

## ASSOCIATION STUDIES ARTICLE

# Epigenome-wide association study identifies *TXNIP* gene associated with type 2 diabetes mellitus and sustained hyperglycemia

Carolina Soriano-Tárraga<sup>1</sup>, Jordi Jiménez-Conde<sup>1,\*</sup>, Eva Giralt-Steinhauer<sup>1</sup>, Marina Mola-Caminal<sup>1</sup>, Rosa M. Vivanco-Hidalgo<sup>1</sup>, Angel Ois<sup>1</sup>, Ana Rodríguez-Campello<sup>1</sup>, Elisa Cuadrado-Godia<sup>1</sup>, Sergi Sayols-Baixeras<sup>2,3</sup>, Roberto Elosua<sup>2</sup> and Jaume Roquer<sup>1</sup>, on behalf of the GENESTROKE Consortium

<sup>1</sup>Department of Neurology, Hospital del Mar, Neurovascular Research Group, IMIM (Institut Hospital del Mar d'Investigacions Mèdiques), Universitat Autònoma de Barcelona/DCEXS-Universitat Pompeu Fabra, Barcelona, Spain, <sup>2</sup>Cardiovascular Epidemiology and Genetics Research Group, IMIM, Barcelona, Spain and <sup>3</sup>Universitat Pompeu Fabra, Barcelona, Spain

\*To whom correspondence should be addressed at: Jordi Jiménez-Conde, IMIM, Hospital del Mar Medical Research Institute, Dr. Aiguader 88, 08003 Barcelona, Spain. Tel: +34 933160611; Fax: +34 933160611; Email: [jjimenez@imim.es](mailto:jjimenez@imim.es)

## Abstract

Type 2 diabetes mellitus (DM) is an established risk factor for a wide range of vascular diseases, including ischemic stroke (IS). Glycated hemoglobin A1c (HbA1c), a marker for average blood glucose levels over the previous 12 weeks, is used as a measure of glycemic control and also as a diagnostic criterion for diabetes (HbA1c levels  $\geq 6.5\%$ ). Epigenetic mechanisms, such as DNA methylation, may be associated with aging processes and with modulation of the risk of various pathologies, such as DM. Specifically, DNA methylation could be one of the mechanisms mediating the relation between DM and environmental exposures. Our goal was to identify new CpG methylation sites associated with DM. We performed a genome-wide methylation study in whole-blood DNA from an IS patient cohorts. Illumina HumanMethylation450 BeadChip array was used to measure DNA methylation in CpG sites. All statistical analyses were adjusted for sex, age, hyperlipidemia, body mass index (BMI), smoking habit and cell count. Findings were replicated in two independent cohorts, an IS cohort and a population-based cohort, using the same array. In the discovery phase ( $N = 355$ ), we identified a CpG site, cg19693031 (located in the *TXNIP* gene) that was associated with DM ( $P = 1.17 \times 10^{-12}$ ); this CpG was replicated in two independent cohorts ( $N = 167$  and  $N = 645$ ). Methylation of *TXNIP* was inversely and intensely associated with HbA1c levels ( $P = 7.3 \times 10^{-16}$ ), specifically related to diabetic patients with poor control of glucose levels. We identified an association between the *TXNIP* gene and DM through epigenetic mechanisms, related to sustained hyperglycemia levels (HbA1c  $\geq 7\%$ ).

## Introduction

Type 2 diabetes mellitus (DM) is an established risk factor for a wide range of vascular diseases, including ischemic stroke (IS), independently of other conventional risk factors (1,2). The prevalence of DM in IS ranges from 15 to 44% (3–5).

In patients with DM, glucose accumulation results in enhanced glycation of many proteins. The glycated hemoglobin A1c (HbA1c) is formed when glucose attaches specifically to the NH<sub>2</sub>-terminal valine of the  $\beta$ -chain of hemoglobin A. Formation of HbA1c is essentially irreversible, and its concentration in the

Received: August 3, 2015. Revised: October 30, 2015. Accepted: November 26, 2015

© The Author 2015. Published by Oxford University Press. All rights reserved. For Permissions, please email: [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

blood depends on both the life span of the red blood cell, which averages ~120 days, and the blood glucose concentration. HbA1c is firmly established as a biomarker of long-term blood glucose concentrations and represents the average of glucose levels over the preceding 8 to 12 weeks. This biomarker has many favorable attributes, including no requirement that the patient be fasting, the possibility of collecting the sample at any time of the day, and the lack of any association between its concentration in the blood and acute factors such as stress or exercise (6). As our group has previously published, HbA1c determination provides prognostic information in IS patients without known DM (7). Elevated blood glucose is common in the acute phase of stroke, making HbA1c determination useful to diagnose new DM cases (8).

Diabetes is a complex disease, and both genetic and environmental factors contribute to its development. However, despite the depth of available research in the genetic area, all the identified DM risk variants explain only ~10% of its estimated heritability (9,10). Environmental factors, such as diet and exposure to hyperglycemia, contribute to the etiology of DM and its associated microvascular and macrovascular complications. Thus, researchers have begun to examine the role of epigenetics in DM (11,12). DNA methylation could be one of the epigenetic mechanisms mediating the relation between DM and environmental exposures. There is great interest in methylation profiling in peripheral blood to find disease-related associations, as specific methylation regions could be used as potent biomarkers (13). Indeed, recent studies have proposed that specific changes in the epigenome are associated with the onset and progression of diabetes and serve as markers of DM risk (14–16).

DNA methylation is an epigenetic mechanism regulating high-order DNA structure and gene expression. It is a heritable but also reversible addition of a methyl group to the 5-carbon position of cytosine in a cytosine-phosphate-guanine (CpG) context. This dinucleotide is quite rare in mammalian genomes (~1%) and clusters in regions known as CpG islands. The methylation of the CpG island is associated with gene silencing (17).

The aim of the present study was to determine the association between DM and DNA methylation. For this purpose, we screened DNA methylation changes using an epigenome-wide approach in our IS cohort; first, comparing non-DM patients with patients with DM; and second, we classified DM patients in two groups depending on their control of glucose levels: well-controlled DM with HbA1c < 7% (DM-good control) or poorly controlled DM with HbA1c ≥ 7% (DM-poor control). We identified and validated a new methylation CpG site associated with DM and HbA1c levels.

## Results

### Discovery analysis: genome-wide effect of diabetes on methylation status

A genome-wide DNA methylation analysis from whole blood was done in the discovery cohort (N = 355) with a DM prevalence of 42.5%. Clinical and demographic differences between DM and non-DM patients are shown in Table 1. Of the initial 485 577 CpG sites, 426 036 passed the quality controls and were included in the discovery analyses (Supplementary Material, Table S1).

There were no differences between the DM and non-DM groups in sex, hypertension, smoking habit, atrial fibrillation, stroke severity (assessed by the National Institutes of Health Stroke Scale, NIHSS) (18) and TOAST classification of IS subtype (19). Differences between patients with and without DM included age (72.6 versus 75.0 years,  $P = 0.04$ ), past history of ischemic heart disease (17.9 versus 10.8%,  $P = 0.06$ ), mean body mass index (BMI)

**Table 1.** Descriptive characteristics of IS patients of BASICMAR\_1 cohort by diabetic status

BASICMAR_1 cohort N = 355	Non-DM N = 204	DM N = 151	P-value
Age, years <sup>a</sup>	75.0 (11.4)	72.6 (4.5)	0.04
Sex, female, n (%)	88 (43.1)	61 (40.4)	0.61
Hypertension, n (%)	141 (69.1)	111 (73.5)	0.37
Smoking habit, n (%)	53 (26.0)	40 (26.5)	0.61
BMI (kg/m <sup>2</sup> ) <sup>a</sup>	27.1 (4.7)	28.0 (4.5)	0.09
Hyperlipidemia, n (%)	82 (40.2)	83 (55.0)	0.006
Blood glucose (mg/dl) <sup>a</sup>	114.2 (26.2)	183.9 (86.4)	<0.001
Hemoglobin A1c % <sup>a</sup>	5.4 (0.5)	7.7 (1.9)	<0.001
DM treatment, n (%)	–	104 (68.9)	<0.001
Insulin treatment, n (%)	–	17 (16.3)	<0.001
Oral DM medication, n (%)	–	73 (70.2)	<0.001
Oral and insulin treatment, n (%)	–	14 (13.5)	<0.001
Atrial fibrillation, n (%)	55 (27.0)	40 (26.5)	0.92
Ischemic heart disease, n (%)	22 (10.8)	27 (17.9)	0.06
NIHSS <sup>b</sup>	5 (2–10)	5 (3–9)	0.90
Ischemic stroke etiology, n (%)			0.57
Large-artery atherosclerosis	63 (31.0)	54 (36.2)	
Small-artery disease	82 (40.4)	54 (36.2)	
Cardioembolism	59 (28.6)	43 (27.5)	

<sup>a</sup>Mean (standard deviation).

