPUTE2 (Howie et al. 2009).

Stringent QC was performed on imputed SNPs that followed: Certainty ≥ 0.90, Info ≥ 0.5 and MAF ≥ 0.01. Imputed SNPs that passed QC were tested for association with inverse normalized BMI levels in Indians using PLINK (Purcell et al. 2007). Age, sex, and the first two principal components were employed as covariates in the model.

DNA methylation analysis

We performed whole-genome DNA methylation in peripheral blood of 236 Indian individuals that were genotyped in discovery phase using Infinium Human Methylation 450K BeadChip. Methylation data was analyzed through ENmix and Minfi packages in R with BMIQ normalization (Xu et al. 2016; Aryee et al. 2014; Teschendorff et al. 2013) as described previously (Giri et al. 2017). For meth-QTL analysis, we selected CpGs present in SNP-related genes to figure out any alterations in methylation level of nearby CpGs due to presence of identified SNPs.

Sample QC involved removal of samples with sex discrepancy, incomplete bisulphite conversion, and samples with > 5% CpG sites missing. CpG QC involved removal of CpGs with bead count less than 3 in 5% of samples and detection *p* value > 0.01 in less than 1% of samples. Additionally, CpGs falling in sex chromosomes, cross-hybridization probes and polymorphic CpGs were excluded from the analysis (Chen et al. 2013). Only CpGs with 100% call rate in all the samples were carried forward for meth-QTL analysis.

CpG outliers were fixed using fixMeth-Outliers command in Minfi (Aryee et al. 2014). Data was regressed for confounders such as cell composition, age, sex, BMI, bisulphite conversion efficiency and plate number. Methylation values for 53 CpGs (19 CpGs present in *BAI3*, 25 CpGs in *SLC22A11* and 9 CpGs in *ZNF45*) were extracted, and SNP-CpG association was executed using linear regression model in PLINK (Purcell et al. 2007).

Integration of gene regulatory data

GTEx-portal-v7 (Lonsdale et al. 2013) was used for retrieving Global expression-QTL (eQTL) and tissue expression profiling data for identified SNPs and genes, respectively. Whole Genome Bisulphite Sequencing data (WGBS) for adipose tissue was obtained from female and male adult subjects aged 30 and 34 years, respectively, from ENCODE dataset (Davis et al. 2018). ATAC-seq data of subcutaneous adipose tissue was derived from an adult female aged 53 years from ENCODE (Davis et al. 2018).

ChIP-Seq data for regulatory histone marks belong to subcutaneous adipose tissue of five adult females aged 25, 41, 49, 59 and 81 years, and was acquired from ENCODE (Davis et al. 2018). ENCODE data for DNase I hypersensitive sites in 95 cell types was also examined. Experimentally defined as well as predicted transcription factor (TF) binding sites were obtained from ENCODE and JASPAR database, respectively (Davis et al. 2018; Khan et al. 2018). Additionally, likely chromatin interaction potential data was retrieved from GeneHancer database that features human regulatory elements (enhancers and promoters) and their target genes (Fishilevich et al. 2017). UCSC genome browser was used to schematically visualize gene regulatory aspects of discovered regions (Haeussler et al. 2019).

Expression analysis of identified genes

We examined correlation between BMI levels and expression level of discovered genes in adipose and skeletal muscle tissue of 6 Indian subjects (3 males + 3 females) using our earlier published data generated through Illumina HumanHT-12 v3 Expression BeadChip arrays (Jain et al. 2013). Each sample was assessed four times to reduce technical variability in gene expression profiles. Pearson correlation was calculated between average of expression values for respective gene probes and BMI.

Gene-based association study

Besides SNP-based GWAS analysis, we also conducted a univariate gene-based genome-wide association scan using

an effective Chi squared method (ECS) employed in knowledge-based mining system for genome-wide genetic studies (KGG v4) accessible at http://statgenpro.psychiatry.hku.hk/limx/kgg/download.php. Gene-based association study may identify novel gene sets for a population based on associated marker buildup in whole genes. Genome-wide markers with association *p* values for BMI were used as input for KGG v4.

As *BAI3, SLC22A11* and *ZNF45* signals were robustly linked to majority of adiposity parameters besides BMI, we also used a multivariate gene-based association test for these genomic regions utilizing extended Simes method (MGAS) (Van der Sluis et al. 2015). Multivariate approach provides gene-based testing of several correlated phenotypes in large number of unrelated subjects. Association *p* values of discovery phase SNPs within 2 Mb regions of *BAI3, SLC22A11* and *ZNF45* for all adiposity parameters and correlation values among these traits were provided in MGAS model using KGG v4.

The 1000 genome phase III data of European, African, American, East Asian, and South Asian populations were used for computing Linkage disequilibrium (LD) among all the markers present within the respective region.

Results

Present study is effectively powered to detect variants with similar effect sizes as observed in earlier GWAS studies for BMI in literature (> 98%) (Fig. S3). Under null hypothesis, QQ plot displayed a good concordance between expected and observed *p* values in discovery phase samples (Fig. S4). The genomic inflation factor (λ) was estimated to be 1.06 signifying genetic homogeneity of study population. The association *p* values for SNPs across chromosomes in discovery phase has been shown in Fig. 1.

Genome-wide association analysis of BMI

Meta-analysis of summary results from discovery ($N = 1142$ adults) and validation phases ($N = 4831$ adults + 1286 adolescents) revealed *BAI3* (rs6913677, $p = 1.08 \times 10^{-8}$) and *SLC22A11* (rs2078267, $p = 4.62 \times 10^{-8}$) as novel GWAS signals for BMI in Indians (Table 1, Fig. 1). A novel near GWAS level association for BMI in *ZNF45* loci (rs8100011, $p = 1.04 \times 10^{-7}$) was also observed (Table 1, Fig. 1). Further analysis revealed that variants in *BAI3* and *ZNF45* also associates with Z-BMI at GWAS significance level (Table S2).

Effect sizes were calculated with respect to risk alleles. Association results presented here have been obtained from meta-analysis of summary results from discovery and validation phases in Indian adult and adolescent cohort. *p* denotes unadjusted *p* values. Dir: direction; Het-P: *p* value for heterogeneity in effect sizes in the meta-analysis; I^2: Chi square value for heterogeneity test. Direction ++/−− represents a concordance between the discovery and replication phase. Proxy SNPs—rs11752858 and rs2277312 have been utilized for rs6913677 and rs2078267, respectively, for analysis in the adolescent cohort. rs2078267 was selected for genotyping in validation phase because it was

