



(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/EP2011/070796

(22) International Filing Date:
23 November 2011 (23.11.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
11151585.4 20 January 2011 (20.01.2011) EP

(71) Applicant (for all designated States except US): **UNIVERSITÉ LIBRE DE BRUXELLES** [BE/BE]; Avenue F.D. Roosevelt 50 CP 161, B-1050 Bruxelles (BE).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **CNOP, Miriam** [BE/BE]; Boudewijnlaan 21, B-1700 Dilbeek (BE). **VOLKMAR, Michael** [DE/DE]; Bahnhofstrasse 52, 99955 Bad Tennstedt (DE). **FUKS, François** [BE/BE]; rue du Docteur 36, B-1160 Bruxelles (BE). **EIZIRIK, Decio** [SE/BE]; Boudewijnlaan 21, B-1700 Dilbeek (BE).

(74) Agents: **VANHALST, Koen Victor Rachel** et al.; E. Gevaertdreef 10a, B-9830 Sint-Martens-Latem (BE).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ,

CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— of inventorship (Rule 4.17(iv))

Published:

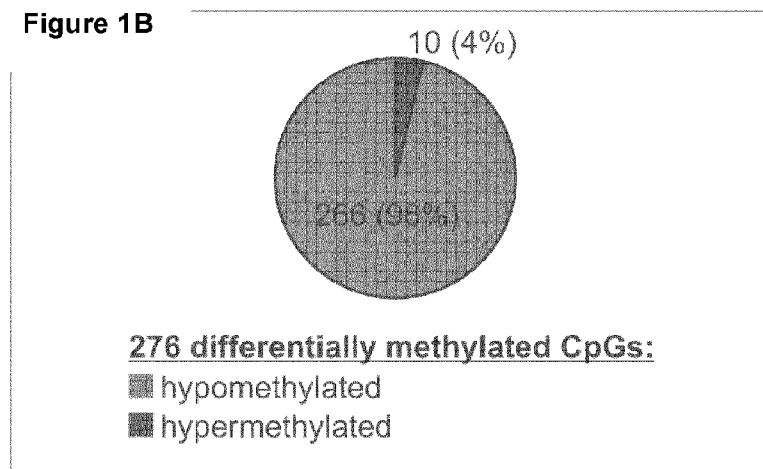
— with international search report (Art. 21(3))

— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

— with sequence listing part of description (Rule 5.2(a))

(54) Title: METHYLATION PATTERNS OF TYPE 2 DIABETES PATIENTS

Figure 1B



(57) Abstract: The present invention provides new target gene regions for use in prediction, prognosis, diagnosis and therapy of T2D, and/or follow up of intervention trials, based on the differential methylation profile of said targets in samples from subjects with type 2 diabetes and healthy subjects.



METHYLATION PATTERNS OF TYPE 2 DIABETES PATIENTS**FIELD OF THE INVENTION**

5 The present invention is situated in the medical diagnostics, therapeutics field, more particular in the field of early prediction of type-2 diabetes, and methods for treating said type-2 diabetes, based on the new diagnostic tools and targets identified herein.

BACKGROUND OF THE INVENTION

10

Type-2 diabetes (T2D) has developed into a major public health problem. While previously considered a problem primarily for Western populations, the disease is rapidly gaining global importance, as today around 285 million people are affected worldwide. Lifestyle and behavioural factors play an important role in determining T2D risk.

15 The pancreatic islets of Langerhans are of central importance in the development of T2D. Under normal conditions, increasing blood glucose levels after a meal trigger insulin secretion from the pancreatic islet beta-cells to regulate glucose homeostasis. Beta-cell failure marks the irreversible deterioration of glucose tolerance and results in T2D (Cnop M et al., 2007 Diabetes Care 30: 677-682; Tabak AG et al., 2009 Lancet 373: 2215-2221; U.K. prospective diabetes study 16. (1995)
20 Diabetes 44: 1249-1258). The unbiased genome-wide search for T2D risk genes has placed the insulin-producing beta-cells residing in the islets at centre stage. These approaches have also inadvertently highlighted the complexity inherent in biological mechanisms critical to T2D development. Most T2D risk genes identified in these genome-wide association studies (GWAS) affect beta-cell mass and/or function (Florez JC et al., 2008 Diabetologia 51: 1100-1110).

25 The majority of studies in the field have characterised diabetes aetiology on the basis of genetics, while little or no progress was made regarding the potential involvement of epigenetic mechanisms in T2D as a crucial interface between the effects of genetic predisposition and environmental influences. Epigenetic changes are heritable yet reversible modifications that occur without alterations in the primary DNA sequence. DNA methylation and histone modifications are
30 the main molecular events that initiate and sustain epigenetic modifications. These modifications may therefore provide a link between the environment, *i.e.* nutrition and lifestyle, and T2D but only few studies so far have documented aberrant DNA methylation events in T2D (Park JH et al., 2008 J Clin Invest 118: 2316-2324; Ling C et al., 2008 Diabetologia 51: 615-622). Park et al. reports on the hypermethylation of the *pdx1* gene promotor, while Ling et al. reports the
35 hypermethylation of the *PPAEGC1A* gene promoter. In both cases, the hypermethylation of the promoter results in gene silencing in T2D patients.

DNA methylation occurs as 5-methyl cytosine mostly in the context of CpG dinucleotides, so-

called CpG sites. It is the best-studied epigenetic modification and governs transcriptional regulation and silencing (for review see Suzuki MM and Bird A 2008 Nat Rev Genet 9: 465-476). Unlike the relatively sturdy genome, the methylome changes in a dynamic way during development, tissue differentiation and aging. Pathologically altered DNA methylation is well described in various cancers (reviewed in Jones PA and Baylin SB 2007 Cell 128: 683-692). About 75% of human gene promoters are associated with CpG islands, which are clusters of 500bp to 2kb length with a comparatively high frequency of CpG dinucleotides. They usually harbour low levels of DNA methylation but can become hypermethylated; this CpG island hypermethylation was demonstrated to abrogate tumour suppressor gene transcription during tumourigenesis. Lately, DNA methylation changes in CpG sites adjoining yet outside of CpG islands, so-called CpG island shores (Irizarry RA et al., 2009 Nat Genet 41: 178-186), are gaining increased attention. Intriguingly, CpG sites in these shore sequences, in addition to those within CpG islands, are proposed to display differential DNA methylation between cancer and normal cells as well as between cells of different tissues.

The goal of the present invention is to clarify the hitherto poorly understood connection between the DNA methylation status in T2D patients, i.e. both hyper- and/or hypomethylation with respect to a healthy subject, and its relation to T2D pathogenesis. The invention aims at providing new prognostic and diagnostic tools for identifying T2D at a very early stage and provides new targets for treatment of T2D. The invention will also provide the tools to evaluate the impact of new intervention therapies aiming to prevent T2D via epigenetic modulation.

SUMMARY OF THE INVENTION

In the context of methods for predicting, diagnosing and/or treating Type 2 Diabetes (T2D), the present invention provides methods for identifying nucleic acid target gene regions that serve as prognostic, diagnostic and therapeutic targets. It also provides the tools to evaluate the impact of new intervention therapies aiming to prevent T2D via epigenetic modulation. A nucleic acid target gene region can also refer to an amplified product of a nucleic acid target gene region, including an amplified product of a treated nucleic acid target gene region, where the nucleotide sequence of such an amplified product reflects the methylation state of the nucleic acid target gene region. One skilled in the art would recognize that the size or length of the nucleic acid target gene region may vary depending on the limitation, or limitations, of the equipment used to perform the analysis. The nucleic acid target gene region may comprise intragenic nucleic acid, a gene of interest, more than one gene of interest, at least one gene of interest or a portion of a gene of interest. Correspondingly a sequential or non-sequential series of nucleic acid target gene regions may be analyzed and exploited to map an entire gene or genome. The nucleic acid targets of the present invention are specified below in Table 4 and the sequence of the identified CpG regions are defined by SEQ ID Nos 1-276.

The present invention thus provides a method for the prognosis, diagnosis or prediction of Type 2 Diabetes (T2D) and/or for the follow up of intervention therapies comprising the steps of:

- a) measuring or analyzing the methylation status of one or more of the CpG site(s) defined in Table 4 (defined by SEQ ID Nos 1-276) in a sample of the subject, and
- 5 b) comparing the methylation status of said one or more CpG site(s) obtained from step a) with the methylation status of said CpG site(s) in a control sample, wherein a difference in methylation status as detected in step b) indicates the subject has or is at risk of developing T2D, optionally comprising the step of:
- c) comparing the methylation status of said one or more CpG site(s) obtained from step a) with
- 10 the methylation status of said CpG site(s) in a sample obtained after an intervention therapy aimed to prevent or treat T2D.

Preferably, said difference in methylation status is due to hypermethylation or hypomethylation.

- 15 In a preferred embodiment, said sample is pancreatic islet tissue, a blood sample, adipose tissue, muscle, or any other biological sample that serve as surrogate material for the pancreatic islet tissue.

In a particular embodiment, the methylation status of up to, or more than: 5, 10, 15, 20, 25, 30, 20 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, 200, 205, 210, 215, 220, 225, 230, 235, 240, 245, 250, 255, 260, 265, 270, or 275 of the CpG sites as defined in Table 4 or by SEQ ID Nos 1 to 276, or of all 276 CpG sites as defined in Table 4 is analyzed.

In a more preferred embodiment, the methylation status of the following CpG sites: 1, 3, 5, 6, 8, 25 15, 17, 23, 26, 28, 50, 65, 71, 92, 114, 129, 142, 178, 210, 223, 235, 263, 270, and 275, all taken from Table 4 (preferably defined by SEQ ID Nos 1, 3, 5, 6, 8, 15, 17, 23, 26, 28, 50, 65, 71, 92, 114, 129, 142, 178, 210, 223, 235, 263, 270, and 275 respectively) is analyzed.

The invention further provides for a method of treating T2D by targeting one or more genes 30 having aberrant methylation in T2D in one or more CpG sites defined by SEQ ID Nos 1-276 taken from Table 4.

Preferably, said targeting implies changing the methylation status by using demethylating or methylating agents, by changing the expression level, or by changing the protein activity of the protein encoded by said one or more genes.

- 35 In a preferred embodiment, said methylating agents are methyl donors such as folic acid, methionine, choline or any other chemicals capable of elevating DNA methylation.

In a more preferred embodiment, one or more of the following CpG sites are targeted: 1, 3, 5, 6, 8, 15, 17, 23, 26, 28, 50, 65, 71, 92, 114, 129, 142, 178, 210, 223, 235, 263, 270, and 275, taken from Table 4.

Said CpG sites are preferably defined by SEQ ID NOs 1, 3, 5, 6, 8, 15, 17, 23, 26, 28, 50, 65, 71, 92, 114, 129, 142, 178, 210, 223, 235, 263, 270, and 275 respectively.

In preferred embodiments of the method according to the invention, the methylation status is analysed by one or more techniques selected from the group consisting of nucleic acid amplification, polymerase chain reaction (PCR), methylation specific PCR (MCP), methylated-CpG island recovery assay (MIRA), combined bisulfite-restriction analysis (COBRA), bisulfite pyrosequencing, single-strand conformation polymorphism (SSCP) analysis, restriction analysis, microarray analysis, or bead-chip technology.

In preferred embodiments of the methods according to the invention, the patient is in a high risk group for developing diabetes or suffering from any beta-cell related disorder such as: type 1 diabetes mellitus, type 2 diabetes mellitus, hyperinsulinemia, obesity, neuroendocrine tumors or occurrence of insulinoma.

The invention further provides a method for identifying an agent that modulates one or more of the genes having aberrant methylation in T2D in any one or more of the CpG site(s) defined by SEQ ID Nos 1-276, taken from Table 4, comprising the steps of:

a) contacting the candidate agent with said one or more genes, and

c) analysing the modulation of said one or more gene by the candidate agent

Preferably, said agent modulates the methylation status, the expression level or the activity of said one or more gene. More preferably, one or more of the CpG site(s) defined by SEQ ID Nos 1, 3, 5, 6, 8, 15, 17, 23, 26, 28, 50, 65, 71, 92, 114, 129, 142, 178, 210, 223, 235, 263, 270, and/or 275 taken from Table 4, are targeted.

The invention further provides for a method for establishing a reference methylation status profile comprising the steps of: measuring the methylation status of one or more CpG site(s) as defined by SEQ ID Nos 1-276, taken from Table 4, having aberrant methylation in T2D, in a sample of subject.

Preferably, said subject is healthy, thereby producing a reference profile of a healthy subject, or wherein said subject is suffering from T2D, thereby producing a T2D reference profile.