<sup>b</sup>Median (interquartile range).

[28 kg/m<sup>2</sup> (SD = 4.5) versus 27.1 kg/m<sup>2</sup> (SD = 4.7),  $P = 0.09$ ], and hyperlipidemia (55.0% versus 40.2%,  $P = 0.006$ ), respectively. As expected, HbA1c values differed significantly [means of 7.7 (SD = 1.9) versus 5.4 (SD = 0.5), respectively;  $P < 0.001$ ]. In the first blood glucose determination during hospitalization, diabetic patients had significantly higher mean values [183.9 mg/dl (SD = 86.4) versus 114.2 mg/dl (SD = 26.2) in patients without DM,  $P < 0.001$ ]. A total of 104 (68.9%) DM patients were receiving treatment (16.3% insulin, 70.2% oral medication and 13.5% both therapies).

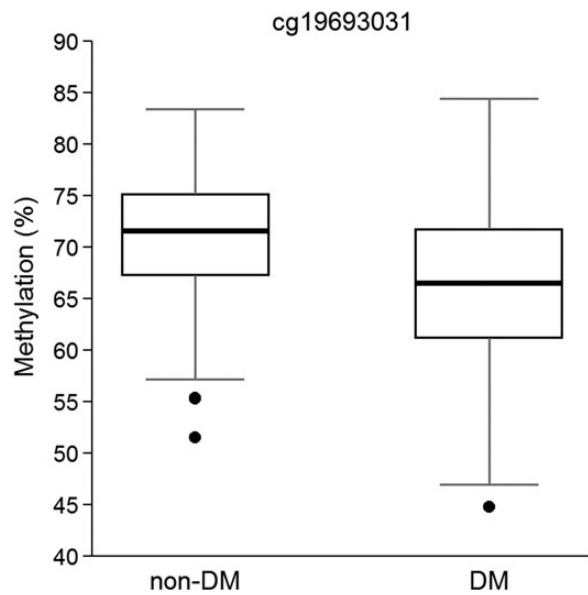
Regression models fitted to test the differences in DNA methylation levels between DM and non-DM patients showed a significant CpG site corresponding to cg19693031 located in TXNIP, which was hypomethylated in the DM group [ $P = 1.17 \times 10^{-12}$ , with a false discovery rate (FDR) correction adjusted  $P$ -value of  $5.0 \times 10^{-7}$ ], and a 5% difference in methylation between the two groups (Fig. 1). A Manhattan plot of the distribution of  $P$ -values and quantile–quantile plot for expected versus observed  $\chi^2$  values are shown in Figure 2.

To further validate our genome-wide DNA methylation pipeline, we also assessed the highly replicated DNA methylation association with smoking habit (current versus non-current smokers), in the genes AHRR, cg05575921 ( $P = 1.21 \times 10^{-33}$ ) and F2RL3, cg03636183 ( $P = 3.03 \times 10^{-24}$ ) within our dataset (20,21).

### Replication analysis of the novel CpG site associated with diabetes

Vascular risk factors and demographic data of the BASICMAR\_2 and REGICOR cohorts, used to replicate the CpG identified in our analysis, are shown in Supplementary Material, Table S2 and S3. BASICMAR\_2 contains 167 samples with methylation data from patients with IS, 108 non-DM and 59 DM (DM prevalence ~35.3%). The REGICOR data are from a population-based cohort with a total of 645 samples with methylation data, 582 non-DM and 63 DM (DM prevalence of ~10%).

The regression analysis between methylation and DM status was adjusted for sex, age, smoking habit, hyperlipidemia, BMI



**Figure 1.** Boxplot of  $\beta$  values of methylation in cg19693031 in *TXNIP* locus in non-DM and DM patients ( $N = 355$ ).  $P$ -value =  $1.17 \times 10^{-12}$ , adjusted for sex, age, hyperlipidemia, BMI, smoking habit and cell count.

and cell count. The significant cg19693031 CpG site in *TXNIP* observed in the discovery analysis was attempted to replication, and it was successfully replicated in both cohorts. The BASICMAR\_2 cohort showed mean methylation values of 0.74 ( $SD = 0.06$ ) in non-DM patients and 0.69 ( $SD = 0.09$ ) in DM patients, with an absolute methylation difference of 0.05 or 5% between groups ( $P = 1.37 \times 10^{-6}$ ). The REGICOR cohort had mean methylation values of 0.71 ( $SD = 0.06$ ) in non-DM patients and 0.68 ( $SD = 0.07$ ) in DM patients, with a 3% methylation difference between groups ( $P = 6.08 \times 10^{-5}$ ) (Table 2).

### Analysis of the role of HbA1c determination in methylation status at *TXNIP*

In order to explore the relationship between *TXNIP* hypomethylation and DM, we tested whether this association is more related to exposure to sustained hyperglycemia than to DM status. Therefore, we analyzed the association between DNA methylation and HbA1c (%), which reflects average glucose levels over the previous 12 weeks.

After FDR correction for multiple testing, we identified four CpG sites with differential methylation levels (Table 3). The most significant site was again cg19693031 in *TXNIP* locus,  $P = 7.3 \times 10^{-16}$ . The Manhattan plot of  $P$ -value distribution and quantile-quantile plot for expected versus observed  $\chi^2$  values are shown in Figure 3. HbA1c (%) were negatively correlated with percentage of methylation in this locus, with a correlation coefficient of  $-0.13$  (95% CI:  $-0.16, -0.88, P < 0.001$ ) (Fig. 4). Of the four CpGs identified, we replicated one of them in the BASICMAR\_2 cohort, the previously mentioned cg19693031 in *TXNIP* locus ( $P = 5.11 \times 10^{-9}$ ) (Table 3).

Ordinal regression analysis between methylation and DM status based on HbA1c values—non-DM, DM-good control (HbA1c  $< 7\%$ ) and DM-poor control (HbA1c  $\geq 7\%$ )—confirmed the significant association in cg19693031 of *TXNIP* locus ( $P$ -value =  $7.93 \times 10^{-11}$ ). Non-DM group was considered the reference group in logistic regression analyses, and the significant association in cg19693031 of *TXNIP* locus was due to DM patients with poor

control of glucose levels ( $P$ -value =  $1.89 \times 10^{-13}$ ) (Table 4). DM patients with poor control of glucose levels showed a mean hypomethylation of 6% in this CpG than the non-DM group (Fig. 5). A modest association was observed between DM-poor control and DM-good control ( $P$ -value =  $2.6 \times 10^{-5}$ ), with a mean methylation difference of 3%, DM-poor control group was hypomethylated compared with DM-good control group. No association was observed between non-DM and DM-good control groups, but a tendency is observed with a mean methylation difference between groups of 3%. All these results were replicated in the BASICMAR\_2 cohort (Table 4).

### Analysis of the role of DM medication

To explore possible effects of DM medication on our findings, we tested the association between methylation and DM medication at the time of blood sampling in DM patients ( $N = 151$ ). We did not find a significant association with cg19693031 in *TXNIP* locus ( $P$ -value = 0.29), demonstrating that DM treatment is not a confounder in our analysis.

