LUND UNIVERSITY

The Human Pancreatic Islet Methylome and Its Role in Type 2 Diabetes

Dayeh, Tasnim

2016

[Link to publication](#)

Total number of authors:
1

ARTICLE

Identification of CpG-SNPs associated with type 2 diabetes and differential DNA methylation in human pancreatic islets

T. A. Dayeh · A. H. Olsson · P. Volkov · P. Almgren ·
T. Rönn · C. Ling

Abstract

Aims/hypothesis To date, the molecular function of most of the reported type 2 diabetes-associated loci remains unknown. The introduction or removal of cytosine–phosphate–guanine (CpG) dinucleotides, which are possible sites of DNA methylation, has been suggested as a potential mechanism through which single-nucleotide polymorphisms (SNPs) can affect gene function via epigenetics. The aim of this study was to examine if any of 40 SNPs previously associated with type 2 diabetes introduce or remove a CpG site and if these CpG-SNPs are associated with differential DNA methylation in pancreatic islets of 84 human donors.

Methods DNA methylation was analysed using pyrosequencing.

Results We found that 19 of 40 (48%) type 2 diabetes-associated SNPs introduce or remove a CpG site. Successful DNA methylation data were generated for 16 of these 19 CpG-SNP loci, representing the candidate genes *TCF7L2*, *KCNQ1*, *PPARG*, *HHEX*, *CDKN2A*, *SLC30A8*, *DUSP9*, *CDKAL1*, *ADCY5*, *SRR*, *WFS1*, *IRS1*, *DUSP8*, *HMGA2*, *TSPAN8* and *CHCHD9*. All analysed CpG-SNPs were associated with differential DNA methylation of the CpG-SNP site in human islets. Moreover, six CpG-SNPs, representing *TCF7L2*, *KCNQ1*, *CDKN2A*, *ADCY5*, *WFS1* and *HMGA2*, were also associated with DNA methylation of surrounding CpG sites. Some of the type 2 diabetes CpG-SNP sites that exhibit differential DNA methylation were further associated with gene expression, alternative splicing events determined by splice index, and hormone secretion in the human islets. The 19 type 2 diabetes-associated CpG-SNPs are in strong linkage disequilibrium ($r^2 > 0.8$) with a total of 295 SNPs, including 91 CpG-SNPs.

Conclusions/interpretation Our results suggest that the introduction or removal of a CpG site may be a molecular mechanism through which some of the type 2 diabetes SNPs affect gene function via differential DNA methylation and consequently contributes to the phenotype of the disease.

Keywords Alternative splicing · CpG-SNP · DNA methylation · Epigenetics · Glucagon · Human pancreatic islets · Insulin secretion · SNP · Splice index · Type 2 diabetes

Abbreviations

CpG	Cytosine–phosphate–guanine
GWAS	Genome-wide association study
LD	Linkage disequilibrium
SI	Stimulation index
SNAP	SNP annotation and proxy search
SNP	Single-nucleotide polymorphism

T. A. Dayeh · A. H. Olsson · P. Volkov · T. Rönn · C. Ling (✉)
Department of Clinical Sciences, Epigenetics and Diabetes,
Lund University Diabetes Centre, Clinical Research Centre,
Scania University Hospital, Jan Waldenströmsgata 35,
205 02 Malmö, Sweden
e-mail: charlotte.ling@med.lu.se

P. Almgren
Department of Clinical Sciences, Diabetes and Endocrinology,
Lund University Diabetes Centre, Clinical Research Centre,
Scania University Hospital, Jan Waldenströmsgata 35,
205 02 Malmö, Sweden

Introduction

Type 2 diabetes is a heterogeneous disease caused by a complex interplay between multiple genetic, epigenetic and environmental factors [1]. Attempts at elucidating the genetic mechanisms behind this disease have identified approximately 40 single-nucleotide polymorphisms (SNPs)

that show genome-wide associations with type 2 diabetes [2]. However, evidence for how these SNPs predispose to type 2 diabetes is still scarce and is only available for a few variants that mainly affect insulin secretion and sensitivity [3]. Understanding the molecular mechanisms through which these SNPs exert an effect on their target genes is not an easy task, especially as most of them are non-coding SNPs [2]. To further understand the pathogenesis of the disease, additional studies are required that link these SNPs with their molecular and biological functions.

Recent studies from our group and others have shown that epigenetic modifications such as DNA methylation and histone modifications may affect the pathogenesis of type 2 diabetes [4–9]. In differentiated mammalian cells, DNA methylation mainly takes place at a cytosine in a cytosine–phosphate–guanine (CpG) dinucleotide [10] and increased DNA methylation has been associated with transcriptional silencing [11]. In a previous study, we showed that genetic, epigenetic and non-genetic factors integrate to influence the expression of a candidate gene for type 2 diabetes in human skeletal muscle [12]. We identified a SNP near *NDUFB6* that introduces a CpG site and this CpG-SNP was associated with increased DNA methylation, decreased gene expression, and insulin resistance in the muscle of elderly, but not young, individuals [12]. Moreover, a recent study has suggested that CpG-SNPs may affect the expression of target genes by interfering with the binding of certain proteins [13]. However, to our knowledge the role of CpG-SNPs associated with type 2 diabetes has not been examined in human pancreatic islets.

Therefore, the aim of this study was to examine if 40 SNPs previously associated with type 2 diabetes [2] introduce or remove possible DNA methylation sites – CpG sites – and if these CpG-SNPs are associated with differential DNA methylation and subsequently affect gene expression, alternative splicing events and hormone secretion in human pancreatic islets.

Methods

Human pancreatic islet donors Pancreatic islets from 84 deceased non-diabetic donors were provided by the Nordic Network for Islet Transplantation, Uppsala University. Clinical characteristics of these islet donors are given in Table 1. The donor before death or her/his relatives on admission to intensive care unit gave their consent to donate organs and the local ethics committees approved the protocols. Islet culture and nucleic acid purification were performed as described previously [7].

SNP selection and identification of CpG-SNPs The selection criterion for SNPs was an association with type 2

Table 1 Characteristics of the human pancreatic islet donors

Number (women/men)	84 (32/52)
Age (years)	57.1 ± 10.3
BMI (kg/m^2)	25.9 ± 3.6
HbA$_{1c}$ (%)	5.6 ± 0.4
HbA$_{1c}$ (mmol/mol)	48.1 ± 5

Data are presented as mean ± SD

HbA$_{1c}$ was measured with the Mono-S method

diabetes identified by major genome-wide association studies (GWASs) as reported by McCarthy [2]. To determine if the selected SNPs were CpG-SNPs, the DNA sequences of the variants were retrieved from the dbSNP database (www.ncbi.nlm.nih.gov/snp/, on or after 03 July 2011). A SNP was considered to be a CpG-SNP if it introduces or removes a CpG site (Table 2). As a result, 19 SNPs fulfilled our criteria of being CpG-SNPs that show genome-wide association with type 2 diabetes. DNA sequences of these 19 CpG-SNPs and surrounding regions were then retrieved from University of California, Santa Cruz Human Genome Browser (UCSC) – hg18 assembly.

