


# BMJ Open Protocol for a prospective, observational, deep phenotyping study on adipose epigenetic and lipidomic determinants of metabolic homeostasis in South Asian Indians: the Indian Diabetes and Metabolic Health (InDiMeT) study

Nikhil Nadiger,<sup>1</sup> Sarita Devi,<sup>1</sup> Tinku Thomas ,<sup>2</sup> Ambily Sivadas,<sup>1</sup> Rebecca Raj-Kuriyan,<sup>1</sup> Sridar Govindaraj,<sup>3</sup> Anura V Kurpad,<sup>1</sup> Arpita Mukhopadhyay <sup>1</sup>

**To cite:** Nadiger N, Devi S, Thomas T, *et al.* Protocol for a prospective, observational, deep phenotyping study on adipose epigenetic and lipidomic determinants of metabolic homeostasis in South Asian Indians: the Indian Diabetes and Metabolic Health (InDiMeT) study. *BMJ Open* 2021;**11**:e043644. doi:10.1136/bmjopen-2020-043644

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2020-043644>).

Received 10 August 2020  
Revised 15 March 2021  
Accepted 12 April 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Dr Arpita Mukhopadhyay; [arpitam@sjri.res.in](mailto:arpitam@sjri.res.in)

## ABSTRACT

**Introduction** We describe the rationale and broad study design of the Indian Diabetes and Metabolic Health (InDiMeT) study, a new prospective, observational study incorporating extensive epigenetic (DNA methylation) and lipidomic signatures to examine their association with the dysregulation of adipose de novo lipogenesis (DNL) in South Asian Indians. The InDiMeT study aims to use a case-control design to identify genetic and modifiable-environmental-lifestyle associated determinants of (1) epigenomic (DNA methylome) dysregulation of adipose DNL in type 2 diabetes mellitus (T2DM) adipose tissue, (2) identify correlates of epigenomic (DNA methylome) dysregulation of adipose DNL in peripheral blood mononuclear cells (PBMCs) from T2DM subjects and (3) elucidate plasma lipidomic correlates of adipose DNL in T2DM that can be used as biomarkers of adipose tissue dysfunction.

**Methods and analysis** The InDiMeT study will involve recruitment of 176 normoglycaemic and T2DM individuals who will be undergoing laparoscopic surgery for clinical conditions. Extensive phenotyping of the subjects will be conducted and DNA methylome and lipidomic measurements will be made. The adipose DNL pathway genes are likely to be hypermethylated in patients with T2DM with corresponding reduction of gene expression. Correlates of epigenomic (DNA methylome) dysregulation of adipose DNL pathway in PBMCs and their adipose and plasma lipidomic signatures in T2DM subjects could act as early markers of development of T2DM.

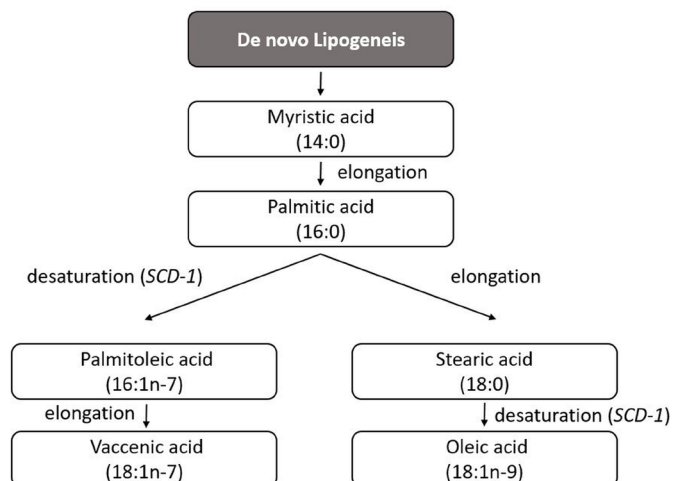
**Ethics and dissemination** For the InDiMeT study, ethical approval for addressing the specific aims has been obtained from the Institutional Ethics Committee, St John's Medical College and Hospital, St John's National Academy of Health Sciences, Bangalore. Findings from this study will be disseminated through scientific publications in peer-reviewed journals, research conferences and via presentations to stakeholders, patients, clinicians, public and policymakers through appropriate channels.

## Strengths and limitations of this study

- The InDiMeT (Indian Diabetes and Metabolic Health) study will assess the epigenome (DNA methylome)-lipidome connection in relation to their contribution to adipose de novo lipogenesis (DNL) and metabolic homeostasis.
- The prospective case-control study design will permit us to assess the blood epigenetic (DNA methylation) and lipidomic markers that exhibit correlations with adipose DNL dysregulation and can also potentially serve as minimally invasive early biomarkers of dysglycaemia.
- Understanding the epigenetic (DNA methylation) mechanisms and their associated lipidomic profiles related to dysregulated adipose DNL in South Asian Indians will provide the rationale for effective interventional strategies targeted to this population and discriminative diagnostic/prognostic biomarkers for prevention of type 2 diabetes mellitus (T2DM) in this high-risk population.
- Though the sample size of 176 individuals will be sufficient to detect differences in adipose DNA methylation between T2DM and normoglycaemic control subjects, the primary specific aim of this study, it is unlikely to provide required sensitivity to detect dysregulation of adipose DNA methylation in specific subgroups of T2DM, such as smokers or those with chronic kidney disease.
- The subjects being recruited are the individuals who come for treatment of their pre-existing clinical conditions and could not be a random representation of the total population.

