



# Increased particle size of triacylglycerol-enriched remnant lipoproteins, but not their plasma concentration or lipid content, augments risk prediction of incident type 2 diabetes

Luiz Sérgio F. Carvalho<sup>1,2,3</sup> · Isabela M. Benseñor<sup>4</sup> · Ana C. C. Nogueira<sup>3</sup> · Bruce B. Duncan<sup>5</sup> · Maria I. Schmidt<sup>5</sup> · Michael J. Blaha<sup>6</sup> · Peter P. Toth<sup>6,7</sup> · Steven R. Jones<sup>6</sup> · Raul D. Santos<sup>4,8</sup> · Paulo A. Lotufo<sup>4</sup> · Andrei C. Sposito<sup>2</sup> · on behalf of the ELSA-Brasil study

Received: 13 July 2020 / Accepted: 21 September 2020 / Published online: 7 November 2020  
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

## Abstract

**Aims/hypothesis** Type 2 diabetes prevention requires the accurate identification of those at high risk. Beyond the association of fasting serum triacylglycerols with diabetes, triacylglycerol-enriched remnant lipoproteins (TRLs) more accurately reflect pathophysiological changes that underlie progression to diabetes, such as hepatic insulin resistance, pancreatic steatosis and systemic inflammation. We hypothesised that TRL-related factors could improve risk prediction for incident diabetes.

**Methods** We included individuals from the Brazilian Longitudinal Study of Adult Health cohort. We trained a logistic regression model for the risk of incident diabetes in 80% of the cohort using tenfold cross-validation, and tested the model in the remaining 20% of the cohort (test set). Variables included medical history and traits of the metabolic syndrome, followed by TRL-related measurements (plasma concentration, TRL particle diameter, cholesterol and triacylglycerol content). TRL features were measured using NMR spectroscopy. Discrimination was assessed using the area under the receiver operating characteristic curve (AUROC) and the area under the precision-recall curve (AUPRC).

**Results** Among 4463 at-risk individuals, there were 366 new cases of diabetes after a mean ( $\pm$ SD) of 3.7 ( $\pm$ 0.63) years of follow-up. We derived an 18-variable model with a global AUROC of 0.846 (95% CI: 0.829, 0.869). Overall TRL-related markers were not associated with diabetes. However, TRL particle diameter increased the AUROC, particularly in individuals with HbA<sub>1c</sub> <39 mmol/mol (5.7%) (hold-out test set [ $n$  = 659]; training-validation set [ $n$  = 2638]), but not in individuals with baseline HbA<sub>1c</sub> 39–46 mmol/mol (5.7–6.4%) (hold-out test set [ $n$  = 233]; training-validation set [ $n$  = 933]). In the subgroup with baseline HbA<sub>1c</sub> <39 mmol/mol (5.7%), AUROC in the test set increased from 0.717 (95% CI 0.603, 0.818) to 0.794 (95% CI 0.731, 0.862), and AUPRC in the test set rose from 0.582 to 0.701 when using the baseline model and the baseline model plus TRL particle diameter, respectively. TRL particle diameter was highly correlated with obesity, insulin resistance and inflammation in those with impaired fasting glucose at baseline, but less so in those with HbA<sub>1c</sub> <39 mmol/mol (5.7%).

**Supplementary Information** The online version of this article (<https://doi.org/10.1007/s00125-020-05322-1>) contains peer-reviewed but unedited supplementary material, which is available to authorised users.

✉ Luiz Sérgio F. Carvalho  
lsergio@clarityhealth.com.br

<sup>1</sup> Data Lab, Clarity Healthcare Intelligence, Jundiaí, SP, Brazil

<sup>2</sup> Cardiology Division, Faculty of Medical Sciences, State University of Campinas (Unicamp), Campinas, SP, Brazil

<sup>3</sup> Laboratory of Data for Quality of Care and Outcomes Research, Institute for Strategic Management in Healthcare DF (IGESDF), Brasília, DF, Brazil

<sup>4</sup> Center for Clinical and Epidemiological Research, University Hospital, University of São Paulo, São Paulo, SP, Brazil

<sup>5</sup> Postgraduate Studies Program in Epidemiology, School of Medicine and Hospital de Clínicas, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

<sup>6</sup> The Johns Hopkins Ciccarone Center for the Prevention of Heart Disease, Baltimore, MD, USA

<sup>7</sup> Preventive Cardiology, CGH Medical Center, Sterling, IL, USA

<sup>8</sup> Lipid Clinic Heart Institute (InCor), University of São Paulo, Medical School Hospital, São Paulo, SP, Brazil

## Research in context

### What is already known about this subject?

- Each year, up to 3% of individuals in the general population who are not in the impaired fasting glucose range develop diabetes worldwide
- Little is known regarding risk prediction of type 2 diabetes before the detection of glucose abnormalities
- Among the pathophysiological processes that play a role in type 2 diabetes, hepatocyte insulin resistance and the loading of triacylglycerol-enriched remnant lipoproteins (TRLs) with lipids (resulting in large particles) may be useful for risk prediction

### What is the key question?

- Could the measurement of TRL particle composition and properties by NMR spectroscopy improve risk prediction for type 2 diabetes?

### What are the new findings?

- This study found that, when included in a risk prediction model, TRL particle diameter improved the risk prediction of diabetes after 3.7 years of follow-up as compared with a model that did not include this variable, particularly in individuals with normal HbA<sub>1c</sub> levels at baseline
- TRL particle size was highly correlated with obesity, insulin resistance and inflammation in individuals with impaired fasting glucose at baseline, but this association was attenuated in those with HbA<sub>1c</sub> <39 mmol/mol (5.7%)
- These data support TRL particle size as a risk factor that is changed early in the pathophysiological processes that lead to the development of type 2 diabetes, even before glucose abnormalities are established

### How might this impact on clinical practice in the foreseeable future?

- Our findings provide a step forward in risk prediction of type 2 diabetes by allowing identification of individuals at high risk, but with normal HbA<sub>1c</sub> values. A potential causal relationship between TRL particle size and incident diabetes should be investigated further

**Conclusions/interpretation** TRL particle diameter improves the prediction of diabetes, but only in individuals with HbA<sub>1c</sub> <39 mmol/mol (5.7%) at baseline. These data support TRL particle diameter as a risk factor that is changed early in the course of the pathophysiological processes that lead to the development of type 2 diabetes, even before glucose abnormalities are established.

**Keywords** Particle diameter · Risk prediction · Triacylglycerol-enriched remnant lipoprotein · Triacylglycerol-rich lipoprotein · Type 2 diabetes

### Abbreviations

AUPRC	Area under the precision-recall curve
AUROC	Area under the receiver operating characteristic curve
ELSA-Brasil	Estudo Longitudinal de Saúde do Adulto (Brazilian Longitudinal Study of Adult Health)
FPG	Fasting plasma glucose
hsCRP	High-sensitivity C-reactive protein
IFG	Impaired fasting glucose
IGT	Impaired glucose tolerance
ROC	Receiver operating characteristic
TRL	Triacylglycerol-enriched remnant lipoprotein

### Introduction

Accurate prediction of the risk of type 2 diabetes permits more targeted allocation of resources for the implementation of preventative strategies. Clinical trials on lifestyle modifications and drug interventions have convincingly shown that these measures can prevent the onset of diabetes [1].

Up to 3% of those without impaired fasting glucose in the general population worldwide may develop diabetes every year [2]. Glycaemic status at baseline is the most important predictor of future diabetes, and individuals characterised by

impaired fasting glucose show a two- to tenfold higher risk for diabetes compared with those without impaired fasting glucose [3]. Conversely, among those without impaired fasting glucose, glycaemic status has a smaller predictive value, and thus the long-term risk prediction of diabetes in this population is much less precise [4].

Aiming to improve risk prediction of type 2 diabetes, several studies evaluated the addition of complex traits and measures, such as fasting insulin, C-peptide, HOMA-IR and the homeostasis model  $\beta$ -cell sensitivity-index to risk prediction models, but these variables have not been conclusively found to improve risk prediction [5]. Simple formulas that account for age, parental diabetes, obesity and traits of the metabolic syndrome are enough to effectively predict diabetes risk in the general population [3, 6], but there is still an unmet need in individuals without impaired fasting glucose.