A flow chart of the study design and SNP selection can be seen in Fig. 1.

DNA methylation assay design To analyse DNA methylation of the 19 selected CpG-SNP sites, pyrosequencing PCR and sequencing primers were designed for the forward strand using the PyroMark Assay design Software 2.0 (Qiagen, Hilden, Germany) (electronic supplementary material [ESM] Table 1). The reverse or forward primer was biotinylated at its 5′ end. When possible, CpG sites surrounding the CpG-SNP of interest were also analysed. Problems caused by sequence characteristics that resulted in amplicons that were too long or too many CpG sites in the primer sequences resulted in failure of the assay design for rs5219 (*KCNJ11*) (Fig. 1).

DNA methylation analysis Bisulfite conversion of 500 ng genomic DNA from human pancreatic islets was performed using the EpiTect Bisulfite Kit (Qiagen) according to the manufacturer's instructions. Bisulfite-converted DNA was amplified by PCR using the PyroMark PCR kit (Qiagen) and the designed primers (ESM Table 1). Pyrosequencing was performed using the PSQ HS96A (Biotage, Uppsala, Sweden) and PyroMark Gold CDT kit (Qiagen) or using PyroMark Q96ID and the PyroMark Gold Q96 kit (Qiagen), all according to the manufacturer's instructions. Data analysis was performed using the Pyro Q-CpG 1.0.9 and PSQ HS96A 1.2 software (Biotage) or the PyroMark Q96 (version 2.5.8) software (Qiagen). Only data that passed a quality check or had medium quality were included in the analysis and as a result the following assay was excluded:

Table 2 SNPs associated with type 2 diabetes identified by GWAS

SNP ID	Candidate gene	Risk allele	Chr.	SNP position (build 37.1)	SNP location	Reference
CpG-SNPs where risk alleles remove a CpG site (n=8)						
rs5945326	DUSP9[a]	A	X	152899922	Intergenic	Voight et al (2010) [20]
rs564398	CDKN2A[a]	A	9	22029547	Intergenic	Zeggini et al (2007) [21]
rs11708067	ADCY5[a]	A	3	123065778	Intron	Dupuis et al (2010) [22]
rs1801214	WFS1[a]	T	4	6303022	Cds-synon	Voight et al (2010) [20]
rs2334499	DUSP8[a]	A	11	1696849	Intergenic	Kong et al (2009) [23]
rs7578326	IRS1[a]	A	2	227020653	Intergenic	Voight et al (2010) [20]
rs5219	KCNJ11[b]	T	11	17409572	Cds-nonsynon	Scott et al (2007) [24]
rs1801282	PPARG[a]	C	3	12393125	Cds-nonsynon	Scott et al (2007) [24]
CpG-SNPs where risk alleles introduce a CpG site (n=11)						
rs13292136	CHCHD9[a]	C	9	81952128	Intergenic	Voight et al (2010) [20]
rs7901695	TCF7L2[a]	G	10	114754088	Intron	Zeggini et al (2007) [21]
rs7754840	CDKAL1[a]	C	6	20661250	Intron	Scott et al (2007) [24]
rs391300	SRR[a]	G	17	2216258	Intron	Tsai et al (2010) [25]
rs5015480	HHEX[a]	G	10	94465559	Intergenic	Voight et al (2010) [20]
rs13266634	SLC30A8[a]	G	8	118184783	Cds-nonsynon	Sladek et al (2007) [26]
rs4457053	ZBED3[c]	G	5	76424949	Intergenic	Voight et al (2010) [20]
rs7961581	TSPAN8[a]	C	12	71663102	Intergenic	Zeggini et al (2008) [27]
rs1531343	HMGA2[a]	C	12	66174894	Intergenic	Voight et al (2010) [20]
rs2237895	KCNQ1[a]	C	11	2857194	Intron	Yasuda et al (2008) [28]
rs12779790	CDC123[c]	G	10	12328010	Intergenic	Zeggini et al (2008) [27]
Non-CpG SNPs (n=21)						
rs243021	BCL11A	A	2	60584819	Intergenic	Voight et al (2010) [20]
rs10830963	MTNR1B	G	11	92708710	Intron	Dupuis et al (2010) [22]
rs7578597	THADA	T	2	43732823	Missense	Zeggini et al (2008) [27]
rs17584499	PTPRD	T	9	8879118	Intron	Tsai et al (2010) [25]
rs10923931	NOTCH2	T	1	120517959	Intron	Zeggini et al (2008) [27]
rs7593730	RBMS1	C	2	161171454	Intron	Qi et al (2010) [29]
rs4607103	ADAMTS9	C	3	64711904	Intergenic	Zeggini et al (2008) [27]
rs1470579	IGF2BP2	C	3	185529080	Intron	Voight et al (2010) [20]
rs864745	JAZF1	T	7	28180556	Intron	Zeggini et al (2008) [27]
rs972283	KLF14	G	7	130466854	Intergenic	Voight et al (2010) [20]
rs896854	TP53INP1	T	8	95960511	Intron	Voight et al (2010) [20]
rs1552224	CENTD2	A	11	72433098	Intron	Voight et al (2010) [20]
rs7957197	HNF1A	T	12	121460686	Intron	Voight et al (2010) [20]
rs11634397	ZFAND6	G	15	80432222	Intergenic	Voight et al (2010) [20]
rs8042680	PRC1	A	15	91521337	Intron	Voight et al (2010) [20]
rs9939609	FTO	A	16	53820527	Intron	Frayling et al (2007) [30]
rs4607517	GCK	A	7	44235668	Intergenic	Dupuis et al (2010) [22]
rs2191349	DGKB	T	7	15064309	Intergenic	Dupuis et al (2010) [22]
rs780094	GCKR	C	2	27741237	Intron	Dupuis et al (2010) [22]
rs340874	PROX1	C	1	214159256	Intergenic	Dupuis et al (2010) [22]
rs4430796	HNF1B	G	17	36098040	Intron	Voight et al (2010) [20]