## INTRODUCTION

India is experiencing a type 2 diabetes mellitus (T2DM) epidemic, attributed to rapid sociodemographic and nutritional changes, making T2DM the 7<sup>th</sup> leading cause of death



**Figure 1** The de novo lipogenesis pathway. SCD-1, Stearoyl-CoA Desaturase-1

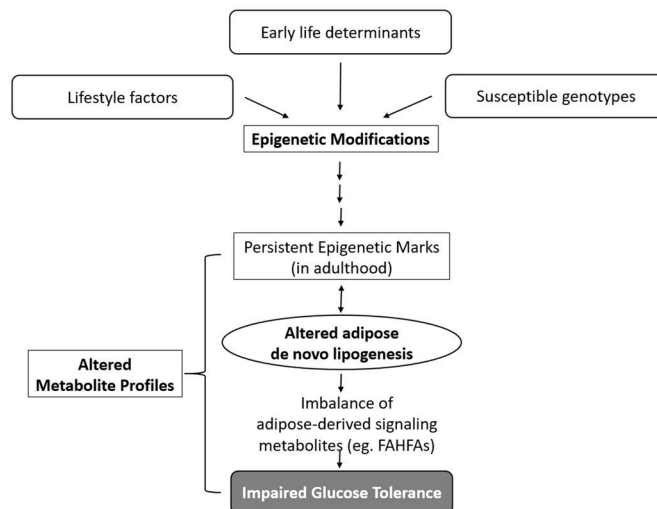
overall and 4<sup>th</sup> amongst the non-communicable diseases in 2019.<sup>1</sup> Younger, non-obese South Asian Indians develop insulin resistance, attributed to inappropriately elevated fat accretion.<sup>2</sup> An important reason for elevated/ectopic fat deposition in South Asian Indians may be their uniquely high carbohydrate diets (~65 energy%), which has been associated with elevated risk of T2DM, likely due to induction of hepatic insulin resistance through increased hepatic de novo lipogenesis (DNL, [figure 1](#)).<sup>3-5</sup>

Unlike hepatic DNL, adipose DNL is understood to be associated with preservation of insulin sensitivity and an overall favourable metabolic phenotype ('metabotype').<sup>6</sup> Strawford *et al* reported negative correlations between adipose DNL and plasma glucose in humans.<sup>7</sup> Mice over-expressing adipose-specific GLUT4 are obese but paradoxically insulin sensitive, likely mediated by elevated adipocyte glucose influx-induced DNL.<sup>8</sup> Further, adipose-specific GLUT4 overexpression can rescue muscle-specific GLUT4 knockout-mediated insulin resistance, in a background of increased adiposity.<sup>9</sup> Therefore, it can be posited that white adipose tissue (AT) versus hepatic DNL may have opposed effects with respect to obesity-associated metabolic complications. Thus, understanding the contribution of adipose DNL in healthy versus T2DM-associated obesity is critical.

Understanding the epigenetic (DNA methylation) mechanisms and their associated lipidome profiles related to exacerbated adipose DNL in South Asian Indians will further provide the rationale for effective interventional strategies targeted to this population and discriminative diagnostic/prognostic biomarkers for prevention of T2DM in this high-risk population.

### Rationale and specific aims

The overarching goal of the Indian Diabetes and Metabolic Health (InDiMeT) study is to identify aetiologically valuable determinants that affect susceptibility to type 2 diabetes at individual, subgroup and population level by testing the hypothesis that the epigenetic (DNA

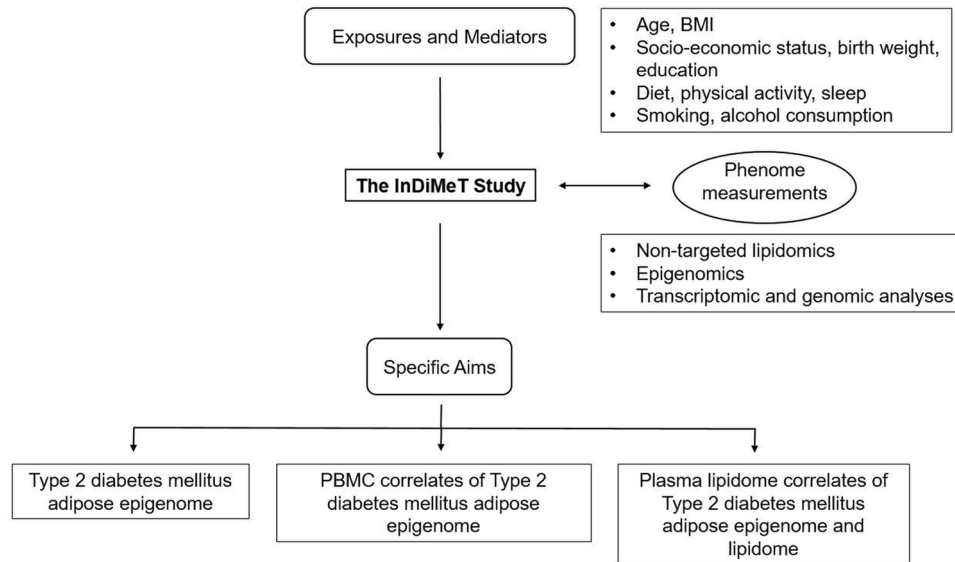


**Figure 2** The proposed link between early life determinants and lifestyle exposures on adipose DNL via epigenetic and metabolomics perturbations leading to impaired glucose tolerance. FAHFAs, Fatty acid esters of hydroxy fatty acids

methylation) dysregulation of adipose DNL induced by high carbohydrate diet and environmental factors is an important step in the development adipose tissue dysfunction leading to type 2 diabetes ([figure 2](#)) in India. The InDiMeT study proposes to achieve this by conducting detailed phenotyping and phenome measurements in a prospective manner in a high-risk population setting. Though animal studies have underlined the epigenetic changes in the adipose associated, few have investigated adipose epigenetics in humans and none in the context of adipose DNL dysregulation and T2DM<sup>10-12</sup> which will help to identify epigenomic dysregulation of adipose DNL in T2DM AT. Although DNA methylation signatures are specific to tissue and cell types, only ~38% of the DNA methylation probes have been reported to be differentially methylated in blood in a tissue-specific fashion, implying likelihood of the rest ~62% probes being conservatively methylated between blood and other tissues.<sup>13</sup> Since blood is an easily available tissue, correlates of epigenomic (DNA methylome) dysregulation of adipose DNL in peripheral blood mononuclear cells (PBMCs) from T2DM subjects will be evaluated to validate the potential use of PBMCs as surrogates of AT dysfunction.