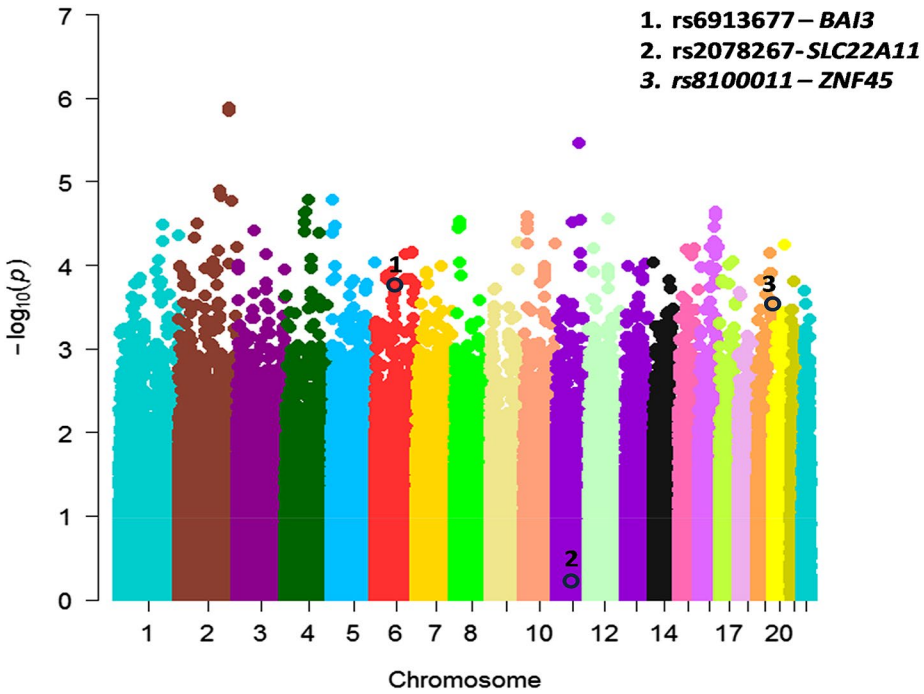

Fig. 1 Manhattan plot for the novel SNPs associated with BMI in discovery phase. The − log10 *p* values for the association of novel SNPs are plotted as function of genomic position (National Center for Biotechnology Information Build 37). The *p* value was calculated using logistic regression adjusted for age, sex, PC1 and PC2 in discovery phase analysis. Each chromosome (CHR) has been represented with a unique color

Table 1 Association status of SNPs with BMI

Trait	SNP (RA/OA) Gene	Proxy SNP (r^2, D')	RAF	Discovery phase in adults ($N=1142$)		Validation phase in adults ($N=4831$)		Validation phase in adolescents ($N=1286$)		Meta-analysis			
				BETA (SE)	p	BETA (SE)	p	BETA (SE)	p	BETA (SE)	p	Het-P (I^2)	Dir
BMI	rs6913677 (G/A) *BAI3*	rs11752858 (0.98, 0.99)	0.57	0.71 (0.19)	1.21×10^{-4}	0.38 (0.11)	3.46×10^{-4}	0.55 (0.20)	5.70×10^{-3}	0.47 (0.08)	1.08×10^{-8}	0.26 (2.87)	+++
	rs2078267 (A/G) *SLC22A11*	rs2277312 (0.86,0.98)	0.41	0.16 (0.19)	0.40	0.52 (0.10)	6.76×10^{-7}	0.53 (0.20)	7.12×10^{-3}	0.45 (0.08)	4.62×10^{-8}	0.23 (23.78)	+++
	rs8100011 (A/G) *ZNF45*		0.55	0.64 (0.19)	7.17×10^{-4}	0.40 (0.11)	1.76×10^{-4}	0.40 (0.20)	0.05	0.44 (0.08)	1.04×10^{-7}	0.51 (1.33)	+++

Effect sizes were calculated with respect to risk alleles. Association results presented here have been obtained from meta-analysis of summary results from discovery and validation phases in Indian adult and adolescent cohort. p denotes unadjusted p values. Direction ++/−− represents a concordance between the discovery and replication phase. Proxy SNPs—rs11752858 and rs2277312 have been utilized for rs6913677 and rs2078267 respectively for analysis in the adolescent cohort. rs2078267 was selected for genotyping in validation phase because it was associated with waist circumference and serum uric acid level at a p value less than 10^{-4} in our data

Dir direction, *Het-P* p value for heterogeneity in effect sizes in the meta-analysis, I^2 Chi-square value for heterogeneity test

associated with waist circumference and serum uric acid level at a p value less than 10^{-4} in our data.

Additionally, among the known variants for BMI that were genotyped in replication phase ($N=4831$), previously reported signals near *MC4R* (rs17782313 and rs12970134) showed replication for BMI levels with GWAS significance (Table S3). Moreover, variants in/near *BDNF, FTO, LOC199899, MTNR1B, TCF7L2, FADS1, KCTD15, TMEM18* and *SH2B1* replicated with nominal significance for BMI (Table S3).

Imputation of novel regions

Detailed imputation pipeline has been shown in Fig. S5. For *BAI3*, we observed seven variants that showed marginally greater association significance with BMI levels than index SNP rs6913677 (Table S4a). At *ZNF45* loci, six variants other than the index SNP rs8100011 showed slightly greater significance with BMI levels (Table S4b). All these imputed variants showed moderate to strong LD ($r^2=0.5$–1) with the index variant in African, American, European, East and South Asian populations (Tables S4a and S4b).

Few of these imputed variants (rs10945151, rs55736013, and rs11880216) overlapped with binding sites for key transcription factors such as PBX1, HOXC9, KLF4, KLF5, KLF9, and IRF3, which are known for their involvement in adipogenesis-related processes. Further, some of these imputed variants of *ZNF45* (rs55736013 and rs11880216) were also observed as robust *cis*-eQTLs in subcutaneous adipose tissue, thyroid, skin and tibial nerve, similar to the genotyped *ZNF45* variant (Lonsdale et al. 2013).

Allele dosage analysis

We analyzed the combined effect of multiple alleles at identified loci—14 established (as listed in Table S3) and 3 novel loci (rs6913677-*BAI3,* rs2078267-*SLC22A11* and rs8100011-*ZNF45*) on BMI levels using weighted and unweighted allele dosage analysis. Results suggested significantly increased levels of BMI by 0.15 units with the rise in each unit of "effective" risk allele ($p=4.23 \times 10^{-21}$) (Fig. S6a). Subjects with > 172 "effective" risk alleles (2%) displayed 3.1-times enhanced risk for overweight/obesity in comparison to subjects having less than 80 "effective" risk alleles (11%) ($p=2.1 \times 10^{-5}$) (Fig. S6a). An unweighted allele dosage analysis revealed that subjects with 25–29 "effective" risk alleles (2.33%) displayed nearly fivefold enhanced risk for overweight/obesity in contrast to subjects having less than 10 "effective" risk alleles (1.94%) ($p=9.35 \times 10^{-7}$) (Fig. S6b).

Discovered variants are novel loci for BMI and associate with obesity-related metabolic traits

We examined former reported associations within *BAI3, SLC22A11* and *ZNF45* for BMI levels. The strongest reported associations included the following SNPs: rs618714 ($p = 2.10 \times 10^{-7}$), rs693591 ($p = 0.04$) and rs454376 ($p = 0.01$), within *BAI3, SLC22A11,* and *ZNF45,* respectively, in Europeans of GIANT UK Biobank GWAS (Figs. S7a, S7b and S7c). Further, another variant in *BAI3* (rs513357) has also been previously associated with BMI at GWAS significance level ($p = 4.4 \times 10^{-8}$) in a multiethnic GWAS (Hoffmann et al. 2018). Conditional analysis in discovery phase samples ($N = 1142$ adults) revealed independent association of presently identified *BAI3* locus—rs6913677 from previously identified locus (Table S5). Further, both the SNPs in *BAI3* show poor linkage disequilibrium in Indians ($r^2 = 0.15$, $D' = 0.73$, physical distance = 245 kb) and other major populations ($r^2 < 0.07$, $D' = 0.48$, https://ldlink.nci.nih.gov) (Machiela and Chanock 2015).