### EpiTYPER validation

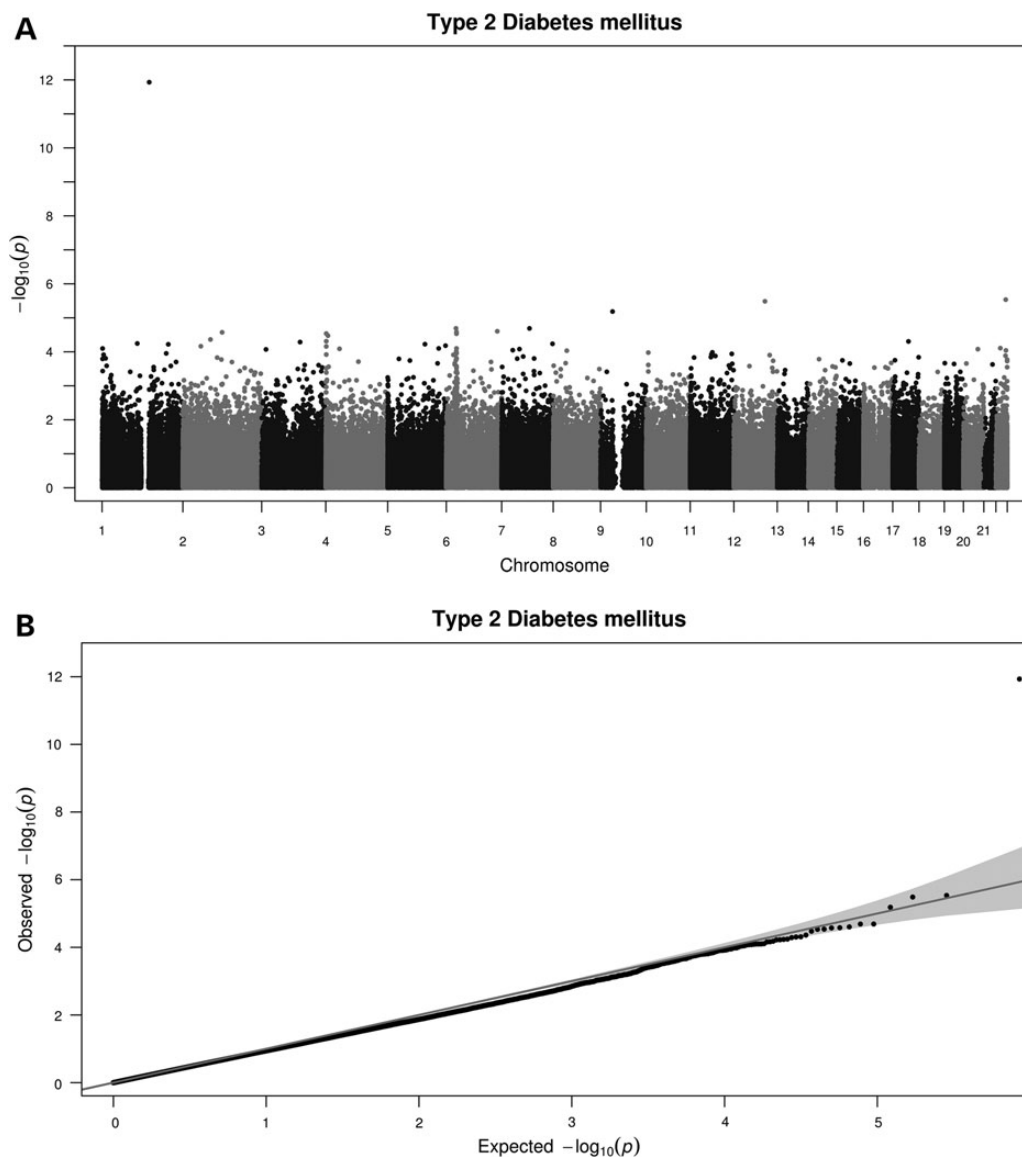
The differential methylation for cg19693031 in *TXNIP* was validated via Sequenom EpiTYPER approach in 283 randomly selected samples with HbA1c (%) available from the discovery cohort. The characteristics are summarized in Supplementary Material, Table S4. Sequenom EpiTYPER confirmed differential methylation in cg19693031 in *TXNIP* between DM status groups and HbA1c (%) as continuous variable ( $P < 0.001$ ). Moreover, an adjacent CpG in the same PCR product (chr1:145441393–145441394) in *TXNIP* locus showed comparable behavior than cg19693031 (chr1:145441552–145441553) (Supplementary Material, Table S5).

The association of patients in the DM-poor control group and cg19693031 from the Illumina HumanMethylation 450 BeadChip array was technically validated with this technique, demonstrating the reliability of the array in general.

### Discussion

In the present study, DM was associated with hypomethylation in the *TXNIP* gene in peripheral blood DNA of patients with IS and in a healthy population. Hypomethylation of this gene is related mainly to the average glucose levels in previous months (measured through HbA1c), suggesting an epigenetic consequence of hyperglycemia exposure rather than a cause of DM.

DNA methylation is an epigenetic mechanism regulating high-order DNA structure and gene expression. Usually DNA hypomethylation is associated with higher gene expression (17). Theoretically, the binding affinity of transcription factors may be affected by either local alteration of their DNA target sequences or through more global changes in chromatin accessibility (22–25). *In silico* analysis of cg19693031, located in the 3' UTR of *TXNIP*, showed that is a likely binding motive associated with glucose import and/or chaperone binding site (5' ACCCCAGAAGGT GA 3'). Thus, methylation differences in this locus may modify the binding affinity of methylation-sensitive transcription regulators, which are related to the mechanism of glucose exposure. In general, regulatory regions tend to be DNase-sensitive; cg19693031 is located in a DNase-hypersensitive area assayed in three blood cell lines (CMK, GM12878 and HL-60), as reported from the ENCODE project (<http://www.epigenomebrowser.org>), which reinforces our assumption.



**Figure 2.** Manhattan plot shows the distribution of the P-values of the association between methylation probes in DM analysis in BASICMAR\_1 discovery cohort (A) and QQ plot (B).

TXNIP encodes for thioredoxin-interacting protein and has multiple functions. It plays an important role in redox homeostasis and as a physiologic regulator of peripheral glucose uptake in humans induced by hyperglycemia and inhibits glucose uptake into fat and muscle (26–28). The phenotype of TXNIP knock-out (KO) mice demonstrated that TXNIP plays a crucial role in several pathological conditions, including metabolic syndrome, inflammation and cancer (29,30). Moreover, these KO mice were markedly more insulin-sensitive than controls and augmented glucose transport was identified in both adipose and skeletal muscles (31). TXNIP expression is glucose-responsive and its expression is consistently elevated in the muscle of pre-diabetic and diabetic patients (28). However, there is no evidence for an association between DM and common genetic variations in TXNIP (28). In a recent GWAS publication (32), some variants at multiple loci were associated with HbA1c. However, these loci seem to be linked to erythrocyte parameters rather than to glucose metabolism, which may explain why TXNIP did not appear to be associated

and also points to a TXNIP role in glycemia regulation. TXNIP expression is widely regulated by nutritional status, signals and enzymes (e.g., feeding-fasting, obesity, high glucose, amino acids and nuclear receptor signals). It has been reported as an important protein in glucose regulation and lipid homeostasis (29,33,34). In a recent publication, cg19693031 in TXNIP was associated with chylomicrons (class A) and known metabolic markers of DM, such as hexose and alpha-hydrobutyrate (35). While finalizing this manuscript, three studies were published and observed an association of DNA methylation at cg19693031 in TXNIP, and type 2 diabetes (36–38). Kulkarni *et al* (36) identified an association between cg19693031 methylation and DM, fasting blood glucose and insulin resistance in 850 pedigreed Mexican-American individuals using the Illumina HumanMethylation450 Beadchip. Chambers *et al* (37) in a large, prospected, nested case–control study ( $N = 25,372$ ), identified an association between differential methylation at five genetic loci, including TXNIP, and risk of future DM incidence among Indian Asians and Europeans. Finally, Florath *et al* (38)

**Table 2.** Methylation differences in the discovery BASICMAR\_1 cohort identified in the discovery analysis and validated in BASICMAR\_2 and REGICOR cohorts in peripheral blood between non-DM and DM in TXNIP gene

cg19693031	Cohorts	N	Non-DM <sup>a</sup>	DM <sup>a</sup>	Effect size (%)	P-value	Coef.
Discovery	BASICMAR_1	355	0.71 (0.06)	0.66 (0.08)	5	$1.17 \times 10^{-12}$	-0.35
Validation	BASICMAR_2	167	0.74 (0.06)	0.69 (0.09)	5	$1.37 \times 10^{-6}$	-0.30
	REGICOR	645	0.71 (0.06)	0.68 (0.07)	3	$6.08 \times 10^{-5}$	-0.53

Effect size (%): absolute difference in the mean beta values between groups expressed in percentage. P-value adjusted by sex, smoking habit, age, hyperlipidemia, BMI and cell count. Coef.: Regression coefficient.

<sup>a</sup>Observed mean  $\beta$  values (standard deviation).