The overall outline of the InDiMeT study is depicted in [figure 3](#). The specific aims of this study are:

1. To identify epigenomic (DNA methylome) dysregulation of adipose DNL in T2DM adipose tissue.
2. To identify correlates of epigenomic (DNA methylome) dysregulation of adipose DNL in PBMCs from T2DM subjects.
3. To elucidate plasma lipidomic correlates of adipose DNL in T2DM that can be used as biomarkers of adipose tissue dysfunction.



**Figure 3** The overall outline of the InDiMeT study. BMI, body mass index; InDiMeT, Indian Diabetes and Metabolic Health; PBMC, peripheral blood mononuclear cell.

**METHODS AND ANALYSIS**

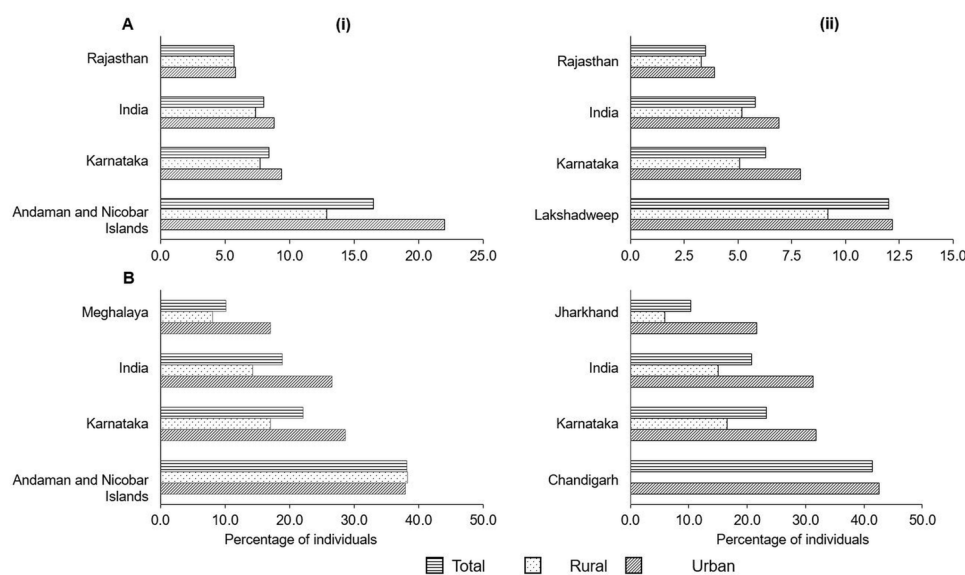
**Study setting**

The InDiMeT study is being conducted at St John’s Medical College and Hospital, a tertiary care hospital in Bangalore district, India. It is surrounded by the state of Tamil Nadu in the southeast, the Bangalore Rural district in the northeast and northwest and the Ramanagara district in the southwest.<sup>14</sup> As per the Indian 2011 Census, the Bangalore district recorded a decadal growth rate increase of 12% in the decade of 2001–2011 compared with the previous decade. This growth was ~4 times higher in the urban areas compared with the rural ones. The sex ratio of the Bangalore district was 916, an 8-point increase compared with the 2001 Census. The

literacy rate was 88% while the work participation rate was 44%, higher for men at 62% compared with women at 25%. As per the National Family Health Survey-4, one-third of the women and 26% of the men in Bangalore (Bangalore urban district) are overweight or obese.<sup>15</sup> In Bangalore urban district, prevalence of elevated random blood sugar measurements stands at 8.3% among women and 10.9% among men which compares to Karnataka state-wide figures of 6.3% for women and 8.4% for men, respectively (figure 4).

**Study design and population**

The InDiMeT study has been designed as a prospective tertiary care hospital-based cohort study. One hundred



**Figure 4** Metabolic indicators. (A) Per cent of surveyed population with random blood glucose levels >140mg/dL (measured in 15–54 year old men and 15–49 year old women) and (B) per cent of surveyed population that are overweight or obese of (i) men and (ii) women in India, Karnataka and the states with minimum and maximum values (adapted from the National Family Health Survey-4).<sup>15</sup>

and seventy-six normoglycaemic and T2DM individuals will be recruited into the study after explanation of details of the study and obtaining informed consent at the department of General Surgery, St John's Medical College and Hospital, St John's National Academy of Health Sciences, Bangalore. Dissemination of information about the broader study goals and likely benefits to the community as well as engagement of potential subjects from the hospital was initiated in August 2020. Recruitment of subjects commenced from 19 November 2020, after ethical approval for the study had been obtained from the Institutional Ethics Committee. Recruitment of subjects will continue until the recruitment target (176 individuals) is achieved. Strategies elaborated below are being employed for recruitment:

- ▶ Chart reviews—From the existing group of patients coming in to the Department of General Surgery, St John's Medical College and Hospital, subjects are screened and then contacted to request for their participation.
- ▶ Databases—From the database of patients coming to the St John's Medical College and Hospital, those who are eligible to participate in the study are contacted for possible recruitment.