Moreover, discovered variants (rs6913677, rs2078267 and rs8100011) in these genes were found nominally associated with BMI and related traits in previous studies (Table S6).

In-silico replication of novel variants

In-silico replication of novel signals in South Asian individuals of UK Bio-Bank did not improve the significance for any of the three loci. It, however, revealed similar direction of association for *SLC22A11* (rs2078267), and *ZNF45* (rs8100011) (Table S7).

Gene expression correlation analysis

Correlation analysis between BMI and expression level of identified genes in Indian subjects revealed a strong negative correlation between *ZNF45* gene expression and BMI ($R = -0.85$, $p = 0.03$) in skeletal muscle of study subjects (Fig. S8).

Publicly available data mining in GTEx portal revealed *ZNF45* variant-rs8100011 as strong *cis*-eQTL in subcutaneous adipose, thyroid, skin and tibial nerve (Table S8, Fig. S9). The eQTL results from GTEx domain suggested that,double risk allele genotype of rs8100011for BMI ("Genotype AA") was associated with reduced expression of *ZNF45* in various tissues (Fig. S9). This is consistent with our observations of increase in BMI with decrease in expression of *ZNF45* in skeletal muscle of obese subjects (Fig. S8).

Further, expression correlation analysis for *BAI3* and *SLC22A11* exhibited modest negative ($R = -0.47$, $p = 0.34$) and positive correlation ($R = 0.48$, $p = 0.33$) with BMI levels in skeletal muscle of Indian subjects (Fig. S8).

Identified variations are regulatory in nature

These genes were observed to be moderately expressed in subcutaneous and visceral adipose tissues with reasonable ubiquitous expression in liver, skeletal muscle, tibial nerve, pancreas, thyroid and whole blood (Fig. S10a, S10b and S10c). In order to understand functional impacts of identified variants, we first examined their open chromatin features, regulatory histone marks and DNA methylation patterns in subcutaneous adipose tissue. We also examined predicted TFs that bind to DNA regions of identified variants. Further, likely chromatin interaction potential of identified variants was also explored.

BAI3 variant (rs6913677) represented repressive chromatin marks assisted by weak ATAC-seq peaks in subcutaneous adipose tissue and also presented absence of DNase I hypersensitivity clusters in multiple cell types (Fig. S11a). This was further supported by enrichment of heterochromatin-specific histone marks H3K27me3 and H3K9me3 around rs6913677 (Fig. S11a).

Interestingly, rs6913677 displayed strong predicted binding sites for several crucial transcription factors implicated in obesity—PPAR-α, PPAR-γ, E2F4, HNF4G, and ZNF263 (Fig. S11b) with highly conserved DNA binding motifs (denoted by nominal to high TF bit-score) for these key transcription factors (Fig. S11c). In addition, WGBS data revealed differential methylation at rs6913677 position in adipose tissue of male and female samples, wherein 50% of sequence reads were found to be methylated in male sample with no methylation in female sample, indicating sex-dependent regulatory potential of this variant (Fig. S11b).

In contrast, *SLC22A11* variant (rs2078267) constituted strong open chromatin marks supported with robust ATAC-seq and DNase I hypersensitivity signals in subcutaneous adipose tissue and various other cell types (Fig. S12a). Strong peaks of H3K4me1 accompanied by feeble peaks of H3K27ac and H3K4me3 around rs2078267 in subcutaneous adipose tissue represent an active enhancer region (Fig. S12a). This was supported by GeneHancer database that demonstrated strong enhancer functionality of rs2078267 variant targeting *TRMT112, CDC42BPG* and *SF1* genes in its vicinity (Fig. S12b).

Interestingly, rs2078267 displayed strong binding sites for obesity and metabolic disease-associated transcription factors such as-NFIC, KAP1, TCF7L2, ZNF263. etc. (Fig. S12b).None of the sequenced reads were found to be methylated in adipose tissue around variant rs2078267 in WGBS dataset (Fig. S12b).

ZNF45 variant (rs8100011) exhibited extensive enrichment of active H3K36me3 and H3K4me3 marks and feeble peaks of repressive H3K27me3 and H3K9me3 marks in subcutaneous adipose tissue that is a signature for active gene transcription (Fig. S13a). The variant region indicated strong chromatin interaction potential of *ZNF45* with neighboring genes—*ZNF222, ZNF230, ZNF283* and *ZNF221* (Fig. S13b).

Methylation quantitative trait loci (meth-QTL) analysis

To investigate likely functional roles of identified variants, we overlapped genetic data with DNA methylation data generated from peripheral blood in Indians and performed meth-QTL analysis. Meth-QTL analysis revealed rs6913677 as *cis*-methylation-QTL for a CpG, cg17094144, in *BAI3* gene (Table 2). Additionally, rs2078267 displayed robust association with differential methylation patterns for 5 CpGs within *SLC22A11* gene (*cis*-methylation-QTL) (Table 2).

Since the identified genes are nominally expressed in blood (Fig. S14), we retrieved gene regulatory features for associated CpGs from K562 leukemia cell line. Associated CpG site for meth-QTL—rs6913677 is located in exon 1 and 5′ UTR of *BAI3* that may serve as strong promoter with enrichment of binding sites for RBBP5, EZH2, MYC, RFX5 and MAX transcription factors (Table S9).

Similarly, meth-QTL—rs2078267 (*SLC22A11*) was robustly associated with five CpGs that resided in important functional regions—gene body, TSS200 (200 bases upstream from transcription start site) or were intergenic with enhancer- or insulator-specific functions (Table S9).

However, we did not observe rs8100011 (*ZNF45*) as meth-QTL even though *ZNF45* is nominally expressed in blood (Fig. S14). We also examined WGBS dataset around identified variants in classical monocytes (CD14-positive), B-cells and T-cells from ENCODE dataset. All these blood cells showed lack of any DNA methylation mark at rs8100011 (*ZNF45*) loci in contrast to rs6913677 (*BAI3*) and rs2078267 (*SLC22A11*) (Fig. S15). Similarly, no methylation mark for rs8100011 (*ZNF45)* was observed in adipose tissue derived from ENCODE dataset (Fig. S13b).

Gene-based GWAS analysis

As SNPs alone cannot fully explain the heritability of complex traits, we also implemented a gene-based GWAS analysis to identify complementary genetic determinants for BMI. This analysis uncovered distinct novel loci that were not detected earlier in SNP-based association method. Protein-coding genes *CPS1* and *UPP2* were strongest signals (p value $\leq 1 \times 10^{-8}$) associated with BMI for the first time (Table S10).

This was followed by additional protein-coding genes *ACOXL, FAM71E2, NRCAM, SLC25A12, PKD1L3, UBA5* and *APBA2,* and non-coding RNA genes *LINC00358, LINC01142* and *SLC16A12-AS1* that associated modestly (p value $\leq 1 \times 10^{-5}$) with BMI in Indians (Table S10). Further, in agreement with SNP-based analysis, *BAI3* and *ZNF45* genes persisted significance in gene-based analysis as well (respective $p = 0.02$ and 3.55×10^{-5}).

Since *BAI3, SLC22A11* and *ZNF45* variants associated robustly with most adiposity traits, we also implemented a multivariate gene-based testing for all three loci using extended Simes method (MGAS). This analysis suggested *BAI3* and *ZNF45* as leading candidates within their respective 2 Mb genomic regions (p value $= 0.01$ and 5.50×10^{-4} respectively) associated with all adiposity measures in Indians (Table S11).