**Table 3.** CpG sites with significantly differential methylation in relation to HbA1c (%) in the discovery cohort

CpG ID	Chr:position	Gene	Discovery BASICMAR_1			Replication BASICMAR_2				
			$\beta$ value <sup>a</sup>	P-value	FDR P-value	Coef.	$\beta$ value <sup>a</sup>	P-value	FDR P-value	Coef.
cg19693031	1: 145441552	TXNIP	0.69 (0.07)	$7.30 \times 10^{-16}$	$3.11 \times 10^{-10}$	-0.129	0.72 (0.08)	$5.11 \times 10^{-9}$	$2.05 \times 10^{-8}$	-0.144
cg01676795	7: 75586348	POR	0.80 (0.06)	$1.05 \times 10^{-7}$	0.019	0.061	0.51 (0.08)	0.413	0.818	0.018
cg07805383	2: 74346210	. . .	0.89 (0.10)	$1.32 \times 10^{-7}$	0.019	-0.162	0.89 (0.08)	0.345	0.783	0.039
cg26262157	10: 6214079	PFKFB3	0.48 (0.04)	$4.41 \times 10^{-7}$	0.047	-0.061	0.28 (0.07)	0.048	0.505	-0.037

The global significance level of 0.05% corrected for multiple comparisons was established at P-value = 0.05/6 = 0.008 to define a difference as statistically significant. CpGID; Chr:position, chromosome location (hg17); Gene: associated gene; P-value adjusted by sex, smoking habit, age, hyperlipidemia, BMI and cell count; FDR adjusted P-value; Coef.: regression coefficient.

<sup>a</sup>Observed mean  $\beta$  value (standard deviation).

showed the same association between cg19693031 methylation and DM in an elderly cohort from Germany (N = 988). Altogether, this information confirms a role for TXNIP in DM pathogenesis. DNA methylation could be the mechanism through which TXNIP expression is controlled and regulates glucose homeostasis.

Epidemiological evidence and data from clinical studies suggest the persistence of a 'metabolic memory' of past exposures to environmental factors or glycemic control (39). Results from two different clinical trials [Diabetes Control and Complications Trial (DCCT) in DM type 1 and Epidemiology of Diabetes Interventions and Complications in DM type 2] showed that early glycemic control after diagnosis contributes to decreased incidence of cardiovascular events, compared with late control. These observations led to the 'metabolic memory' hypothesis (40). DNA methylation may be one of the mechanisms involved in this phenomenon.

In this study, we analyzed one of the largest available series of participants with genome-wide methylation data, the first epigenome-wide association study analysis of IS patients. We replicated our results in two independent cohorts and validated them with a different technique. Moreover, we identified a CpG site related to type 2 DM patients and sustained hyperglycemia levels and validated it with different approaches. Additionally, we identified a new association of a nearby CpG in TXNIP (chr1:145441393–145441394). Recent publications have observed the same association between cg19693031 methylation and DM, validating and confirming our own research.

Some limitations of the study should be considered. First, we measured DNA methylation in peripheral blood cells. Methylation levels of some CpGs/regions are tissue-specific (41), and we might have lost some signals by not choosing specific tissues where glucose concentration could have a higher impact on DNA methylation. However, methylation patterns of whole blood have been reported as a good proxy for methylation levels from a specific site of action (13,21). Second, the design of the study is cross-sectional and we can only suspect but not infer causality of the reported associations.

In conclusion, hypomethylation of the TXNIP gene is related to type 2 DM. The inverse relationship between TXNIP

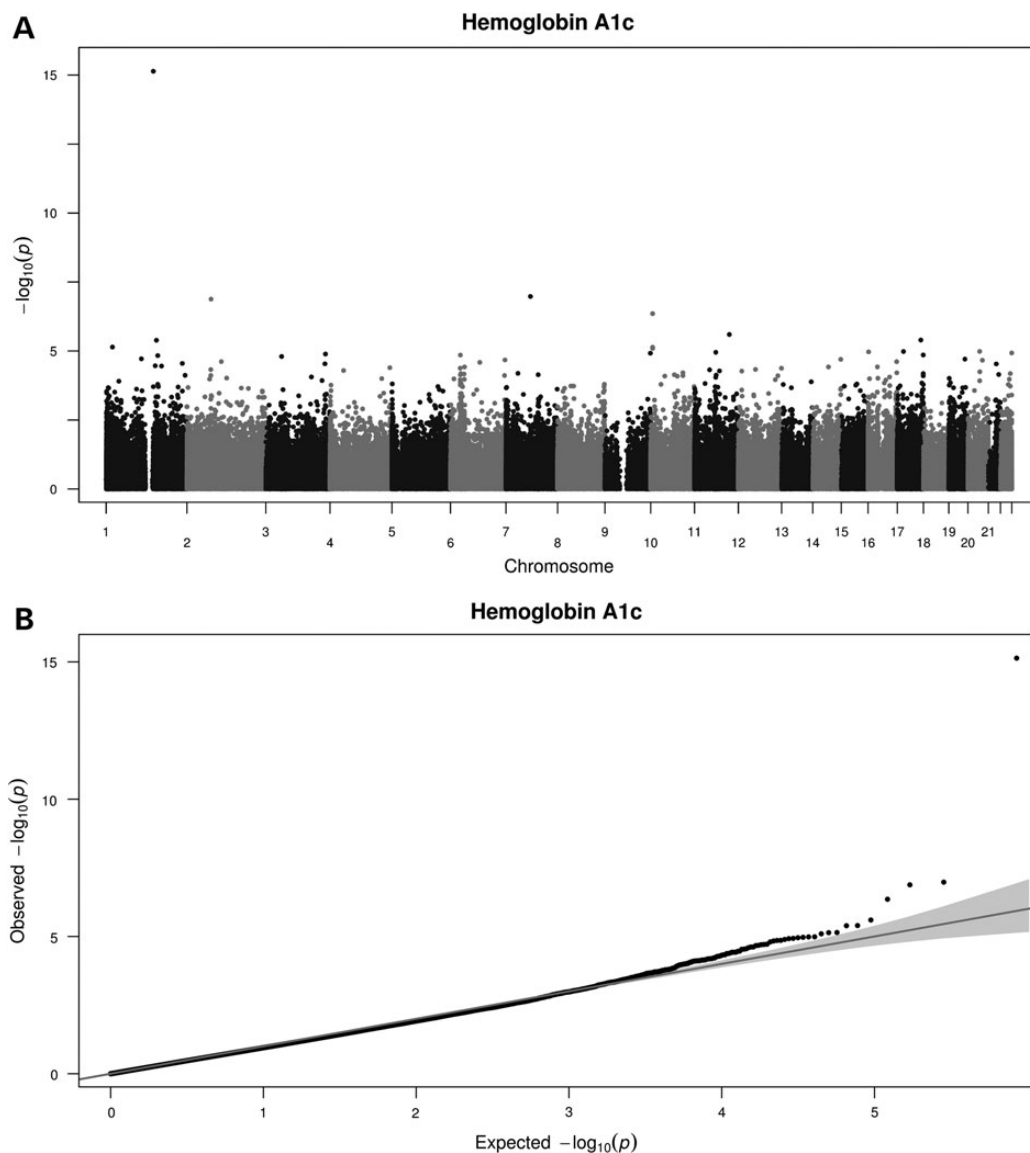
methylation and HbA1c values suggests that TXNIP hypomethylation is a consequence of sustained hyperglycemia levels. It has been reported that TXNIP expression is highly sensitive to glucose concentrations, and DNA methylation may be a key mechanism to modulate its expression and likely to be an early biomarker for impaired glucose homeostasis.

## Materials and Methods

### Study participants

The study included three independent prospective cohorts of Caucasian patients recruited in Spain (N = 355, N = 167 and N = 645), analyzed retrospectively.

The discovery cohort, BASICMAR\_1 cohort, was recruited in Hospital del Mar in Barcelona, Spain, and included a subsample of IS patients (N = 355), recruited from 2005 to 2009, from those enrolled in BasicMar Register (Ministerio de Sanidad y Consumo, Instituto de Salud Carlos III; FIS No. PI051737), an ongoing prospective registry of stroke patients (42). The BASICMAR\_2 cohort, a second subsample of IS patients (N = 167) from those enrolled in the BasicMar Register, was used as one of the replication cohorts recruited from 2009 to 2012. The BasicMar Register prospectively recruited all consenting patients who were admitted to our hospital from 2005 to 2012 (n = 4291; response rate, 80.8%) with a diagnosis of stroke fulfilling World Health Organization criteria. Inclusion criteria in BASICMAR cohorts were as follows: (1) first-ever IS, (2) brain imaging with CT or MRI, (3) availability of the clinical data supporting the assigned stroke subtype according to TOAST classification (19), (4) availability of the clinical data of HbA1c and BMI and (5) absence of intracranial hemorrhage, neoplasms, demyelinating and autoimmune diseases and vasculitides. All patients were assessed and classified by a neurologist and were included in the study by consecutive order of recruitment. The decision of using an IS cohort as the discovery cohort to study DM was based on several reasons. It has a higher prevalence of DM (~44%), compared with the healthy population, which may improve the discovery power, and also the well



**Figure 3.** Manhattan plot shows the distribution of the P-values of the association between methylation probes in Hemoglobin A1c analysis in BASICMAR\_1 discovery cohort (A) and QQ plot (B).

characterized and accurate phenotyped individuals add value to the discovery phase.