[a] Analysed CpG-SNPs

[b] Assay design problems

[c] Failed assay

Chr., chromosome; Cds-synon., coding-synonymous; Cds-nonsynon., coding-nonsynonymous
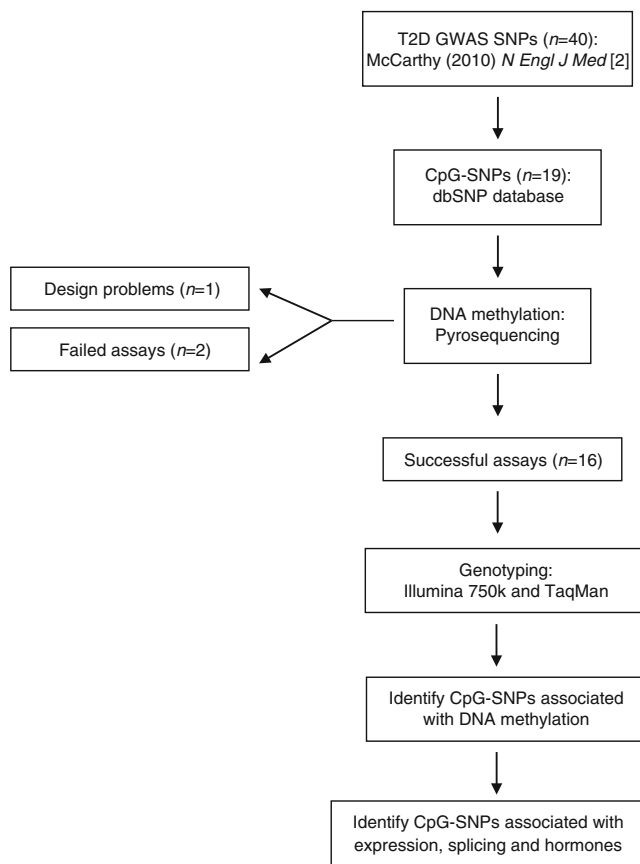
Fig. 1 Flow chart of the study design and SNP selection process. T2D, type 2 diabetes

rs12779790 (*CDC123*) (Fig. 1). Moreover, the assay for rs4457053 (*ZBED3*) failed PCR amplification and hence was also excluded.

Genotyping analysis Islet donors were genotyped for the 16 CpG-SNPs for which we retrieved DNA methylation data. The following 12 SNPs were genotyped using the HumanOmniExpressBeadChip and iScan system (Illumina, San Diego, CA, USA) according to the manufacturer's instructions: rs5015480, rs7901695, rs391300, rs11708067, rs7754840, rs13266634, rs564398, rs5945326, rs7961581, rs2237895, rs1801282 and rs2334499. Four SNPs (rs13292136, rs7578326, rs1531343 and rs1801214) were not included on the Illumina chip and were therefore genotyped using TaqMan allelic discrimination assays on the ABI7900 platform (assay IDs: C__11871167_10, C__29270128_10, C__1692218_10 and C__8841076_10; Applied Biosystems, Foster City, CA, USA). Furthermore, three CpG-SNPs, rs4712523, rs3821943 and rs12778642, which are in strong linkage disequilibrium (LD) with the three type 2 diabetes CpG-SNPs rs7754840, rs1801214 and rs5015480, respectively, were genotyped using the HumanOmniExpressBeadChip and iScan system (Illumina).

Genotype frequencies were in Hardy–Weinberg equilibrium. The genotype data analysed with either the Illumina chip or TaqMan assays showed 100% concordance with the pyrosequencing genotype data.

Associations between CpG-SNPs and DNA methylation The effects of the SNPs on DNA methylation of the CpG-SNP sites and surrounding CpG sites were analysed using the Kruskal–Wallis H test. The association between CpG-SNPs and DNA methylation was based on an additive genetic model and *p* values were adjusted for multiple testing using Bonferroni correction and significance threshold <0.05 after Bonferroni correction was considered significant. Statistical calculations were performed using IBM SPSS statistics 20 (SPSS, Chicago, IL, USA).

Analysis of mRNA expression in human islets Expression of mRNA was analysed in pancreatic islets from the human donors using the GeneChip Human Gene 1.0 ST array from Affymetrix (Santa Clara, CA, USA) as described previously by Olsson et al [14]. The array data were summarised and normalised using the robust multi-array analysis method and the software 'Expression Console' (Affymetrix), and were corrected for batch effects within the array data using COMBAT [15].

Associations between type 2 diabetes CpG-SNPs that exhibit differential DNA methylation and mRNA expression in human islets Associations between type 2 diabetes CpG-SNPs that exhibit differential DNA methylation and mRNA expression in the human islets were analysed using linear regression assuming an additive model. In the linear model, mRNA expression values were used as quantitative traits, SNP genotypes were encoded as 0, 1 or 2 according to their minor allele, and sex, age, BMI, islet purity and islet culture days were included as covariates. The analysis was restricted to genes located within a maximum distance of 500 kb from a CpG-SNP site. The analysis was performed by using the Matrix expression quantitative trait loci (eQTL) R package [16].

Associations between type 2 diabetes CpG-SNPs that exhibit differential DNA methylation and alternative mRNA splicing The Gene Array Analyzer (GAA) [17] was used together with mRNA expression data from the GeneChip Human Gene 1.0 ST array (Affymetrix) to test if type 2 diabetes CpG-SNPs that exhibit differential DNA methylation are associated with alternative splicing events in human islets. Each exon on the GeneChip Human Gene 1.0 ST array is covered by at least two probes, which is why the expression data from these arrays can be used to analyse alternative splicing events, as previously described [17]. The

gene-level-normalised intensity was calculated for each exon and splice index and the Student's *t* test was applied for identification of differentially expressed exons between different type 2 diabetes CpG-SNP groups.

Associations between type 2 diabetes CpG-SNPs that exhibit differential DNA methylation and measures of islet function Insulin and glucagon secretion as well as insulin content were used as measures of islet function. Glucose-stimulated insulin release was measured as stimulation index (SI) in static incubations of human islets as previously described [18]. Glucagon secretion from human islets was measured after stimulation with either 8.3 or 16.7 mmol/l glucose for 1 h. Twelve human islets in eight replicates per culture condition were pre-incubated in HEPES-balanced salt solution (HBSS). Thereafter, glucose was added to a final concentration and islets incubated for 1 h. The supernatant fraction was immediately removed and the glucagon concentration in the medium was measured by RIA (Millipore, Billerica, MA, USA). Insulin content in homogenised human islets was assessed by ELISA (Mercodia, Uppsala, Sweden) and then normalised to the total DNA in each sample as determined by fluorometric assay (Quant-iT Picogreed, Invitrogen Molecular Probes, Stockholm, Sweden). Associations between type 2 diabetes CpG-SNPs that exhibit differential DNA methylation and insulin secretion, glucagon secretion and insulin content in the human islets were analysed using linear regression, assuming an additive model adjusted for the covariates age, sex and BMI.