Table 2 Meth-QTL analysis for associated signals in 236 adult subjects who have been genotyped in discovery phase of study

SNP					CpG				BETA (95% CI)	p
Name	CHR	BP	Gene	A1/A2	Name	CHR	BP	Gene		
rs6913677	6	69860183	*BAI3*	G/A	cg17094144	6	69345881	*BAI3*	0.01 (0.00–0.01)	0.03
rs2078267	11	64090690	*SLC22A11*	A/G	cg00270878	11	64334216	*SLC22A11*	− 0.03 (− 0.04 to − 0.01)	4.44×10^{-4}
rs2078267	11	64090690	*SLC22A11*	A/G	cg07086353	11	64257659	*SLC22A11*	− 0.01 (− 0.02 to 0.00)	2.20×10^{-3}
rs2078267	11	64090690	*SLC22A11*	A/G	cg08822897	11	64258102	*SLC22A11*	− 0.05 (− 0.07 to − 0.04)	4.96×10^{-10}
rs2078267	11	64090690	*SLC22A11*	A/G	cg09337943	11	64335732	*SLC22A11*	0.01 (0.00–0.01)	4.09×10^{-3}
rs2078267	11	64090690	*SLC22A11*	A/G	cg18158438	11	64322993	*SLC22A11*	− 0.01 (− 0.01 to 0.00)	0.01

p value has been obtained from association of SNPs with methylation level at corresponding CpG sites using PLINK. CpG sites have been annotated using Illumina 450K BeadChip manifest file

CHR chromosome, *BP* base position, *CI* Confidence interval

Signals with p value < 0.05 have only been shown

Discussion

Present study investigated genome-wide markers regulating body-mass index in Indians. We discovered novel GWAS significant loci in—*BAI3* (brain-specific angiogenesis inhibitor) and *SLC22A11* (solute carrier family 22 member 11) followed by *ZNF45* (zinc finger protein 45) locus with near GWAS significance. Similar effect size of identified variants in adult and adolescent cohort suggests role of identified variants since early phase of life in Indian population. However, lack of significant association in UKBB South Asian samples could be because of the following: (1) different genetic background of UKBB samples than samples used in the study. The South Asian samples collected in UKBB comprises of diverse ethnic groups from Bangladeshi, Indian, and Pakistani origin. On a global population map, these samples may look similar. However, they differ from themselves in intra-population analysis. For example, Indian population itself is comprised of 4693 diverse communities and various endogamous groups (Bhasin et al. 1994). The samples included in our study are Indians of Indo-European ethnicity. The mixture of samples from different ethnic background in UKBB can mask the association. (2) Moreover, this also suggests an environment-specific influence of identified variants on BMI. Since obesity is a complex trait with strong impact from environmental factors (e.g. physical activity, food habit, alcohol consumption and socioeconomic status) (Andersen et al. 2017; Qi and Cho 2008), the lack of association could be because of difference in lifestyle (e.g. food habit, physical activity) between study participants and South Asian samples in UKBB. However, replication studies in other populations with different environmental exposures are needed to understand the influence of the identified loci on BMI.

Intriguingly, multiple intergenic GWAS loci for BMI were observed in GIANT UK Bio-bank GWAS conducted in ~ 700,000 individuals of European ancestry (Yengo et al. 2018). Moreover, only one locus within *BAI3* (rs513357) has been identified as GWAS signal for BMI in a multi-ethnic GWAS (Hoffmann et al. 2018). Further, the gene has been associated with triglyceride levels in Amerindian population at nominal significance (Ko et al. 2014). We identified a novel locus tagged by rs6913677 in earlier reported *BAI3* gene as GWAS signal for BMI. However, the independence of these identified loci in conditional analysis suggests that Indian population harbors a different genomic region in *BAI3* that influences obesity.

BAI3 is a seven-span transmembrane G protein-coupled receptor (Rebhan et al. 1997) which is widely expressed in multiple regions of brain along with its ligands (cerebral hemisphere, cerebellum, and pituitary) (Lonsdale et al. 2013). One of its four ligands, C1ql4 (Complement C1q-Like Protein 4), has been identified as a negative regulator of adipogenesis in mouse model via inhibiting p42/44-mitogen-activated protein kinase signaling pathway (p42/44-MAPK) (Wei et al. 2013; Bolliger et al. 2011). This pathway plays a critical role in regulating the balance between adipocyte growth and differentiation (Aubert et al. 1999). Interestingly, over-expression of C1ql4 repressed the differentiation of 3T3-L1 adipocytes marked up by parallel reduction in transcript as well as protein levels of PPAR-γ and C/EBP-α, major transcription factors that drive adipogenesis (Wei et al. 2013).

C1ql4 protein is a small secreted signaling molecule expressed variably in brain (Lonsdale et al. 2013), wherein globular C1q domains bind strongly with extracellular thrombospondin-repeat domain of BAI3 receptor and may direct synapse formation and maintenance in human brain (Bolliger et al. 2011). Owing to consideration that both *BAI3* and *C1ql4* genes are nominally expressed in subcutaneous and visceral adipose tissues (Lonsdale et al. 2013), we speculate that BAI3 receptor can mediate the inhibitory effect of C1ql4 ligands on adipocyte differentiation plausibly via p42/44-MAPK pathway.

A modest enrichment for gene repression histone marks (H3K27me3 and H3K9me3) surrounding *BAI3* locus in subcutaneous adipose tissue indicates it as a facultative heterochromatin zone. We observed that alternate alleles of *BAI3* variant rs6913677 significantly influenced the methylation levels of a CpG site—cg17094144. The CpG site resided within active promoter region of *BAI3,* and is surrounded by regulatory histone modifications with specific binding sites for key transcription factors. Altered DNA methylation can modulate the binding of these transcription factors and methyl CpG associated proteins and thereby may influence the underlying gene expression. For instance, DNA methylation in intron 1 of *HIF3A* presents robust association with BMI levels and significant negative correlation with *HIF3A* mRNA levels in adipose tissue (Dick et al. 2014). Additionally, DNA methylation marks in blood at four CpG sites residing in *LGALS3BP, RORC, SOCS3* and *ANGPT4* genes retained robust association with BMI levels in American Women (Wilson et al. 2017).

Identification of *BAI3* variant rs6913677 as *cis* meth-QTL demonstrates that variant may fine-tune adiposity via affecting *BAI3* gene expression. In addition, this variant functions as strong binding motif for known transcription factors like PPAR-α, PPAR-γ, E2F4, HNF4G and ZNF263 (Khan et al. 2018) that are known to play a major role in obesity. Likewise, imputed variant of *BAI3*-rs10945151 also marked a strong seat for adiposity-related transcription factors such as PBX1 and HOXC9 (Khan et al. 2018). PBX1 controls the process of adipogenesis in stage-dependent manner, in both mouse and human (Monteiro et al. 2011). Additionally,

HOXC9 is involved in adverse fat distribution, metabolism, adipose tissue function and development of obesity (Brune et al. 2016). Alternate alleles of *BAI3* variant may efficiently modulate the binding of these transcription factors to their respective binding motifs and thereby alter the *BAI3* gene expression levels. Variable gene expression due to altered binding of transcription factors may subsequently modify levels of expressed BAI3 protein that may influence downstream receptor–ligand interactions and consequently adipocyte differentiation in adipose tissue.