Girona Heart Registry (REGICOR, which stands for REGistre Gironi del COR) is a population-based cohort recruited in the province of Girona, in northeast Spain, ~100 km from Hospital del Mar (Barcelona). This register includes a randomized representative sample of men and women of the province of Girona. We used follow-up data from a population-based cohort originally enrolled in 2003–2005 ( $n = 6352$ ; response rate, 71.5%) from towns that represent the urban and rural diversity of Girona Province (43). During 2009–2013, participants still residing in these towns were invited to participate in a follow-up visit; institutionalized residents were excluded. The response rate was 78.4%. A subsample of those attending their follow-up visit was selected for a replication cohort, 645 individuals (582 controls and 63 DM). The selection of a population-based cohort for replication warrants the external validity of the results to the general population.

All subjects were of European descent. The study was approved by the local ethics committee, CEIC-Parc de Salut Mar,

and participants gave written informed consent. The study was conducted according to the principles expressed in the Declaration of Helsinki and relevant legislation in Spain.

#### Demographic and vascular risk factor variables

As defined by international guidelines, data on vascular risk factors analyzed were obtained from direct interview of the patient, relatives and caregivers, and from medical records. Examinations were performed and standardized questionnaires administered by a team of neurologists and reviewer by an additional neurologist.

We recorded age, sex and vascular risk factors, which were collected in a structured questionnaire during hospitalization, as follows: arterial hypertension, defined as systolic blood pressure (SBP)  $\geq 140$  mmHg or diastolic (DPB)  $\geq 90$  mmHg previous to the acute event in more than two determinations, a physician's diagnosis, or use of medication (collected during hospitalization); hyperlipidemia, defined as a physician's diagnosis, use of

medication, serum cholesterol concentration >220 mg/dl, low-density lipoprotein cholesterol (LDL) > 130 mg/dl, or serum triglyceride concentration > 150 mg/dl (collected during hospitalization); first measure glucose value (non-fasting) during hospitalization; ischemic heart disease (documented history of angina pectoris or myocardial infarction); and atrial fibrillation (documented history or diagnosis during hospitalization) and self-reported smoking habit. During hospitalization lymphocyte count, BMI, stroke severity was measured by the NIHSS (18), and TOAST criteria was used to classify IS subtype (19), is recorded through standardized protocol.

No missing data were observed for the variables age, sex, hypertension, hyperlipidemia, atrial fibrillation and ischemic heart disease. The percentage of participants in BASICMAR\_1 cohort with missing values for smoking habit was 2% and BMI was 12.1%. The percentage of participants in BASICMAR\_2 cohort with missing values for BMI was 9.7%.

### Diagnosis of diabetes mellitus type 2

DM diagnosis was based on evidence of two or more fasting blood glucose values  $\geq 126$  mg/dl, use of diabetes medication or

previous physician diagnosis. Hemoglobin A1c (HbA1c) values were obtained during hospitalization, only from BASICMAR cohorts, measured by high-performance liquid chromatography. The percentage of participants in BASICMAR\_1 cohort with missing values for HbA1c was 25%, and 1.6% in BASICMAR\_2 cohort. Patients with unknown DM and HbA1c values of  $\geq 6.5\%$  were classified as DM patients independently of current blood glucose levels. On the other hand, DM patients were classified in two groups depending on their control of glucose levels: HbA1c < 7% (DM-good control) or HbA1c  $\geq 7\%$  (DM-poor control).

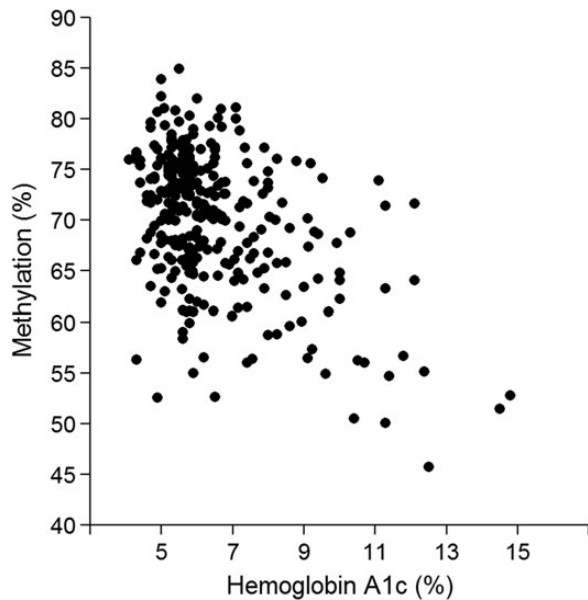
### Peripheral blood collection and DNA extraction

DNA samples were extracted from whole peripheral blood collected in 10 ml EDTA tubes. The Chemagic Magnetic Separation Module I system (Chemagen) was used for DNA isolation in BASICMAR cohorts, and The Autopure LS (Qiagen) in the REGICOR cohort. Genome-wide DNA methylation was assessed using the Illumina HumanMethylation450 Beadchip. DNA extractions were performed at the same time and stored together at  $-20^{\circ}\text{C}$ . DNA concentrations were quantified using Picogreen assay and Nanodrop technology. The quality of DNA samples was visualized in agarose gels.

### Array-based DNA methylation analysis with Infinium Human Methylation450k

Genomic DNA (1  $\mu\text{g}$ ) was bisulfite-converted using the EZ-96 DNA Methylation Kit (Zymo Research, Orange, CA, USA) according to the manufacturer's procedure, with the alternative incubation conditions recommended when using the Illumina Methylation Assay.

Genome-wide DNA methylation was assessed using the Illumina HumanMethylation450 Beadchip (Illumina Netherlands, Eindhoven, Netherlands) following the manufacturer's protocol with no modifications. This array covers 485 577 methylation CpG sites in 99% of RefSeq genes (21 231 genes). The arrays were scanned with the Illumina HiScan SQ scanner. These processes were carried out in Progenika Biopharma in Bizkaia, Spain, for the BASICMAR cohorts and in two different laboratories of the Spanish National Genotyping Centre for the REGICOR cohort: the Centre for Genomic Regulation in Barcelona and the Centro Nacional de Investigaciones Oncológicas in Madrid. Two repeated samples were included in all the plates to take into account batch effects. In the REGICOR cohort was performed a sensitivity analysis, normalizing the data by laboratory batch effect (Supplementary Material, Table S6).

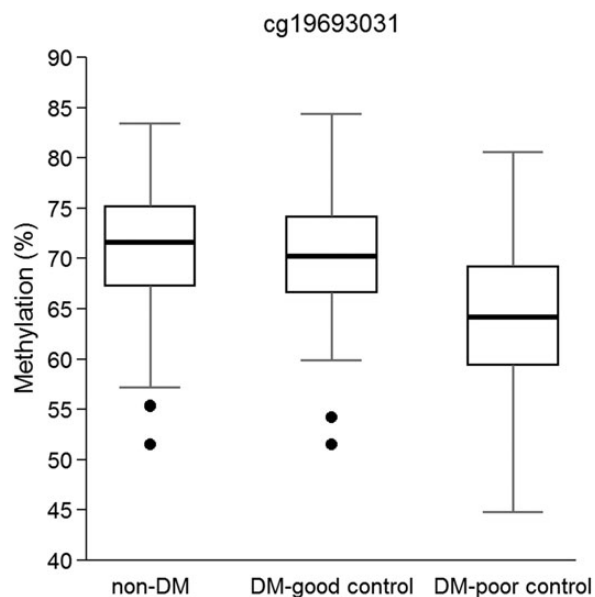


**Figure 4.** Scatterplot of  $\beta$  values of methylation in cg19693031 in TXNIP locus and Hemoglobin A1c in BASICMAR\_1 discovery cohort (%).  $P$ -value =  $7.30 \times 10^{-16}$ , adjusted for sex, age, hyperlipidemia, BMI, smoking habit and cell count.