Beyond the association between fasting serum triacylglycerols and diabetes, evidence suggests that plasma concentrations of triacylglycerol-enriched remnant lipoproteins (TRLs) and the cholesterol content in TRL more accurately reflect the pathophysiological processes that underlie progression to diabetes [7–9]. Events such as hepatic and pancreatic steatosis and heightened systemic inflammation are strongly related to the presence of large VLDL particles, namely VLDL1 [10, 11]. Therefore, we hypothesised TRL-related measures could improve risk prediction for new-onset diabetes.

To investigate the association of TRL-related particles with incident type 2 diabetes, we focused on developing a parsimonious prediction model with participants from the Estudo Longitudinal de Saúde do Adulto (Brazilian Longitudinal Study of Adult Health; ELSA-Brasil).

## Methods

**Study design and participants** ELSA-Brasil is a large, racially mixed, longitudinal cohort, with baseline data collected in 2008–2010, that recruited civil servants from a broad range of sociodemographic and clinical backgrounds from six cities of Brazil (Belo Horizonte, Porto Alegre, Rio de Janeiro, Salvador, São Paulo and Vitoria). The study included active or retired civil servants of public universities or research institutions who were 35–74 years old. Deidentified data were used, and the research protocol was approved by the ethics committee at each participating institution and by the National Research Ethics Committee [12]. This study is registered with [clinicaltrials.gov](https://clinicaltrials.gov) (ClinicalTrials.gov registration no. NCT02320461).

The ELSA-Brasil cohort encompasses a total of 15,105 men and women, 2505 of whom presented with diabetes. As seen in electronic supplementary material (ESM) Fig. 1, after exclusion of individuals who were taking glucose-lowering

therapies at baseline or who did not attend visit 2, full data were available from 10,729 participants.

The primary analyses in this study aimed to evaluate the impact of TRL components measured by NMR spectroscopy on predicting risk for developing type 2 diabetes. A total of 598 individuals from the São Paulo site were excluded because of the presence of diabetes at baseline ( $n = 557$ ), the use of glucose-lowering therapies at baseline ( $n = 3$ ), or death or loss during follow-up ( $n = 38$ ). Thus, of the 5061 participants of ELSA-Brasil from the São Paulo Research Center who had full lipid NMR spectroscopy data, we analysed data from 4463 participants without diabetes.

To investigate potential relationships between TRL-related markers and inflammation or adipokines, we measured several biomarkers in a random sample of 970 individuals without diabetes or CVD from the 4463 participants of ELSA-Brasil in São Paulo.

**Measurements and protocols** Blood samples were collected by certified technicians and measurements were obtained during visits 1 and 2 followed standardised protocols [13]. We recorded age, medication use and past medical history using a questionnaire. Medication use was verified against medications or prescriptions brought to the clinic. We measured weight and height, and calculated BMI ( $\text{kg}/\text{m}^2$ ).

Participants were deemed to be fasting if they had no oral intake other than water or medication 12 h prior to sample collection. We obtained blood samples by venepuncture and followed standardised procedures for a 2 h 75 g OGTT [14]. Plasma samples were frozen and stored in liquid nitrogen for up to 5 years. Plasma glucose concentrations were measured using the hexokinase method; insulin concentrations with an immunoenzymatic assay (Siemens, USA); triacylglycerol and HDL-cholesterol concentrations with enzymatic methods (Siemens); C-reactive protein (hsCRP) was measured with a high-sensitivity assay by immunochemistry-nephelometry (BNII; Siemens); and HbA<sub>1c</sub> with high-pressure liquid chromatography (Bio-Rad, USA). Intraclass correlation coefficients by analysis of replicate pairs of samples drawn at baseline were 94% (95% CI 0.86, 0.97) for HbA<sub>1c</sub> and 99% (95% CI 0.95, 1.0) for plasma glucose concentration [15].

High-field  $^1\text{H}$ -NMR spectroscopy was performed by LabCorp (NMR LipoProfile test spectra, USA) to assess TRL plasma concentration, particle size, cholesterol and triacylglycerol content. NMR spectroscopy methods are described elsewhere [16]. Briefly,  $^1\text{H}$ -NMR-based lipoprotein profiling is a unique platform for investigating lipoprotein particle distributions, primarily because different lipoprotein fractions and subfractions have different magnetic susceptibilities that will broadcast different signals, with amplitudes reflecting particle concentrations [16].

ELISA kits were used for the determination of adiponectin and leptin (Enzo Life Sciences, USA) and asymmetric

dimethylarginine (Affinity Biologicals, Canada) concentrations. IL-6 and TNF- $\alpha$  levels were determined using the Bio-Plex Pro Human Cytokine four-plex assay panel (Bio-Rad, Brazil). Intra-assay coefficients of variation ranged from 1.8% to 7.2%, while inter-assay coefficients of variation varied from 0.9% to 9.1% [13].

**Characterisation of diabetes and intermediate hyperglycaemia at baseline** At visit 1, we characterised diabetes on the basis of self-reported information and laboratory measurements. We did not consider reports of diabetes diagnosed only during pregnancy. Those without a previous diagnosis were assessed for undiagnosed diabetes on the basis of their laboratory values, and ascertained as having newly diagnosed diabetes if they reached the thresholds for fasting plasma glucose (FPG) of 7.0 mmol/l, 2 h plasma glucose during an OGTT of 11.1 mmol/l or HbA<sub>1c</sub>  $\geq$ 48 mmol/mol (6.5%).

Among participants who did not meet the criteria for diabetes, impaired glucose tolerance (IGT) was defined as a 2 h plasma glucose of  $\geq$ 7.8 mmol/l, impaired fasting glucose (IFG) was based on the ADA criteria (ADA-IFG criteria) of FPG  $\geq$ 5.5 mmol/l, and high HbA<sub>1c</sub> was defined according to the ADA definition (ADA-HbA<sub>1c</sub> criteria) as a value of  $\geq$ 39 mmol/mol (5.7%) [14].

As described below, we used the ADA-HbA<sub>1c</sub> definition for grouping participants with and without intermediate hyperglycaemia at baseline. This choice was based on the simplicity of a single measurement and the common use of HbA<sub>1c</sub> in clinical practice. In clinical models for predicting incident diabetes, ADA-HbA<sub>1c</sub> criteria did not differ from the IGT criteria or ADA-IFG criteria (ESM Table 4).

**Follow-up** Participants were followed from the baseline visit 1 to visit 2 (2012–2014) for a mean ( $\pm$ SD) of 3.7( $\pm$ 0.6) years. Participants were characterised as developing new diabetes during follow-up if they: (1) started receiving oral hypoglycaemic agents or insulin; or (2) met the criteria described above at the follow-up examination conducted ~4 years after the baseline examination.

**Grouping** In the absence of glucose abnormalities, the risk prediction of diabetes is usually less accurate, and a large body of evidence indicates that predictors that make pathophysiological sense in this subpopulation differ from predictors in populations with impaired fasting glucose [2, 17]. Using this rationale, we split the cohort into two subgroups: (1) individuals with HbA<sub>1c</sub> <39 mmol/mol (5.7%) at baseline ( $n = 3297$ ); and (2) individuals with intermediate hyperglycaemia at baseline ( $n = 1166$ ), according to the ADA-HbA<sub>1c</sub> definition (ESM Fig. 1) [14]. Furthermore, to evaluate if TRL-related measurements improved predictive capacity for incident type 2 diabetes, we randomly split both subgroups into two sets (training-validation and test sets: HbA<sub>1c</sub> <39 mmol/mol

[5.7%],  $n = 2638$ ; HbA<sub>1c</sub> 39–46 mmol/mol [5.7–6.4%],  $n = 933$ ) using a tenfold cross-validation framework (ESM Fig. 2).

**Statistical analyses** Logistic regression models were used to evaluate the relationship between variables and incident diabetes, using the OR and 95% CI. Two models were created following two levels of information: the first level was based on characteristics known to the participant, simple clinical measurements that included traits of the metabolic syndrome (age, sex, parental history of diabetes, hypertension, waist circumference, HbA<sub>1c</sub>) and hsCRP, while the second level also included TRL-related measurements. Among the selected variables, none showed statistically significant multicollinearity across each other. In logistic regression models, non-categorical variables were normalised using  $z$  scores.