We also observed strong association of variant rs2078267 in *SLC22A11*, a known GWAS signal for uric acid in several populations including Indians (Giri et al. 2016; Huffman et al. 2015). Variant rs2078267 overlies an active enhancer region in subcutaneous adipose tissue and is potential binding site for numerous transcription factors implicated in obesity and associated metabolic disorders (Davis et al. 2018). *SLC22A11* expresses in renal membranes and fetal-facing basal membrane of placenta and is a transporter for glutamate (Skwara et al. 2017; Zhou et al. 2010). Higher glutamate uptake has been associated with higher BMI in Chinese adults (He et al. 2008). Further, glutamate release facilitates leptin action on energy expenditure in mice (Xu et al. 2013). These findings suggest that change in expression due to genetic variations in *SLC22A11* may affect adiposity by glutamate-mediated leptin signaling.

Interestingly, we also discovered novel association of variant residing in active chromatin region—rs8100011 in *ZNF45* gene with BMI at near genome-wide significance level. ZNF45 is a predicted transcription factor (Rebhan et al. 1997) that binds with CEBPA, a major transcription factor for adipocyte maturation and CBX5, a histone modifier involved in proliferation and differentiation of cellular lineages including adipocytes (Zada et al. 2006; Nozawa et al. 2010). The variant (rs8100011) was identified as explicit *cis*-eQTL for *ZNF45* expression in subcutaneous adipose tissue and identified risk allele (A) has been associated with lower expression of gene (Lonsdale et al. 2013). GTEx data is in agreement with our expression correlation analysis in Indians that show decreased expression of *ZNF45* in skeletal muscle of obese individuals and opens up further exploration into mechanistic insight of *ZNF45* variant in scheming obesity features. Although the effect of rs8100011 on obesity by modulating *ZNF45* expression seems a more plausible explanation, alternative modes of action of rs8100011 in obesity biology cannot be negated as rs8100011 was also identified as eQTL for other genes such as *ZNF155, ZNF283, ZNF404, AC084219.4, KCNN4, RP11-15A1.3*.

Further, imputed *ZNF45* variant—rs55736013 constituted strong binding site for several members of Krüppel-Like adipogenesis-related transcription factors—KLF4, KLF5 and KLF9, that promote adipogenesis via stimulating CCAAT-enhancer-binding proteins (C/EBPs) (Khan et al. 2018; Pollak et al. 2018). Additionally, another imputed variant—rs11880216, constitutes binding site for IRF3, a major transcriptional regulator of adipose inflammation, involved in maintaining systemic glucose and energy homeostasis (Khan et al. 2018; Kumari et al. 2016).

We also found convincing evidence that discovered variants display strong associations with numerous metabolic phenotypes (Type 2 Diabetes Knowledge Portal 2019). This include BMI, weight, WC, HC, WHR, body fat percentage, fasting glucose, fasting insulin, total cholesterol, LDL cholesterol, HDL cholesterol, leptin, adiponectin and type 2 diabetes. This essentially pinpoints their dynamic metabolic influence in diverse populations.

Further, our gene-based analysis revealed additional new loci—*CPS1* (Carbamoyl-Phosphate Synthase 1) and *UPP2* (Uridine Phosphorylase 2) that powerfully dictate BMI levels in Indians and have never been accounted for obesity in earlier GWAS studies.

Nevertheless, our study has one major limitation. Initially we performed the two-staged GWAS for Indian adults and identified the 3 SNPs as lead hits. Further, we validated these lead SNPs in another adolescent cohort available at our laboratory and performed the meta-analysis. However, a better design would have been to validate all the SNPs selected in discovery phase of adults in adolescent cohort and perform the meta-analysis. Nonetheless, due to economic constraint, we could not genotype all the selected SNPs in adolescent individuals. Yet, the significant associations of identified SNPs from adult cohort in adolescent cohort and concordance in direction of effect sizes in both cohorts suggests that identified genes have role in both adult and adolescent obesity.

In conclusion, we identified three novel loci in two novel genes, namely- *SLC22A11* and *ZNF45* and one previously reported gene-*BAI3* governing BMI levels in Indians. The discovered loci feature crucial gene-regulatory signatures that likely pinpoints their definite mechanistic role in obesity. The identification of novel genes entails for population-specific genetic studies in diverse populations. Discovered genetic leads exclusively opens up a new biology and therapeutic considerations for obesity phenotype in genetically diverse Indians.

Acknowledgements The authors are greatly thankful to all the study participants. We acknowledge the support and participation of members of the INDICO consortium in data generation. We thank Prof. Mark I McCarthy for providing data for South Asian subjects of United Kingdom Biobank (UKBB) and Prof. Pawan Dhar (School of Biotechnology, Jawaharlal Nehru University, India) for critical evaluation of the manuscript. GP and AKG acknowledge University Grants Commission (UGC), Government of India for Senior Research Fellowship. KB acknowledges Council of Scientific and Industrial Research (CSIR), Government of India for Senior Research Fellowship. Members of INDICO: Viswanathan Mohan, Saurabh Ghosh, Abhay Sharma,

Rubina Tabassum, Ganesh Chauhan, Anubha Mahajan, Om Prakash Dwivedi, Lakshmi Ramakrishnan, Radha Venkatesan, M. Chidambaram, D. Prabhakaran, K. S. Reddy, Monisha Banerjee, Madhukar Saxena, Sandeep Kumar Mathur, Anil Bhansali, Viral Shah, S. V. Madhu, R. K. Marwaha, Pradeep Venkatesh, S. K. Aggarwal, Shantanu Sengupta, Sreenivas Chavali, Amitabh Sharma, Analabha Basu, Khushdeep Bandesh, Anil K. Giri, Shraddha Chakraborty, Yasmeen Kauser, B. Abitha, Aditya Undru, Donaka Rajashekar, Vaisak Parekatt, Suki Roy, Anjali Singh, Priyanka Banerjee, Gauri Prasad, Punam Jha, Nikhil Tandon and Dwaipayan Bharadwaj (Coordinator).

Author contributions AKG and GP: assembled and analyzed data; contributed to discussions and wrote the manuscript; KB: provided intellectual inputs and critically reviewed the manuscript; AKG, GP, KB, VP, PB, YK, SC, DR, and INDICO: helped in sample preparation and data generation; AM and AB: helped in statistical analysis; AM, AS, SKM, AB, MIM and NT: critically reviewed the manuscript; DB: is the guarantor of this work who conceived, supervised, obtained financial support and oversaw the entire study.

Funding This work was supported by Council of Scientific and Industrial Research [CSIR], Government of India through Centre for Cardiovascular and Metabolic Disease Research [CARDIOMED] project [Grant no: BSC0122-(12)] and Department of Biotechnology (DBT), Government of India through BIOCHILD Grant given to CSIR-Institute of Genomics and Integrative Biology, New Delhi, India. This study was also partially funded by Department of Science and Technology, Government of India through PURSE II CDST/SR/PURSE PHASE II/11 provided to Jawaharlal Nehru University, New Delhi, India. Part of this work was conducted using the UK Biobank Resource under Application Number 9161.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Ethical approval The study was approved by Human Ethics Committees of All India Institute of Medical Sciences, New Delhi and CSIR-Institute of Genomics and Integrative Biology, New Delhi, India and was conducted in accordance with the principles of Helsinki Declarations.