#### Inclusion criteria

Inclusion criteria are 35–60 years age; normoglycaemic and T2DM; normal and high body mass index (BMI); men and women who are undergoing laparoscopic surgeries for their clinical conditions such as cholecystectomies, hernia or splenectomy are invited to participate in the study.

#### Exclusion criteria

Exclusion criteria are age outside the range of 35–60 years, individuals who are unwilling to participate in the study, individuals participating in any other study and those who tested positive for hepatitis (HBsAg: Hepatitis B surface antigen), HIV or syphilis (VDRL test: Venereal disease research laboratory test) infections. Those who have serious pre-existing medical condition, those undergoing laparoscopic surgery for clinical conditions other than cholecystectomies, hernia or splenectomy will be excluded, and these will be defined as conditions that require chronic or daily medical therapy, including medications, for T2DM, connective tissue diseases, inflammatory bowel disease, active tuberculosis and symptomatic heart disease.

#### Sample size calculation

Based on the study by Nilsson *et al* the sample size required to observe a mean difference of 7% in DNA methylation in AT with an SD of 10% between T2DM and normoglycaemic controls with 90% power and 5% level of significance is 44 per group. As AT DNA methylation differences could be associated with BMI of the subjects, 44 individuals will be recruited in the normal (BMI <25 kg/m<sup>2</sup>) and

high BMI (BMI ≥25 kg/m<sup>2</sup>) groups of normoglycaemic controls and T2DM (35–60 years, men and women).<sup>10</sup>

Findings of this study could be influenced by selection bias in recruitment of the study subjects. Known sources of selection bias, such as geographical location of the study, which influences wide ranging subject parameters such as dietary diversity, levels of sunlight and air pollutant exposures, as well as recruitment of subjects from a hospital, and not from the community, are being addressed by ensuring recruitment of case and control subjects in tandem from a single study site.

#### Data collection

##### Socio-demographic data collection

Detailed socio-demographic data will be collected by trained personnel using well-structured questionnaires as elaborated in [box 1](#). The questionnaire will be provided and explained in English or local language in which the subject is most comfortable.

Socioeconomic status will be calculated based on scoring relevant for the Indian scenario.<sup>16</sup>

##### Medical history and medication and supplement use data collection

Participants will be asked regarding their birth weight, recent morbidity conditions and the medications and supplements they are currently taking. Details of the type and usage of these medications and supplements will be recorded by trained staff ([box 1](#)). Participants will be requested to provide these details from the prescriptions or their hospital records will be sought after.

##### Vital sign measurements

Participants will be requested to relax and then blood pressure, heart rate and pulse will be measured and recorded by trained staff.

##### Body composition measurement by DXA scanning and anthropometry

Participants will be weighed in minimal clothing using a digital scale, to a precision of 0.1 kg. The height of the subjects will be recorded to the nearest 0.1 cm. Whole body and regional body composition will also be estimated using dual-energy X-ray absorptiometry (DXA) (DPXMD 7254, Lunar Corporation, Madison, Wisconsin, USA). The total body fat (BF) will be measured and both fat and muscle (ALST mass: Appendicular Lean Soft Tissue mass) will be expressed as a percentage of body weight (% BF and % ALST).

##### Laboratory assessments

Blood samples before and after their surgery will be collected from the participants. The blood sample collected before surgery will be a fasting sample as the participants abstain from eating anything 10 hours before surgery, as a preoperative requirement for surgery. These samples will be used for assessments as elaborated in [box 1](#). Muscle insulin sensitivity and  $\beta$ -cell function will be

## Box 1 Exposure and outcome measurements within the InDiMeT (Indian Diabetes and Metabolic Health) study

### General questionnaires

#### Socio-demographic

Age, ethnicity and religion, family composition, education, employment status and type, household income, socioeconomic status.

#### Medical history

Medical history related to birth weight, chronic and infectious diseases, last 1 month's morbidity conditions, current and past usage of tobacco (cigarettes, beedis or chewing tobacco) and alcohol, family history of diabetes and cardiovascular disease, details on most recent routine clinical investigations (haemoglobin, fasting/random/postprandial glucose, OGTT glucose, HbA1C, serum insulin and C-peptide, urea, creatinine, calcium, phosphorus, TSH, uric acid, serum cholesterol, HDL, LDL, VLDL, triglycerides, serum amylase and lipase, total protein, albumin, total bilirubin, direct and indirect bilirubin, SGOT, SGPT, alkaline phosphatase, GGT, sodium, potassium, chloride, routine urine test, urine albumin).

#### Medication and supplement use

Type, brand, dose and duration of medication, vitamin and mineral supplements (clinician prescribed or over the counter).

#### Dietary habits

3 months food-frequency questionnaire.

#### Physical activity

Daily, weekly and monthly frequency of sports activities, manual labour related hobbies and household chores, sedentary activities; occupational classification in terms of physical activity; sleep habits.

### Anthropometry, body composition measurements and laboratory assessments

#### Vital sign measurements

Blood pressure, heart rate and pulse.

#### Body composition and anthropometry measurements

DXA scanning, measurement of height and weight.

#### Laboratory assessments

Blood investigations: Blood glucose level, HbA1C, serum insulin, C-peptide, C reactive protein, glucagon-like peptide and glucose-dependent insulinotropic polypeptide.

### Biobanking

#### Collection of blood samples

Separation and storage of plasma for lipidomic analyses.