Models were compared using receiver operating characteristic (ROC) curve plots and by calculating the area under the ROC curve (AUROC), the area under the precision-recall curve (AUPRC), the F1 score and the Matthews correlation coefficient score [18] for their performance on the test datasets. We built models with the training-validation set using a tenfold cross-validation framework with down-sampling to mitigate outcome imbalance, and then evaluated the model on the test set to obtain an estimate of performance. Finally, to compare the AUROC of linear models with different predictors, we used Stata's *comproc* method [19].

We performed three series of sensitivity analyses. First, we investigated the effect of excluding all incident cases of diabetes characterised only on the basis of a single FPG abnormality, but not according to HbA<sub>1c</sub> or the OGTT. Second, we evaluated the impact of using different criteria to ascertain glucose abnormalities at baseline, such as OGTT/IGT and the ADA-IFG criteria. Third, we evaluated the effect of ascertaining glucose abnormalities at baseline if any test was positive (ADA-HbA<sub>1c</sub>, IGT or ADA-IFG). Because multiple comparisons were used, only  $p$  values of 0.01 or less were considered statistically significant. Statistical analyses were performed using R v3.5.1 ([www.r-project.org/](http://www.r-project.org/); packages *caret*, *pROC*, *MLmetrics*, *mccr*, *ggplot2*) and Stata v13 (StataCorp, Texas, TX, USA) for Mac.

## Results

**Participant characteristics at baseline** Our final sample included 4463 participants after exclusions because of missing data and individuals lost to follow-up (ESM Fig. 1). After a mean ( $\pm$ SD) of 3.7( $\pm$ 0.63) years of follow-up, 366 (8.2%; 2.23 per 100 patient-year) of the individuals at risk developed diabetes. Among these, 33.4% reported a diabetes diagnosis between visits 1 and 2, and the remaining 66.6% were diagnosed at

**Table 1** Participant characteristics at baseline according to TRL particle size quartiles in the ELSA-Brasil study

Variable	TRL particle size				P value <sup>a</sup>
	1st quartile (<40 nm)	2nd quartile (40–45 nm)	3rd quartile (45–52 nm)	4th quartile (>52 nm)	
<b>HbA<sub>1c</sub> &lt;39 mmol/mol (5.7%)</b>					
<i>n</i>	922	866	783	726	
Age, years	49.51 ± 8.51	50.48 ± 8.78	51.77 ± 9.16	51.97 ± 9.31	<0.001
Sex, male	341 (37.0)	368 (42.5)	384 (49.0)	410 (56.5)	<0.001
Hypertension	188 (20.4)	212 (24.5)	226 (28.9)	265 (36.5)	<0.001
Dyslipidaemia	203 (22.0)	229 (26.4)	263 (33.6)	298 (41.0) <sup>b</sup>	<0.001
AMI	8 (0.9)	8 (0.9)	9 (1.1) <sup>b</sup>	14 (1.9)	0.194
CAD	10 (1.1)	15 (1.7)	12 (1.5)	25 (3.4)	0.004
Heart failure	6 (0.7)	6 (0.7)	11 (1.4)	4 (0.6)	0.219
Stroke	7 (0.8)	10 (1.2)	8 (1.0)	12 (1.7)	0.390
Smoking					<0.001
Never	552 (59.9)	500 (57.7)	406 (51.9)	355 (48.9)	
Former	227 (24.6)	254 (29.3)	261 (33.3)	253 (34.8)	
Present	144 (15.6)	112 (12.9)	116 (14.8)	120 (16.5)	
Alcohol intake	67 (7.3)	64 (7.4)	79 (10.1)	105 (14.5)	<0.001
Parental history of diabetes	341 (37.0)	345 (39.8)	317 (40.5)	309 (42.6)	0.386
Lipid-lowering therapy	42 (4.6)	71 (8.2)	94 (12.0)	92 (12.7)	<0.001
Glucose-lowering therapy	0 (0)	0 (0)	0 (0)	0 (0)	1.000
BMI, kg/m <sup>2</sup>	25.13 ± 3.88	26.19 ± 4.27	27.46 ± 4.56	28.75 ± 4.64	<0.001
Waist circumference, cm	83.13 ± 10.38	86.81 ± 11.39	90.95 ± 11.68	95.31 ± 11.79	<0.001
Systolic BP, mmHg	114.0 ± 15.0	116.95 ± 14.78	119.21 ± 15.67	123.12 ± 16.42	<0.001
Diastolic BP, mmHg	71.39 ± 10.1	73.55 ± 9.97	74.76 ± 9.92	77.89 ± 10.86	<0.001
GFR, ml min <sup>-1</sup> (1.73 m) <sup>-2</sup>	101 (86–117)	101 (87–119)	101 (85–117)	100 (85–119)	0.368
Fasting blood glucose, mmol/l	5.58 ± 0.5	5.74 ± 0.5	5.84 ± 0.6	5.96 ± 0.6	<0.001
Fasting blood triacylglycerol, mmol/l	0.9 (0.7–1.1)	1.1 (0.9–1.4)	1.4 (1.1–1.8)	2.1 (1.5–2.8)	<0.001
HbA <sub>1c</sub> , mmol/mol	31 ± 1.3	32 ± 1.3	32 ± 1.3	32 ± 1.3	0.040
HbA <sub>1c</sub> , %	5.00 ± 0.40	5.04 ± 0.39	5.04 ± 0.40	5.06 ± 0.40	0.040
hsCRP, mg/l	1.0 (0.5–2.3)	1.3 (0.7–3.1)	1.3 (0.7–3.0)	1.6 (0.9–3.5)	<0.001
HOMA-IR	1.1 (0.6–1.7)	1.4 (0.8–2.3)	1.8 (1.1–2.9)	2.5 (1.5–3.7)	<0.001
Incident type 2 diabetes	17 (1.8)	27 (3.1)	59 (7.5)	69 (9.5)	<0.001
<b>HbA<sub>1c</sub> 39–46 mmol/mol (5.7–6.4%)</b>					
<i>n</i>	293	289	289	295	
Age	50.23 ± 8.27	51.08 ± 8.55	52.27 ± 8.67	51.63 ± 8.60	0.030
Sex, male	104 (35.5)	122 (42.2)	118 (40.8)	161 (54.6)	<0.001
Hypertension	76 (25.9)	98 (33.9)	120 (41.5)	131 (44.4)	<0.001
Dyslipidaemia	74 (25.3)	77 (26.6)	107 (37.2) <sup>b</sup>	120 (41.0) <sup>b</sup>	<0.001
AMI	2 (0.7)	5 (1.7)	7 (2.4)	10 (3.4)	0.128
CAD	6 (2.0)	8 (2.8)	9 (3.1)	11 (3.8) <sup>b</sup>	0.667
Heart failure	2 (0.7)	8 (2.8)	5 (1.7)	7 (2.4)	0.271
Stroke	1 (0.3)	2 (0.7)	5 (1.7)	4 (1.4)	0.336
Smoking					0.008
Never	166 (56.7)	125 (43.3)	141 (48.8)	128 (43.4)	
Former	84 (28.7)	92 (31.8)	85 (29.4)	102 (34.6)	
Present	43 (14.7)	72 (24.9)	63 (21.8)	65 (22.0)	
Alcohol intake	22 (7.5)	22 (7.6)	28 (9.7)	45 (15.3)	0.005
Parental history of diabetes <sup>b</sup>	102 (36.7)	100 (34.5)	156 (49.2)	153 (50.0)	<0.001
Lipid-lowering therapy	15 (5.1)	18 (6.2)	39 (13.5)	43 (14.6)	<0.001