**Table 4.** Genome-wide differentially methylated CpG cg19693031 in TXNIP locus between non-DM and DM patients based on HbA1c values, DM-good control (HbA1c < 7%) and DM-poor control (HbA1c  $\geq 7\%$ ), from discovery BASICMAR\_1 cohort and replication BASICMAR\_2 cohort

	cg19693031	N	HbA1c (SD)	$\beta$ values (SD)	Effect size (%)	P-value	Coef.
Discovery BASICMAR_1	Non-DM	151	5.41 (0.53)	0.70 (0.01)	Ref	Ref	Ref
	DM-good control	52	6.22 (0.49)	0.67 (0.01)	3	0.460	-0.05
	DM-poor control	65	8.83 (1.81)	0.64 (0.01)	6	$1.89 \times 10^{-13}$	-0.45
Validation BASICMAR_2	Non-DM	103	5.26 (0.53)	0.74 (0.02)	Ref	Ref	Ref
	DM-good control	22	6.25 (0.49)	0.71 (0.02)	3	0.570	0.057
	DM-poor control	37	8.44 (1.35)	0.67 (0.02)	7	$3.80 \times 10^{-8}$	-0.256

Non-DM was considered the reference group in the statistical analysis (Ref). HbA1c% (standard deviation, SD): mean HbA1c values (standard deviation);  $\beta$  values: observed mean  $\beta$  value in this CpG site (SD); effect size (%): absolute difference in the mean beta values with respect to reference group expressed in percentage. P-value, adjusted by sex, smoking habit, age, hyperlipidemia, BMI and cell count. Coef.: regression coefficient.



**Figure 5.** Boxplot of  $\beta$  values of methylation in cg19693031 in TXNIP locus in BASICMAR\_1 discovery cohort. Non-DM group ( $N = 151$ ), diabetics with a good control of glucose levels (DM-good control,  $N = 52$ ) and diabetics with a poor control of glucose levels (DM-poor control,  $N = 65$ ).  $P$ -values =  $7.93 \times 10^{-11}$ , adjusted for sex, age, hyperlipidemia, BMI, smoking habit and cell count.

### Data preprocessing and normalization

Data were preprocessed using standardized pipelines (44,45).

#### Sample quality control

Initial quality control of sample data was conducted using GenomeStudio version 2011.1 (Illumina, San Diego, CA, USA) with the methylation module (version 1.9.0) to determine the status of staining, extension, hybridization, target removal, bisulfite conversion, specificity, non-polymorphic and negative controls (without background correction or normalization). We used all the samples that had a detection rate over 95%. Then, we tested whether we could group samples by sex according to their DNA methylation levels on the X-chromosome using the *methylumi* R package (46). Samples that performed poorly in these quality controls were excluded from further analysis (Supplementary Material, Table S1).

#### Cpg quality control

We excluded all probes that were represented by a bead count under three in at least 5% of the samples. CpG sites having 1% of samples with a detection  $P$ -value of  $>0.05$  were removed and cross-reactive probes were excluded (47) using the *wateRmelon* R package (44). To avoid SNP (single-nucleotide polymorphism) effects on methylation measures and sex bias, we excluded CpGs with SNPs present inside the probe body and at the CpG site, and all probes associated with an autosomal position (Supplementary Material, Table S1).

Before analysis of BASICMAR cohorts, methylation values were corrected for background values and then normalized by SWAN using *minfi* Bioconductor package (48,49). REGICOR cohort normalized by *dasen* method (44,50) using the *RnBeads* package (51).

In statistical models,  $\beta$  values were transformed using a variance stabilization transformation to methylation  $M$ -values (52).  $\beta$  value ranges between 0 (completely unmethylated) and 1 (completely methylated), expressed as percentages (range 0–100%). Effect size refers to the absolute difference in the mean  $\beta$  values

between the groups analyzed; it can be expressed as percentage. Owing to its good statistical properties,  $M$ -value was the main outcome variable used in our analyses (51,53). We used a previously published algorithm to infer white blood cell counts from DNA methylation data (53). Prior to statistical analyses,  $M$ -values were adjusted to remove batch effect using the *SVA* R package (54). We used the array annotations provided by Illumina to assign probes to the corresponding genes. Finally, a total of 355 samples and 426 036 autosomal probes in the BASICMAR\_1 cohort were included in the analysis, 167 samples from the BASICMAR\_2 cohort and 645 samples from the REGICOR cohort. The significant CpG site was analyzed in each cohort.

### Sequenom EpiTYPER analysis: technical validation

The differential methylation for the most significant locus, cg19693031 in TXNIP (chr1:145 438 438–145 442 635), was validated with EpiTYPER assay (Sequenom, San Diego, CA) on 283 samples of the BASICMAR\_1 cohort with HbAc1 data available. Following the manufacturer's instructions, 10 ng of bisulfite-treated DNA were PCR-amplified and processed. At least two replicate amplifications were performed in all instances. The primers used were as follows: forward, 5' TGATTTTTTAATGTAGGAGGTGGTT 3' and reverse, 5' AATTCAAATCCAAAATTCCTATCA 3'.

### Analysis of in silico binding affinities

The MEME Suite web server was used to discover and analyze the sequence motifs representing features such as DNA binding sites and protein interaction domains in TXNIP locus (55). Conserved motif analyses were performed using the online MEME system version 4.10. The maximum motif width, the minimal motif width and the number of different motifs were set to 50, 5 and 8, respectively (56). The remaining parameters are the default values. Transcription factor motifs (discovered using MEME) were further analyzed for putative function by association with Gene Ontology (GO) terms using the motif-GO term association tool GOMO (57).

### Statistical analysis

Baseline characteristics were compared using  $t$ -test for continuous, and chi-squared for categorical variables.

First, we analyzed the association between DNA methylation at all the individual CpG sites and DM in BASICMAR\_1 cohort. We included all the individuals and CpG sites that passed quality controls and had the adjustment variables available (BASICMAR\_1 cohort  $N = 355$ ). The global significance level of 0.05% in the discovery cohort, corrected for multiple comparisons, was established at  $P$ -value  $< 1 \times 10^{-7}$  to define a difference as statistically significant. We analyzed for differences in methylation at the CpG sites between the group of individuals without DM and the group with diagnosed DM, using a multivariate linear regression. The CpGs with  $P$ -values  $< 5 \times 10^{-7}$  were analyzed in BASICMAR\_2 and REGICOR cohorts to replicate the results of the discovery analysis, to include significant CpGs in the EWAS analysis ( $P$ -value  $< 1 \times 10^{-7}$ ) and CpGs close to significance.

Second, after the replication of the results, and in order to better understand the relationship between DM and methylation changes, we analyzed the association between DNA methylation and HbAc1 values in BASICMAR cohorts. HbAc1 values were available in 268 patients in BASICMAR\_1 and in 162 patients in BASICMAR\_2.



We relied on methylation  $\beta$  values for the presentation of the scatterplots and boxplots, as they allow an easier interpretation of the results. In the linear models with covariates, we used the M-value because it offers better statistical capacity.

The validation of methylation values in cg19693031 obtained with Sequenom EpiTYPER assay included comparison of DM status using two-tailed Student t-test and Spearman correlation test with HbA1c as continuous variable.

The analyses were performed using the R statistical package, version 3.1.2 (58). The following packages were utilized: *minfi*, *sva* and *limma* (48,54,59). The same statistical methods were used in the replication phase. Methylation M-values were analyzed using multivariate linear regression. All the analyses were adjusted for age, sex, smoking habit, hyperlipidemia, BMI and cell count. The variables age, sex, smoking habit and cell count are known confounders in DNA methylation analysis and must be considered in the statistical analysis (Supplementary Material, Fig. S1) (60–66).

### Accession numbers

Methylation data have been submitted to the Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE69138.

### Supplementary Material

Supplementary Material is available at HMG online.

### Acknowledgements

We thank the GeneStroke ‘The Spanish Stroke Genetics Consortium’ led by Jordi Jiménez-Conde (jjimenez@imim.es), and formed by Carolina Soriano-Tárraga, Eva Giralte-Steinhauer, Marina Mola, Ángel Ois, Ana Rodríguez-Campello, Elisa Cuadrado-Godia and Jaume Roquer from Hospital del Mar-IMIM (Barcelona, Spain), Jurek Krupinski and Israel Fernández-Cadenas from Mútua de Terrassa (Terrassa, Barcelona), Caty Carrera from Hospital Vall d’Hebrón (Barcelona), Kelly Rabionet from CRG (Barcelona), Víctor Obach Hospital Clínic (Barcelona) Tófol Vivas and Rosa María Díaz Navarro from Hospital Son Espases (Mallorca). The authors wish to thank Elaine M. Lilly, Ph.D., Writers First Aid, provided assistance in the English language revision of the manuscript.