Informed consent The adult participants were informed about objectives of study and written consents were taken from all of them. For adolescent study population, prior official permission from school authorities, informed written consent from parents and verbal consent from participants themselves were obtained before their participation in the study.

References

Andersen RM, Karlsson T, Ek WE, Johansson Å (2017) Gene–environment interaction study for BMI reveals interactions between genetic factors and physical activity, alcohol consumption and socioeconomic status. PLoS Genet 13:e1006977

Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA (2014) Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 30:1363–1369

Aubert J, Belmonte N, Dani C (1999) Role of pathways for signal transducers and activators of transcription, and mitogen-activated protein kinase in adipocyte differentiation. Cell Mol Life Sci 56:538–542

Bandesh K, Jha P, Giri AK, Marwaha RK, Scaria V, Tandon N, Bharadwaj D, INDICO (2019a) Normative range of blood biochemical parameters in urban Indian school-going adolescents. PLoS One 14:e0213255

Bandesh K, Prasad G, Giri AK, Kauser Y, Upadhyay M, Basu A, Tandon N, Bharadwaj D, INDICO (2019b) Genome-wide association study of blood lipids in Indians confirms universality of established variants. J Hum Genet 64:573–587

Bandesh K, Prasad G, Giri AK, Voruganti SV, Butte NF, Cole SA, Comuzzie AG, Tandon N, Bharadwaj D, INDICO (2019c) Genome-wide association study of C-peptide surfaces key regulatory genes in Indians. J Genet 98:8

Bhasin MK, Walter H, Danker-Hopfe H (1994) People of India: an investigation of biological variability in ecological ethnoeconomic and linguistic groups. Kamla-Raj Enterprises, Delhi. https://doi.org/10.1002/ajhb.1310070224

Bolliger MF, Martinelli DC, Sudhof TC (2011) The cell-adhesion G protein-coupled receptor BAI3 is a high-affinity receptor for C1q-like proteins. Proc Natl Acad Sci 108:2534–2539

Brune JE, Kern M, Kunath A, Flehmig G, Schön MR, Lohmann T, Dressler M, Dietrich A, Fasshauer M, Kovacs P, Stumvoll M, Blüher M, Klöting N (2016) Fat depot-specific expression of HOXC9 and HOXC10 may contribute to adverse fat distribution and related metabolic traits. Obesity 24:51–59

Chauhan G, Spurgeon CJ, Tabassum R, Bhaskar S, Kulkarni SR, Mahajan A, Chavali S, Kumar MV, Prakash S, Dwivedi OP, Ghosh S, Yajnik CS, Tandon N, Bharadwaj D, Chandak GR (2010) Impact of common variants of PPARG, KCNJ11, TCF7L2, SLC30A8, HHEX, CDKN2A, IGF2BP2, and CDKAL1 on the risk of type 2 diabetes in 5164 Indians. Diabetes 59:2068–2074

Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R (2013) Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation 450 microarray. Epigenetics 8:203–209

Collaborators GBDO, Afshin A, Forouzanfar MH, Reitsma MB, Sur P, Estep K, Lee A, Marczak L, Mokdad AH, Moradi-Lakeh M, Naghavi M, Salama JS, Vos T, Abate KH, Abbafati C, Ahmed MB, Al-Aly Z, Alkerwi A, Al-Raddadi R, Amare AT, Amberbir A, Amegah AK, Amini E, Amrock SM, Anjana RM, Ärnlöv J, Asayesh H, Banerjee A, Barac A, Baye E, Bennett DA, Beyene AS, Biadgilign S, Biryukov S, Bjertness E, Boneya DJ, Campos-Nonato I, Carrero JJ, Cecilio P, Cercy K, Ciobanu LG, Cornaby L, Damtew SA, Dandona L, Dandona R, Dharmaratne SD, Duncan BB, Eshrati B, Esteghamati A, Feigin VL, Fernandes JC, Fürst T, Gebrehiwot TT, Gold A, Gona PN, Goto A, Habtewold TD, Hadush KT, Hafezi-Nejad N, Hay SI, Horino M, Islami F, Kamal R, Kasaeian A, Katikireddi SV, Kengne AP, Kesavachandran CN, Khader YS, Khang YH, Khubchandani J, Kim D, Kim YJ, Kinfu Y, Kosen S, Ku T, Defo BK, Kumar GA, Larson HJ, Leinsalu M, Liang X, Lim SS, Liu P, Lopez AD, Lozano R, Majeed A, Malekzadeh R, Malta DC, Mazidi M, McAlinden C, McGarvey ST, Mengistu DT, Mensah GA, Mensink GBM, Mezgebe HB, Mirrakhimov EM, Mueller UO, Noubiap JJ, Obermeyer CM, Ogbo FA, Owolabi MO, Patton GC, Pourmalek F, Qorbani M, Rafay A, Rai RK, Ranabhat CL, Reinig N, Safiri S, Salomon JA, Sanabria JR, Santos IS, Sartorius B, Sawhney M, Schmidhuber J, Schutte AE, Schmidt MI, Sepanlou SG, Shamsizadeh M, Sheikhbahaei S, Shin MJ, Shiri R, Shiue I, Roba HS, Silva DAS, Silverberg JI, Singh JA, Stranges S, Swaminathan S, Tabarés-Seisdedos R, Tadese F, Tedla BA, Tegegne BS, Terkawi AS, Thakur JS, Tonelli M, Topor-Madry R, Tyrovolas S, Ukwaja KN, Uthman OA, Vaezghasemi M, Vasankari T, Vlassov VV, Vollset SE, Weiderpass E, Werdecker A, Wesana J, Westerman R, Yano Y, Yonemoto N, Yonga G, Zaidi Z, Zenebe ZM, Zipkin B, Murray CJL (2017)

Health effects of overweight and obesity in 195 countries over 25 years. N Engl J Med 377:13–27

Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, Hilton JA, Jain K, Baymuradov UK, Narayanan AK, Onate KC, Graham K, Miyasato SR, Dreszer TR, Strattan JS, Jolanki O, Tanaka FY, Cherry JM (2018) The encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Res 46:D794–D801

Delaneau O, Zagury JF, Marchini J (2013) Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods 10:5–6

Dick KJ, Nelson CP, Tsaprouni L, Sandling JK, Aïssi D, Wahl S, Meduri E, Morange PE, Gagnon F, Grallert H, Waldenberger M, Peters A, Erdmann J, Hengstenberg C, Cambien F, Goodall AH, Ouwehand WH, Schunkert H, Thompson JR, Spector TD, Gieger C, Trégouët DA, Deloukas P, Samani NJ (2014) DNA methylation and body-mass index: a genome-wide analysis. Lancet 383:1990–1998

Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Iny Stein T, Rosen N, Kohn A, Twik M, Safran M, Lancet D, Cohen D (2017) GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. Database. https://doi.org/10.1093/database/bax028

Giri AK, Banerjee P, Chakraborty S, Kauser Y, Undru A, Roy S, Parekatt V, Ghosh S, Tandon N, Bharadwaj D (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 AK, Bharadwaj S, Banerjee P, Chakraborty S, Parekatt V, Rajashekar D, Tomar A, Ravindran A, Basu A, Tandon N, Bharadwaj D (2017) DNA methylation profiling reveals the presence of population-specific signatures correlating with phenotypic characteristics. Mol Genet Genom 292:655–662