**Table 1** (continued)

Variable	TRL particle size				P value <sup>a</sup>
	1st quartile (<40 nm)	2nd quartile (40–45 nm)	3rd quartile (45–52 nm)	4th quartile (>52 nm)	
Glucose-lowering therapy	0 (0)	0 (0)	0 (0)	0 (0)	1.000
BMI, kg/m <sup>2</sup>	25.77 ± 4.29	26.88 ± 5.32	28.54 ± 5.07	30.43 ± 5.38	<0.001
Waist circumference, cm	84.54 ± 10.6	88.61 ± 12.5	92.88 ± 11.99	98.32 ± 12.3	<0.001
Systolic BP, mmHg	116.6 ± 17.4	120.5 ± 18.9	122.0 ± 15.9	125.7 ± 15.7	<0.001
Diastolic BP, mmHg	73.5 ± 11.1	75.9 ± 12.3	77.0 ± 10.3	79.99 ± 11.04	<0.001
GFR, ml min <sup>-1</sup> (1.73 m) <sup>-2</sup>	104 (90–121)	102 (84–119)	98 (87–118)	104 (89–123)	0.076
Fasting blood glucose, mmol/l	5.83 ± 0.6	5.99 ± 1.1	6.22 ± 1.2	6.72 ± 2.0	<0.001
Fasting blood triacylglycerol, mmol/l	0.9 (0.7–1.1)	1.2 (0.9–1.6)	1.5 (1.2–1.9)	2.2 (1.5–3.2)	<0.001
HbA <sub>1c</sub> , %	6.08 ± 0.36	6.07 ± 0.47	6.12 ± 0.50	6.21 ± 0.80	0.011
HbA <sub>1c</sub> , mmol/mol	43 ± 1.2	43 ± 1.3	43 ± 1.5	44 ± 1.7	
hsCRP, mg/l	1.2 (0.6–3.1)	1.7 (0.8–3.7)	1.9 (0.9–4.6)	2.6 (1.3–5.3)	<0.001
HOMA-IR	1.1 (0.6–1.8)	1.5 (0.8–2.4)	2.1 (1.3–3.2)	3.0 (1.9–4.9)	<0.001
Incident type 2 diabetes <sup>b</sup>	15 (5.4)	31 (10.4)	60 (18.9)	88 (28.8)	<0.001

Data are presented for  $n = 4463$  participants enrolled at the São Paulo site

Data are means ± SD for normally distributed variables, medians (IQR) for nonparametric data or  $n$  (%)

<sup>a</sup>  $p$  values are for group comparisons. Statistical analysis:  $\chi^2$  for categorical data; ANOVA or Kruskal–Wallis for continuous variables

<sup>b</sup> Minor mismatches between percentages and  $n$  values are due to missing data

AMI, acute myocardial infarction; CAD, coronary artery disease

visit 2 on the basis of laboratory values. To better predict the risk of type 2 diabetes, we split the cohort twice: (1) based on the presence of glucose abnormalities at baseline according to the ADA-HbA<sub>1c</sub> definition; and (2) randomly in training-validation and test sets.

Among the 3297 individuals with baseline HbA<sub>1c</sub> <39 mmol/mol (5.7%), the incidence of diabetes during follow-up was 1.40 cases per 100 patient-years (172 new cases). Participants with baseline HbA<sub>1c</sub> ≥39 mmol/mol (5.7%) ( $n = 1166$ ) had an incidence of diabetes during follow-up of 4.50 cases per 100 patient-years (194 new cases).

The impact of TRL-related measurements on incident diabetes is described in ESM Tables 1 and 2, but only TRL particle diameter improved risk discrimination. We further divided the cohort according to baseline TRL particle size quartiles in order to better characterise the relationship between TRL and diabetes. As seen in Table 1, there was an imbalance in the characteristics of participants in different quartiles. While participants in the first quartile had a lower mean BMI, a lower frequency of hypertension and lower levels of inflammatory markers, participants in the fourth quartile had more dysfunction in glucose homeostasis and lipid metabolism.

**Impact of TRL plasma concentration, particle size and cholesterol/triacylglycerol content on incident diabetes** In bivariate analyses (ESM Table 2), the majority of TRL-

related markers were positively associated with incident type 2 diabetes. ORs for type 2 diabetes were higher in the fourth quartile compared with the first quartile of TRL particle size (data not shown). The magnitude of association with incident type 2 diabetes was similar between individuals with HbA<sub>1c</sub> <39 mmol/mol (5.7%) and those with HbA<sub>1c</sub> 39–46 mmol/mol (5.7–6.4%) for continuous TRL particle size, as well as for TRL plasma concentrations, TRL-cholesterol per particle, total triacylglycerols and total cholesterol content in the pool of TRL particles.

Although total triacylglycerol content in the pool of TRL particles was weakly associated with incident type 2 diabetes, triacylglycerol content per TRL particle (ratio of total triacylglycerol in TRL to TRL particle concentration) was more robustly associated with diabetes (data not shown). Since the majority of the lipid cargo of TRL particles is triacylglycerol, the size of the particles is highly correlated with the lipid content and, of course, triacylglycerol content per TRL particle. TRL particle size strongly correlated with TRL triacylglycerol per particle (overall  $R^2$  0.43) and weakly correlated with TRL-cholesterol per particle (data not shown).

### Multivariate models to predict incident diabetes after 3.7 years' follow-up: the role of TRL-related markers

Commonly available clinical features and laboratory variables were used to build baseline models. In ESM Table 1, we show that in a model with 18 variables the magnitude of ORs and

**Table 2** Multivariate logistic regression to predict incident diabetes in the ELSA-Brasil study

Cohort and variable	OR (95% CI)	<i>p</i>
<b>HbA<sub>1c</sub> &lt;39 mmol/mol (5.7%) at baseline<sup>a</sup></b>		
Baseline model		
Age (per 1 year)	1.0280 (1.0124, 1.0398)	<0.0001
Sex (male vs female)	0.9644 (0.7775, 1.1980)	0.7436
Hypertension (yes vs no)	1.5837 (1.2933, 1.9366)	<0.0001
Dyslipidaemia (yes vs no)	1.1612 (0.9003, 1.5039)	0.2317
Prior CAD (yes vs no)	0.7791 (0.3711, 1.4829)	0.4704
Prior stroke (yes vs no)	0.8979 (0.4315, 1.7216)	0.6945
Alcohol intake (yes vs no)	1.1243 (0.8692, 1.4527)	0.3459
Lipid-lowering therapy (yes vs no)	1.1759 (0.8196, 1.5914)	0.3836
Parental history of diabetes (yes vs no)	2.3074 (1.9141, 2.7711)	<0.0001
Creatinine clearance (per ml min <sup>-1</sup> 1.73 m <sup>-2</sup> )	1.0001 (0.9996, 1.0005)	0.3948
Waist circumference (per 1 cm)	1.0301 (1.0219, 1.0381)	<0.0001
Systolic BP (per 1 mmHg)	1.0124 (1.0064, 1.0169)	<0.0001
Diastolic BP (per 1 mmHg)	1.0072 (0.9881, 1.0261)	0.7411
HbA <sub>1c</sub> (per 3 mmol/mol)	2.6231 (1.9887, 3.4569)	<0.0001
hsCRP (per 1 mg/l)	0.9917 (0.9662, 1.0145)	0.4465
Fasting blood triacylglycerols (per 1.13 mmol/l)	1.2117 (1.1349, 1.2881)	<0.0001
HDL-cholesterol (per 0.026 mmol/l)	0.9841 (0.9750, 0.9911)	<0.0001
LDL-cholesterol (per 0.026 mmol/l)	0.9979 (0.9949, 1.0001)	0.0667
Baseline model + TRL particle size or TRL concentration or TRL triacylglycerol or TRL-cholesterol content		
TRL particle size (continuous, per 10 nm) <sup>b</sup>	1.5689 (1.3183, 1.8317)	<0.0001
Plasma TRL concentration (continuous, per 10 nmol/l) <sup>b</sup>	0.9983 (0.9952, 1.0013)	0.2623
Triacylglycerol per TRL particle (continuous, per 1 mg/nmol) <sup>b</sup>	1.8972 (1.0935, 3.2933)	0.0225
Cholesterol per TRL particle (continuous, per 0.1 mg/nmol) <sup>b</sup>	9.3899 (0.6991, 102.02)	0.2092
Total triacylglycerol in the pool of TRL particles (continuous, per 0.113 mmol/l) <sup>b</sup>	0.9990 (0.9941, 1.0054)	0.7635
Total cholesterol in the pool of TRL particles (continuous, per 0.259 mmol/l) <sup>b</sup>	0.9896 (0.9714, 1.0080)	0.2382
<b>HbA<sub>1c</sub> 39–46 mmol/mol (5.7–6.4%) at baseline<sup>c</sup></b>		
Baseline model		
Age (per 1 year)	1.0258 (1.0136, 1.0387)	<0.0001
Sex (male vs female)	0.9723 (0.7772, 1.2171)	0.7602
Hypertension (yes vs no)	1.0344 (0.8343, 1.2775)	0.7641
Dyslipidaemia (yes vs no)	1.1418 (0.8739, 1.4348)	0.2917
Prior CAD (yes vs no)	0.4576 (0.2171, 0.8839)	0.0238
Prior stroke (yes vs no)	0.9271 (0.4848, 1.9116)	0.4385
Alcohol intake (yes vs no)	0.9544 (0.7012, 1.2641)	0.6608
Lipid-lowering therapy (yes vs no)	1.2651 (0.8889, 1.7101)	0.2383
Parental history of diabetes (yes vs no)	1.5631 (1.2803, 1.9091)	<0.0001
Creatinine clearance (per ml min <sup>-1</sup> 1.73 m <sup>-2</sup> )	1.0001 (0.9993, 1.0008)	0.5781
Waist circumference (per 1 cm)	1.0297 (1.0207, 1.0378)	<0.0001
Systolic BP (per 1 mmHg)	1.0154 (1.0107, 1.0225)	<0.0001
Diastolic BP (per 1 mmHg)	1.0096 (0.9903, 1.0293)	0.5418
HbA <sub>1c</sub> (per 3 mmol/mol)	2.9811 (2.0027, 4.4722)	<0.0001
hsCRP (per 1 mg/l)	0.9989 (0.9773, 1.0186)	0.7772
Fasting blood triacylglycerols (per 1.13 mmol/l)	1.1698 (1.0767, 1.2821)	0.0002
HDL-cholesterol (per 0.026 mmol/l)	0.9873 (0.9769, 0.9957)	0.0039
LDL-cholesterol (per 0.026 mmol/l)	0.9961 (0.9931, 0.9990)	0.0130
Baseline model + TRL particle size or concentration or TRL triacylglycerol or cholesterol content		
TRL particle size (continuous, per 10 nm) <sup>b</sup>	1.2488 (1.0191, 1.4803)	0.0299