Identification of SNPs in LD with 19 type 2 diabetes CpG-SNPs SNP annotation and proxy (SNAP) search [19] was used to identify SNPs in LD with the 19 type 2 diabetes CpG-SNPs. The search was based on genotype data from the 1000 Genomes project with the Centre d'Etude du Polymorphisme Humain (Utah residents with northern and western European ancestry) (CEU) population panel. An r^2 threshold >0.8 from pairwise LD calculations and a distance limit of 500 kb from the query SNP were set as inclusion criteria for LD SNPs. To determine if the identified LD SNPs were CpG-SNPs, the DNA sequences were retrieved from the dbSNP database (www.ncbi.nlm.nih.gov/snp/). A SNP was considered to be a CpG-SNP if it introduces or removes a CpG site.

Results

Identification of CpG-SNPs associated with type 2 diabetes Of the 40 type 2 diabetes genome-wide-associated SNPs [2, 20–30], 19 (48%) of the risk alleles introduce or remove a CpG site and are considered CpG-SNPs (Table 2). Of these 19 CpG-SNPs associated with type 2 diabetes, ten SNPs are intergenic, five are located in introns and four are coding SNPs (Table 2).

CpG-SNPs associated with type 2 diabetes affect DNA methylation in human pancreatic islets The impact of these 19 type 2 diabetes-associated CpG-SNPs on DNA methylation in pancreatic islets from 84 human donors was examined (Fig. 1 and Table 2). Successful DNA methylation data were generated for 16 of the 19 type 2 diabetes-associated CpG-SNP loci representing the candidate genes *CDKAL1*, *CHCHD9*, *DUSP8*, *IRS1*, *SRR*, *TSPAN8*, *CDKN2A*, *HHEX*, *PPARG*, *TCF7L2*, *DUSP9*, *SLC30A8*, *HMGA2*, *KCNQ1*, *ADCY5* and *WFS1* in human islets (Fig. 1). DNA methylation data were also generated for a total of 22 CpG sites located near the *CDKN2A*, *HHEX*, *PPARG*, *TCF7L2*, *DUSP9*, *SLC30A8*, *HMGA2*, *KCNQ1*, *ADCY5* and *WFS1* type 2 diabetes-risk SNPs. As a result of either design or technical problems, DNA methylation data could not be generated for the remaining three type 2 diabetes CpG-SNP loci representing *KCNJ11*, *ZBED3* and *CDC123* (Fig. 1).

We then tested if the type 2 diabetes-associated CpG-SNPs are associated with differential DNA methylation of the CpG-SNP sites as well as the surrounding CpG sites. All 16 type 2 diabetes CpG-SNPs that generated successful data were associated with differential DNA methylation of the CpG-SNP site in the human islets ($p \leq 2.3 \times 10^{-5}$, Fig. 2a–p and ESM Table 2). In addition, six CpG-SNP loci representing *CDKN2A*, *TCF7L2*, *HMGA2*, *KCNQ1*, *ADCY5* and *WFS1* were associated with differential DNA methylation of surrounding CpG sites ($p < 0.05$, Fig. 2a–p and ESM Table 2).

Associations between type 2 diabetes CpG-SNPs that exhibit differential DNA methylation and mRNA expression in human islets Increased DNA methylation of certain genome regions may silence gene transcription [7, 11]. We therefore tested if carriers of CpG-SNP genotypes associated with increased DNA methylation have decreased mRNA expression of genes located in regions 500 kb up- and downstream of the CpG-SNPs using microarray expression data from the human islets. We found that the CpG-SNP genotype carriers of rs391300 and rs5945326 that have increased DNA methylation also have decreased mRNA expression of *RPA1* and *AVPR2*, respectively, in human islets (Table 3).

Associations between type 2 diabetes CpG-SNPs that exhibit differential DNA methylation and alternative splicing events in human islets There are further data supporting a role for DNA methylation in alternative splicing events [31]. We hence tested if the type 2 diabetes CpG-SNPs that exhibit differential DNA methylation and are located in gene introns or exons are associated with differentially expressed exons assessed by splice index using Gene Array Analyzer (http://GAA.mpi-bn.mpg.de) [17]. We found that type 2
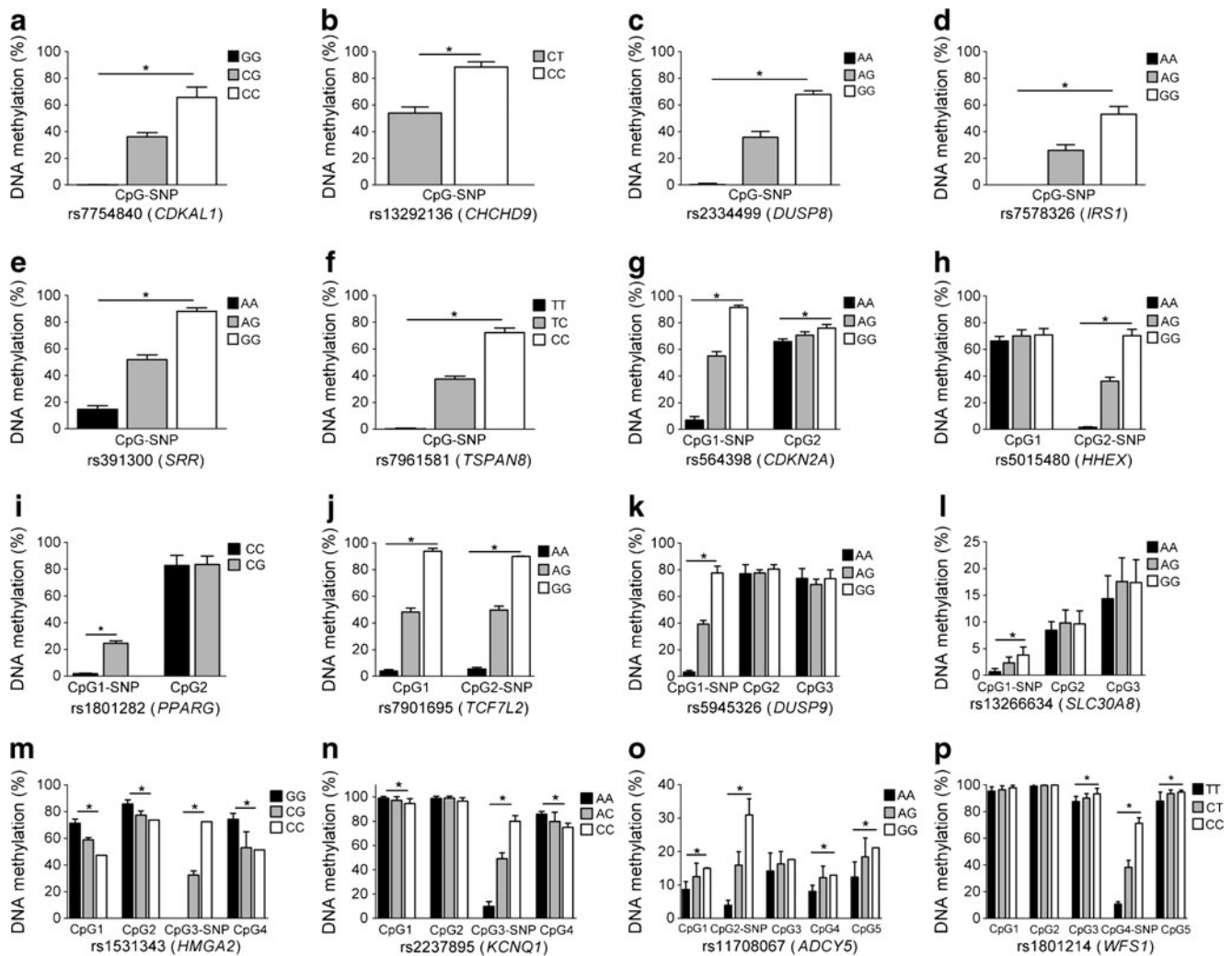
Fig. 2 Histograms showing the mean degree of DNA methylation for each genotype of 16 type 2 diabetes-associated CpG-SNPs and surrounding CpG sites in human pancreatic islets of 84 donors (**a–p**). [*]*p*<0.05 for associations between the CpG-SNP and DNA methylation after Bonferroni correction (*n*=16). Data are presented as mean ± SD