More preferably, said reference profile concerns the methylation status profile of up to, or more than: 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, 200, 205, 210, 215, 220, 225, 230, 235, 240, 245, 250, 255, 260, 265, 270, or 275 of the CpG site(s) defined by SEQ ID Nos 1 to 276, taken from Table 4.

Even more preferably, said reference profile concerns the methylation status profile of the CpG site(s) defined by SEQ ID Nos 1, 3, 5, 6, 8, 15, 17, 23, 26, 28, 50, 65, 71, 92, 114, 129, 142, 178, 210, 223, 235, 263, 270, and/or 275, taken from Table 4.

The invention further provides a microarray or chip comprising one or more T2D-specific CpG sites defined by SEQ ID Nos 1 to 276, taken from Table 4.

Preferably, said microarray or chip comprises up to, or more than: 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, 200, 205, 210, 215, 220, 225, 230, 235, 240, 245, 250, 255, 260, 265, 270, or 275 of the CpG site(s) defined by SEQ ID Nos 1 to 276 taken from Table 4.

More preferably, said microarray comprises the CpG site(s) defined by SEQ ID Nos 1, 3, 5, 6, 8, 15, 17, 23, 26, 28, 50, 65, 71, 92, 114, 129, 142, 178, 210, 223, 235, 263, 270, and/or 275, taken from Table 4.

The invention also provides for the use of the methylation status of one or more of the CpG site(s) defined by SEQ ID Nos 1- 276, taken from Table 4, in the prognosis, diagnosis or prediction of Type 2 Diabetes (T2D).

Preferably, the methylation status of one or more of the CpG sites defined by SEQ ID Nos 1, 3, 5, 6, 8, 15, 17, 23, 26, 28, 50, 65, 71, 92, 114, 129, 142, 178, 210, 223, 235, 263, 270, and/or 275, taken from Table 4 is used.

Finally, the invention provides a method for identifying T2D specific SNPs in comprising the step of comparing the sequence of one or more of the differentially methylated genes corresponding to one or more CpG site(s) defined by SEQ ID Nos 1-276 as defined in Table 4, in a sample from a healthy versus a T2D subject, wherein a difference or polymorphism in said one or more gene regions between the healthy and T2D subject sample is identified as a T2D-specific SNP.

Preferably, the differentially methylated CpG site(s) are selected from the group consisting of those defined by SEQ ID Nos 1, 3, 5, 6, 8, 15, 17, 23, 26, 28, 50, 65, 71, 92, 114, 129, 142, 178, 210, 223, 235, 263, 270, and/or 275, taken from Table 4.

DESCRIPTION OF THE DRAWINGS

Figure 1. Hierarchical profile-based clustering and evaluation of T2D islet DNA methylation. (A) Supervised clustering based on the differentially methylated CpG loci separates T2D from control samples. As an indication of statistical significance, bootstrap values (>0.7) are shown next to the branches. (B) Pie chart depicting the 276 CpG sites showing differential methylation between T2D and control samples. Note the high proportion of hypomethylation events as compared to hypermethylation events. (C) LINE1 repetitive element DNA methylation for CTL and T2D samples by BPS.

Figure 2. Validation of aberrantly methylated loci in T2D islets. (A) Shown is an exemplary locus, *ALDH3B1*. Methylation data for the indicated CpGs obtained with the Infinium assay as well as by conventional bisulfite sequencing (BS), and by bisulfite pyrosequencing (BPS), are compared. Detected DNA methylation levels at other loci (B-D) are also in good agreement between the Infinium methylation assay (bar charts above the gene scale) and BPS used for validation (bar and line charts below gene scale) (see Figure 6 for further examples). CpG islands are indicated by a green line below gene scale. (E) Validation of the Infinium DNA methylation data by BPS. Methylation values obtained by Infinium assay and BPS show high correlation (Spearman correlation coefficient $R=0.873$).

Figure 3. Classification of differentially methylated CpG sites and regulatory element analysis of affected genes. (A) Classification of the CpG sites according to their location relative to CpG islands. Most of the differentially methylated CpG sites are affiliated to genes not possessing a CpG island or are >2kb away from the nearest CpG island (termed "other CpGs" in the legend); only 7% of the affected CpGs are located inside a CGI ("CGI") and about one quarter is located in CGI shores ("CGI shore"), *i.e.* distance to the CGI is between 1 and 2000bp. (B) Classification of the promoters affiliated to the CpG sites based on the promoter CpG content. (for numerical representation of classifications in A and B: *cf.* Table 2) (C) Prediction of putative TF binding sites using the set of low-CpG differentially methylated gene promoters (CpG ratio <0.5) with the Pscan transcription factor motif analysis software. Statistically significant over-represented binding sites have been found for members of the GATA transcription factor family.

Figure 4. Biological pathways associated with differentially methylated loci. (A) Ingenuity Pathway Analysis reveals canonical pathways significantly enriched in T2D pancreatic islets. Measure of significance is indicated by Benjamini-Hochberg-corrected p value (abbreviated as 'B-H p value' in the x axis label of the depicted chart). (B) Manual curation of the biological functions associated with the differentially methylated genes unravelled three broad categories of cellular responses that might be affected in T2D islets. Some of these genes are part of processes leading to beta-cell dysfunction and cell death while others seem to facilitate beta-cell survival and adaptation to the T2D environment. (C) Biological functions of a selection of the differentially methylated loci are highlighted that we hypothesise may play a critical role of the respective gene in the diabetic islets.

Figure 5. Sample dendrogram resulting from unsupervised hierarchical clustering. For unsupervised clustering, 24,421 CpGs were used that satisfy Infinium assay detection p value of $p<0.05$ in every sample. Type 2 diabetic samples (indicated by vertical blue line) cluster together as a self-contained group distinct from the control samples (vertical yellow lines). The dendrogram was derived using the cluster analysis routine in the Methylation module of the GenomeStudio™ software (Illumina, Inc.) applying Euclidean distance metrics.

Figure 6. Validation of differential DNA methylation in T2D. Conventional bisulfite sequencing (A) and bisulfite pyrosequencing (A-M) confirm differential DNA methylation identified by Infinium assay. CpGs affiliated to *ADCY7*, *GLP2R* and *RUNX3* (K-M) served as controls for high, intermediate and low DNA methylation, respectively. Their methylation levels are unaffected by T2D and thus can serve as an additional measure of the concordance between methylation values obtained by Infinium assay and BPS. Generally, the diagram in the upper part of each subfigure depicts the group-wise averaged DNA methylation percentage as assessed by the Infinium assay (yellow: non-diabetic controls, CTL; blue: T2D), while in the lower part, results of bisulfite pyrosequencing (BPS) for the respective CpG position are shown. BPS results are presented as a bar chart only, when the BPS reaction covered a single CpG site and as a line graph when 2 or more sites were covered. Error bars indicate standard deviation of the respective values. A drawn to scale representation of the gene is given for perspective. CpG positions are represented by "lollipop" markers, the gene body appears as a grey bar and the TSS is indicated by an angled arrow. The CpG site analysed by Infinium assay and validated by BPS is fenced with dotted red lines for easier tracing. Green bars below the scale indicate CpG islands. Note that high (>80%) and low (<20%) methylation levels are consistent between Infinium and BPS analyses.

Figure 7. GATA transcription factor binding site prediction and test of significance.

(A) Results of the transcription factor binding site search among 172 CpG-poor differentially methylated gene promoters (CpG ratio <0.5) using Pscan (only the 3 most significant predictions are shown). The most significant motifs detected are affiliated to GATA factor proteins. (B) Comparison of significance for TF motif enrichment from random datasets with significance of GATA motif enrichment from the differentially methylated dataset. As a negative control, Pscan was run on 100 random sets of 172 gene promoters with same properties as the differentially methylated promoter dataset (CpG ratio <0.5; genes represented on Infinium but excluding the differentially methylated ones). For each random set the p value of most significant motif predicted by Pscan was kept. The distribution of p values was then computed and plotted (histogram in lower panel in B). For comparison, the p values of GATA motifs (*cf.* subfigure A) from the differentially methylated dataset are shown (vertical lines in upper panel of B).

Figure 8. Physiological role of selected differentially methylated genes in beta-cell apoptosis.

(A) NIBAN mRNA expression in human islets treated for 24h with oleate or palmitate (left panel) and with synthetic ER stressors thapsigargin (THA), tunicamycin (TUN) or brefeldin (BRE). CTL denotes control. Results represent mean $\pm$ S.E. of 3 - 5 independent experiments. (B) Niban mRNA expression in rat INS-1E cells transfected with negative siRNA (siCTL) or Niban siRNA (siNiban) and treated for 16h with palmitate or cyclopiazonic acid (CPA). Results represent mean $\pm$ S.E. of 3 - 4 independent experiments. (C) Percentage of apoptosis in siRNA-transfected INS-

1E cells after 16h of palmitate or CPA treatment. Lower panel: Western blot quantification of cleaved caspase 3 as an additional apoptosis marker. The blot is representative of 4 independent experiments. (D) CHAC1 mRNA expression in human islets treated for 24 h with oleate or palmitate (left panel) and with synthetic ER stressors. Results represent mean $\pm$ S.E. of 3 – 5 independent experiments. (E) Chac1 mRNA expression in rat INS-1E cells transfected with negative siRNA (siCTRL) or Chac1 siRNA (siChac). (F) Apoptosis in siRNA-transfected INS-1E cells after 16h of palmitate or CPA treatment. Results represent mean $\pm$ S.E. of 4 - 10 independent experiments. Lower panel: Western blot quantification of cleaved caspase 3 as an additional apoptosis marker. The blot is representative of 4 independent experiments. * $p < 0.05$ against untreated (CTL) cells, # $p < 0.05$ by paired ratio t-test.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

As used herein, the singular forms "a", "an", and "the" include both singular and plural referents unless the context clearly dictates otherwise. By way of example, "an antibody" refers to one or more than one antibody; "an antigen" refers to one or more than one antigen.

The terms "comprising", "comprises" and "comprised of" as used herein are synonymous with "including", "includes" or "containing", "contains", and are inclusive or open-ended and do not exclude additional, non-recited members, elements or method steps.

The term "and/or" as used in the present specification and in the claims implies that the phrases before and after this term are to be considered either as alternatives or in combination.

As used herein, the term "level" or "expression level" refers to the expression level data that can be used to compare the expression levels of different genes among various samples and/or subjects.

The term "amount" or "concentration" of certain proteins refers respectively to the effective (i.e. total protein amount measured) or relative amount (i.e. total protein amount measured in relation to the sample size used) of the protein in a certain sample.

All documents cited in the present specification are hereby incorporated by reference in their entirety. In particular, the teachings of all documents herein specifically referred to are incorporated herein by reference.

The term "CpG islands" is a region of genome DNA which shows higher frequency of 5'-CG-3' (CpG) dinucleotides than other regions of genome DNA. Methylation of DNA at CpG

dinucleotides, in particularly, the addition of a methyl group to position 5 of the cytosine ring at CpG dinucleotides, is one of the epigenetic modifications in mammalian cells. CpG islands often occur in the promoter regions of genes and play a pivotal role in the control of gene expression. In normal tissues CpG islands are usually unmethylated, but a subset of islands becomes differentially methylated (hyper- or hypomethylated) during the development of a disease.

Detection of methylation state of CpG islands can be done by any known assay currently used in scientific research. Some non-limiting examples are: Methylation-Specific PCR (MSP), which is based on a chemical reaction of sodium bisulfite with DNA, converting unmethylated cytosines of CpG dinucleotides to uracil (UpG), followed by traditional PCR. Methylated cytosines will not be converted by the sodium bisulfite, and specific nucleotide primers designed to overlap with the CpG site of interest will allow determining the methylation status as methylated or unmethylated, based on the amount of PCR product formed. Alternatively, the HELP assay can be used, which is based on the differential ability of restriction enzymes to recognize and cleave methylated and unmethylated CpG DNA sites. Furthermore, ChIP-on-chip assays, based on the ability of commercially prepared antibodies to bind to DNA methylation-associated proteins like MCP2, can be used to determine the methylation status. Also restriction landmark genomic scanning, also based upon differential recognition of methylated and unmethylated CpG sites by restriction enzymes can be used. Methylated DNA immunoprecipitation (MeDIP), analogous to chromatin immunoprecipitation, can be used to isolate methylated DNA fragments for input into DNA detection methods such as DNA microarrays (MeDIP-chip) or DNA sequencing (MeDIP-seq). The unmethylated DNA is not precipitated. Further, methylated-CpG island recovery assay (MIRA) can be used.

These techniques require the presence of methylated cytosine residues within the recognition sequence that affect the cleavage activity of restriction endonucleases (e.g., HpaII, HhaI) (Singer et al. (1979)). Southern blot hybridization and polymerase chain reaction (PCR)-based techniques can be used with along with this approach.