*Conflict of Interest statement.* None declared.

### Funding

This project was funded in part by the following sources: Agència de Gestió Ajuts Universitaris de Recerca (2014 SGR 1213); the Spanish Ministry of Economy through the Carlos III Health Institute (ISCIII-FIS-FEDER-ERDF, PI12/01238); the Red de Investigación Cardiovascular (RD12/0042/0020, RD12/0042/0013); and a RecerCaixa 2013 research grant (JJ086116). S.S.-B. is funded by an iPFIS contract from Instituto de Salud Carlos III (IFI14/00007).

### References

- Sarwar, N., Gao, P., Seshasai, S.R.K., Gobin, R., Kaptoge, S., Di Angelantonio, E., Ingelsson, E., Lawlor, D.A., Selvin, E., Stampfer, M. et al. (2010) Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. *Lancet*, **375**, 2215–2222.
- Spencer, E.A., Pirie, K.L., Stevens, R.J., Beral, V., Brown, A., Liu, B., Green, J. and Reeves, G.K. (2008) Diabetes and modifiable risk factors for cardiovascular disease: the prospective Million Women Study. *Eur. J. Epidemiol.*, **23**, 793–799.
- Jamrozik, K., Broadhurst, R.J., Anderson, C.S. and Stewart-Wynne, E.G. (1994) The role of lifestyle factors in the etiology of stroke. A population-based case-control study in Perth, Western Australia. *Stroke.*, **25**, 51–59.
- Kernan, W.N., Viscoli, C.M., Inzucchi, S.E., Brass, L.M., Bravata, D.M., Shulman, G.I. and McVeety, J.C. (2005) Prevalence of abnormal glucose tolerance following a transient ischemic attack or ischemic stroke. *Arch. Intern. Med.*, **165**, 227–233.
- Khoury, J.C., Kleindorfer, D., Alwell, K., Moomaw, C.J., Woo, D., Adeoye, O., Flaherty, M.L., Khatri, P., Ferioli, S., Broderick, J.P. et al. (2013) Diabetes mellitus: a risk factor for ischemic stroke in a large biracial population. *Stroke.*, **44**, 1500–1504.
- Sacks, D.B. (2013) Hemoglobin A1c in diabetes: panacea or pointless? *Diabetes*, **62**, 41–43.
- Roquer, J., Rodríguez-Campello, A., Cuadrado-Godia, E., Giralte-Steinhauer, E., Jiménez-Conde, J., Soriano, C. and Ois, A. (2014) The role of HbA1c determination in detecting unknown glucose disturbances in ischemic stroke. *PLoS One*, **9**, e109960.
- Lindsberg, P.J. and Roine, R.O. (2004) Hyperglycemia in acute stroke. *Stroke.*, **35**, 363–364.
- Voight, B.F., Scott, L.J., Steinthorsdottir, V., Morris, A.P., Dina, C., Welch, R.P., Zeggini, E., Huth, C., Aulchenko, Y.S., Thorleifsson, G. et al. (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.*, **42**, 579–589.
- McCarthy, M.I. (2010) Genomics, type 2 diabetes, and obesity. *N. Engl. J. Med.*, **363**, 2339–2350.
- Yuan, W., Xia, Y., Bell, C.G., Yet, I., Ferreira, T., Ward, K.J., Gao, F., Loomis, A.K., Hyde, C.L., Wu, H. et al. (2014) An integrated epigenomic analysis for type 2 diabetes susceptibility loci in monozygotic twins. *Nat. Commun.*, **5**, 5719.
- Hidalgo, B., Irvin, M.R., Sha, J., Zhi, D., Aslibekyan, S., Absher, D., Tiwari, H.K., Kabagambe, E.K., Ordovas, J.M. and Arnett, D. K. (2014) Epigenome-wide association study of fasting measures of glucose, insulin, and HOMA-IR in the Genetics of Lipid Lowering Drugs and Diet Network study. *Diabetes*, **63**, 801–807.
- Heyn, H. and Esteller, M. (2012) DNA methylation profiling in the clinic: applications and challenges. *Nat. Rev. Genet.*, **13**, 679–692.
- De Mello, V.D.F., Pulkkinen, L., Lalli, M., Kolehmainen, M., Pihlajamäki, J. and Uusitupa, M. (2014) DNA methylation in obesity and type 2 diabetes. *Ann. Med.*, **46**, 103–113.
- Pinney, S.E. and Simmons, R.A. (2010) Epigenetic mechanisms in the development of type 2 diabetes. *Trends Endocrinol. Metab.*, **21**, 223–229.
- Pirola, L., Balcerczyk, A., Okabe, J. and El-Osta, A. (2010) Epigenetic phenomena linked to diabetic complications. *Nat. Rev. Endocrinol.*, **6**, 665–675.
- Portela, A. and Esteller, M. (2010) Epigenetic modifications and human disease. *Nat. Biotechnol.*, **28**, 1057–1068.
- Goldstein, L.B. and Samsa, G.P. (1997) Reliability of the National Institutes of Health Stroke Scale. Extension to non-neurologists in the context of a clinical trial. *Stroke.*, **28**, 307–310.
- Adams Jr, H.P., Bendixen, B.H., Kappelle, L.J., Biller, J., Love, B. B., Gordon, D.L. and Marsh, E.E. 3rd. (1993) Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in acute stroke treatment. *Stroke.*, **24**, 35–41.