Table 3 Associations between type 2 diabetes CpG-SNPs that exhibit differential DNA methylation and mRNA expression in human pancreatic islets

Chr.	SNP ID	Gene	SNP location	Genotype	*n*	Gene expression		
						Mean ± SD	β (SEM)	*p* value
17	rs391300	*RPA1*	~413.4 kb down	A A	12	574.3±54.3	−0.056 (0.027)	0.040
				A G	47	562.1±57.4		
				G G	22	537.1±60.5		
X	rs5945326	*AVPR2*	~268.1 kb up	A A	50	66.8±9.0	−0.052 (0.024)	0.034
				A G	9	64.1±5.1		
				G G	22	62.3±8.7		

Data are expressed as mean ± SD of unlogged gene expression values

β (SEM) and *p* values are based on a linear regression of $\log_2$ values of gene expression adjusted for age, sex, BMI, islet purity and islet culture days

Chr., chromosome

Table 4 Associations between CpG-SNPs that exhibit differential DNA methylation and alternative splicing events, assessed by splice index in human pancreatic islets

SNP ID	Gene	Affymetrix transcript cluster ID	Affymetrix probe set ID	Exon annotation for probe set	AA vs Aa		AA vs aa		Aa vs aa	
					Splice index	p value	Splice index	p value	Splice index	p value
rs13266634	*SLC30A8*	8148003	8148008	Exon 3	0.066	0.39	0.34	0.031	0.28	0.085
		8148003	8148011	Exon 7	−0.005	0.88	−0.14	0.030	−0.13	0.039
		8148003	8148015	Exon 11	−0.056	0.21	−0.26	0.0048	−0.21	0.028
rs1801214	*WFS1*	8093906	8093909	Exon 2	−0.029	0.70	0.20	0.040	0.23	0.011
rs7754840	*CDKAL1*	8117140	8117152	Exon 10	−0.039	0.40	−0.15	0.035	−0.12	0.10
rs7901695	*TCF7L2*	7930537	7930538	Exon 1	−0.38	0.045	−0.34	0.56	0.033	0.96

AA, homozygotes for major allele; Aa, heterozygotes; aa, homozygotes for minor allele; p value, based on t test of gene-level-normalised intensity of exon

diabetes CpG-SNPs located in *SLC30A8*, *WFS1*, *CDKAL1* and *TCF7L2* were associated with alternative splicing events in the respective gene in the human islets (Table 4).

Associations between type 2 diabetes CpG-SNPs that exhibit differential DNA methylation and measures of islet function We further examined if the type 2 diabetes CpG-SNPs that exhibit differential DNA methylation were associated with measures of islet function. As type 2 diabetes is characterised by decreased glucose-stimulated insulin secretion and elevated glucagon levels [32], we tested if the risk alleles of the type 2 diabetes CpG-SNPs were associated with lower insulin content and/or secretion as well as elevated glucagon secretion from the human islets. The CpG-SNPs rs11708067 (*ADCY5*) and rs5015480 (*HHEX*) were associated with insulin secretion, while rs11708067 (*ADCY5*) and rs2237895 (*KCNQ1*) were associated with glucagon secretion (Table 5). Moreover, the type 2 diabetes CpG-SNP rs564398 (*CDKN2A*) was associated with insulin content in the human islets (Table 5).

Identification of CpG-SNPs in LD with the 19 type 2 diabetes-associated CpG-SNPs We next performed a systematic analysis to examine if the 19 type 2 diabetes CpG-SNPs are in LD with any nearby SNPs and if any of these LD SNPs are also CpG-SNPs using SNAP search [19] together with genotype data from the 1000 Genomes project. We identified a total of 295 SNPs, of which 91 are CpG-

Table 5 Associations between type 2 diabetes CpG-SNPs that exhibit differential DNA methylation and measures of islet function including glucose-stimulated insulin secretion (SI), glucagon secretion and insulin content in human pancreatic islets

Phenotype	SNP ID	Candidate gene	Genotype	n	Mean ± SD	β (SEM)	p value
Insulin secretion (SI)	rs11708067	*ADCY5*	AA	48	7.1±7.8	0.18 (0.08)	0.031
			AG	28	13.2±20.1		
			GG	2	16.2±17.7		
	rs5015480	*HHEX*	GG	21	5.4±5.9	0.14 (0.08)	0.036
			AG	44	9.5±10.1		
			AA	13	16.1±27.1		
Glucagon (pg islet^{-1}h^{-1}, 8.3 mmol/l glucose)[a]	rs11708067	*ADCY5*	AA	3	7.0±2.6	−0.64 (0.15)	0.027
			AG	5	3.2±2.7		
Glucagon (pg islet^{-1}h^{-1}, 16.7 mmol/l glucose)	rs2237895	*KCNQ1*	AA	17	8.6±10.6	0.26 (0.12)	0.042
			AC	19	23.6±26.1		
			CC	8	35.2±42.2		
Insulin content (ng/ng DNA)	rs564398	*CDKN2A*	AA	24	4.5±2.5	0.10 (0.04)	0.023
			AG	36	6.9±3.9		
			GG	11	6.5±4.1		