In another embodiment, a bisulfite dependent methylation assay is known as a combined bisulfite-restriction analysis (COBRA assay) whereas PCR products obtained from bisulfite-treated DNA can also be analyzed by using restriction enzymes that recognize sequences containing 5'CG, such as TaqI (5'TCGA) or BstUI (5'CGCG) such that methylated and unmethylated DNA can be distinguished.

In another embodiment, a methylation detection technique is based on the ability of the MBD domain of the MeCP2 protein to selectively bind to methylated DNA sequences. The bacterially expressed and purified His-tagged methyl-CpG-binding domain is immobilized to a solid matrix and used for preparative column chromatography to isolate highly methylated DNA sequences.

Restriction endonuclease-digested genomic DNA is loaded onto the affinity column and methylated-CpG island-enriched fractions are eluted by a linear gradient of sodium chloride. PCR or Southern hybridization techniques are used to detect specific sequences in these fractions. In addition, one can make use of MALDI-TOF for DNA methylation analysis. Using a combination of
5 four base specific cleavage reactions, each CpG of a target region can be analyzed individually and is represented by multiple indicative mass signals. Another exemplary method for detecting the methylation status of a gene makes use of a bead chip such as the Infinium® bead chip sold by Illumina Inc. San Diego (US).

10 In selected embodiments, the methods for determining the methylation state of (one or more) target gene regions may include treating a target nucleic acid molecule with a reagent that modifies nucleotides of the target nucleic acid molecule as a function of the methylation state of the target nucleic acid molecule, amplifying treated target nucleic acid molecule, fragmenting amplified target nucleic acid molecule, and detecting one or more amplified target nucleic acid
15 molecule fragments, and based upon the fragments, such as size and/or number thereof, identifying the methylation state of a target nucleic acid molecule, or a nucleotide locus in the nucleic acid molecule, or identifying the nucleic acid molecule or a nucleotide locus therein as methylated or unmethylated. Fragmentation can be performed, for example, by treating amplified products under base specific cleavage conditions. Detection of the fragments can be effected by
20 measuring or detecting a mass of one or more amplified target nucleic acid molecule fragments, for example, by mass spectrometry such as MALDI-TOF mass spectrometry. Detection also can be affected, for example, by comparing the measured mass of one or more target nucleic acid molecule fragments to the measured mass of one or more reference nucleic acid, such as measured mass for fragments of untreated nucleic acid molecules. In an exemplary method, the
25 reagent modifies unmethylated nucleotides, and following modification, the resulting modified target is specifically amplified. In some embodiments, the methods for determining the methylation state of (one or more) target gene regions may include treating a target nucleic acid molecule with a reagent that modifies a selected nucleotide as a function of the methylation state of the selected nucleotide to produce a different nucleotide. In particular embodiments, the
30 reagent that modifies unmethylated cytosine to produce uracil is bisulfite. In certain embodiments, the methylated or unmethylated nucleic acid base is cytosine. In another embodiment, a non-bisulfite reagent modifies unmethylated cytosine to produce uracil.

As used herein, a "nucleic acid target gene region" is a nucleic acid molecule that is examined
35 using the methods disclosed herein. For the purposes of the application, "nucleic acid target gene region", "target gene", "target region", "region" and "gene" may be used interchangeably. A nucleic acid target gene region includes genomic DNA or a fragment thereof, which may or may not be part of a gene, a segment of mitochondrial DNA of a gene or RNA of a gene and a segment of RNA of a gene. Examples of "targets" as defined herein are listed in Table 4A by

means of their gene name or Gene ID number. A nucleic target gene region may be further defined by its chromosome position range as is e.g. done in Table 4B for each target sequence identified herewith. The chromosome position ranges provided herein were gathered from the human reference sequence (genome Build hg18/NCBI36, March 2006), which was produced by the International Human Genome Sequencing Consortium.

As used herein, a "nucleic acid target gene molecule" is a molecule comprising a nucleic acid sequence of the nucleic acid target gene region. The nucleic acid target gene molecule may contain less than 10%, less than 20%, less than 30%, less than 40%, less than 50%, greater than 50%, greater than 60%, greater than 70% greater than 80%, greater than 90% or up to 100% of the sequence of the nucleic acid target gene region. A "target peptide" refers to a peptide encoded by a nucleic acid target gene.

As used herein, the "methylation state" or "methylation status" of a nucleic acid target gene region refers to the presence or absence of one or more methylated nucleotide bases or the ratio of methylated cytosine to unmethylated cytosine for a methylation site in a nucleic acid target gene region as defined herein.

For example, a nucleic acid target gene region containing at least one methylated cytosine can be considered methylated (i.e. the methylation state of the nucleic acid target gene region is methylated). A nucleic acid target gene region that does not contain any methylated nucleotides can be considered unmethylated.

Similarly, the methylation state of a nucleotide locus in a nucleic acid target gene region refers to the presence or absence of a methylated nucleotide at a particular locus in the nucleic acid target gene region.

For example, the methylation state of a cytosine at the 10th nucleotide in a nucleic acid target gene region is methylated when the nucleotide present at the 10th nucleotide in the nucleic acid target gene region is 5-methylcytosine. Similarly, the methylation state of a cytosine at the 10th nucleotide in a nucleic acid target gene region is unmethylated when the nucleotide present at the 10th nucleotide in the nucleic acid target gene region is cytosine (and not 5-methylcytosine).

Correspondingly the ratio of methylated cytosine to unmethylated cytosine for a methylation site(s) or locus can provide a methylation state of a nucleic acid target gene region. In certain embodiments the methylation state or status may be expressed as a percentage of methylatable nucleotides (e.g., cytosine) in a nucleic acid (e.g., amplicon or gene region) that are methylated (e.g., about 5%, about 10%, about 15%, about 20%, about 25%, about 30%, about 35%, about 40%, about 45%, about 50%, about 55%, about 60%, about 65%, about 70%, about 75%, about

80%, about 85%, about 90%, about 95% or about 100% methylated; greater than 80% methylated, between 20% to 80% methylated, or less than 20% methylated). A nucleic acid may be "hypermethylated," which refers to the nucleic acid having a greater number of methylatable nucleotides that are methylated relative to a control or reference. A nucleic acid may be
5 "hypomethylated," which refers to the nucleic acid having a smaller number of methylatable nucleotides that are methylated relative to a control or reference. The methylation status or state is determined in a CpG island in certain embodiments. Examples of target CpG islands according to the present invention are listed in Table 4B and in SEQ ID Nos 1-276.

10 As used herein, a "characteristic methylation state" refers to a unique, or specific data set comprising the methylation state of at least one of the methylation sites of one or more nucleic acid(s), nucleic acid target gene region(s), gene(s) or group of genes of a sample obtained from a subject. It can be the combined data of the methylation state of a panel of multiple target genes according to the present invention in said sample, as compared to a reference sample from e.g. a
15 healthy subject.

As used herein, "methylation ratio" refers to the number of instances in which a molecule or locus is methylated relative to the number of instances the molecule or locus is unmethylated.

20 Methylation ratio can be used to describe a population of individuals or a sample from a single individual.

For example, a nucleotide locus having a methylation ratio of 50% is methylated in 50% of instances and unmethylated in 50% of instances. Such a ratio can be used, for example, to
25 describe the degree to which a nucleotide locus or nucleic acid region is methylated in a population of individuals. Thus, when methylation in a first population or pool of nucleic acid molecules is different from methylation in a second population or pool of nucleic acid molecules, the methylation ratio of the first population or pool will be different from the methylation ratio of the second population or pool. Such a ratio also can be used, for example, to describe the degree to
30 which a nucleotide locus or nucleic acid region is methylated in a single individual. For example, such a ratio can be used to describe the degree to which a nucleic acid target gene region of a group of cells from a tissue sample are methylated or unmethylated at a nucleotide locus or methylation site.

35 As used herein, a "methylated nucleotide" or a "methylated nucleotide base" refers to the presence of a methyl moiety on a nucleotide base, where the methyl moiety is not present in a recognized typical nucleotide base. Cytosine does not contain a methyl moiety on its pyrimidine ring, however 5-methylcytosine contains a methyl moiety at position 5 of its pyrimidine ring. In this respect, cytosine is not a methylated nucleotide and 5-methylcytosine is a methylated nucleotide.

As used herein, a "methylation site" is a nucleotide within a nucleic acid, nucleic acid target gene region or gene that is susceptible to methylation either by natural occurring events in vivo or by an event instituted to chemically methylate the nucleotide in vitro.

- 5 As used herein, a "methylated nucleic acid molecule" refers to a nucleic acid molecule that contains one or more methylated nucleotides that is/are methylated.

As used herein "CpG island" refers to a G:C-rich region of genomic DNA containing a greater number of CpG dinucleotides relative to total genomic DNA, as defined in the art. It should be
10 noted that differential methylation of the target genes according to the invention is not limited to CpG islands only, but can be in so-called "shores" or can be lying completely outside a CpG island region.

As used herein, a first nucleotide that is "complementary" to a second nucleotide refers to a first
15 nucleotide that base-pairs, under high stringency conditions to a second nucleotide. An example of complementarity is Watson-Crick base pairing in DNA (e.g., A to T and C to G) and RNA (e.g., A to U and C to G). Thus, for example, G base-pairs, under high stringency conditions, with higher affinity to C than G base-pairs to G, A or T, and, therefore, when C is the selected nucleotide, G is a nucleotide complementary to the selected nucleotide.

20 As used herein, the term "correlates" as between a specific diagnosis or a therapeutic outcome of a sample or of an individual and the changes in methylation state of a nucleic acid target gene region refers to an identifiable connection between a particular diagnosis or therapy of a sample or of an individual and its methylation state.

25 As used herein, a "subject" includes, but is not limited to, an animal, plant, bacterium, virus, parasite and any other organism or entity that has nucleic acid. Among animal subjects are mammals, including primates, such as humans. As used herein, "subject" may be used interchangeably with "patient" or "individual".

30 As used herein, a "methylation" or "methylation state" correlated with a disease, disease outcome or outcome of a treatment regimen refers to a specific methylation state of a nucleic acid target gene region or nucleotide locus that is present or absent more frequently in subjects with a known disease, disease outcome or outcome of a treatment regimen, relative to the methylation state of
35 a nucleic acid target gene region or nucleotide locus than otherwise occur in a larger population of individuals (e.g., a population of all individuals).

As used herein, "sample" refers to a composition containing a material to be detected, and includes e.g. "biological samples", which refer to any material obtained from a living source, for example an animal such as a human or other mammal that can suffer from diabetes or related

disorders. The biological sample can be in any form, including a solid material such as a tissue, cells, a cell pellet, a cell extract, or a biopsy, or it can be in the form of a biological fluid such as urine, whole blood, plasma, or serum, or any other fluid sample produced by the subject. In addition, the sample can be solid samples of tissues or organs, such as collected tissues, including pancreatic tissues, more specifically pancreatic island of Langerhans cells. Samples can include pathological samples such as a formalin- fixed sample embedded in paraffin. If desired, solid materials can be mixed with a fluid or purified or amplified or otherwise treated. Samples examined using the methods described herein can be treated in one or more purification steps in order to increase the purity of the desired cells or nucleic acid in the sample, Samples also can be examined using the methods described herein without any purification steps to increase the purity of desired cells or nucleic acid. In particular, herein, the samples include a mixture of matrix used for mass spectrometric analyses and a biopolymer, such as a nucleic acid. Preferably, said sample is a pancreatic island of Langerhans cell extract or biopsy, more preferably an extract or pool of pancreatic beta-cells, or is whole blood, plasma or serum of a subject.

The term "beta cell related disorder" described in the methods or uses or kits of the invention encompasses all disorders related to beta cells such as: type 1 diabetes mellitus, type 2 diabetes mellitus, hyperinsulinemia, obesity, neuroendocrine tumors or occurrence of insulinoma.

The present invention identifies 276 differentially methylated CpG sites that are affiliated to 254 genes performing the to our knowledge first comprehensive DNA methylation profiling of human T2D pancreatic islets. The present invention shows that the uncovered DNA methylation changes in diabetic islets may reflect epigenetic adaptations acquired over time by long-lived beta-cells that were exposed to the stressful environment of T2D for many years. Without wanting to be bound by any specific theory, hypomethylation has for example been observed in oxidative stress, ER stress and apoptotic pathways, that may result from chronic exposure to high concentrations of free fatty acids and glucose in T2D. It can be assumed that beta-cells utilise different pathways to adapt to different stresses. The present invention therefore suggest a new concept: demethylation of key CpGs may play a more important role in T2D-related inflammatory and signal transduction responses than in metabolic pathways. Moreover, the identification of methylation changes in T2D islets outlines an unexpected level of epigenetic regulation in beta-cell function, which is very important for the understanding of the pathogenesis of T2D.