#### Collection of abdominal visceral adipose tissue

Done during laparoscopic surgeries; collection and storage of adipose tissue for lipidomic, epigenomic, genomic and gene expression analyses.

### Phenome measurements

#### Non-targeted lipidomic analyses

On adipose tissue and plasma samples.

#### Epigenomic analysis

On adipose tissue samples and PBMCs.

#### Gene expression analysis

On adipose tissue samples.

#### Genomic analysis

On PBMCs.

DXA, dual-energy X-ray absorptiometry; GGT, gamma-glutamyl transferase; HbA1C, glycosylated haemoglobin; HDL, high-density lipoprotein; LDL, low-density lipoprotein; OGTT, oral glucose tolerance test; PBMC, peripheral blood mononuclear cells; SGOT, serum glutamic oxaloacetic transaminase; SGPT, serum glutamic pyruvic transaminase; TSH, thyroid-stimulating hormone; VLDL, very low-density lipoprotein.

assessed by the homeostasis model assessment-of insulin resistance and of  $\beta$ -cell function, respectively.<sup>17 18</sup>

### Lifestyle factors

#### Dietary data collection

The dietary intake will be assessed using a validated food frequency questionnaire (FFQ). The FFQ will be administered by trained personnel to obtain information about the habitual dietary intake for the preceding 3 months.<sup>19</sup> The nutrients and food groups will be estimated for all the foods listed in the FFQ and summed to obtain the total nutrient or food group intake per day for an individual. Nutrient information will be assessed for 27 macronutrients and micronutrients.

#### Physical activity data collection

The habitual physical activity and sleep habits of the subjects will be assessed by detailed and validated physical activity questionnaire.<sup>20</sup>

### Biobanking

#### Blood collection and storage

Blood samples will be collected before and after the surgery by phlebotomy in EDTA/serum tubes, kept on ice and immediately processed for aliquoting and storage at  $-80^{\circ}\text{C}$ .

#### Abdominal visceral adipose tissue collection and storage

Intra-abdominal omental adipose tissue will be collected during laparoscopic surgery with clinical indications of cholecystectomy, meshplasty (hernia) or splenectomy. The collected tissue will be appropriately stored for phenome measurements (epigenomic and non-targeted lipidomic analyses) as described below.

### Phenome measurements

#### Epigenomic (DNA methylome) analysis

Nucleic acids will be extracted from PBMCs and adipose tissue. Epigenome-wide assessment of DNA methylation levels in CpG islands will be done by reduced representation bisulfite sequencing (RRBS) after bisulfite conversion.<sup>10 21 22</sup>

Gene expression by quantitative real-time PCR will be performed with SYBR Green chemistry for the top 10 differentially methylated loci and for the adipose DNL pathway related genes: *MLXIPL*, *SREBP1*, *ACLY*, *ACCI*, *FASN*, *CPT1A*, *CPT1B*, *SCD* and *ELOVL6*. While lysis of citrate to acetyl CoA, the initial substrate of lipogenesis, is catalysed by ATP-citrate lyase (*ACLY*) in the cytosol, cytosolic acetyl-CoA carboxylase (*ACCI*) condenses acetyl CoA into malonyl CoA.<sup>23</sup> Fatty acid synthase (*FASN*) then sequentially condenses two-carbon units from malonyl CoA to form myristic acid and palmitic acid. Palmitic acid can then either undergo further elongation by elongase 6 (*ELOVL6*) or desaturation by stearoyl-CoA desaturase (*SCD*). Carnitine palmitoyltransferase 1A and 1B (*CPT1A* and *CPT1B*), *SREBP1* and *MLXIPL* are involved in regulation of the DNL pathway. *CPT1A* and *CPT1B* are required for transport of long-chain fatty acyl-CoAs from

cytosol to mitochondria for the beta-oxidation pathway. MLX Interacting Protein Like (MLXIPL), also known as Carbohydrate-Responsive Element-Binding Protein primarily and Sterol Regulatory Element Binding Transcription Factor 1 (SREBP1), to a lesser extent, control and activate transcription of multiple DNL genes such as ACC1 and ELOVL6 in adipocytes.<sup>24</sup>

Genotyping of the subjects will be done using Metabochip (a custom genotyping array for metabolic and cardiovascular traits) or equivalent single nucleotide polymorphism array, such as the Infinium Global Screening Array (Illumina, USA).<sup>25 26</sup>

### Non-targeted lipidomic analysis

Plasma and omental AT samples will be used for non-targeted lipidomic analysis. Samples will be spiked with SPLASH Lipidomix Mass Spec Standard (Avanti Polar Lipids, USA) before lipid extraction and reconstitution. Non-targeted lipidomic analysis will be performed on a high-resolution analytical platform consisting a Vanquish Flex Binary UHPLC coupled to a Q Exactive Orbitrap, high-resolution accurate-mass spectrometer (LC-HRAM-MS, Thermo Scientific, USA). High-resolution accurate mass data generated by Thermo Scientific Q Exactive Orbitrap will be identified and quantified by Thermo Scientific Lipid Search software. The following lipid species are of special interest: palmitic acid, stearic acid, oleic acid, myristic acid, palmitoleic acid, acyl carnitines, palmitic acid esters of hydroxyl stearic acids, furan fatty acid metabolite 3-carboxy-4-methyl-5-propyl-2-furan propanoic acid.<sup>27-30</sup>