Data are expressed as mean ± SD. β (SEM) and p values are from linear regression adjusted for age, sex and BMI as covariates

[a] Data available from a subset of the cohort

Table 6 The number of SNPs and CpG-SNPs in LD with the 19 identified type 2 diabetes CpG-SNPs

Candidate SNP	Candidate gene	LD SNPs (n)	LD CpG-SNPs (n)
rs11708067	ADCY5	14	2
rs12779790	CDC123	2	1
rs13266634	SLC30A8	5	3
rs13292136	CHCHD9	22	3
rs1531343	HMGA2	26	5
rs1801214	WFS1	94	41
rs1801282	PPARG	21	7
rs2237895	KCNQ1	0	0
rs2334499	DUSP8	4	0
rs391300	SRR	2	0
rs4457053	ZBED3	4	0
rs5015480	HHEX	11	7
rs5219	KCNJ11	15	5
rs564398	CDKN2A	22	6
rs5945326	DUSP9	1	0
rs7578326	IRS1	9	1
rs7754840	CDKAL1	30	7
rs7901695	TCF7L2	7	2
rs7961581	TSPAN8	6	1
		$\sum$295	$\sum$91

Identification of LD SNPs based on 1000 Genome project in the CEU population with a distance limit of ±500 kb from candidate SNP. CpG-SNPs identified based on sequence data from dbSNP

LD, $r^2 > 0.8$

SNPs in strong LD ($r^2 > 0.8$) with the 19 type 2 diabetes CpG-SNPs (Table 6). The 295 LD SNPs can be found in ESM Table 3. We further tested if CpG-SNPs that are in LD with type 2 diabetes CpG-SNPs also give rise to differential DNA methylation in the CpG-SNP site. Three LD CpG-SNPs were selected for this analysis based on their location near a type 2 diabetes CpG-SNP with a large number of LD CpG-SNPs. DNA methylation was subsequently analysed in the three CpG-SNPs rs4712523, rs3821943 and rs12778642,

which are in strong LD with the type 2 diabetes CpG-SNPs rs7754840 (**CDKAL1**), rs1801214 (WFS1) and rs5015480 (HHEX), respectively. The three analysed LD CpG-SNPs were associated with differential DNA methylation in the CpG-SNP site in human islets (Fig. 3a–c, ESM Table 4) and the differences in methylation were similar to the data seen in respective type 2 diabetes CpG-SNPs (Fig. 2).

Discussion

Approximately 40 type 2 diabetes-associated loci have been identified by GWASs [2]. Furthermore, a number of studies that have aimed to unravel the physiological mechanisms through which these loci predispose to diabetes have shown that most of the identified loci act through impaired insulin secretion while some affect insulin sensitivity [3, 33]. However, the molecular mechanisms through which these SNPs influence gene function and disease pathology are largely unknown. In this study, we found that 19 of 40 SNPs associated with type 2 diabetes introduce or remove CpG sites, which are potential sites for DNA methylation. Many of these CpG-SNPs have previously been associated with impaired insulin secretion, hence the focus on these CpG-SNPs in human pancreatic islets [2]. Interestingly, we show that all analysed type 2 diabetes CpG-SNPs are associated with differential DNA methylation of the CpG-SNP sites in human pancreatic islets. We also found that six of the CpG-SNPs are associated with differential DNA methylation of surrounding CpG sites. Moreover, some of the type 2 diabetes CpG-SNPs that exhibit differential DNA methylation are associated with gene expression, alternative splicing events and hormone secretion in the human islets. The 19 type 2 diabetes CpG-SNPs are in strong LD with 91 additional CpG-SNPs that also seem to exhibit differential DNA methylation at the SNP site. Our data suggest that interactions between genetic and epigenetic factors may affect type 2 diabetes as several risk SNPs that influence islet function, such as TCF7L2, HHEX, CDKN2A, SLC30A8, CDKAL1,
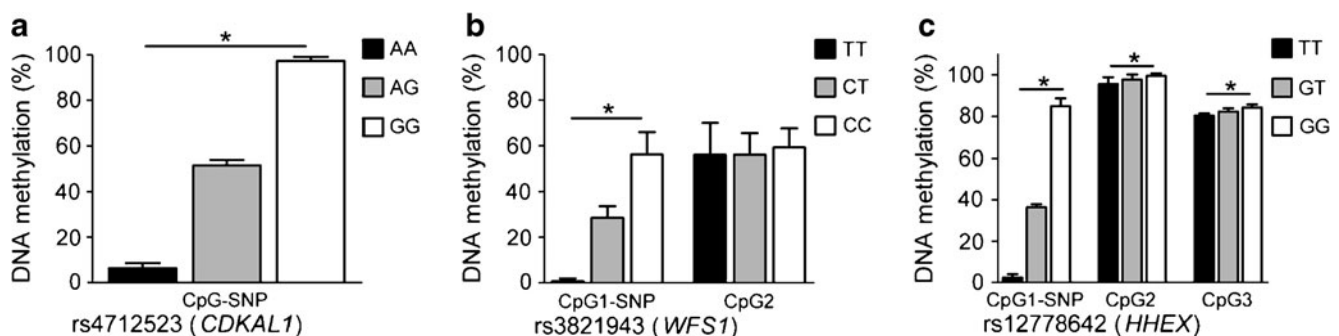

Fig. 3 Histograms showing the mean degree of DNA methylation in human islets for each genotype of CpG-SNPs in LD with three type 2 diabetes-associated CpG-SNPs: (**a**) rs7754840 (CDKAL1); (**b**) rs1801214 (WFS1); and (**c**) rs5015480 (HHEX). $^*p < 0.05$ for associations between the CpG-SNP and DNA methylation. Data are presented as mean ± SD

ADCY5 and *WFS1* [2, 3], are CpG-SNPs that exhibit differential DNA methylation in human islets.

An important characteristic of CpG sites is their mutability, which makes them hotspots for human genetic disease [34]. Indeed, more than 30% of all point mutations in humans are a consequence of DNA methylation [35]. While hydrolytic deamination of cytosines produces uracil, which is a foreign base in the DNA sequence that is identified and fixed by the DNA repair system, methylated cytosines are hydrolytically deaminated to thymines, which are not identified as foreign and hence less likely to be repaired, resulting in a C→T transition [34], which is the most common point mutation in humans [34, 36]. Indeed, of the approximately 2.3 million C/T or G/A autosomal SNPs identified by the second-phase HapMap database, 34% are located within a CpG site [37] and some have also been associated with disease [12, 13, 36, 38]. However, to our knowledge no previous study has examined if the 40 type 2 diabetes-associated SNPs are CpG-SNPs.