The present invention postulates that the prevalent hypomethylation in T2D islets is indicative of processes involved in adaptation to the diabetic environment as well as biological pathways associated to beta-cell dysfunction and apoptosis are also activated (Figure 4C).

In terms of genomic features, T2D-associated differential DNA methylation was mainly detected

in low and intermediate CpG promoters (LCP, ICP), high CpG promoters (HCP) are underrepresented. Upon analysing LCP and a subset of ICP genes (CpG ratio <0.5; Saxonov S et al., 2006 Proc Natl Acad Sci U S A 103: 1412-1417) the inventors identified the GATA family transcription factors that are predicted to regulate a significant subset of these genes as being prone to differential methylation in healthy versus T2D subjects.

The physiological roles of the differentially methylated loci in T2D can coarsely be described as genes responding to (external) stimuli and to stress. Of note, Saxonov *et al.* found that a disproportionately high percentage of genes affiliated to these biological functions possess promoters with a low CpG density (Saxonov S et al., 2006 Proc Natl Acad Sci U S A 103: 1412-1417). This might indicate a general principle with regard to the promoter class of the differentially methylated gene loci: chronic diseases such as T2D and SLE (Javierre BM, et al., 2010 Genome Res 20: 170-179) (LCP genes overrepresented) on one hand are distinct from diseases associated with cellular overgrowth (e.g. cancer; prevalence of HCP, relatively few LCP genes; (Martin-Subero JI et al., 2009 Blood 113: 2488-2497; Martin-Subero JI et al., 2009 PLoS One 4: e6986; Richter J et al., 2009 BMC Cancer 9: 455) on the other hand.

As acknowledged by McCarthy and Zeggini, the >20 gene variants of T2D susceptibility genes known to date cannot fully explain T2D predisposition (McCarthy and Zeggini 2009 Curr Diab Rep 9: 164-171). Our study points to the involvement of epigenetic alterations in T2D thus underscoring the previously established contribution of habit and lifestyle to its development. Combining the advantages of genome-scanning techniques and epigenome analyses will pave the way to better comprehend the pathogenesis of T2D.

In addition, the invention provides new tools for identifying T2D linked small nucleotide polymorphisms (SNPs) in the epigenetically regulated genes or CpG's according to the invention (cf. Table 4), based on the known interplay between SNPs and differential (allele-specific) DNA methylation (ASM) as e.g. described in Shoemaker R et al., Genome Res 2010 Jul;20(7):883-889. The invention thus provides a method for identifying T2D-specific SNPs, based on the identification of new differentially methylated gene regions in Table 4, using techniques known in the art.

The invention thus provides an altered DNA methylation profile in the pancreatic islets of T2D patients with a major preponderance of hypomethylation in sequences outside of CpG islands. These aberrant methylation events affect over 250 genes, whose dysregulation in T2D may notably be linked to beta-cell dysfunction, cell death and adaptation to metabolic stress. In particular, the present invention highlights genes belonging to biological processes whose involvement in T2D is not yet fully understood, such as inflammation and ion transporters/channels/sensors, thereby further unravelling the biological complexity of T2D and

providing new diagnostic and therapeutic possibilities. A similar molecular signature is present in other tissues (e.g. blood, adipose tissue, muscle etc) of easier access than the islets of Langerhans, thus providing means for disease prediction.

- 5 The invention is illustrated by the following non-limiting examples.

EXAMPLES

Materials and Methods

10

Isolation of pancreatic islets. From September 2004 to November 2009, pancreatic islets of Langerhans were isolated from pancreata of 5 type 2 diabetic and 11 non-diabetic male cadaveric donors (Table 1) as described previously (Del Guerra S et al., 2005 Diabetes 54: 727-735). Glucose-induced insulin secretion was measured as described. The diagnosis of type 2 diabetes
15 was based on previously described clinical criteria (ADA (1997) Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Diabetes Care 20: 1183-1197; Genuth S et al., 2003 Diabetes Care 26: 3160-3167).

Table 1: Patient data of pancreatic islet donors. While age and BMI are comparable between
20 both groups, *in vitro* islet insulin secretion upon stimulation with 16.7mM glucose is significantly lower in islets of T2D donors, as expected (*p* values in brackets indicate results of two-tailed t-tests). CTL denotes non-diabetic controls.

	CTL	T2D
number of samples	11	5
age (years)	57.2 ± 15.6	64.4 ± 5.3 (<i>p</i> ~0.33)
BMI (kg m ⁻²)	27.0 ± 1.1	26.5 ± 2.2 (<i>p</i> ~0.8)
induced insulin secretion (ratio high/standard conc.)		
glucose 16.7mM/3.3mM	3.59 ± 1.05	1.24 ± 0.30 (<i>p</i> <0.001)

25 **Methylation Profiling Using the Infinium® Assay.** Genomic DNAs were isolated from the above isolated pancreatic islets using Wizard® SV Genomic DNA kit (Promega Corp.). 1µg of genomic DNA was treated with sodium bisulfite using the EZ DNA Methylation™ kit (Zymo Research) according to the manufacturer's procedure, respecting the recommended alterations in

protocol for consecutive Infinium® methylation analysis. Methylation status of 27,578 distinct CpG sites was analysed using the Infinium® HumanMethylation27 BeadChip array (Illumina, Inc., San Diego) according to the manufacturer's protocol. Data acquisition was done with the Illumina BeadArrayReader and quality control, data handling, comparisons etc. were performed with the Methylation module of the GenomeStudio™ software package (Illumina), MS Excel2003, R 2.9.0 software, Openstat and MicroarrayAnalyse v1.0 (Graessler J (2008) MicroarrayAnalyse v1.0.).

CpG island and promoter class annotation. Annotations for the Infinium HumanMethylation27 provided by Illumina were augmented with respect to (i) the position of the analysed CpG relative to the nearest CpG island (inside a CpG island, in CpG island shore or more than 2kb away from an island) and (ii) the promoter class of the gene affiliated to the evaluated CpG (high/intermediate/low CpG content promoter). For all annotations, the human genome build 36.1 (hg18, March 2006) provided the basis.

For classification of the CpG position relative to CpG islands, the CpG island map provided by Bock *et al.* (combined epigenetic score >0.5; genome assembly hg18/NCBI36) was used as reference (Bock C et al., 2007 PLoS Comput Biol 3: e110); the CpGs were classified into 3 categories according to (Irizarry RA et al., 2009 Nat Genet 41: 178-186). Designation of the CpGs is as follows: "inside CGI" if the CpG was inside a CpG island, "CGI shore" if the CpG was located within a 2kb region around a CpG island, and "other CpG" otherwise (distance to closest CpG island >2kb).

Promoters of the gene loci affiliated to the analysed CpG sites were classified according to their CpG content. First, sequences ranging from positions -700 to +500 relative to the transcription start site (TSS) from the UCSC genome browser database were extracted, then the CpG ratio and the GC content of these sequences in sliding windows of 500bp with 5bp offsets was calculated. For classification criteria the definition by Weber *et al.* (Weber M et al., 2007 Nat Genet 39: 457-466) was followed. In short, promoters were defined as high CpG promoters (HCP) if at least one 500bp window contained a CpG ratio >0.75 together with a GC content >0.55 whereas in low CpG promoters (LCP) no 500bp window reached a CpG ratio of at least 0.48. All promoters not fitting in either of the above promoter classes were termed intermediate CpG promoters (ICP). 5 differentially methylated gene promoters (and a total of 54 gene promoters on the Infinium array) could not be classified due to great distance to TSS or lack of annotation.

Hierarchical Clustering. For unsupervised hierarchical clustering, the data sets were filtered for probes/CpG sites with a detection p value <0.05. CpGs not fitting into this criterion in at least one sample were excluded from further analysis. The clustering was performed with the average beta

values (equals methylation percentage $\div$ 100) of 24,421 CpGs for each sample using the cluster analysis function in GenomeStudio™ applying Euclidean distance metrics.

For UPGMA (Unweighted Pair Group Method using Arithmetic averages) clustering, the data sets for computation were assembled as follows: A group-wise $|\Delta\beta| > 0.15$ and $p < 0.01$ (Mann-Whitney) were set as filtering criteria. 276 probes fitted into these criteria. The methylation percentage of the CpG site corresponding to each probe was extracted for each sample. Then, the methylation values were categorised into 10 equal classes and imported into MEGA4 (Tamura K et al., 2007 Mol Biol Evol 24: 1596-1599) in which the phylogenetic analysis was conducted. The dendrogram was computed using the UPGMA method applying the 'number of differences' model (Sneath PHA et al., (1973) Numerical taxonomy : the principles and practice of numerical classification: W. H. Freeman, San Francisco). To determine the validity of the sample clustering based on the methylation data a bootstrap test (10,000 sampling steps) was used to calculate the percentage of replicate trees in which the associated samples clustered together (Felsenstein J 1985 Evolution 39: 783-791). Bootstrap values of 0.7 or higher were considered significant and are shown next to the branches in Figure 1A.

Conventional Bisulfite Sequencing and Bisulfite Pyrosequencing. 750ng genomic DNA was subjected to bisulfite conversion using the Epitect® Bisulfite Kit (Qiagen) or the EZ DNA Methylation™ kit (Zymo Research) according to the manufacturer's protocol. Elution of the converted DNA was generally performed with 26µl elution buffer and 8µl of the eluate was used as template in subsequent PCRs. To ensure sufficient amount of product, amplifications were generally performed as nested PCRs.

PCR and sequencing primers for bisulfite pyrosequencing were deduced using the PyroMark® Assay Design 2.0 software (Qiagen). Primers for pre-amplification and conventional bisulfite sequencing were designed manually or with the help of BiSearch primer design tool (<http://bisearch.enzim.hu>) and evaluated using the GeneRunner software (v3.05 Hastings Software, Inc.). Primers were obtained from Eurogentec S.A. or Sigma-Aldrich Corp. Biotinylated primers were ordered HPLC purified, all other primers desalted.

The pre-amplification PCR was conducted with primers (see EF, ER primers in Table 3) amplifying 400-720bp spanning the CpG of interest and additionally as many as possible neighbouring CpG sites. CpG sites in the annealing positions of the PCR primers were avoided where possible; otherwise primers were ordered with ambiguities at the respective positions. PCR was conducted with 3mM MgCl₂, 1mM of each dNTP, 12% (v/v) DMSO, 500nM of each primer and optionally 500mM Betaine in heated-lid thermocyclers under the following conditions: 95°C 3:00; 25x[94°C 0:30; 51°C 0:40; 72°C 1:30]; 72°C 5:00 and cooled afterwards to 10°C.

For conventional bisulfite sequencing (BS), nested PCRs were conducted as described above

using 35 PCR cycles and a decreased elongation time of 1 minute (for PCR primers *cf.* Table 3). PCR products were separated on a 1% agarose gel and single bands were cut and eluted from the gel. Cloning of the nested PCR products was performed with TOPO TA Cloning® kit (Invitrogen Corp.) and the plasmids were sequenced by Genoscreen.

5

For pyrosequencing, a nested PCR was performed with primers designed by the PyroMark® Assay Design software (Qiagen) using HotStarTaq PCR kit (Qiagen) according to the manufacturer's recommendations. Reactions were performed in heated-lid thermocyclers under the following conditions: 95°C 15:00; 45x[94°C 0:30; 55°C 0:30; 72°C 0:30]; 72°C 10:00 and finally cooled to 8°C. Sample preparation and pyrosequencing reactions were performed with the Pyromark™ Q24 system (Qiagen). For validation of Infinium assay-derived DNA methylation by BPS, usually three to five randomly chosen samples from each group (CTL and T2D) were analysed and DNA methylation degrees were averaged.

15 **Example 1: Identification of the T2D-related Differential DNA Methylation Profile.**

A comprehensive DNA methylation profiling was performed to analyse the methylomes of freshly isolated islets from 16 human cadaveric donors (5 diabetic and 11 non-diabetic donors; *cf.* Table 1). For this purpose, the recently developed Infinium Methylation Assay (Illumina® Infinium® HumanMethylation27 BeadChip; Bibikova M et al., 2009 Epigenomics 1: 177-200) was used to interrogate the methylation status of more than 27,000 CpGs corresponding to over 14,000 genes. As an initial step, it was evaluated whether DNA methylation changes in T2D were sufficient to distinguish the diabetic from the control samples when comparing complete methylation profiles. For this, an unsupervised hierarchical clustering was performed which placed the diabetic islet samples as one self-contained group distinct from the control samples in the resulting dendrogram (Figure 5). This outcome highlights two facts: firstly, diabetic DNA methylation profiles are more similar to each other than to the methylation profile of any control sample, hinting towards a T2D-specific DNA methylation profile; secondly, the existence of a single branch containing the five diabetic samples shows that DNA methylation changes are sufficiently pronounced (even in the unfiltered datasets) to distinguish diabetic from control samples. This is indicative of distinct, T2D-specific changes in the epigenome of pancreatic islets. Following this initial analysis, T2D-related methylation changes were identified by filtering the datasets for CpG sites showing significant differences in DNA methylation levels between control and T2D groups (*cf.* Material and Methods and Table 4). The results of the filtering show that there are pronounced DNA methylation changes in T2D islets.