**Table 2** (continued)

Cohort and variable	OR (95% CI)	<i>p</i>
Plasma TRL concentration (continuous, per 10 nmol/l) <sup>b</sup>	0.9987 (0.9953, 1.0020)	0.4806
Triacylglycerol per TRL particle (continuous, per 1 mg/nmol) <sup>b</sup>	1.3157 (0.8565, 1.8523)	0.1431
Cholesterol per TRL particle (continuous, per 0.1 mg/nmol) <sup>b</sup>	12.311 (0.1274, 90.131)	0.4241
Total triacylglycerol in the pool of TRL particles (continuous, per 0.113 mmol/l) <sup>b</sup>	0.9981 (0.9917, 1.0050)	0.6003
Total cholesterol in the pool of TRL particles (continuous, per 0.259 mmol/l) <sup>b</sup>	0.9896 (0.9683, 1.0096)	0.3406

Data are presented for  $n = 4463$  participants enrolled at the São Paulo site

<sup>a</sup>  $n = 3297$  and 172 new cases of type 2 diabetes after 3.7 years

<sup>b</sup> Adjusted by age, sex, hypertension, dyslipidaemia, prior coronary artery disease, prior stroke, alcohol intake, lipid-lowering therapy, parental history of diabetes, creatinine clearance, waist circumference, systolic BP, diastolic BP, HbA<sub>1c</sub>, hsCRP, fasting blood triacylglycerols, HDL-cholesterol, LDL-cholesterol

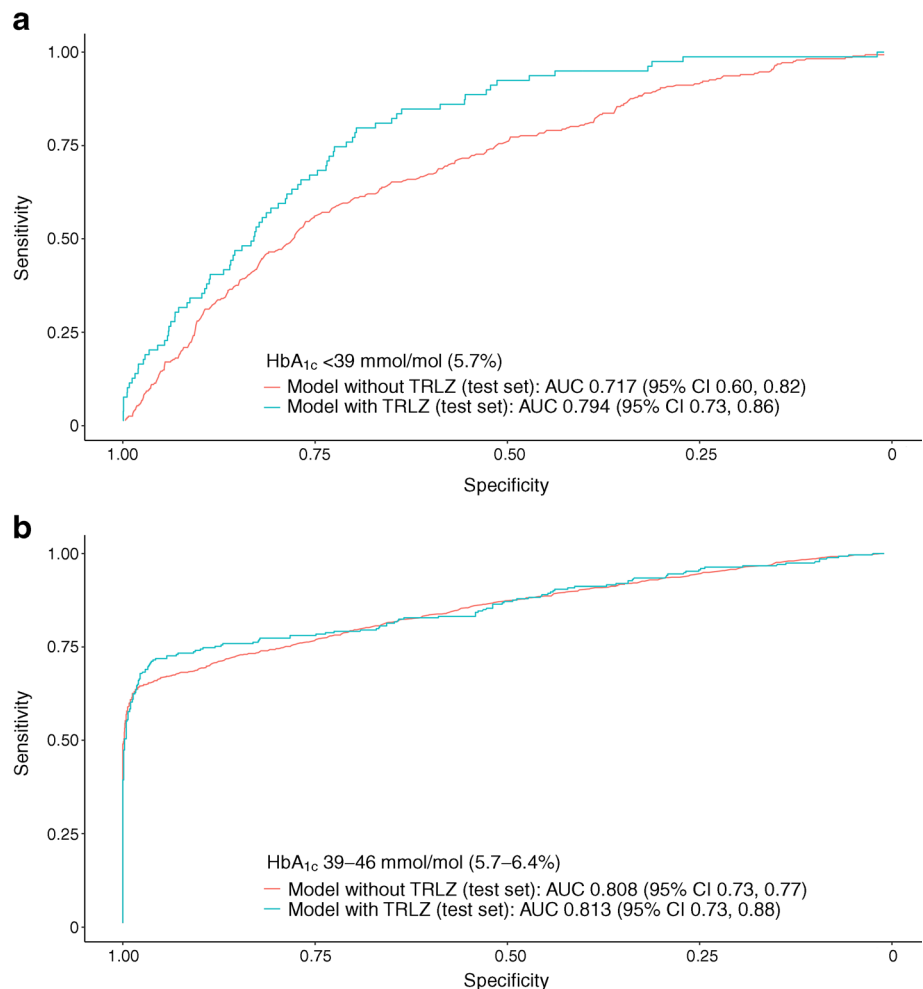
<sup>c</sup>  $n = 1166$  and 194 new cases of type 2 diabetes after 3.7 years

CAD, coronary artery disease

overall model discriminatory capacity (AUROC) were similar in the original ELSA-Brasil cohort ( $n = 10,729$ ) and the subsample of the São Paulo site cohort ( $n = 4463$ ) (global AUROC: 0.857 (95% CI 0.848, 0.866) and 0.846 (95% CI 0.829, 0.869), respectively).