### Assessment of adipose DNL

Assessment of adipose DNL will be done by determining AT triglyceride (TG) content of the fatty acids myristic acid (14:0), palmitic acid (16:0) and stearic acid (18:0) that have been reported to be correlated with plasma non-esterified fatty acids (NEFA) 14:0 and 18:0 but not with plasma NEFA 15:0, that is exclusively derived from dietary dairy sources.<sup>31</sup> Further, AT lipid species (from non-targeted lipidomic data) that positively correlate with transcript abundances of adipose DNL pathway genes will be identified as indicators of adipose DNL. It should be noted here that though adipose tissue myristic acid, palmitic acid and stearic acid contents are not direct measures of adipose DNL, they represent end products of DNL and have been reported to be correlated to expression of genes involved in adipose DNL.<sup>31</sup>

### Plasma NEFA correlates of adipose DNL

Plasma NEFA correlates of adipose DNL will be identified in the T2DM and normoglycaemic subjects by correlation analysis of plasma lipidomic data with adipose DNL measures (AT TG content of the fatty acids myristic acid (14:0), palmitic acid (16:0) and stearic acid (18:0)). This is a novel strategy for estimating adipose DNL and identifying plasma correlates of adipose DNL as such indirect methods do not exist at present and the direct

stable isotope based method of measuring DNL primarily measures postprandial hepatic DNL rates.<sup>32</sup>

### Data analysis plan

Continuous data will be presented as either mean $\pm$ SD or median (quartile 1 and quartile 3). The normality of distribution will be examined using the Shapiro-Wilk test and Q-Q plots. Demographic and clinical characteristics will be compared between cases of T2DM and controls by Student's t-test, if normally distributed or Mann-Whitney U test, if non-normally distributed, for continuous variables and by  $\chi^2$  test for categorical variables.

For analysis of RRBS data, the read sequences from the RRBS libraries will be adapted/quality trimmed using Trim Galore<sup>33</sup> and aligned to the reference human genome using a wildcard mapper like BSMAP.<sup>34</sup> DNA methylation scores ( $\beta$  values) will be calculated from the aligned reads.<sup>35</sup> Differentially methylated position (DMP) analysis will be conducted using limma software while adjusting for cell proportions (for PBMCs) and including age and sex as covariates.<sup>36</sup> Statistical significance will be set at  $p < 0.05$  following adjustment for multiple testing using the Benjamini-Hochberg false discovery rate (FDR).<sup>37</sup> Differentially methylated regions will be defined as three or more contiguous probes within a 2 kb distance, each sharing the same direction of methylation change, each achieving a Benjamini-Hochberg FDR corrected  $p < 0.05$  in DMP analysis. Enrichment analysis will be performed to understand the top significantly deregulated gene ontologies and pathways. Additionally, in-depth analysis of methylation status of the CpG islands in the promoter and coding regions of genes associated with the DNL pathway will be conducted.

Association studies between the CpG methylated regions and clinical traits will be performed using linear mixed-model package pyLMM and significant associations will be determined using Bonferroni corrected  $p$  value cut-offs.

Gene expression values will be estimated by normalising target gene expression against expression of house-keeping genes followed by comparison of target gene expression between the groups.

Methylation quantitative trait loci will be identified by testing for association between methylation levels and genotypes of the samples using FDR correction.

Correlation-based clustering analysis of the lipid profiles will be performed and visualised as heatmaps to evaluate the inter-relationships between the lipid species and for identifying signatures that differentiate between T2DM and control samples using supervised and unsupervised approaches. For heatmap analysis, only significantly altered features (significant alterations will be defined) will be analysed and data will be organised using a Euclidean distance map and complete clustering algorithm. Random forest will be used to assess the classification performance of the lipidomics signatures. The significantly perturbed pathways will be identified based on the differentially regulated lipids. For random forest

and pathway impact analyses, data will be filtered using the IQR function and scaled using auto-scaling, in which values are mean-centred and divided by the SD of each variable. For the random forest analysis, the number of predictors to try for each node will be set to the square root of the total number of variables.

To identify the adipose lipid species that correlate with transcript abundances of genes in the adipose DNL pathway, first correlation analysis (Pearson's correlation, for normally distributed data and Spearman's correlation, for non-normally distributed data) will be conducted. This will be followed by regression analysis, with levels of lipid species exhibiting significant correlations as dependent variable and transcript abundance of individual genes in the adipose DNL pathway as independent variables, adjusting for probable confounders such as BMI and sex, if found to be significantly correlated with the dependent variables.

All statistical analysis will be processed and analysed using a combination of SPSS V.25 (IBM), Stata V.14 (StataCorp) and R (R Foundation for Statistical Computing, Vienna).

### Patient and public involvement

Patients and general public were not invited to comment on the study design and were not consulted to formulate study outcomes. Results for the InDiMeT study will be communicated to the patients and public in two complementary ways. Data related to individual participants of the study will be shared with them, if deemed beneficial for improving their health status, in consultation with clinician collaborators on this study and specific participant or their primary care providers. This is mandated by our Institutional Ethics Committee as well. The overall study results will be disseminated to participating individuals and public via appropriate forums for public science outreach.

## ETHICS AND DISSEMINATION

### Ethical considerations and informed consent

Institutional ethical approval for the study has been obtained from the Institutional Ethics Committee (IEC), St John's Medical College and Hospital, St John's National Academy of Health Sciences, Bangalore (IEC Study Ref No. 246/2020, IEC Study title: 'De novo lipogenesis in adiposity and glycaemic control: Epigenetic mechanisms and associated metabolite profiles'). Subject recruitment has been initiated from 19 November 2020.

The objectives and aims of the study are being explained in subject's preferred local language to interested volunteers who fulfil the inclusion criteria. The patient information sheet contains relevant and specific information about the study, collection and storage of samples, genetic analysis and safe storage of subject data. A written, informed consent is being obtained from the subjects before the start of the protocol.