35

We then set out to evaluate the descriptive power of the CpG sites in the filtered dataset to differentiate diabetic from non-diabetic specimens in sample-wise comparisons. We therefore extracted the methylation values for each sample (*cf.* Materials and Methods) and performed a supervised clustering (*cf.* Figure 1A). As expected, the resulting dendrogram shows that samples

group together in two clusters containing exclusively control (CTL) or diabetic (T2D, grey bar) samples, indicating that class identity (CTL,T2D) is the most important separation criterion (Figure 1A, left-most branch). To assess clustering confidence, a bootstrapping analysis was carried out additionally after dendrogram computation. The obtained bootstrap of 0.85 indicates significant statistical support for the bipartite distribution between diabetic and non-diabetic samples based on the analysis of the CpG contained in the filtered dataset. The occasional high bootstrap values adjoined to sample pairs illustrate similarities in the DNA methylation profiles of these samples. These data demonstrate that human pancreatic islets undergo DNA methylation alterations in T2D that are discernable by means of DNA methylation profiles.

10 **Example 2: T2D-related Aberrations Encompass Mostly Promoter-specific DNA Hypomethylation.**

The above experiments enabled us to collect the first comprehensive DNA methylation dataset for T2D human islets. 276 CpG sites were identified, affiliated to 254 gene promoters, showing differential methylation between normal and diseased samples (Figure 1B and Table 4). Strikingly, 266 of these 276 CpGs (96%) showed decreased methylation levels, while only 10 (4%) were hypermethylated (Figure 1B). This unexpected finding contrasts with the well-known DNA methylation changes observed in cancers, where gene-specific hypo- and hypermethylation are more or less evenly distributed (Jones PA and Baylin SB 2007 Cell 128: 683-692). With respect to global DNA methylation, cancers generally display hypomethylation (Tost J 2010 Mol Biotechnol 44: 71-81), primarily in repetitive DNA, while they generally show gene-specific hypermethylation.

To test whether the observed T2D-related changes are gene-specific or whether they reflect global hypomethylation in the genome of islet cells, we measured DNA methylation levels of the repetitive LINE1 element in control and diabetic samples with bisulfite pyrosequencing. Analysing DNA methylation of LINE-1, which makes up approximately 20% of human genome, has been shown to serve as an accurate estimate of global DNA methylation changes (Yang AS et al., 2004 Nucleic Acids Res 32: e38). Figure 1C shows that repetitive elements are not differentially methylated in T2D as substantiated by the strong overlap between CTL and T2D samples, indicating the absence of global hypomethylation in T2D islets. This indicates that the detected hypomethylation of the target CpG areas of the present invention (cf. Table 4), are not due to global hypomethylation in T2D, but are due to gene-specific hypomethylation in T2D patients.

Example 3: Bisulfite Sequencing Validation of T2D-related Differential DNA Methylation

To corroborate the observed Infinium measurements (cf. Figure 1 and Table 4), bisulfite pyrosequencing (BPS) and in some cases conventional bisulfite genomic sequencing (BS) was applied to randomly selected differentially methylated CpG sites. In all 17 cases tested,

differential DNA methylation at the respective CpG sites was confirmed by BPS (Figure 2 and Figure 5). Where implemented, bisulfite genomic sequencing (BS) also confirmed the DNA methylation profiling data (Figure 2A and Figure 6A). Figure 2A depicts an exemplary analysed gene, *ALDH3B1*, for which the Infinium data were confirmed by BS and BPS. Additional validated genes are shown in Figure 2B (*CASP10*) and 2C (*PPP2R4* alias *PP2A*). Two differentially methylated CpG sites inside a CpG island in the *IGF2/IGF2AS* locus were discovered. The differential methylation of one of the CpGs in this region was tested and confirmed by BPS (Figure 2D). Further examples are shown in Figure 6. A direct comparison of methylation percentages obtained by the Infinium Methylation assay and BPS (Figure 2E) yielded a highly positive correlation (Spearman correlation $R=0.873$) confirming their validity. BPS analysis of three negative controls constituting high (>90%), intermediate (~40%) and low (<10%) levels of DNA methylation showed expectedly no differential DNA methylation between sample groups (Figure 6K-M). Of note, the BPS experiments designed for validation of the methylation profiles encompassed neighbouring CpGs to the Infinium-assayed site in the majority of cases (*cf.* Figure 2 and Figure 6); the adjacent CpG sites often displayed similar DNA methylation levels. Hence, the validation experiments indicate that individual CpGs from the Infinium Methylation array can be used as informative markers for the methylation status of the respective surrounding regions. Taken together, bisulfite pyrosequencing and genomic bisulfite sequencing results confirm the methylation values obtained from the Infinium-based assay, thus validating the T2D-specific DNA methylation profiles.

Example 4: Genomic Features Associated with Differential DNA Methylation in T2D.

Having shown that the observed differential DNA methylation in T2D is gene-specific rather than global (*cf.* Figure 1C), it was further determined whether this aberrant DNA methylation is (a) located within or outside of CpG islands, (b) prevalent in distinct promoter classes (that are based on CpG density) or (c) correlated with specific regulatory elements.

Previous work that investigated the localisation of differential DNA methylation in cancer and in different tissues has suggested that substantial DNA methylation differences occur in CpG island (CGI) shores (Irizarry RA et al., 2009 Nat Genet 41: 178-186). Bioinformatic tools were used to determine the distance of differentially methylated CpGs to the nearest CGI. Utilising CGI prediction (Bock C et al., 2007 PLoS Comput Biol 3: e110), it was shown that approximately 65% of the differentially methylated CpG sites is located >2kb from the nearest CGI ("other CpGs" in Figure 3A and Table 2). Additionally, about one quarter of CpGs resides in CGI shores (1 to 2000bp from a CGI border) while only a small percentage is located inside CGIs (Figure 3A and Table 2). Thus, these results seem to distinguish the localisation of the DNA methylation profile in T2D islets from those found in tumours.

Next, the occurrence of these differentially methylated CpG sites in relation to CpG density of the affiliated gene promoters was analysed. Saxonov *et al.* discovered a bipartite distribution of gene

promoters with minor overlap between both classes when categorising promoter sequences by means of their CpG content (Saxonov S et al., 2006 Proc Natl Acad Sci U S A 103: 1412-1417). They discovered that promoters are either relatively depleted of CpG sites (low CpG promoters) or they are enriched in CpG sites (high CpG promoters), preferentially around the transcription start site (TSS). Weber *et al.* introduced a third class of promoters called intermediate CpG promoters (ICP) (Weber et al., 2007 Nat Genet 39: 457-466), to account for the overlap between classes mentioned above. They also developed precise classification criteria for the three classes which were utilised here in an adapted form: positions -700 to +500 relative to the TSS were used for the promoter classification (*cf.* Material and Methods), since CpG density as well as differential DNA methylation are distributed symmetrically around the TSS (Saxonov S et al., 2006 Proc Natl Acad Sci U S A 103: 1412-1417). Figure 4B shows that most of the differentially methylated CpG sites from T2D islets are located in LCP and ICP class promoters. ICP class promoters have been described as regions of dynamic DNA methylation changes (Weber et al., 2007 Nat Genet 39: 457-466), while LCP class promoters have seldomly been investigated at all. Their role as sites of hypomethylation in T2D therefore remains to be explored.

In terms of genomic features, it was evaluated whether the observed aberrant DNA methylation in T2D islets correlated with specific regulatory elements. The use of computational approaches to extract functional meaning from the annotated genome has gained importance in recent years. Roeder *et al.* underscored the need to first separate gene promoters on the basis of their CpG content before analysing the presence of enriched transcription factor-binding motifs (Roeder HG et al., 2009 Nucleic Acids Res 37: 6305-6315). These authors found that promoters depleted of CpG sites often contain tissue-specific transcription factor binding sites. Therefore, CpG-depleted promoters were first extracted by selecting all differentially methylated promoters with a CpG ratio lower than 0.5 as performed by Saxonov *et al.* the extracted set of 172 CpG-poor promoters was then used as a starting point to detect putative transcriptional regulatory signals using the Pscan (Zambelli F et al., 2009 Nucleic Acids Res 37: W247-252) software and the TRANSFAC transcription factor motifs database (Matys V et al., 2006 Nucleic Acids Res 34: D108-110). As shown in Figure 3C and Figure 7A, significant enrichment of GATA transcription factor family binding sites was identified, namely GATA1 and GATA2 binding sites as well as a common binding sites assigned to all GATA proteins (p values for enrichment ranged from 2.38×10^{-15} to 3.28×10^{-13}). This analysis was then repeated with a set of all differentially methylated LCP and ICP promoters (*cf.* Table 4) to enlarge the above mentioned set of 172 CpG ratio <0.5 promoters. This analysis yielded similar results (data not shown). To increase precision and to assess validity of our predictions, a negative control was performed by repeating the detection routine described above with 100 sets of 172 promoters randomly sampled from CpG-depleted genes (CpG ratio <0.5) that are not differentially methylated but present on the Infinium array. Importantly, the highest significance of enrichment (best p value) encountered in the searches with these random sets never exceeded the best p value obtained with the set of differentially methylated genes

(Figure 7B). Combined, these data reveal a statistically significant enrichment of GATA transcription factor binding motifs in our set of promoters (CpG ratio <0.5) differentially methylated in T2D islets.

5 Example 5: Biological Pathways Associated with the Differential DNA Methylation in Diabetic Islets

To determine the biological relevance of the differential DNA methylation in pancreatic islets, the affected genes were analysed with regard to their reported functions and the biological pathways they are part of. The obtained data was compared with the list of known T2D risk genes to find out whether these loci are also targets of epigenetic dysregulation. Of the approximately 20 known T2D risk genes, *GRB10* was found to be differentially methylated in diabetic islets; other established T2D susceptibility loci revealed no significant differential DNA methylation in the analyses. Instead, aberrantly methylated genes with similar or identical biological functions to these known T2D risk genes were found (Table 4). For example, *KCNQ1* and *KCNJ11* (SNP variants of which are associated with higher T2D risk; (Scott LJ et al., 2007 Science 316: 1341-1345)) were not significantly altered in their methylation levels but three other potassium channel genes, *KCNE2*, *KCNJ1* and *KCNK16*, were changed in their promoter methylation state (Table 4). Other examples of T2D susceptibility loci for which genes with related or identical functions were identified are *SLC30A8* and *CDKAL1*. In the datasets of the present invention, *SLC30A8* was not differentially methylated but two other zinc transporter genes *SLC39A5* and *ZIM2* were hypomethylated. For the T2D risk gene *CDKAL1*, its methylation state was found to be unchanged in T2D islets, while its target gene *CDK5R1* exhibited pronounced hypomethylation (Figure 6). In summary, although the promoter methylation of established T2D risk loci remained unchanged in the present profiling approach (with *GRB10* as an exception), other genes with the same biological function (i.e. potassium, zinc transporters) or genes in the same regulatory networks (i.e. *CDK5* pathway) as established T2D risk genes displayed aberrant DNA methylation.

Interestingly *GRB10*, the only established T2D susceptibility gene from GWAS that was characterised as differentially methylated in this study (cf. Table 4), is also known to be imprinted (Morison IM et al., 2001 Nucleic Acids Res 29: 275-276). Furthermore, *IGF2AS* which is located inside the imprinted *IGF2* locus, was found to display profound hypomethylation (cf. Figure 2D). This led to the more rigorous investigation of the number of imprinted genes that display altered DNA methylation in the present T2D samples. The conducted search resulted in the identification of *GABRB3*, *ZIM2* and *PER2* besides *IGF2AS* and *GRB10*. This finding suggests a partial loss of imprinting in the islets during pathogenesis of T2D.