Giri AK, Parekatt V, Dwivedi OP, Banerjee P, Bandesh K, Prasad G, Tandon N, Bharadwaj D (2018) Common variants of ARID1A and KAT2B are associated with obesity in Indian adolescents. Sci Rep 8:3964

Haeussler M, Zweig AS, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, Lee CM, Lee BT, Hinrichs AS, Gonzalez JN, Gibson D, Diekhans M, Clawson H, Casper J, Barber GP, Haussler D, Kuhn RM, Kent WJ (2019) The UCSC genome browser database: 2019 update. Nucleic Acids Res 47:D853–D858

He K, Zhao L, Daviglus ML, Dyer AR, Van Horn L, Garside D, Zhu L, Guo D, Wu Y, Zhou B, Stamler J, INTERMAP Cooperative Research Group (2008) Association of monosodium glutamate intake with overweight in Chinese adults: the INTERMAP study. Obesity 16:1875–1880

Hoffmann TJ, Choquet H, Yin J, Banda Y, Kvale MN, Glymour M, Schaefer C, Risch N, Jorgenson E (2018) A large multiethnic genome-wide association study of adult body mass index identifies novel loci. Genetics 210(2):499–515

Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 5:e1000529

Huffman JE, Albrecht E, Teumer A, Mangino M, Kapur K, Johnson T, Kutalik Z, Pirastu N, Pistis G, Lopez LM, Haller T, Salo P, Goel A, Li M, Tanaka T, Dehghan A, Ruggiero D, Malerba G, Smith AV, Nolte IM, Portas L, Phipps-Green A, Boteva L, Navarro P, Johansson A, Hicks AA, Polasek O, Esko T, Peden JF, Harris SE, Murgia F, Wild SH, Tenesa A, Tin A, Mihailov E, Grotevendt A, Gislason GK, Coresh J, D'Adamo P, Ulivi S, Vollenweider P, Waeber G, Campbell S, Kolcic I, Fisher K, Viigimaa M, Metter JE, Masciullo C, Trabetti E, Bombieri C, Sorice R, Döring A, Reischl E, Strauch K, Hofman A, Uitterlinden AG, Waldenberger M, Wichmann HE, Davies G, Gow AJ, Dalbeth N, Stamp L, Smit JH, Kirin M, Nagaraja R, Nauck M, Schurmann C, Budde K, Farrington SM, Theodoratou E, Jula A, Salomaa V, Sala C, Hengstenberg C, Burnier M, Mägi R, Klopp N, Kloiber S, Schipf S, Ripatti S, Cabras S, Soranzo N, Homuth G, Nutile T, Munroe PB, Campbell H, Hastie N, Rudan I, Cabrera C, Haley C, Franco OH, Merriman TR, Gudnason V, Pirastu M, Penninx BW, Snieder H, Metspalu A, Ciullo M, Pramstaller PP, Van Duijn CM, Ferrucci L, Gambaro G, Deary IJ, Dunlop MG, Wilson JF, Gasparini P, Gyllensten U, Spector TD, Wright AF, Hayward C, Watkins H, Perola M, Bochud M, Kao WH, Caulfield M, Toniolo D, Völzke H, Gieger C, Köttgen A, Vitart V (2015) Modulation of genetic associations with serum urate levels by body-mass-index in humans. PLoS One 10:e0119752

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

Jain P, Vig S, Datta M, Jindel D, Mathur AK, Mathur SK, Sharma A (2013) Systems biology approach reveals genome to phenome correlation in type 2 diabetes. PLoS One 8:e53522

Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, van der Lee R, Bessy A, Chèneby J, Kulkarni SR, Tan G, Baranasic D, Arenillas DJ, Sandelin A, Vandepoele K, Lenhard B, Ballester B, Wasserman WW, Parcy F, Mathelier A (2018) JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res 46:D260–D266

Ko A, Cantor RM, Weissglas-Volkov D, Nikkola E, Reddy P, Sinsheimer JS, Pasaniuc B, Brown R, Alvarez M, Rodriguez A, Rodriguez-Guillen R, Bautista IC, Arellano-Campos O, Muñoz-Hernández LL, Salomaa V, Kaprio J, Jula A, Jauhiainen M, Heliövaara M, Lehtimäki T, Raitakari O, Eriksson JG, Perola M, Lohmueller KE, Matikainen N, Taskinen M, Rodriguez-Torres M, Riba L, Tusie-Luna T, Aguilar-Salinas CA, Pajukanta P (2014) Amerindian-specific regions under positive selection harbour new lipid variants in Latinos. Nat Commun 5:3983

Kumari M, Wang X, Lantier L, Lyubetskaya A, Eguchi J, Kang S, Tenen D, Roh HC, Kong X, Kazak L, Ahmad R, Rosen ED (2016) IRF3 promotes adipose inflammation and insulin resistance and represses browning. J Clin Investig 126:2839–2854

Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, Hasz R, Walters G, Garcia F, Young N, Foster B, Moser M, Karasik E, Gillard B, Ramsey K, Sullivan S, Bridge J, Magazine H, Syron J, Fleming J, Siminoff L, Traino H, Mosavel M, Barker L, Jewell S, Rohrer D, Maxim D, Filkins D, Harbach P, Cortadillo E, Berghuis B, Turner L, Hudson E, Feenstra K, Sobin L, Robb J, Branton P, Korzeniewski G, Shive C, Tabor D, Qi L, Groch K, Nampally S, Buia S, Zimmerman A, Smith A, Burges R, Robinson K, Valentino K, Bradbury D, Cosentino M, Diaz-Mayoral N, Kennedy M, Engel T, Williams P, Erickson K, Ardlie K, Winckler W, Getz G, DeLuca D, MacArthur D, Kellis M, Thomson A, Young T, Gelfand E, Donovan M, Meng Y, Grant G, Mash D, Marcus Y, Basile M, Liu J, Zhu J, Tu Z, Cox NJ, Nicolae DL, Gamazon ER, Im HK, Konkashbaev A, Pritchard J, Stevens M, Flutre T, Wen X, Dermitzakis ET, Lappalainen T, Guigo R, Monlong J, Sammeth M, Koller D, Battle A, Mostafavi S, McCarthy M, Rivas M, Maller J, Rusyn I, Nobel A, Wright F, Shabalin A, Feolo M, Sharopova N, Sturcke A, Paschal J, Anderson JM, Wilder EL, Derr LK, Green ED, Struewing JP, Temple G, Volpi S, Boyer JT, Thomson EJ, Guyer MS, Ng C, Abdallah A, Colantuoni D, Insel TR, Koester SE, Little AR, Bender PK, Lehner T, Yao Y, Compton CC, Vaught JB, Sawyer S, Lockhart NC, Demchok J, Moore HF et al (2013) The genotype-tissue expression (GTEx) project. Nat Gene 5:580–585

Machiela MJ, Chanock SJ (2015) LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics 31:3555–3557