CpG-SNPs may provide a possible molecular mechanism through which SNPs can influence a phenotype. One suggested scenario is that CpG-SNPs can affect local DNA methylation which, in turn, affects the expression of a gene by allowing or preventing the binding of CpG methyl-binding proteins [13]. Another scenario is that CpG-SNPs may affect transcriptional splicing via differential DNA methylation. Indeed, DNA methylation has been shown to play a role in the regulation of splicing and it has been suggested that exon DNA methylation may aid the spliceosome in distinguishing exons from introns [31]. Interestingly, recent studies of *TCF7L2* variants have suggested that SNPs located in introns can affect gene function through splicing [39]. Intragenic DNA methylation has also been suggested to play a direct role in reducing or enhancing transcription elongation efficiency via alternative promoters located within gene bodies [40] or indirectly by regulating non-coding transcripts that then affect the transcription of the associated genes [41, 42]. Another possible mechanism is by affecting the recombination rate [37]. Our data suggest that differential DNA methylation in CpG-SNP sites of human islets may influence gene expression of nearby genes as well as alternative splicing when the CpG-SNP is located within a gene. This may subsequently affect islet function. Two of the analysed type 2 diabetes CpG-SNPs that show differential DNA methylation (rs391300 and rs5945326) are associated with differential gene expression and 50% of the analysed type 2 diabetes CpG-SNPs located in introns and/or exons affect alternative splicing events in the human islets. Moreover, several type 2 diabetes CpG-SNPs are associated with measures of islet function. It should be taken into account that our study included 84 donors of human islets and although this is one of the largest human islet cohorts existing, for some of the type 2 diabetes CpG-SNPs there are only a few or no donors homozygous for the rare

variant, which may explain the lack of power to identify associations between more CpG-SNPs and gene expression or hormone measures. Other reasons for the lack of associations between some of the type 2 diabetes CpG-SNPs and the analysed phenotypes could be that some of the type 2 diabetes CpG-SNPs do not affect gene expression or hormone secretion but possibly cause downstream modifications of mRNA or protein translation, or other phenotypes such as genome stability and recombination. Future studies are needed to fully elucidate the role of differential DNA methylation in CpG-SNPs.

SNPs within *TCF7L2* confer the highest genetic risk of developing type 2 diabetes [21, 24, 43–45]. While rs7903146 within intron 4 of *TCF7L2* is the most studied SNP, there are multiple SNPs in strong LD with rs7903146, including rs7901695, that also show strong associations with type 2 diabetes [46, 47]. The aim of our study was to examine if known type 2 diabetes SNPs introduce or remove CpG sites. The SNP rs7903146 is not a CpG-SNP. However, other *TCF7L2* SNPs associated with type 2 diabetes, including rs7901695, delete CpG sites and we hence included rs7901695 data in our study. While rs7901695 and rs7903146 are in full LD ($r^2 = 1$) in the islet donors included in our study, these two SNPs are in strong, but not full, LD in the CEU population based on genotype data from the 1000 genome project ($r^2 = 0.892$).

The causal variants remain to be identified for a number of the identified type 2 diabetes SNPs [2]. It should further be noted that many of the type 2 diabetes SNPs have been connected to a gene name based on location and/or biological relevance, rather than showing association with expression of that particular gene. Several of the type 2 diabetes SNPs may hence be associated with expression of genes other than the one to which they have been assigned. Indeed, we found that rs391300 and rs5945326 are associated with mRNA expression of *RPA1* and *AVPR2*, respectively, in the human islets. *RPA1* encodes replication protein A1, which is involved in DNA replication, and *AVPR2* encodes the vasopressin receptor, type 2, which belongs to the seven-transmembrane-domain G-protein-coupled receptor (GPCR) superfamily that stimulates adenylate cyclase. However, whether these two genes affect islet function remains to be tested.

We found that a CpG-SNP may not only be associated with differential DNA methylation of the CpG-SNP site itself, but may also be associated with DNA methylation of surrounding CpG sites. These results are in agreement with another study that has reported CpG-SNPs to affect DNA methylation of other, nearby, CpG sites and not only within the CpG-SNP site [48]. The observed correlation in DNA methylation between the CpG-SNP site and the adjacent CpG sites may be due to the presence of two active sites on DNMT3a, which allow for the simultaneous methylation of two CpG sites separated by one DNA helical turn [49].

We also found 91 CpG-SNPs out of a total of 295 SNPs in strong LD with the 19 type 2 diabetes-associated CpG-SNPs. The LD CpG-SNPs seem to affect the DNA methylation pattern in a similar fashion to the type 2 diabetes CpG-SNP. Genome regions with a large number of CpG-SNPs in LD are therefore likely to affect the chromatin structure and subsequently gene function.

Together, this study demonstrates how genetic (SNPs) and epigenetic (DNA methylation) mechanisms can interact and our data provide a possible molecular mechanism through which type 2 diabetes-associated SNPs that introduce or remove CpG sites may affect gene function through differential DNA methylation and consequently the phenotype of the disease.

Duality of interest The authors declare that there is no duality of interest associated with this manuscript.

Contribution statement TAD and CL designed the study and experiments. AHO, PV, PA and TR contributed to the design of the study and the experiments. TAD, AHO and TR performed the research. All authors analysed, interpreted and discussed the data. TAD and CL wrote the manuscript. All authors reviewed and revised the manuscript critically. All authors approved the final version of the manuscript. TAD and CL are the guarantors of this work and, as such, had full access to all of the study data and take responsibility for the integrity of the data.