The analyses described above found only few common T2D-candidate genes among the differentially methylated genes uncovered in this study. This could imply that T2D pathogenesis in

islets is partially mediated by unexpected and thus previously unappreciated genes. To decipher their roles in the context of T2D islets, as a first step an Ingenuity Pathway Analysis (IPA) was performed to determine which canonical pathways were over-represented in our set of genes (Figure 4A). Inflammation-related processes were highly enriched, in particular the acute phase response and IL-8 signalling. Other pathways, such as apoptosis and death receptor signalling, emphasise the role of beta-cell loss in T2D. Enrichment for pathways involved in metabolism and internal and external cell structure (e.g. actin cytoskeleton and integrin signalling) may be indicative of altered islet function and architecture.

Secondly, we performed an extensive manual curation according to a previously described beta-cell-targeted annotation (Kutlu B et al., 2003 Diabetes 52: 2701-2719; Ortis F et al., 2010 Diabetes 59: 358-374). In partial agreement with the IPA analysis, these genes were found to fall into three broad categories: (1) genes related to beta-cell dysfunction and death, (2) genes potentially facilitating the adaptation of the pancreatic islets to the altered metabolic situation in T2D and (3) genes whose role in disease pathogenesis remains to be unearthed (Figure 4B). In the first category there were hypomethylated genes related to DNA damage and oxidative stress (e.g. *GSTP1*, *ALDH3B1*), endoplasmic reticulum (ER) stress responses (*NIBAN*, *PPP2R4*, *CHAC1*), and apoptosis (*CASP10*, *NR4A1*, *MADD*) (Figure 4B, left part). The second category, which comprises adaptation-related genes, contains few metabolism-associated genes (e.g. *HK1*, *FBP2*; Figure 3C, right part) and many more genes involved in signal transduction or encoding hormones, growth factors (e.g. *EGF*, *FGF1*, *IGF2AS*), or transcription factors involved in important regulatory networks (for instance *FOXA2/HNF3B*, *PAX4* and *SOX6*) (Figure 4B, right part).

Some genes of interest from the highlighted categories are depicted below, providing more functional background and a possible explanation of how these genes are connected to T2D pathogenesis. The identified differentially methylated genes can be classified according to their function as follows:

1. Dysfunction / cell death:

DNA damage / oxidative stress

ALDH3B1 encodes an aldehyde dehydrogenase that can protect cells from lipid peroxidation-induced cytotoxicity (Marchitti SA et al., 2007 Biochem Biophys Res Commun 356: 792-798).

GSTP1 plays an important role in detoxification by catalysing the conjugation of many hydrophobic and electrophilic compounds with reduced glutathione.

Endoplasmic Reticulum (ER) stress

NIBAN is induced during ER stress and may counteract the suppression of protein

translation that occurs under this condition (Sun GD et al., 2007 Biochem Biophys Res Commun 360: 181-187).

CHAC1 is also induced by ER stress and may trigger apoptosis (Mungrue IN et al., 2009 J Immunol 182: 466-476).

5

Apoptosis

NR4A1 is involved in ER stress-induced apoptosis and can interact with the anti-apoptotic protein BCL2 (Contreras JL et al., 2003 Transpl Int 16: 537-542; Liang B et al., 2007 Exp Cell Res 313: 2833-2844).

10 MADD encodes a MAP-kinase activating death domain-containing protein with anti-apoptotic function. It may also play a role in glucose homeostasis (Dupuis J et al., 2010 Nat Genet 42: 105-116).

CASP10 is involved in advanced-glycation-endproduct-induced apoptosis (Lecomte M et al., 2004 Biochim Biophys Acta 1689: 202-211; Obrenovich ME et al., 2005 Sci Aging Knowledge Environ 2005: pe3) and can activate NF-kappa-B (Wang H et al., 2007 Biochim Biophys Acta 15
1770: 1528-1537). The CASP10 gene is absent from the mouse and rat genome; both species are frequently used as T2D model organisms (Reed JC et al., 2003 Genome Res 13: 1376-1388).

2. Adaptation:

20

Signal transduction

MAPK1 is an important regulator of beta-cell function (Lawrence M et al., 2008 Acta Physiol (Oxf) 192: 11-17), e.g. contributing directly to short- vs. long-term insulin response and regulation of pro-apoptotic CHOP10 (Lawrence MC, et al., 2007 Proc Natl Acad Sci U S A 104:
25 11518-11525). MAPK1 constitutes the centre of a regulatory network activated by elevated free fatty acid levels (Sengupta U et al., 2009 PLoS One 4: e8100) common in T2D patients. MAPK/ERK signalling leads to dephosphorylation of cascade proteins by PP2A/PPP2R4 (Guo J et al., 2010 Mol Cell Biochem: (Epub; DOI: 10.1007)) pointing towards an interaction between the identified processes, in this case signal transduction (adaptation category) and ER stress
30 (dysfunction/cell death category) (cf. Figure 5B).

CDK5R1 acts as an activator of CDK5 (Ubada M et al., 2004 Endocrinology 145: 3023-3031) whose expression is regulated by glucose and which inhibits insulin secretion (Wei FY et al., 2005 Nat Med 11: 1104-1108). Hyperglycemia-caused overactivation of CDK5 may contribute to beta-cell glucotoxicity (Ubada M et al., 2006 J Biol Chem 281: 28858-28864).

35 Hormones, growth factors

The growth factor EGF has been shown to increase beta-cell mass in human islets in vitro and in vivo (Suarez-Pinzon WL et al., 2005 Diabetes 54: 2596-2601) and protect against oxidative stress (Maeda H et al., 2004 Transplant Proc 36: 1163-1165). FGF1 stimulates beta-cell differentiation (Oberberg-Welsh C and Welsh M 1996 Pancreas 12: 334-339) while its experimental

attenuation has been shown to induce diabetes (Hart AW et al., 2000 Nature 408: 864-868).

The hypomethylated CpG detected in the insulin-like growth factor 2 (IGF2) locus, which is situated in a large open chromatin domain specific for pancreatic islets (Mutskov V, and Felsenfeld G 2009 Proc Natl Acad Sci U S A 106: 17419-17424), is located inside a CpG island between exons 3 and 4, which translates to a position inside IGF2AS according to the Infinium array probe annotation (Figure 3D). Interestingly, this hypomethylated CpG island is also present in the mouse homolog juxtaposed to mouse Igf2 DMR1. We speculate that this differential DNA methylation may influence the regulation of one or several isoforms of IGF2. Moreover, as IGF2AS has been described as a paternally imprinted antisense transcript of IGF2, there is a possibility that the differential DNA methylation could lead to altered IGF2AS expression thereby dysregulating IGF2.

Transcription factors

FOXA2 (HNF3B) is part of a transcription factor network regulating beta-cell differentiation (Lee CS et al., 2002 Diabetes 51: 2546-2551; Sund NJ et al., 2001 Genes Dev 15: 1706-1715).

PAX4 promotes beta-cell proliferation and is anti-apoptotic in human islets (Brun T et al., 2004 J Cell Biol 167: 1123-1135), while SOX6 inhibits proliferation (Iguchi H et al., 2007 J Biol Chem 282: 19052-19061); both transcription factors can affect beta-cell function (Iguchi H et al., 2005 J Biol Chem 280: 37669-37680)

SIRT6 is a H3K9 and H3K56 histone deacetylase (HDAC) that is induced upon nutrient deprivation (Kanfi Y et al., 2008 FEBS Lett 582: 543-548). It is implicated in attenuating of NF-kappa-B signalling (Kawahara TL et al., 2009 Cell 136: 62-74) (cf. CASP10) as well as telomere chromatin regulation (Michishita E et al., 2008 Nature 452: 492-496) and might, via both mechanisms, influence beta-cell lifespan.

Metabolism

Among the few differentially methylated metabolism-related genes were HK1 and FBP2, both of which show elevated expression in diabetic islets and negatively affect glucose-induced insulin secretion (Kebede M et al., 2008 Diabetes 57: 1887-1895; Malmgren S et al., 2009 J Biol Chem 284: 32395-32404).

CDO1 catalyses the first step in the major cysteine catabolism pathway (Stipanuk MH et al., 2006 J Nutr 136: 1652S-1659S). Its hypermethylation could lead to an elevated intracellular cysteine concentration which inhibits insulin secretion (Kaneko Y et al., 2006 Diabetes 55: 1391-1397) but might promote glutathione synthesis (Williamson JM et al., 1982 Proc Natl Acad Sci U S A 79: 6246-6249) (cf. GSTP1) thus protecting cells from oxidative stress. However, inhibition of CDO1 pathway-facilitated taurine synthesis may counterbalance that effect (Anuradha et al., 1999 Can J Physiol Pharmacol 77: 749-754).

One interesting example is *CASP10*: significant hypomethylation was found in its promoter (*cf.* Figure 2B) and since caspase 10 is inducible by advanced glycation endproducts (Lecomte M et al., 2004 Biochim Biophys Acta 1689: 202-211; Obrenovich ME and Monnier VM 2005 Sci Aging Knowledge Environ 2005: pe3), this hypomethylation may be indicative of gene activation caused by elevated blood glucose levels that result in heightened non-enzymatic glycosylation events. Furthermore, at least two *CASP10* isoforms have been demonstrated to be able to activate NF- κ B (Wang H et al., 2007 Biochim Biophys Acta 1770: 1528-1537). This could explain the number of NF- κ B targets in the present dataset (*cf.* Table 4) and the emergence of a NF- κ B-centred network from the analysis of said dataset with the Ingenuity Pathway Analysis software (data not shown). Another compelling example with potentially far-reaching implications for the islets in T2D is the observed differential methylation inside a CGI of the *IGF2* locus (*cf.* Figure 1C). The localisation might bear importance as it corresponds to a locus control region (DMR1, differentially methylated region 1) of the mouse homolog *m/gf2* (Constancia M et al., 2000 Nat Genet 26: 203-206). It remains presently unclear whether the hypomethylation indicates partial imprinting failure (which would probably lead to a gene-dosage-like effect, *i.e.* affect all *IGF2* isoforms) or whether DNA hypomethylation in this region will consequentially cause changes in active *IGF2* isoform composition.

Collectively, the above analyses looking at biological pathways and affected genes that are associated with differential DNA methylation in diabetic islets provides novel insights into the molecular and cell-biological pathogenesis of T2D.

Table 2. Classification of differentially methylated CpG positions according to their position relative to CpG islands (CGI).

CpG class	changed methylation	hypomethylated CpGs	hypermethylated CpGs	total on Infinium array
CGI	20	19	1	11581
CGI shore 1-2000bp away from CGI	75	68	7	8320
other CpG >2kb away from CGI	181	179	2	7677
sum	276	266	10	27578
promoter class	number of genes	hypomethylated promoters	hypermethylated promoters	total on Infinium array
LCP	38	38	0	1188

ICP	141	137	4	4038
HCP	70	64	6	9196
not categorised	5	5	0	54
sum	254	266	10	14477

CGI shores were considered 1-2000bp from CGIs; CpGs located further away from a CGI were designated 'other CpGs'. The genes to which the differentially methylated CpGs are affiliated were classified according to their promoter CpG content class (*cf.* Material and Methods). For graphical representations see Figure 8.

Table 3. PCR primers used for pre-amplification, conventional bisulfite sequencing and bisulfite pyrosequencing. For PCR conditions see Material and Methods.