When we incorporated TRL-related markers into multivariate models, only TRL particle size and triacylglycerol content per TRL particle increased discriminatory efficacy (Table 2). Since the AUROC for TRL particle size was slightly better than that for triacylglycerol content per TRL particle, we kept

**Fig. 1** ROC curve in the test sets for predicting incident type 2 diabetes after 4 years' follow-up, based on multivariate models with and without TRL particle size (TRLZ). **(a)** AUCs for individuals with baseline HbA<sub>1c</sub> <39 mmol/mol (5.7%) (hold-out test set,  $n = 659$ ) show that addition of TRLZ to the prediction model improved the AUROC for incident type 2 diabetes in comparison with a model using age, sex, hypertension, dyslipidaemia, history of coronary artery disease, prior stroke, alcohol intake, family history of diabetes, use of lipid-lowering therapies, creatinine clearance, waist circumference, systolic BP, diastolic BP, HbA<sub>1c</sub>, hsCRP, plasma triacylglycerols, HDL-cholesterol and LDL-cholesterol. **(b)** AUCs for individuals with baseline HbA<sub>1c</sub> 39–46 mmol/mol (5.7–6.4%) (hold-out test set,  $n = 233$ ) show that TRLZ numerically improved the AUROC for incident type 2 diabetes, but this was not statistically significant in this subgroup





**Table 3** Prediction performance in test sets (hold-out cohorts) after tenfold cross-validation in the ELSA-Brasil study

HbA <sub>1c</sub> cohort and model	AUROC (95% CI)	AUPRC	F1 score <sup>a</sup>	Matthews correlation coefficient score
HbA <sub>1c</sub> <39 mmol/mol (5.7%): test set, <i>n</i> = 659				
Baseline model <sup>b</sup>	0.717 (0.603, 0.818)	0.582	0.19	0.208
Baseline model <sup>b</sup> + TRL particle size	0.794 (0.731, 0.862)	0.701	0.34	0.350
HbA <sub>1c</sub> 39–46 mmol/mol (5.7–6.4%): test set, <i>n</i> = 233				
Baseline model <sup>b</sup>	0.808 (0.726, 0.875)	0.750	0.55	0.409
Baseline model <sup>b</sup> + TRL particle size	0.813 (0.730, 0.879)	0.755	0.56	0.412

Data are presented for *n* = 4463 participants enrolled at the São Paulo site

<sup>a</sup> F1 score =  $(2 \times \text{precision} \times \text{recall}) / (\text{precision} + \text{recall})$

<sup>b</sup> The baseline model included the following covariates: age, sex, hypertension, dyslipidaemia, prior coronary artery disease, prior stroke, alcohol intake, lipid-lowering therapy, parental history of diabetes, creatinine clearance, waist circumference, systolic BP, diastolic BP, HbA<sub>1c</sub>, hsCRP, fasting blood triacylglycerols, HDL-cholesterol and LDL-cholesterol

the particle size in further analyses. TRL plasma concentrations, total triacylglycerol and total cholesterol content in the pool of TRL particles did not change discriminatory capacity (data not shown). As expected, when we added TRL particle size into the model, the strong relationship between fasting serum triacylglycerols and incident diabetes disappeared (data not shown).

**Baseline HbA<sub>1c</sub> interacts with TRL particle size** As expected, TRL particle size was highly and positively correlated with insulin resistance (HOMA-IR), waist circumference and hsCRP. However, these associations were stronger in individuals with a baseline HbA<sub>1c</sub> of 39–46 mmol/mol (5.7–6.4%) vs <39 mmol/mol (5.7%) (ESM Table 3, ESM Fig. 3a–c). Although hsCRP correlated with TRL particle size, other inflammatory markers such as IL-6 and TNF- $\alpha$  did not.

Based on these observations, we hypothesised that individuals with baseline HbA<sub>1c</sub> of 39–46 mmol/mol (5.7–6.4%) are more prone to presenting with a number of risk factors (obesity, insulin resistance and heightened inflammation) that increase the hepatic release of large VLDL particles or impair their catabolism (i.e. reduced lipoprotein lipase activity).

ESM Fig. 3d strengthens this hypothesis by showing that individuals with impaired fasting glucose present with a stronger linear correlation between TRL particle size and the combination of HOMA-IR, waist circumference and hsCRP, measured by the sum of their *z* scores. Indeed, in these individuals with impaired fasting glucose the sum of *z* scores for HOMA-IR, waist circumference and hsCRP was also sharply correlated with the probability of diabetes (ESM Fig. 4).

**TRL particle size is associated with incident type 2 diabetes in individuals without impaired fasting glucose at baseline** To further investigate the association between TRL particle size

and the risk of incident diabetes during follow-up and its additive predictive ability, we generated multivariate models for individuals with impaired fasting glucose (HbA<sub>1c</sub> 39–46 mmol/mol [5.7–6.4%]) or HbA<sub>1c</sub> <39 mmol/mol (5.7%).

As shown in Table 2, among TRL-related measurements, only TRL particle size was associated with incident diabetes in multivariate models (with 18 covariates) both in individuals with impaired fasting glucose and those with HbA<sub>1c</sub> <39 mmol/mol (5.7%). Indeed, as seen in ROC curves (Fig. 1), the addition of TRL particle size numerically improved the AUROC for incident diabetes during follow-up in both subgroups.

**TRL particle size adds predictive value for incident type 2 diabetes in individuals without impaired fasting glucose at baseline** To avoid potential overfitting and to evaluate whether adding TRL particle size improved the predictive ability of the logistic regression model on top of the other 18 predictors, we split the subgroups into a training-validation set and a test (hold-out) set. After developing the training and validation models with a tenfold cross-validation framework using 80% of the cohort, we tested the models in 20% of the sample (hold-out cohort).

Table 3 shows prediction performance metrics in the hold-out cohort in individuals with baseline HbA<sub>1c</sub> <39 mmol/mol (5.7%) (test set, *n* = 659) and HbA<sub>1c</sub> 39–46 mmol/mol (5.7–6.4%) (test set, *n* = 233). Compared with a baseline model with 18 predictors, the model including TRL particle size appeared to improve AUROC, AUPRC, F1 and Matthews correlation coefficient scores only in the group with baseline HbA<sub>1c</sub> <39 mmol/mol (5.7%).

**Sensitivity analyses** As a sensitivity analysis, we excluded participants whose diagnosis of incident diabetes was made

based on a single fasting glucose level of  $>7.0$  mmol/l. By doing this in the São Paulo site cohort, we eliminated 54 new cases, leaving 312 new cases, but it did not change the AUROC for incident diabetes as a dependent variable in the baseline model (AUROC 0.832 [95% CI 0.813, 0.855]), nor did it change the discriminative ability of the TRL particle size model (AUROC 0.852 [95% CI 0.830, 0.881]). This suggests that individuals diagnosed with a single FPG test could be excluded without affecting the main results.

As we chose ADA-HbA<sub>1c</sub> as the criterion for ascertaining abnormal glucose homeostasis at baseline, we carried out sensitivity analyses regarding the use of different criteria. In clinical models for predicting incident diabetes, the AUROC with the ADA-HbA<sub>1c</sub> criterion did not differ from that with the OGTT/IGT criterion (ADA-HbA<sub>1c</sub> vs IGT,  $p=0.12$ ) or the ADA-IFG criterion (ADA-HbA<sub>1c</sub> vs ADA-IFG criteria,  $p=0.27$ ) (ESM Table 4).

We also evaluated the impact of ascertaining abnormal glucose homeostasis at baseline on the basis of a combined any positive test approach (ADA-HbA<sub>1c</sub>, IGT or ADA-IFG). With this combined criterion, we excluded an additional 347 individuals with type 2 diabetes at baseline (previously unknown who had criteria for type 2 diabetes only on FPG or 2 h glucose on OGTT), leaving 4116 individuals at risk for type 2 diabetes. Among these participants, 2469 met the criteria for intermediate hyperglycaemia at baseline and there were 138 incident cases of type 2 diabetes during follow-up in this subgroup. In these participants, using a model limited to nine variables in order to avoid model saturation, TRL particle size did not improve risk prediction for type 2 diabetes (baseline model [age, sex, hypertension, parental history of type 2 diabetes, waist circumference, systolic BP, HbA<sub>1c</sub>, triacylglycerol, HDL-cholesterol] AUROC 0.708 [95% CI 0.685, 0.732] vs baseline model plus TRL particle size AUROC 0.729 [95% CI 0.696, 0.768],  $p=0.0647$ ). Each 10 nm increment in TRL particle size increased the risk for incident diabetes by 32% (95% CI 7.3, 57.2,  $p=0.0103$ ) in multivariate analyses.

Conversely, among the remaining 1647 participants who had no glycaemic abnormalities at baseline (from the 4116 individuals at risk for type 2 diabetes), there were 30 incident cases of diabetes during follow-up. In this subgroup, TRL particle size improved AUROC ( $p=0.00042$ ). A baseline model was set with two variables (age and systolic BP) by using a stepwise approach. The AUROC was 0.696 (95% CI 0.602, 0.790) with this baseline model vs 0.7885 (95% CI 0.6871, 0.8899) with the baseline model plus TRL particle size.

## Discussion

The findings of the present study indicate that TRL particle size is a novel marker for incident type 2 diabetes that adds predictive power to established risk factors in individuals with

normal HbA<sub>1c</sub> values. This particular population is at substantial risk for diabetes (up to 3% develop the condition each year), and current models are inadequate for predicting risk burden as compared with impaired fasting glucose individuals [2].