Malik VS, Willett WC, Hu FB (2013) Global obesity: trends, risk factors and policy implications. Nat Rev Endocrinol 9:13–27
Misra A, Shrivastava U (2013) Obesity and dyslipidemia in South Asians. Nutrients 5:2708–2733
Monteiro MC, Sanyal M, Cleary ML, Sengenès C, Bouloumié A, Dani C, Billon N (2011) PBX1: a novel stage-specific regulator of adipocyte development. Stem Cells 29:1837–1848
Nozawa RS, Nagao K, Masuda HT, Iwasaki O, Hirota T, Nozaki N, Kimura H, Obuse C (2010) Human POGZ modulates dissociation of HP1α from mitotic chromosome arms through Aurora B activation. Nat Cell Biol 12:719–727
Pigeyre M, Yazdi FT, Kaur Y, Meyre D (2016) Recent progress in genetics, epigenetics and metagenomics unveils the pathophysiology of human obesity. Clin Sci 130:943–986
Pollak NM, Hoffman M, Goldberg IJ, Drosatos K (2018) Krüppel-like factors: crippling and uncrippling metabolic pathways. JACC Basic Transl Sci 3:132–156
Popejoy AB, Fullerton SM (2016) Genomics is failing on diversity. Nature 538:161–164
Prakash Dwivedi O, Tabassum R, Chauhan G, Kaur I, Ghosh S, Marwaha RK, Tandon N, Bharadwaj D (2013) Strong influence of variants near MC4R on adiposity in children and adults: a cross-sectional study in Indian population. J Hum Genet 58:27–32
Prasad G, Bandesh K, Giri AK, Kauser Y, Chanda P, Parekatt V, Mathur S, Madhu SV, Venkatesh P, Bhansali A, Marwaha RK, Basu A, Tandon N, Bharadwaj D (2019a) Genome-wide association study of metabolic syndrome reveals primary genetic variants at CETP locus in Indians. Biomolecules 9:E321
Prasad G, Giri AK, Basu A, Tandon N, Bharadwaj D, INDICO1 (2019b) Genomewide association study for C-reactive protein in Indians replicates known associations of common variants. J Genet 98:20
Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575
Qi L, Cho YA (2008) Gene–environment interaction and obesity. Nutr Rev 66:684–694
Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D (1997) GeneCards: integrating information about genes, proteins and diseases. Trends Genet 13:163
Skwara P, Schömig E, Gründemann D (2017) A novel mode of operation of SLC22A11: membrane insertion of estrone sulfate versus translocation of uric acid and glutamate. Biochem Pharmacol 128:74–82
Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, Liu B, Matthews P, Ong G, Pell J, Silman A, Young A, Sprosen T, Peakman T, Collins R (2015) UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 12:e1001779
Tabassum R, Mahendran Y, Dwivedi OP, Chauhan G, Ghosh S, Marwaha RK, Tandon N, Bharadwaj D (2012) Common variants of IL6, LEPR, and PBEF1 are associated with obesity in Indian children. Diabetes 61:626–631
Tabassum R, Chauhan G, Dwivedi OP, Mahajan A, Jaiswal A, Kaur I, Bandesh K, Singh T, Mathai BJ, Pandey Y, Chidambaram M, Sharma A, Chavali S, Sengupta S, Ramakrishnan L, Venkatesh P, Aggarwal SK, Ghosh S, Prabhakaran D, Srinath RK, Saxena M, Banerjee M, Mathur S, Bhansali A, Shah VN, Madhu SV, Marwaha RK, Basu A, Scaria V, McCarthy MI, DIAGRAM; INDICO, Venkatesan R, Mohan V, Tandon N, Bharadwaj D (2013) Genome-wide association study for type 2 diabetes in Indians identifies a new susceptibility locus at 2q21. Diabetes 62:977–986
Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S (2013) A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics 29:189–196
Tremmel M, Gerdtham UG, Nilsson PM, Saha S (2017) Economic burden of obesity: a systematic literature review. Int J Environ Res Public Health 14:E435
Turner SD (2014) qqman: an R package for visualizing GWAS results using Q–Q and Manhattan plots. bioRxiv. https://doi.org/10.1101/005165
Type 2 Diabetes Knowledge Portal (2019) http://www.type2diabetesgenetics.org/. Accessed 03 Mar 2019
Van Der Sluis S, Dolan CV, Li J, Song Y, Sham P, Posthuma D, Li MX (2015) MGAS: a powerful tool for multivariate gene-based genome-wide association analysis. Bioinformatics 31:1007–1015
Wei Z, Seldin MM, Natarajan N, Djemal DC, Peterson JM, Wong GW (2013) C1q/tumor necrosis factor-related protein 11 (CTRP11), a novel adipose stroma-derived regulator of adipogenesis. J Biol Chem 288:10214–10229
Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26:2190–2191
Willyard C (2014) Heritability: the family roots of obesity. Nature 508:S58–S60
Wilson LE, Harlid S, Xu Z, Sandler DP, Taylor JA (2017) An epigenome-wide study of body mass index and DNA methylation in blood using participants from the Sister Study cohort. Int J Obes 41:194–199
Xu Y, Kim ER, Zhao R, Myers MG Jr, Munzberg H, Tong Q (2013) Glutamate release mediates leptin action on energy expenditure. Mol Metab 2:109–115
Xu Z, Niu L, Li L, Taylor JA (2016) ENmix: a novel background correction method for Illumina HumanMethylation 450 BeadChip. Nucleic Acids Res 44:e20
Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88:76–82
Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, Frayling TM, Hirschhorn J, Yang J, Visscher PM, GIANT Consortium (2018) Meta-analysis of genome-wide association studies for height and body mass index in ~ 700 000 individuals of European ancestry. Hum Mol Genet 27:3641–3649
Zada AA, Pulikkan JA, Bararia D, Geletu M, Trivedi AK, Balkhi MY, Hiddemann WD, Tenen DG, Behre HM, Behre G (2006) Proteomic discovery of Max as a novel interacting partner of C/EBPα: a Myc/Max/Mad link. Leukemia 20:2137–2146
Zhou F, Zhu L, Cui PH, Church WB, Murray M (2010) Functional characterization of nonsynonymous single nucleotide polymorphisms in the human organic anion transporter 4 (hOAT4). Br J Pharmacol 159:419–427

Affiliations

Anil K. Giri[1,2] · Gauri Prasad[1,2] · Khushdeep Bandesh[1,2] · Vaisak Parekatt[1] · Anubha Mahajan[3] · Priyanka Banerjee[1] · Yasmeen Kauser[1,2] · Shraddha Chakraborty[1,2] · Donaka Rajashekar[1] · INDICO · Abhay Sharma[1,2] · Sandeep Kumar Mathur[4] · Analabha Basu[5] · Mark I. McCarthy[3] · Nikhil Tandon[6] · Dwaipayan Bharadwaj[2,7]

✉ Nikhil Tandon
nikhil_tandon@hotmail.com

✉ Dwaipayan Bharadwaj
db@jnu.ac.in

1 Genomics and Molecular Medicine Unit, CSIR-Institute of Genomics and Integrative Biology, New Delhi, India

2 Academy of Scientific and Innovative Research, CSIR-Institute of Genomics and Integrative Biology Campus, New Delhi, India

3 Wellcome Trust Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK

4 Department of Endocrinology, Sawai Man Singh Medical College, Jaipur, Rajasthan, India

5 National Institute of Biomedical Genomics, Kalyani, West Bengal, India

6 Department of Endocrinology and Metabolism, All India Institute of Medical Sciences, New Delhi, India

7 Systems Genomics Laboratory, School of Biotechnology, Jawaharlal Nehru University, New Delhi, India