Code	Description	SEQ ID No.
EF	forward primer for pre-amplification	
ER	reverse primer for pre-amplification	
IF	forward primer for conventional BS sequencing	
IR	reverse primer for conventional BS sequencing	
F	forward primer for BPS PCR	
R	reverse primer for BPS PCR	
S	sequencing primer (BPS)	
Bio	primer is biotinylated to facilitate PCR product purification prior to BPS	
primer name	sequence (5' - 3')	
ADCY7_EF	TTGTTGGTTATGGAGTAGTAGGTTTAGAG	277
ADCY7_ER	ATCATACCACTATACTCCAACCTAAATAAC	278
ADCY7_F1	GAGGGTGTTGGTAGATAGATG	279
ADCY7_R1Bio	[Btn]ACCCATACTAACTAATAAACTAATACTTC	280
ADCY7_S1	GTTAAGGGGAGTTATTTTTT	281
ADCY7_S2	TTTAGGTTTTGGGGTT	282
ADCY7_newS1	GGAGGGTTTTGATTAAGA	283
ALDH3B1_EF	TTTAGGTTGTTTTTTAGTTYGGAGTTTAG	284
ALDH3B1_ER	ACCTCRCTACCACTCCCCAAATC	285

Code	Description	SEQ ID No.
ALDH3B1_IF	GGAGTTTAGTTTTGTTGTGTGGTTTTAG	286
ALDH3B1_IR	CTACCACTCCCCAAATCTACCTCTC	287
ALDH3B1_F1	TTGGTAGAGTTTTAGGTAGAGTATTT	288
ALDH3B1_R1-Bio	[Btn]AACCAAAACACAATAACCTCTAAATACA	289
ALDH3B1_S1	GGTTAGAGTTTAGTTTAGTAATTT	290
ALDH3B1_F2	GTGGTTTGGGAGGAGATG	291
ALDH3B1_R2-Bio	[Btn]ACTACCCTTCCTCCTAACTATC	292
ALDH3B1_S2	TTTGGGAGGAGATGT	293
CASP10_EF	GGTTAAGGAGGGTGGATTATAAGTTTAGG	294
CASP10_ER	TCTACCTTTTTCCCTCCCCTTTTCC	295
CASP10_F1	GGTTGAGGTAGGAGAATTATTTGAATAT	296
CASP10_F2	AAGGAGGGTGGATTATAAGTTTAG	297
CASP10_R1Bio	[Btn]ATTAACCCTTTCTTATATCCACATACAAAT	298
CASP10_S1	ATAGGAGGTAGAGGTT	299
CASP10_S2	GTTTTTATTAAAAATATAAAAAATT	300
CASP10_2newS	ATTA AAAAATATAAAAAATTAGT	301
CDK5R1_EF	GAGTGGGAATTTAGAGGTTATATTTGTG	302
CDK5R1_ER	CCCAAACACRTACTACCTTTATTCCC	303
CDK5R1_F1	AGTGGGAATTTAGAGGTTATATTTGT	304
CDK5R1_R1Bio	[Btn]ACCTCCCTTCCCTCATCATAAATAAATAAC	305
CDK5R1_S1	GGGTTTAGGTTTGGT	306
CDK5R1_S2	TTGGTTTGGATTTTGTAG	307
CDO1_EF	AGTTTTGGATTTATTTTTATTTAGTTTGG	308
CDO1_ER	CAAATTCAAATCTATAAAATTCATCCTCC	309
CDO1_F1	GGATTTATTTTTATTTAGTTTGGGGTAT	310
CDO1_F2	ATTTTTTGGTTTAGGAGTGGAATTTA	311
CDO1_R1Bio	[Btn]AATCTCTCCCCCACTTTAAC	312
CDO1_S1	ATTTTTATTTAGTTTGGGGTATAT	313
CDO1_S2	GGAGTGGAATTTATTTTAAATTT	314
CDO1_S3	GTTAGTTTTAGTAGTTATTTTTT	315
CDO1_S4	TGGAGAGGGGAGAGG	316
FCN2_EF	TGAAGTAAAGATTAGAAGAGATGGAGTTGG	317
FCN2_ER	TACTTATCCCCACCTCACACCCTATCTC	318
FCN2_F1	GGAAATTTTTAGTTTTTGAAAGTTGGTAGT	319

Code	Description	SEQ ID No.
FCN2_R1Bio	[Btn]AAAAAACTCCCTCTCTAAAATACCACATCC	320
FCN2_S1	GGTTTTTAGGTTTTGTATTAGG	321
GABRB3_EF	GGAAAAAAAATGAGTTAATATAGGAAAGTAG	322
GABRB3_ER	ACAACACTCCTTAACTACACCTCTTACC	323
GABRB3_F1	TGTGTATTGGTATATTAGGGTTTTTGTA	324
GABRB3_R1Bio	[Btn]CCAACCAACTTAAAACTAAATTATCAT	325
GABRB3_S1	ATTAGGGTTTTGTATTAGTG	326
GABRB3_S2	GGAAGTAAGGATTTTTGTTTATA	327
GCK, for conventional pyrosequencing:		
GCK_EF652	GGTTTTAGGGGTTTGTGTTTTGAGTTA	328
GCK_ER1324	TCACAATTCCTCCTTTTCATTAAA	329
GCK_IF660	GGGTTTGTGTTTTGAGTTAYGTTAAGTTG	330
GCK_IR1321	CAATTCCTCCTTTTCATTATTCTCC	331
GCK, for bisulfite pyrosequencing:		
GCK_EF	TGAGATATTGTTTTAGGATTTGAATAGGTGG	332
GCK_ER	CCTAAATCCCTAAAAAATATAATTCTCC	333
GCK_F1	GTTGTTTTTAGGTTATAGAAGGGAGAGG	334
GCK_R1Bio	ATCACAATTCCTCCTTTTCATTATTC	335
GCK_S1	ATAGTTTAATATAATTAGGAGAGA	336
GCK_F2	GTTATTATGGTGATGGGGATGGAG	337
GCK_R2Bio	CTCTCACATCCTAACCTACTTC	338
GCK_S2	GGTTGGAGTAGGAAATG	339
GLP2R_EF	GAAAGTATAGTTGATTTAGGGAAGGTTTG	340
GLP2R_ER	CTTACTTACTTAATAAAAACCAACAAAACC	341
GLP2R_F1Bio	[Btn]AGGGTAGAGAAGGAATTTGAAGATTT	342
GLP2R_R1	ACCTCCTCTTACATTCCTCTTAATC	343
GLP2R_S1	ACACAAATCCTCTCCA	344
GP1BB_737EF	GTTTTGTTTTGGTGATAGGAGAATAA	345
GP1BB_1349ER	CCAACAACAAAACAATAAACTCAAC	346
GP1BB_IF	TTGGTGATAGGAGAATAATGTTGGTG	347
GP1BB_IR	CAAAAACAATAAACTCAACRCCC	348
GP1BB_F1	GGTATTAGGGGTTGGATGGA	349
GP1BB_R1-Bio	[Btn]CCAACAACAAAACAATAAACTCAAC	350
GP1BB_S1	GTTTGGGTATTTAGAGATG	351

Code	Description	SEQ ID No.
IGF2AS_EF	AATTTTGTTTTYGTTCCTTGGGGTT	352
IGF2AS_ER	CACAAATCCCCTTAAACCACTACC	353
IGF2AS_F1	GGGTGTAAGGAAGAAATTTAAGG	354
IGF2AS_F2	GGGTGTAAGGAAGAAATTTAAGG	355
IGF2AS_R1Bio	[Btn]ACCAAACCCAAAACCTACCC	356
IGF2AS_S1	AAGGAAGAAATTTAAGGG	357
IGF2AS_S2	GGGAGGTAGTAGGTTTTTT	358
LINE1-EF	ATTTTATATTTGGTTTAGAGGG	359
LINE1-ER	ATCAAAAATCAAAAACCCACTT	360
LINE-IF	TTTTATATTTGGTTTAGAGGG	361
LINE-IR-Bio	[Btn]TCAAAAATCAAAAACCCACTT	362
LRR15_EF	GTTTGGGGTAGTTTAGGTTTTGTTGAGG	363
LRR15_ER	CAAAACACATTTTCCTCTTTCCTTCTTC	364
LRR15_F1	TGTGGGTTGTAATGTAGAGG	365
LRR15_R1Bio	[Btn]ATACTAAATCTTACCATTACTACTC	366
LRR15_S1	AGGGGTATAGGGAAG	367
MADD_EF	GATGGTTGAGGTTTGGTGGTTGT	368
MADD_ER	AATCCAACACTTCCTCCTCCACC	369
MADD_F1	GGATGGGTTTGGGGGTTA	370
MADD_F2	AATTTTGGATTGGGGGTGG	371
MADD_R1Bio	[Btn]CTATTCAAACAATACCAAACCTATCAATAA	372
MADD_R2Bio	[Btn]ATTATACTCTCCCTCCTCTAAAC	373
MADD_S1	ATTTTGTGGTATGGTTATAAT	374
MADD_S2	ATTGGGGGTGGGGAT	375
NIBAN_EF	TTTAAGTATTATTTGGGAGTTTGTTAGAAAT	376
NIBAN_ER	CCCTCTACCCAAATAAATCTTATTC	377
NIBAN_F1	GGAAAGAGAAAAAAATGGAAAGATAGG	378
NIBAN_F2	GTTTTTGGGGAGTTTTTTTAGGTT	379
NIBAN_R1Bio	[Btn]CCCTCTACCCAAATAAATCTTATTC	380
NIBAN_R2Bio	[Btn]ATCCACTTCCTCCAAAATATTTCTTA	381
NIBAN_S1	TTTGAAATTTTTTAAGATTTGATGA	382
NIBAN_S2	GTTTGTAGTTAGTTGGTGAGTT	383

Code	Description	SEQ ID No.
NR4A1_EF	TTTGGGTATATAGTAGTTTTGGTGGGTAG	384
NR4A1_ER	ACCCAACCTCCCATTCCTCAATATCC	385
NR4A1_F1	GGTTGTTAATAGGGGTTTTATGAGTGTT	386
NR4A1_R1Bio	[Btn]CTTTAAACCCTCTTAACTAAACCTAACTT	387
NR4A1_S1	ATGAGTGTTAGAGTTGT	388
NR4A1_S2	GTTTGTTAGGTTTGGG	389
distNR4A1_F1	GGTTTAGTTAAGAGGGTTTAAAGTGG	390
distNR4A1_R1Bio	[Btn]ATCCCCAAATTAATTAATAAACTCTTCCTA	391
distNR4A1_S1	GGTTTGGAGGTAGTATTATA	392
distNR4A1_altS1	ATAGGTTGGTTGGGT	393
distNR4A1_S2	GGTTTTTTTTATTTTTAGAGGT	394
PP2A_EF	GATGGTTTTGGAGGATGATTTAGAGg	395
PP2A_ER	TCAATTAACACCCCCACCCCTTAAAC	396
PP2A_F1	AGGGTGTTGTAATATTTAAAGAGTT	397
PP2A_R1Bio	[Btn]CTAACTATATCTACCCCTCAATCCT	398
PP2A_S1	GTTTGAGTTTGGATGTATTTT	399
PP2A_S2	GTAATATATGTTTGGATTAGGGA	400
PP2Adist_F1	GGAGATTATGGGGTTGTTATTTGT	401
PP2Adist_R1Bio	[Btn]AACTCACTCTCCCAATTATTCT	402
PP2Adist_S1	AGGTTTTTTTAGTTTGATGT	403
PP2Adist_S2	GGAGTGTGATTAATTGGAT	404
PP2Adist_S3	AGTGGTTTTGGTTTAAG	405
RUNX3_EF	GAGGAGGTTTTAGTGTTATAGTTTAGGGTT	406
RUNX3_ER	AACCCTTAACTTTACAACCACTACTATTTTTCT	407
RUNX3_newF	GAGTTTTTTAGGGATTTTAAGTAGTTTGG	408
RUNX3_R1Bio	[Btn]ACCCTTAACTTTACAACCACTACTATTT	409
RUNX3_S1	TGGAGGTTTGTGGTTTTTTGA	410
RUNX3_S2	GTTTTAGGATTTTGTAGGT	411
SIRT6_EF2	GGAGGTGGGAATAAATATATTTGGAG	412
SIRT6_ER	AACAAATACCTATAATCTCATCTACTCTACTCA	413
SIRT6_F1	GAGGTGAAGATGGTTTTATTTTATAAGG	414
SIRT6_F2	GAGGTGGGAATAAATATATTTGGAGA	415
SIRT6_R1Bio	[Btn]ACAAAAAACTCCATCTCAAAAATATAAT	416
SIRT6_R2Bio	[Btn]ACTCCCCAACCTAACATTTAA	417
SIRT6_S1	ATTTTGATTTATGTATTTAATGAG	418

Code	Description	SEQ ID No.
SIRT6_S2	GTGGGAATAAATATATTTGGAGAA	419
TPM3_EF	TGAAGTTTTGGGATAGTTTTTAAGGATAGG	420
TPM3_ER	CCTAATATAATCTTTCCAACCTCCCTCCACC	421
TPM3_F1	GGGAGAATTGTTAAGGATTATGAGT	422
TPM3_R1Bio	[Btn]CAATTTACATCCAAAAATCCTAATTACC	423
TPM3_S1	ATTAGTATTTAATATAGTTTGGGG	424
TPM3_S2	ATATTAGGAATTAAAAGTTAAAT	425

Table 4: 276 CpGs showing significant differential DNA methylation between non-diabetic (CTL) and T2D pancreatic islet DNA. Filtering criteria were group-wise methylation difference of >15% and a Mann-Whitney test p value <0.01 between CTL and T2D samples. The table lists probe ID on the Infinium Human Methylation27 array, the corresponding gene symbol and the averaged methylation of the CTL and T2D groups (columns B, C) expressed as average 'beta' values representing methylation percentage ÷ 100. Column D displays the Mann-Whitney DiffScore that can be converted to a p value by the formula given below; negative values indicate hypomethylation while positive numbers indicate hypermethylation. The difference of DNA methylation ('delta beta') is given in the next column (E) followed by annotation (as provided by Illumina Inc.) of the position of the respective CpG site (columns F to AE). Additional annotation (*cf.* Material and Methods) of CpG position relative to CpG islands and promoter class of affiliated gene locus is given in columns AF to AN.