### Data storage, management and protection

Data and resources generated as part of this study will be managed, stored and shared in line with institutional, state and national policies. In brief, specifics are as follows:

#### Organisation

Conventions for naming and storage organisation of data files as well as samples from human subjects (adipose and blood) will be set up, documented and circulated among study team members at the beginning of the study.

#### Storage

Data will be stored in a secure manner in three copies in two different media. Back-up copies of data will be accompanied with metadata that describe the stored data in a standard, understandable format. Data storage and protection support will be obtained from the Institutional Information Technology team.

#### Physical samples

Samples from the study subjects will be processed and stored as per the experimental protocols. Samples in excess of the requirements of the study objectives will be stored appropriately as per the institutional policies as well as policies of the peer-reviewed journals where findings from the study will be submitted for possible publications.

#### Dissemination of data

Sequence data generated from this study will be deposited in GenBank and epigenetic data in Gene Expression Omnibus, while raw and processed lipidomic data will be deposited in appropriate repositories such as Metabolomics Workbench or MetaboLights during the time of publication. Findings from the InDiMeT study will be appropriately communicated in refereed journals and forums for scientific and community science outreach.

## SIGNIFICANCE AND OUTLOOK

The InDiMeT study will provide the base and act as a robust tool for mechanistic interpretations of public health and epidemiological aspects of developing T2DM and its complications in the Indian setting. Though a few large, prospective cohort studies to identify the determinants of T2DM are being conducted in India, to the best of our knowledge, none of them involve stated deep phenotyping or extensive phenome measurements of the participants.<sup>38 39</sup> This study was conceptualised to fill this niche in this country. The carefully monitored and quality-controlled collection and archiving of biological samples, along with availability of anthropometric, socio-demographic, nutritional and medical data on the participants will aid toward planning and implementation of an array of hypothesis-driven mechanistic studies.

The InDiMeT study will not be without limitations. Notable ones will include the low probability of participation by individuals of specific subsets of the population due to the voluntary nature of participation. The

subjects being recruited are the individuals who come for treatment of their pre-existing clinical conditions could not be a random representation of the total population. Further, the study setting being a tertiary care hospital in a large, urban area in India, a country known for its large regional variations in cuisine and culture-associated dietary habits, composition and amounts of habitual dietary intakes of the study subjects are likely to be heterogeneous. These factors could act as confounders and limit our ability to detect differences in DNA methylation and lipid profiles that are of low effect sizes. Detailed recording of these factors as proposed, will however help us in understanding if these factors do act as confounders in our study and allow us and others to improve design of similar studies in future.

#### Author affiliations

<sup>1</sup>Division of Nutrition, St John's National Academy of Health Sciences, Bangalore, Karnataka, India

<sup>2</sup>Department of Biostatistics, St. John's Medical College and Hospital, Bangalore, Karnataka, India

<sup>3</sup>Department of General Surgery, St John's Medical College and Hospital, Bangalore, Karnataka, India

**Acknowledgements** The contribution of research fellow Bajana Nongkhlaw, research assistant Beena Bose and nutritionist J Jayakumar is gratefully acknowledged.

**Contributors** The authors of the manuscript and their individual contributions towards the manuscript are as follows. AM, TT and AVK: designed the research; NN, SD and SG: conduct of the study; AM, TT, AS, RR-K and AVK: analysis of the data, performing statistical analyses and interpretation of the results; AM and AVK: obtained funding; AM: had the idea for and led writing of the manuscript and had primary responsibility for the final content of the manuscript; and all authors critically reviewed the manuscript, read and approved the final version of the manuscript.

**Funding** This study is supported by the Wellcome Trust/DBT India Alliance Fellowship awarded to AM (grant number IA/CPHI/19/1/504593) and to AVK (grant number IA/M/14/1/501681) and by the Department of Biotechnology, Government of India (Grant Sanction order BT/PR24342/PFN/20/1314/2017 to AM and AVK).