In this study, TRL particle size strongly associated with the incidence of type 2 diabetes in the univariate model, especially in individuals with a HbA<sub>1c</sub> of 39–46 mmol/mol (5.7–6.4%). However, this effect was clearly offset by the presence of other metabolic risk factors (i.e. waist circumference, HOMA-IR and hsCRP). This was expected since an overproduction of large VLDL particles has been described as a product of obesity, hepatic steatosis and insulin resistance [10, 11]. TRL typically refers to the mixture of VLDL and chylomicrons. However, in most cases chylomicrons would be absent in a 12 h fasting sample. It is possible that some of the participants with normoglycaemia had fasting chylomicron remnants. In individuals without diabetes, it is noticeable that increasing HOMA-IR and liver fat are strongly and positively correlated with production rates of VLDL1 (larger particles) and decreased secretion of VLDL2 (smaller particles) [11, 20]. Several mechanisms might be implicated in this phenotype.

First, insulin-resistant states induced by the proinflammatory cytokine TNF- $\alpha$  are associated with overproduction of intestinally derived VLDL1 driven by induction of p38-mitogen-activated protein kinase by TNF- $\alpha$  [21]. In parallel, there is increased expression of sterol regulatory element binding protein 1c because of insulin resistance, combined with increased production of glycerol 3-phosphate via increased gluconeogenesis and triacylglycerol synthesis via de novo lipogenesis [22]. Possibly, the overlap between changes in metabolism with the change in TRL physiology in insulin-resistant states might explain why TRL particle size brings additive predictive value in addition to classic risk factors in this subpopulation.

In participants with HbA<sub>1c</sub>  $<39$  mmol/mol (5.7%), TRL particle size had a sustained univariate and multivariate effect on the risk of type 2 diabetes, and increased the AUROC for predicting diabetes on top of the best available model in a large and multi-ethnic cohort. In aggregate, these findings suggest that among individuals with a lower degree of insulin resistance, those who present with a higher proportion of VLDL1 are exposed to diabetogenic particles or may share polymorphisms that increase their risk for diabetes [11, 20].

There are several lines of evidence supporting a causal relationship between VLDL1 levels and diabetes risk [23], potentially due to increased pancreatic steatosis and cytotoxicity, which are hallmarks of type 2 diabetes. In animal models, glycosylphosphatidylinositol HDL-binding protein-1 knockout (*Gpihbp1*<sup>-/-</sup>) and apolipoprotein C3-transgenic (*ApoC3*-tg) mice show increased mean TRL particle size, an

augmented proportion of VLDL1 and amplified cytotoxic effects in pancreatic cells [24]. In humans, removal of excess lipid from the pancreas at the same time as decreasing plasma VLDL1-triacylglycerol allows the return of normal insulin secretion in early type 2 diabetes [25, 26].

Another hypothesis is that the polymorphisms disrupting TRL metabolism may share a phenotype of increased TRL particle size and increased risk for type 2 diabetes because of pancreatic beta cell lipotoxicity or via indirect mechanisms. Several pathways are associated with both increased TRL particle size and diabetes, including the capacity to hydrolyse TRL particles by the endothelium (e.g. lipoprotein lipase T495G HindIII and rs343 polymorphisms [27]) or in the liver (e.g. hepatic lipase rs6083 polymorphism [28]), or preventing TRL hydrolysis (e.g. *APOC3* 482TT genotype [29]). Therefore, it is reasonable to hypothesise that TRL particle size could be used as a proxy for genetic mutations that increase the risk of diabetes.

Some limitations of our study must be acknowledged. First, our characterisation of incident diabetes included participants with a single abnormal glycaemic test. As participants frequently reverse the abnormal glycaemic status after a few weeks, we carried out a sensitivity analysis that showed that individuals diagnosed by a single glycaemic test could be excluded without affecting our main results. Second, our follow-up period was relatively short, and therefore the observed incidence of diabetes could be over- or underestimated. However, the annual incidence rate of diabetes in at-risk individuals (1.98% in ELSA-Brasil and 2.14% at the São Paulo site) was similar to that reported in other cohorts [30, 31] and the robust number of new cases (802 in ELSA-Brasil and 366 at the São Paulo site) suggests that the period was long enough to provide statistical power. Third, we did not classify incident diabetes in terms of type. However, it is highly unlikely that new cases could be classified as type 1 diabetes since we did not include individuals aged 34 years or younger. Finally, since this study included an occupational cohort of civil servants, our findings cannot be taken to be representative of all Brazilian adults. Major strengths of this study are: (1) the consistency of our findings in several sensitivity analyses; (2) validation of the full model in a large cohort; and (3) concordance between TRL particle size and TRL triacylglycerol content per particle in predicting incident type 2 diabetes. Though particle size and triacylglycerol per particle are proxies and both are modified by insulin resistance [10, 11], they are measured through different signals in NMR spectroscopy [16], and the consistent results must be acknowledged.

In summary, we found a biomarker that improves discrimination capacity for the risk of diabetes in individuals who do not show dysglycaemia at baseline. TRL particle size was highly correlated with obesity, insulin resistance and inflammation in individuals with impaired fasting glucose at

baseline, but this association was attenuated in those with  $HbA_{1c} < 39$  mmol/mol (5.7%). These findings indicate TRL particle size as a risk factor that is changed early in the course of the pathophysiological processes that lead to type 2 diabetes, even before the occurrence of glucose abnormalities. A causal relationship between TRL particle size and incident diabetes should be further investigated.

**Data availability** The datasets analysed during the current study are not publicly available due to our sponsor's policy but are available from the corresponding author on reasonable request.

**Funding** ELSA-Brasil is supported by the Brazilian Ministry of Health (Department of Science and Technology), Ministry of Science, Technology and Innovation, the National Council for Scientific and Technological Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico [CNPq]) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes). This study was funded by the Brazilian Ministry of Health (Science and Technology) and the Brazilian Ministry of Science, Technology and Innovation (Financiadora de Estudos e Projetos [Finep] and CNPq). Grant numbers: 01 06 0010.00 and 01.10.0643.03 (RDS); 01 06 0212.00 and 01.10.0742-00 (BA, Brazil); 01 06 0300.00 and 01.12.0284.00 (ES, Brazil); 01 06 0278.00 and 01 10 0746 00 (MG, Brazil); 01 06 0115.00 and 01.10.0773-00 (SP, Brazil); and 01 06 0071.00 and 01.11-0093.01 (RJ, Brazil). LSFC and RDS are recipients of scholarships from the CNPq grants nos. 437413/2018-7 and 303734/2018-3, respectively. ACS is a recipient of a scholarship from the CNPq (grant no. 301465/2017-7) and the Fundação de Amparo a Pesquisa do Estado de São Paulo (Fapesp; grant no. 13/07607-8).

**Authors' relationships and activities** RDS has received honoraria related to consulting, research and or speaker activities from Amgen, Aché, Astra Zeneca, Esperion, Kowa, Merck, Novo-Nordisk, PTC, Pfizer and Sanofi/Regeneron. The other authors declare that there are no relationships or activities that might bias, or be perceived to bias, their work.

**Contribution statement** IMB, PAL, BBD and MIS were responsible for the acquisition of data. LSFC, ACCN, BBD, MIS, MJB, PPT, SRJ, RDS and ACS were responsible for conception and design. LSFC was a central person in data analyses. LSFC and ACCN prepared the manuscript. IMB, PAL, BBD, MIS, MJB, PPT, SRJ, RDS and ACS reviewed the manuscript critically for important intellectual content. All authors made substantial contributions to interpretation of data and all authors gave final approval of the version to be published. LSFC is the guarantor of this work, had full access to all study data, and assumes responsibility for data integrity and analytical accuracy.