20. Zeilinger, S., Kühnel, B., Klopp, N., Baurecht, H., Kleinschmidt, A., Gieger, C., Weidinger, S., Lattka, E., Adamski, J., Peters, A. et al. (2013) Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS One*, **8**, e63812.
21. Shenker, N.S., Polidoro, S., van Veldhoven, K., Sacerdote, C., Ricceri, F., Birrell, M.A., Belvisi, M.G., Brown, R., Vineis, P. and Flanagan, J.M. (2013) Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. *Hum. Mol. Genet.*, **22**, 843–851.
22. Zentner, G.E. and Henikoff, S. (2014) High-resolution digital profiling of the epigenome. *Nat. Rev. Genet.*, **15**, 814–827.
23. Medvedeva, Y.A., Khamis, A.M., Kulakovskiy, I.V., Ba-Alawi, W., Bhuyan, M.S.I., Kawaji, H., Lassmann, T., Harbers, M., Forrest, A.R.R. and Bajic, V.B. (2014) Effects of cytosine methylation on transcription factor binding sites. *BMC Genomics*, **15**, 119.
24. Blattler, A. and Farnham, P.J. (2013) Cross-talk between site-specific transcription factors and DNA methylation states. *J. Biol. Chem.*, **288**, 34287–34294.
25. Choy, M.-K., Movassagh, M., Goh, H.-G., Bennett, M.R., Down, T.A. and Foo, R.S.Y. (2010) Genome-wide conserved consensus transcription factor binding motifs are hyper-methylated. *BMC Genomics*, **11**, 519.
26. Zhou, J. and Chng, W.-J. (2013) Roles of thioredoxin binding protein (TXNIP) in oxidative stress, apoptosis and cancer. *Mitochondrion*, **13**, 163–169.
27. Turturro, F., Friday, E. and Welbourne, T. (2007) Hyperglycemia regulates thioredoxin-ROS activity through induction of thioredoxin-interacting protein (TXNIP) in metastatic breast cancer-derived cells MDA-MB-231. *BMC Cancer*, **7**, 96.
28. Parikh, H., Carlsson, E., Chutkow, W.A., Johansson, L.E., Storgaard, H., Poulsen, P., Saxena, R., Ladd, C., Schulze, P.C., Mazzini, M.J. et al. (2007) TXNIP regulates peripheral glucose metabolism in humans. *PLoS Med.*, **4**, e158.
29. Yoshihara, E., Masaki, S., Matsuo, Y., Chen, Z., Tian, H. and Yodoi, J. (2014) Thioredoxin/Txnip: redoxosome, as a redox switch for the pathogenesis of diseases. *Front. Immunol.*, **4**, 514.
30. Spindel, O.N., World, C. and Berk, B.C. (2012) Thioredoxin interacting protein: redox dependent and independent regulatory mechanisms. *Antioxid. Redox Signal.*, **16**, 587–596.
31. Chutkow, W.A., Birkenfeld, A.L., Brown, J.D., Lee, H.-Y., Frederick, D.W., Yoshioka, J., Patwari, P., Kursawe, R., Cushman, S.W., Plutzky, J. et al. (2010) Deletion of the alpha-arrestin protein Txnip in mice promotes adiposity and adipogenesis while preserving insulin sensitivity. *Diabetes*, **59**, 1424–1434.
32. Chen, P., Takeuchi, F., Lee, J.-Y., Li, H., Wu, J.-Y., Liang, J., Long, J., Tabara, Y., Goodarzi, M.O., Pereira, M.A. et al. (2014) Multiple nonglycemic genomic loci are newly associated with blood level of glycated hemoglobin in East Asians. *Diabetes*, **63**, 2551–2562.
33. Wu, N., Zheng, B., Shaywitz, A., Dagon, Y., Tower, C., Bellinger, G., Shen, C.-H., Wen, J., Asara, J., McGraw, T.E. et al. (2013) AMPK-dependent degradation of TXNIP upon energy stress leads to enhanced glucose uptake via GLUT1. *Mol. Cell*, **49**, 1167–1175.
34. Oka, S.-i. (2005) Impaired fatty acid utilization in thioredoxin binding protein-2 (TBP-2)-deficient mice: a unique animal model of Reye syndrome. *FASEB J.*, **20**, 121–123.
35. Petersen, A.-K., Zeilinger, S., Kastenmüller, G., Römisch-Margl, W., Brügger, M., Peters, A., Meisinger, C., Strauch, K., Hengstenberg, C., Pagel, P. et al. (2014) Epigenetics meets metabolomics: an epigenome-wide association study with blood serum metabolic traits. *Hum. Mol. Genet.*, **23**, 534–545.
36. Kulkarni, H., Kos, M.Z., Neary, J., Dyer, T.D., Kent, J.W., Göring, H.H., Cole, S.A., Comuzzie, A.G., Almasy, L., Mahaney, M.C. et al. (2015) Novel epigenetic determinants of type 2 diabetes in Mexican-American families. *Hum. Mol. Genet.*, **24**, 5330–5344.
37. Chambers, J.C., Loh, M., Lehne, B., Drong, A., Kriebel, J., Motta, V., Wahl, S., Elliott, H.R., Rota, F., Scott, W.R. et al. (2015) Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. *Lancet. Diabetes Endocrinol.*, **3**, 526–534.
38. Florath, I., Butterbach, K., Heiss, J., Bewerunge-Hudler, M., Zhang, Y., Schöttker, B. and Brenner, H. (2015) Type 2 diabetes and leucocyte DNA methylation: an epigenome-wide association study in over 1,500 older adults. *Diabetologia*, doi:10.1007/s00125-015-3773-7.
39. Waki, H., Yamauchi, T. and Kadowaki, T. (2012) The epigenome and its role in diabetes. *Curr. Diab. Rep.*, **12**, 673–685.
40. Ceriello, A. (2012) The emerging challenge in diabetes: the ‘metabolic memory’. *Vascul. Pharmacol.*, **57**, 133–138.
41. Lowe, R., Slodkowitz, G., Goldman, N. and Rakyan, V.K. (2015) The human blood DNA methylome displays a highly distinctive profile compared with other somatic tissues. *Epigenetics*, doi:10.1080/15592294.2014.1003744.
42. Roquer, J., Rodríguez-Campello, A., Gomis, M., Jiménez-Conde, J., Cuadrado-Godia, E., Vivanco, R., Giralt, E., Sepúlveda, M., Pont-Sunyer, C., Cucurella, G. et al. (2008) Acute stroke unit care and early neurological deterioration in ischemic stroke. *J. Neurol.*, **255**, 1012–1017.
43. Grau, M., Subirana, I., Elosua, R., Solanas, P., Ramos, R., Masiá, R., Cerdón, F., Sala, J., Juvinyà, D., Cerezo, C. et al. (2007) Trends in cardiovascular risk factor prevalence (1995–2000–2005) in northeastern Spain. *Eur. J. Cardiovasc. Prev. Rehabil.*, **14**, 653–659.
44. Pidsley, R., Wong, C.C.Y., Volta, M., Lunnon, K., Mill, J. and Schalkwyk, L.C. (2013) A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics*, **14**, 293.
45. Touleimat, N. and Tost, J. (2012) Complete pipeline for Infinium® Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics*, **4**, 325–341.
46. Davis, S., Du, P., Bilke, S., Triche, T. and Bootwalla, M. (2015) methylumi: Handle Illumina methylation data. *R Packag. version 2.12.0*.
47. Chen, Y., Lemire, M., Choufani, S., Butcher, D.T., Grafodatskaya, D., Zanke, B.W., Gallinger, S., Hudson, T.J. and Weksberg, R. (2013) Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics*, **8**, 203–209.
48. Aryee, M.J., Jaffe, A.E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A.P., Hansen, K.D. and Irizarry, R.A. (2014) Minfi: a flexible and comprehensive bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*, **30**, 1363–1369.
49. Maksimovic, J., Gordon, L. and Oshlack, A. (2012) SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome Biol.*, **13**, R44.
50. Dedeurwaerder, S., Defrance, M., Bizet, M., Calonne, E., Bontempi, G. and Fuks, F. (2013) A comprehensive overview of Infinium HumanMethylation450 data processing. *Brief. Bioinform.*, doi:10.1093/bib/bbt054.
51. Assenov, Y., Müller, F., Lutsik, P., Walter, J., Lengauer, T. and Bock, C. (2014) Comprehensive analysis of DNA methylation data with RnBeads. *Nat. Methods*, **11**, 1138–1140.

52. Du, P., Zhang, X., Huang, C.-C., Jafari, N., Kibbe, W.a, Hou, L. and Lin, S.M. (2010) Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*, **11**, 587.
53. Houseman, E.A., Accomando, W.P., Koestler, D.C., Christensen, B.C., Marsit, C.J., Nelson, H.H., Wiencke, J.K. and Kelsey, K.T. (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*, **13**, 86.
54. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E. and Storey, J.D. (2012) The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*, **28**, 882–883.
55. Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W. and Noble, W.S. (2009) MEME SUITE: tools for motif discovery and searching. *Nucl. Acids Res.*, **37**, W202–W208.
56. Bailey, T.L. and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, **2**, 28–36.
57. Buske, F.A., Bodén, M., Bauer, D.C. and Bailey, T.L. (2010) Assigning roles to DNA regulatory motifs using comparative genomics. *Bioinformatics*, **26**, 860–866.
58. R Core Team (2014) *R: A Language and Environment for Statistical Computing*. R Found. Stat. Comput, Vienna, Austria.
59. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W. and Smyth, G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucl. Acids Res.*, doi:10.1093/nar/gkv007.
60. Boks, M.P., Derks, E.M., Weisenberger, D.J., Strengman, E., Janson, E., Sommer, I.E., Kahn, R.S. and Ophoff, R.A. (2009) The relationship of DNA methylation with age, gender and genotype in twins and healthy controls. *PLoS One*, **4**, e6767.
61. Terry, M.B., Delgado-Cruzata, L., Vin-Raviv, N., Wu, H.C. and Santella, R.M. (2011) DNA methylation in white blood cells: association with risk factors in epidemiologic studies. *Epigenetics*, **6**, 828–837.
62. Langevin, S.M., Houseman, E.A., Christensen, B.C., Wiencke, J.K., Nelson, H.H., Karagas, M.R., Marsit, C.J. and Kelsey, K.T. (2011) The influence of aging, environmental exposures and local sequence features on the variation of DNA methylation in blood. *Epigenetics*, **6**, 908–919.
63. Horvath, S., Zhang, Y., Langfelder, P., Kahn, R.S., Boks, M.P.M., van Eijk, K., van den Berg, L.H. and Ophoff, R.A. (2012) Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biol.*, **13**, R97.
64. Zhang, F.F., Cardarelli, R., Carroll, J., Fulda, K.G., Kaur, M., Gonzalez, K., Vishwanatha, J.K., Santella, R.M. and Morabia, A. (2011) Significant differences in global genomic DNA methylation by gender and race/ethnicity in peripheral blood. *Epigenetics*, **6**, 623–629.
65. Tsaprouni, L.G., Yang, T.-P., Bell, J., Dick, K.J., Kanoni, S., Nisbet, J., Viñuela, A., Grundberg, E., Nelson, C.P., Meduri, E. et al. (2014) Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. *Epigenetics*, **9**, 1382–1396.
66. Zhu, Z.-Z., Hou, L., Bollati, V., Tarantini, L., Marinelli, B., Cantone, L., Yang, A.S., Vokonas, P., Lissowska, J., Fustinoni, S. et al. (2012) Predictors of global methylation levels in blood DNA of healthy subjects: a combined analysis. *Int. J. Epidemiol.*, **41**, 126–139.