**Competing interests** None declared.

**Patient and public involvement** Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

#### ORCID iDs

Tinku Thomas <http://orcid.org/0000-0002-1786-6076>

Arpita Mukhopadhyay <http://orcid.org/0000-0001-8260-5385>

#### REFERENCES

- World Health Organization. Global health estimates: leading causes of death, 2011. World Health Organization, Geneva. Available: <https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-leading-causes-of-death> [Accessed May 4th 2021].
- Misra A, Madhavan M, Vikram NK, et al. Simple anthropometric measures identify fasting hyperinsulinemia and clustering of cardiovascular risk factors in Asian Indian adolescents. *Metabolism* 2006;55:1569–73.
- Mohan V, Radhika G, Sathya RM, et al. Dietary carbohydrates, glycaemic load, food groups and newly detected type 2 diabetes among urban Asian Indian population in Chennai, India (Chennai urban rural epidemiology study 59). *Br J Nutr* 2009;102:1498–506.
- Acheson KJ, Schutz Y, Bessard T, et al. Carbohydrate metabolism and de novo lipogenesis in human obesity. *Am J Clin Nutr* 1987;45:78–85.
- Perry RJ, Samuel VT, Petersen KF, et al. The role of hepatic lipids in hepatic insulin resistance and type 2 diabetes. *Nature* 2014;510:84–91.
- Yilmaz M, Claiborn KC, Hotamisligil GS. De Novo lipogenesis products and endogenous Lipokines. *Diabetes* 2016;65:1800–7.
- Strawford A, Antelo F, Christiansen M, et al. Adipose tissue triglyceride turnover, de novo lipogenesis, and cell proliferation in humans measured with  $^2\text{H}_2\text{O}$ . *Am J Physiol Endocrinol Metab* 2004;286:E577–88.
- Shepherd PR, Gnudi L, Tozzo E, et al. Adipose cell hyperplasia and enhanced glucose disposal in transgenic mice overexpressing GLUT4 selectively in adipose tissue. *J Biol Chem* 1993;268:22243–6.
- Carvalho E, Kotani K, Peroni OD, et al. Adipose-specific overexpression of GLUT4 reverses insulin resistance and diabetes in mice lacking GLUT4 selectively in muscle. *Am J Physiol Endocrinol Metab* 2005;289:E551–61.
- Nilsson E, Jansson PA, Perfilyev A, et al. Altered DNA methylation and differential expression of genes influencing metabolism and inflammation in adipose tissue from subjects with type 2 diabetes. *Diabetes* 2014;63:2962–76.
- Orozco LD, Farrell C, Hale C, et al. Epigenome-wide association in adipose tissue from the METSIM cohort. *Hum Mol Genet* 2018;27:1830–46.
- Broholm C, Olsson AH, Perfilyev A, et al. Epigenetic programming of adipose-derived stem cells in low birthweight individuals. *Diabetologia* 2016;59:2664–73.
- Lowe R, Slodkowitz G, Goldman N, et al. The human blood DNA methylome displays a highly distinctive profile compared with other somatic tissues. *Epigenetics* 2015;10:274–81.
- Register General I. *Census of India 2011: provisional population totals-India data sheet*, 2011.
- Indian Institute for Population Sciences. *National family health survey - 4*. Mumbai, 2017. [http://rchiips.org/NFHS/factsheet\\_NFHS-4.shtml](http://rchiips.org/NFHS/factsheet_NFHS-4.shtml)
- Kumar N, Kishore J, Gupta N, Kuppaswamy's socioeconomic scale: updating income ranges for the year 2012. *Indian J Public Health* 2012;56:103.
- Matthews DR, Hosker JP, Rudenski AS, et al. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* 1985;28:412–9.
- Wallace TM, Levy JC, Matthews DR. Use and abuse of HOMA modeling. *Diabetes Care* 2004;27:1487–95.
- Bharathi AV, Kurpad AV, Thomas T, et al. Development of food frequency questionnaires and a nutrient database for the prospective urban and rural epidemiological (PURE) pilot study in South India: methodological issues. *Asia Pac J Clin Nutr* 2008;17:178–85.
- Bharathi AV, Kuriyan R, Kurpad AV, et al. Assessment of physical activity using accelerometry, an activity diary, the heart rate method and the Indian migration study questionnaire in South Indian adults. *Public Health Nutr* 2010;13:47–53.
- Harris RA, Wang T, Coarfa C, et al. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nat Biotechnol* 2010;28:1097–105.
- Meissner A, Gnirke A, Bell GW, et al. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res* 2005;33:5868–77.
- Strable MS, Ntambi JM. Genetic control of de novo lipogenesis: role in diet-induced obesity. *Crit Rev Biochem Mol Biol* 2010;45:199–214.
- Song Z, Xiaoli AM, Yang F. Regulation and Metabolic Significance of De Novo Lipogenesis in Adipose Tissues. *Nutrients* 2018;10:1383.
- Morris AP, Voight BF, Teslovich TM, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* 2012;44:981–90.
- Voight BF, Kang HM, Ding J, et al. The MetaboChip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet* 2012;8:e1002793.
- Ma W, Wu JHY, Wang Q, et al. Prospective association of fatty acids in the de novo lipogenesis pathway with risk of type 2 diabetes: the cardiovascular health study. *Am J Clin Nutr* 2015;101:153–63.
- Mihalik SJ, Michaliszyn SF, de las Heras J, et al. Metabolomic profiling of fatty acid and amino acid metabolism in youth with obesity and type 2 diabetes: evidence for enhanced mitochondrial oxidation. *Diabetes Care* 2012;35:605–11.



- 29 Yore MM, Syed I, Moraes-Vieira PM, *et al.* Discovery of a class of endogenous mammalian lipids with anti-diabetic and anti-inflammatory effects. *Cell* 2014;159:318–32.
- 30 Schooneman MG, Vaz FM, Houten SM, *et al.* Acylcarnitines: reflecting or inflicting insulin resistance? *Diabetes* 2013;62:1–8.
- 31 Roberts R, Hodson L, Dennis AL, *et al.* Markers of de novo lipogenesis in adipose tissue: associations with small adipocytes and insulin sensitivity in humans. *Diabetologia* 2009;52:882–90.
- 32 Lee JJ, Lambert JE, Hovhannisyan Y, *et al.* Palmitoleic acid is elevated in fatty liver disease and reflects hepatic lipogenesis. *Am J Clin Nutr* 2015;101:34–43.
- 33 Krueger F. Trim galore. A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. 2015:516–7.
- 34 Xi Y, Li W. BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinformatics* 2009;10:232.
- 35 Adusumalli S, Mohd Omar MF, Soong R, *et al.* Methodological aspects of whole-genome bisulfite sequencing analysis. *Brief Bioinform* 2015;16:369–79.
- 36 Ventham NT, Kennedy NA, Adams AT, *et al.* Integrative epigenome-wide analysis demonstrates that DNA methylation may mediate genetic risk in inflammatory bowel disease. *Nat Commun* 2016;7:13507.
- 37 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* 1995;57:289–300.
- 38 Chatterjee S, Majumder PP. Kalyani cohort – the first platform in Eastern India for longitudinal studies on health and disease parameters in peri-urban setting. *Glob. Health Epidemiol.* 2017;2:e2.
- 39 Nair M, Ali MK, Ajay VS, *et al.* CARRS surveillance study: design and methods to assess burdens from multiple perspectives. *BMC Public Health* 2012;12:701.