## References

- Haw JS, Galaviz KI, Straus AN et al (2017) Long-term sustainability of diabetes prevention approaches: a systematic review and meta-analysis of randomized clinical trials. *JAMA Intern Med* 177(12):1808–1817. <https://doi.org/10.1001/jamainternmed.2017.6040>
- Gerstein HC, Santaguida P, Raina P et al (2007) Annual incidence and relative risk of diabetes in people with various categories of dysglycemia: a systematic overview and meta-analysis of prospective studies. *Diabetes Res Clin Pract* 78(3):305–312. <https://doi.org/10.1016/j.diabres.2007.05.004>
- Richter B, Hemmingsen B, Metzendorf MI, Takwoingi Y (2018) Development of type 2 diabetes mellitus in people with

- intermediate hyperglycaemia. *Cochrane Database Syst Rev*. Issue 10. Art No.:CD012661. <https://doi.org/10.1002/14651858.CD012661.pub2>
4. Abdul-Ghani MA, DeFronzo RA (2009) Plasma glucose concentration and prediction of future risk of type 2 diabetes. *Diabetes Care* 32(Suppl 2):S104–S198
  5. Wilson PW, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB Sr (2007) Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. *Arch Intern Med* 167(10):1068–1074. <https://doi.org/10.1001/archinte.167.10.1068>
  6. Kahn HS, Cheng YJ, Thompson TJ, Imperatore G, Gregg EW (2009) Two risk-scoring systems for predicting incident diabetes mellitus in U.S. adults age 45 to 64 years. *Ann Intern Med* 150(11):741–751. <https://doi.org/10.7326/0003-4819-150-11-200906020-00002>
  7. Mensenkamp AR, Havekes LM, Romijn JA, Kuipers F (2001) Hepatic steatosis and very low density lipoprotein secretion: the involvement of apolipoprotein E. *J Hepatol* 35(6):816–822. [https://doi.org/10.1016/S0168-8278\(01\)00249-5](https://doi.org/10.1016/S0168-8278(01)00249-5)
  8. Unger RH, Zhou YT (2001) Lipotoxicity of beta-cells in obesity and in other causes of fatty acid spillover. *Diabetes* 50(Suppl 1):S118–S121. <https://doi.org/10.2337/diabetes.50.2007.S118>
  9. van den Berg EH, Corsetti JP, Bakker SJL, Dullaart RPF (2019) Plasma ApoE elevations are associated with NAFLD: the PREVENT Study. *PLoS One* 14(8):e0220659. <https://doi.org/10.1371/journal.pone.0220659>
  10. Adiels M, Taskinen MR, Packard C et al (2006) Overproduction of large VLDL particles is driven by increased liver fat content in man. *Diabetologia* 49(4):755–765. <https://doi.org/10.1007/s00125-005-0125-z>
  11. Adiels M, Westerbacka J, Soro-Paavonen A et al (2007) Acute suppression of VLDL1 secretion rate by insulin is associated with hepatic fat content and insulin resistance. *Diabetologia* 50(11):2356–2365. <https://doi.org/10.1007/s00125-007-0790-1>
  12. Aquino EM, Barreto SM, Bensenor IM et al (2012) Brazilian Longitudinal Study of Adult Health (ELSA-Brasil): objectives and design. *Am J Epidemiol* 175(4):315–324. <https://doi.org/10.1093/aje/kwr294>
  13. Almeida-Pittito B, Ribeiro-Filho FF, Barreto S et al (2016) Circulating early biomarkers of atherogenesis in participants of the Longitudinal Study of Adult Health (ELSA-Brasil) without diabetes or cardiovascular disease. *Arch Endocrinol Metab* 60(6):573–581. <https://doi.org/10.1590/2359-399700000205>
  14. American Diabetes Association (2018) 2. Classification and diagnosis of diabetes: standards of medical care in diabetes—2018. *Diabetes Care* 41(Suppl 1):S13–S27. <https://doi.org/10.2337/dc18-S002>
  15. Schmidt MI, Bracco PA, Yudkin JS et al (2019) Intermediate hyperglycaemia to predict progression to type 2 diabetes (ELSA-Brasil): an occupational cohort study in Brazil. *Lancet Diabetes Endocrinol* 7(4):267–277. [https://doi.org/10.1016/S2213-8587\(19\)30058-0](https://doi.org/10.1016/S2213-8587(19)30058-0)
  16. Jeyarajah EJ, Cromwell WC, Otvos JD (2006) Lipoprotein particle analysis by nuclear magnetic resonance spectroscopy. *Clin Lab Med* 26(4):847–870. <https://doi.org/10.1016/j.cll.2006.07.006>
  17. Zhang X, Gregg EW, Williamson DF et al (2010) A1C level and future risk of diabetes: a systematic review. *Diabetes Care* 33(7):1665–1673. <https://doi.org/10.2337/dc09-1939>
  18. Chicco D, Jurman G (2020) The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* 21(1):6. <https://doi.org/10.1186/s12864-019-6413-7>
  19. Pepe MS, Longton G (2005) Standardizing diagnostic markers to evaluate and compare their performance. *Epidemiology* 16(5):598–603. <https://doi.org/10.1097/01.ede.0000173041.03470.8b>
  20. Pavlic M, Xiao C, Szeto L, Patterson BW, Lewis GF (2010) Insulin acutely inhibits intestinal lipoprotein secretion in humans in part by suppressing plasma free fatty acids. *Diabetes* 59(3):580–587. <https://doi.org/10.2337/db09-1297>
  21. Qin B, Qiu W, Avramoglu RK, Adeli K (2007) Tumor necrosis factor- $\alpha$  induces intestinal insulin resistance and stimulates the overproduction of intestinal apolipoprotein B48-containing lipoproteins. *Diabetes* 56(2):450–461. <https://doi.org/10.2337/db06-0518>
  22. Browning JD, Horton JD (2004) Molecular mediators of hepatic steatosis and liver injury. *J Clin Invest* 114(2):147–152. <https://doi.org/10.1172/JCI200422422>
  23. Qi Q, Liang L, Doria A, Hu FB, Qi L (2012) Genetic predisposition to dyslipidemia and type 2 diabetes risk in two prospective cohorts. *Diabetes* 61(3):745–752. <https://doi.org/10.2337/db11-1254>
  24. Zhang Y, He W, He C et al (2019) Large triglyceride-rich lipoproteins in hypertriglyceridemia are associated with the severity of acute pancreatitis in experimental mice. *Cell Death Dis* 10(10):728. <https://doi.org/10.1038/s41419-019-1969-3>
  25. Lim EL, Hollingsworth KG, Aribisala BS, Chen MJ, Mathers JC, Taylor R (2011) Reversal of type 2 diabetes: normalisation of beta cell function in association with decreased pancreas and liver triacylglycerol. *Diabetologia* 54(10):2506–2514. <https://doi.org/10.1007/s00125-011-2204-7>
  26. Steven S, Hollingsworth KG, Small PK et al (2016) Weight loss decreases excess pancreatic triacylglycerol specifically in type 2 diabetes. *Diabetes Care* 39(1):158–165. <https://doi.org/10.2337/dc15-0750>
  27. Ma YQ, Thomas GN, Ng MC, Critchley JA, Chan JC, Tomlinson B (2003) The lipoprotein lipase gene HindIII polymorphism is associated with lipid levels in early-onset type 2 diabetic patients. *Metabolism* 52(3):338–343. <https://doi.org/10.1053/meta.2003.50053>
  28. Long T, Lu S, Li H et al (2018) Association of APOB and LIPC polymorphisms with type 2 diabetes in Chinese Han population. *Gene* 672:150–155. <https://doi.org/10.1016/j.gene.2018.06.010>
  29. Onat A, Erginel-Unaltuna N, Coban N, Cicek G, Yuksel H (2011) APOC3-482C>T polymorphism, circulating apolipoprotein C-III and smoking: interrelation and roles in predicting type-2 diabetes and coronary disease. *Clin Biochem* 44(5–6):391–396. <https://doi.org/10.1016/j.clinbiochem.2010.12.009>
  30. Abbas HT, Alic L, Erraguntla M et al (2019) Predicting long-term type 2 diabetes with support vector machine using oral glucose tolerance test. *PLoS One* 14(12):e0219636. <https://doi.org/10.1371/journal.pone.0219636>
  31. Tabak AG, Herder C, Rathmann W, Brunner EJ, Kivimaki M (2012) Prediabetes: a high-risk state for diabetes development. *Lancet* 379(9833):2279–2290. [https://doi.org/10.1016/S0140-6736\(12\)60283-9](https://doi.org/10.1016/S0140-6736(12)60283-9)

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.