References

1. Ling C, Groop L (2009) Epigenetics: a molecular link between environmental factors and type 2 diabetes. Diabetes 58:2718–2725
2. McCarthy M (2010) Genomics, type 2 diabetes, and obesity. N Engl J Med 363:2339–2350
3. Ruchat SM, Elks CE, Loos RJ et al (2009) Association between insulin secretion, insulin sensitivity and type 2 diabetes susceptibility variants identified in genome-wide association studies. Acta Diabetol 46:217–226
4. Ling C, del Guerra S, Lupi R et al (2008) Epigenetic regulation of PPARGC1A in human type 2 diabetic islets and effect on insulin secretion. Diabetologia 51:615–622
5. Ronn T, Poulsen P, Hansson O et al (2008) Age influences DNA methylation and gene expression of COX7A1 in human skeletal muscle. Diabetologia 51:1159–1168
6. Yang BT, Dayeh TA, Kirkpatrick CL et al (2011) Insulin promoter DNA methylation correlates negatively with insulin gene expression and positively with HbA(1c) levels in human pancreatic islets. Diabetologia 54:360–367
7. Yang BT, Dayeh TA, Volkov PA et al (2012) Increased DNA methylation and decreased expression of PDX-1 in pancreatic islets from patients with type 2 diabetes. Mol Endocrinol 26:1203–1212
8. Volkmar M, Dedeurwaerder S, Cunha DA et al (2012) DNA methylation profiling identifies epigenetic dysregulation in pancreatic islets from type 2 diabetic patients. EMBO J 31:1405–1426
9. Nitert M, Dayeh T, Volkov P et al (2012) Impact of exercise training on DNA methylation in skeletal muscle of first degree relatives of type 2 diabetes. Diabetes 61:3322–3332
10. Lister R, Pelizzola M, Dowen RH et al (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462:315–322
11. Bird A (2007) Perceptions of epigenetics. Nature 447:396–398
12. Ling C, Poulsen P, Simonsson S et al (2007) Genetic and epigenetic factors are associated with expression of respiratory chain component NDUFB6 in human skeletal muscle. J Clin Invest 117:3427–3435
13. Taqi MM, Bazov I, Watanabe H et al (2011) Prodynorphin CpG-SNPs associated with alcohol dependence: elevated methylation in the brain of human alcoholics. Addict Biol 16:499–509
14. Olsson AH, Yang BT, Hall E et al (2011) Decreased expression of genes involved in oxidative phosphorylation in human pancreatic islets from patients with type 2 diabetes. Eur J Endocrinol 165:589–595
15. Johnson WE, Li C, Rabinovic A (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 8:118–127
16. Shabalin AA (2012) Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics 28:1353–1358
17. Gellert P, Teranishi M, Jenniches K et al (2012) Gene Array Analyzer: alternative usage of gene arrays to study alternative splicing events. Nucleic Acids Res 40:2414–2425
18. Stahle MU, Brandhorst D, Korsgren O, Knutson F (2011) Pathogen inactivation of human serum facilitates its clinical use for islet cell culture and subsequent transplantation. Cell Transplant 20:775–781
19. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI (2008) SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics 24:2938–2939
20. Voight BF, Scott LJ, Steinthorsdottir V et al (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet 42:579–589
21. Zeggini E, Weedon MN, Lindgren CM et al (2007) Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science 316:1336–1341
22. Dupuis J, Langenberg C, Prokopenko I et al (2010) New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet 42:105–116
23. Kong A, Steinthorsdottir V, Masson G et al (2009) Parental origin of sequence variants associated with complex diseases. Nature 462:868–874
24. Scott LJ, Mohlke KL, Bonnycastle LL et al (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 316:1341–1345
25. Tsai FJ, Yang CF, Chen CC et al (2010) A genome-wide association study identifies susceptibility variants for type 2 diabetes in Han Chinese. PLoS Genet 6:e1000847

26. Sladek R, Rocheleau G, Rung J et al (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 445:881–885

27. Zeggini E, Scott LJ, Saxena R et al (2008) Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet 40:638–645

28. Yasuda K, Miyake K, Horikawa Y et al (2008) Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. Nat Genet 40:1092–1097

29. Qi L, Cornelis MC, Kraft P et al (2010) Genetic variants at 2q24 are associated with susceptibility to type 2 diabetes. Hum Mol Genet 19:2706–2715

30. Frayling TM, Timpson NJ, Weedon MN et al (2007) A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science 316:889–894

31. Shukla S, Kavak E, Gregory M et al (2011) CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature 479:74–79

32. Solomon TP, Knudsen SH, Karstoft K, Winding K, Holst JJ, Pedersen BK (2012) Examining the effects of hyperglycemia on pancreatic endocrine function in humans: evidence for in vivo glucotoxicity. J Clin Endocrinol Metab 97:4682–4691

33. Rosengren AH, Braun M, Mahdi T et al (2012) Reduced insulin exocytosis in human pancreatic beta-cells with gene variants linked to type 2 diabetes. Diabetes 61:1726–1733

34. Bird A (2011) The dinucleotide CG as a genomic signalling module. J Mol Biol 409:47–53

35. Cooper DN, Krawczak M (1990) The mutational spectrum of single base-pair substitutions causing human genetic disease: patterns and predictions. Hum Genet 85:55–74

36. Harlid S, Ivarsson MI, Butt S et al (2011) A candidate CpG SNP approach identifies a breast cancer associated ESR1-SNP. Int J Cancer 129:1689–1698

37. Sigurdsson MI, Smith AV, Bjornsson HT, Jonsson JJ (2009) HapMap methylation-associated SNPs, markers of germline DNA methylation, positively correlate with regional levels of human meiotic recombination. Genome Res 19:581–589

38. Mill J, Petronis A (2007) Molecular studies of major depressive disorder: the epigenetic perspective. Mol Psychiatry 12:799–814

39. Osmark P, Hansson O, Jonsson A, Ronn T, Groop L, Renstrom E (2009) Unique splicing pattern of the TCF7L2 gene in human pancreatic islets. Diabetologia 52:850–854

40. Maunakea AK, Nagarajan RP, Bilenky M et al (2010) Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature 466:253–257

41. Deaton AM, Webb S, Kerr AR et al (2011) Cell type-specific DNA methylation at intragenic CpG islands in the immune system. Genome Res 21:1074–1086

42. Deaton AM, Bird A (2011) CpG islands and the regulation of transcription. Genes Dev 25:1010–1022

43. Grant SF, Thorleifsson G, Reynisdottir I et al (2006) Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat Genet 38:320–323

44. Saxena R, Voight BF, Lyssenko V et al (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 316:1331–1336

45. Wen J, Ronn T, Olsson A et al (2010) Investigation of type 2 diabetes risk alleles support CDKN2A/B, CDKAL1, and TCF7L2 as susceptibility genes in a Han Chinese cohort. PLoS One 5:e9153

46. Nemr R, Turki A, Echtay A et al (2012) Transcription factor-7-like 2 gene variants are strongly associated with type 2 diabetes in Lebanese subjects. Diabetes Res Clin Pract 98:e23–e27

47. Tong Y, Lin Y, Zhang Y, Yang J, Liu H, Zhang B (2009) Association between TCF7L2 gene polymorphisms and susceptibility to type 2 diabetes mellitus: a large Human Genome Epidemiology (HuGE) review and meta-analysis. BMC Med Genet 10:15

48. Gertz J, Varley KE, Reddy TE et al (2011) Analysis of DNA methylation in a three-generation family reveals widespread genetic influence on epigenetic regulation. PLoS Genet 7:e1002228

49. Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X (2007) Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature 449:248–251