$$p = 10^{-\frac{|DiffScore|}{10}}$$

(equation to derive p value from DiffScore)

Table 4A

Marker/ SEQ ID No.	TargetID	CT.AVG	T2D.AVG	T2D.DiffScore	T2D.Delta	SYMBOL
1	cg07359545	0,50844	0,20006	-33,393	-0,30839	GP1BB
2	cg24147596	0,73354	0,44351	-30,382	-0,29003	ARL14
3	cg00668685	0,61704	0,33155	-33,393	-0,28549	SLC39A5
4	cg24315815	0,37857	0,10068	-22,601	-0,27789	PLSCR4
5	cg07730301	0,34266	0,07697	-27,372	-0,26569	ALDH3B1
6	cg20792294	0,3798	0,11663	-27,372	-0,26317	IGF2AS
7	cg17568996	0,46368	0,2054	-27,372	-0,25828	NFAM1
8	cg16745604	0,57836	0,32286	-33,393	-0,2555	CASP10
9	cg16179125	0,3028	0,04927	-27,372	-0,25353	CTSZ
10	cg14611112	0,35456	0,10252	-22,601	-0,25204	LCN6
11	cg24512973	0,53361	0,28407	-30,382	-0,24954	MUC1
12	cg05413282	0,37317	0,12758	-24,942	-0,24559	PRSS3

13	cg03380645	0,70899	0,46709	-30,382	-0,2419	SFTPB
14	cg15916061	0,41106	0,16919	-33,393	-0,24187	SLC17A4
15	cg04106785	0,55006	0,30863	-30,382	-0,24144	CDK5R1
16	cg17267907	0,6813	0,44244	-20,605	-0,23887	DEFA1
17	cg25587233	0,6334	0,39737	-30,382	-0,23603	PPP2R4
18	cg06690548	0,73288	0,49708	-24,942	-0,23581	SLC7A11
19	cg02992647	0,3779	0,14309	-24,942	-0,23482	DDX52
20	cg10929387	0,57792	0,34389	-30,382	-0,23404	ITIH4
21	cg27635271	0,42305	0,19026	-20,605	-0,23279	SH3PX3
22	cg10737521	0,41657	0,1853	-27,372	-0,23127	KIAA0676
23	cg25182523	0,54546	0,31443	-33,393	-0,23102	NIBAN
24	cg04413148	0,64878	0,41816	-20,605	-0,23062	CTRL
25	cg11682508	0,55279	0,3238	-22,601	-0,22899	FLJ30313
26	cg26062856	0,56203	0,33509	-22,601	-0,22694	ATP10A
27	cg22224704	0,32503	0,09831	-20,605	-0,22672	GSTP1
28	cg05261299	0,68178	0,45677	-33,393	-0,22501	NR4A1
29	cg00447208	0,68948	0,46771	-33,393	-0,22177	TMC8
30	cg05342835	0,39494	0,17406	-30,382	-0,22089	SYNC1
31	cg05221264	0,72585	0,50503	-27,372	-0,22081	ELA2A
32	cg03112869	0,58591	0,36775	-22,601	-0,21815	FBXW12
33	cg18575221	0,55587	0,33778	-33,393	-0,21809	C14orf162
34	cg20777437	0,39448	0,17644	-20,605	-0,21805	CDCP2
35	cg06352750	0,33574	0,11775	-22,601	-0,21799	SDPR
36	cg02092466	0,45725	0,23987	-33,393	-0,21738	NTSR2
37	cg15952725	0,54778	0,33047	-22,601	-0,2173	ELA2A
38	cg00845900	0,76385	0,5468	-30,382	-0,21705	CPA4
39	cg11884243	0,68865	0,47451	-30,382	-0,21414	FCN2
40	cg16787352	0,317	0,10354	-27,372	-0,21346	ANKRD9
41	cg22324153	0,64482	0,43174	-27,372	-0,21309	SLC35D2
42	cg13718960	0,55576	0,34328	-30,382	-0,21248	RNASE1
43	cg07125166	0,65444	0,44212	-22,601	-0,21232	FBXW12
44	cg14238120	0,45349	0,24173	-24,942	-0,21176	ELA3A
45	cg00895324	0,66636	0,4557	-33,393	-0,21066	PCP4
46	cg20585500	0,69486	0,48427	-27,372	-0,21059	GPHA2
47	cg00931491	0,3316	0,12308	-22,601	-0,20852	SULT1A2
48	cg23732182	0,47595	0,268	-22,601	-0,20795	C21orf84
49	cg17890764	0,38584	0,17822	-27,372	-0,20762	ITIH4
50	cg12322132	0,27389	0,06684	-22,601	-0,20704	IGF2AS
51	cg12473775	0,4092	0,20221	-24,942	-0,20698	RHOD
52	cg17100200	0,82506	0,61903	-30,382	-0,20602	GUCA2B
53	cg26417554	0,51972	0,31379	-33,393	-0,20593	RPUSD3
54	cg24818418	0,66515	0,45992	-24,942	-0,20522	EGF
55	cg26728422	0,48429	0,27918	-27,372	-0,20511	C16orf28
56	cg24691461	0,344	0,13898	-24,942	-0,20503	C20orf160
57	cg06003187	0,68053	0,47565	-33,393	-0,20488	GUCA2A
58	cg14975238	0,76419	0,5601	-33,393	-0,20409	BHLHB8
59	cg11504740	0,57628	0,37394	-33,393	-0,20234	GPR152
60	cg20870362	0,69413	0,49189	-33,393	-0,20224	CCIN
61	cg10347418	0,57328	0,37135	-24,942	-0,20193	ZNF436
62	cg10779183	0,53323	0,33219	-20,605	-0,20104	ELA3A
63	cg18674980	0,52072	0,31974	-33,393	-0,20098	CA3
64	cg15937958	0,66304	0,46211	-30,382	-0,20093	UNQ473
65	cg09936839	0,5613	0,3604	-27,372	-0,20089	SIRT6
66	cg01617750	0,56394	0,36308	-33,393	-0,20086	CMTM8
67	cg05958352	0,53619	0,33591	-27,372	-0,20028	RNASE1

WO 2012/097903

PCT/EP2011/070796

68	cg05722906	0,38694	0,18718	-33,393	-0,19976	CYP4F12
69	cg19391527	0,51673	0,31825	-22,601	-0,19847	PNOC
70	cg05985767	0,55194	0,3535	-22,601	-0,19844	ANPEP
71	cg03533058	0,72106	0,52315	-24,942	-0,19791	NR4A1
72	cg25994725	0,51807	0,3202	-22,601	-0,19787	C6orf81
73	cg14209518	0,62442	0,42683	-33,393	-0,19759	NNMT
74	cg22213042	0,5146	0,31729	-24,942	-0,19731	CPA2
75	cg07611925	0,59727	0,4012	-30,382	-0,19607	KCNE2
76	cg20226764	0,54159	0,34596	-27,372	-0,19563	MLN
77	cg10516359	0,6921	0,49654	-27,372	-0,19556	SLC35C1
78	cg16003913	0,44194	0,24672	-30,382	-0,19522	MPG
79	cg11566244	0,3693	0,17421	-22,601	-0,19509	GSTP1
80	cg20356482	0,68667	0,49202	-22,601	-0,19465	FBP2
81	cg27465569	0,62656	0,43268	-22,601	-0,19388	PUS7L
82	cg11161417	0,55474	0,36125	-20,605	-0,19349	SPACA3
83	cg09467501	0,28048	0,08724	-22,601	-0,19324	PYY
84	cg25406518	0,60539	0,41235	-22,601	-0,19305	DAK
85	cg25021182	0,67542	0,48257	-24,942	-0,19285	ELA2B
86	cg19826026	0,59605	0,40325	-20,605	-0,1928	ARHGDIB
87	cg04749372	0,63931	0,44678	-33,393	-0,19253	INMT
88	cg06812844	0,62316	0,43069	-24,942	-0,19246	TRPM2
89	cg08300860	0,83084	0,63852	-30,382	-0,19232	LDB3
90	cg19101893	0,52598	0,33411	-27,372	-0,19186	CCDC23
91	cg24101578	0,59174	0,40051	-22,601	-0,19123	CDH22
92	cg15940569	0,43065	0,23981	-33,393	-0,19084	GABRB3
93	cg04577715	0,70734	0,5166	-30,382	-0,19074	SPINK1
94	cg24901042	0,28693	0,09625	-30,382	-0,19068	TMPRSS2
95	cg12379145	0,70335	0,51285	-24,942	-0,19051	PRDM2
96	cg13424229	0,64759	0,45728	-24,942	-0,19031	CPA3
97	cg16863382	0,70664	0,51714	-24,942	-0,1895	CTRB1
98	cg17941312	0,78193	0,59244	-30,382	-0,18949	PSF1
99	cg19058765	0,56107	0,3722	-24,942	-0,18887	APOH
100	cg18669381	0,28368	0,095	-24,942	-0,18868	ARHGEF19
101	cg10287137	0,55443	0,36597	-30,382	-0,18846	P2RY2
102	cg06711560	0,65068	0,46235	-22,601	-0,18833	MAMDC4
103	cg12456510	0,63271	0,4445	-27,372	-0,18821	TFF2
104	cg11310496	0,70759	0,51992	-22,601	-0,18767	PNLIPRP2
105	cg12205230	0,41466	0,22753	-33,393	-0,18713	TPM3
106	cg05659947	0,38449	0,19785	-30,382	-0,18664	C4BPB
107	cg21330703	0,29139	0,10525	-30,382	-0,18614	DRD2
108	cg03668539	0,63595	0,45134	-27,372	-0,18461	PEX11G
109	cg13192155	0,64721	0,46264	-27,372	-0,18457	ERAS
110	cg14396117	0,44127	0,25705	-22,601	-0,18422	MYR8
111	cg27069753	0,57448	0,39056	-30,382	-0,18392	ELA3B
112	cg27442349	0,70032	0,5165	-33,393	-0,18382	NFKBIB
113	cg16245261	0,54923	0,36543	-33,393	-0,1838	PRKCDBP
114	cg15322932	0,25714	0,07434	-22,601	-0,1828	ALDH3B1
115	cg21846903	0,33219	0,15069	-22,601	-0,1815	VTN
116	cg16173067	0,47713	0,29612	-24,942	-0,18102	SDCBP2
117	cg05473677	0,31299	0,13203	-22,601	-0,18096	OSTalpha
118	cg25839766	0,66351	0,48283	-33,393	-0,18068	PSG1
119	cg05546044	0,68086	0,50022	-27,372	-0,18065	MAPK1
120	cg17964955	0,63541	0,45504	-33,393	-0,18037	RBP3
121	cg12061236	0,53357	0,35351	-20,605	-0,18007	AKAP12
122	cg15241708	0,61441	0,43482	-22,601	-0,1796	CNOT6

WO 2012/097903

PCT/EP2011/070796

123	cg17795586	0,68094	0,50162	-22,601	-0,17932	P2RY13
124	cg13316191	0,21331	0,035	-33,393	-0,17831	CDCA7L
125	cg23032612	0,55341	0,37514	-33,393	-0,17827	POSTN
126	cg26934034	0,72113	0,54333	-20,605	-0,1778	SDSL
127	cg14430151	0,73371	0,5561	-22,601	-0,17761	FLJ35725
128	cg15201291	0,62099	0,4436	-22,601	-0,17739	CYP2C8
129	cg09837943	0,62908	0,45203	-33,393	-0,17705	ZIM2
130	cg14992108	0,6733	0,49628	-30,382	-0,17702	SNTB1
131	cg04958389	0,64846	0,47169	-24,942	-0,17676	PRSS2
132	cg26750002	0,49005	0,31332	-24,942	-0,17673	RASIP1
133	cg23106559	0,59618	0,41964	-30,382	-0,17654	METTLL1
134	cg04929865	0,58822	0,412	-27,372	-0,17621	BGN
135	cg21030400	0,55299	0,37687	-20,605	-0,17612	MKNK2
136	cg26164184	0,70699	0,5309	-30,382	-0,1761	FCN2
137	cg14659547	0,55295	0,377	-20,605	-0,17595	RETNLB
138	cg11081833	0,36515	0,18923	-20,605	-0,17592	LGALS2
139	cg24821554	0,71684	0,5412	-33,393	-0,17564	GUCY1B2
140	cg10883352	0,65332	0,47807	-27,372	-0,17525	
141	cg14930674	0,70229	0,52707	-24,942	-0,17521	SLC12A1
142	cg25787984	0,67256	0,49737	-22,601	-0,17518	DKK3
143	cg22215192	0,52312	0,34802	-27,372	-0,1751	AMY2A
144	cg04584523	0,79824	0,62326	-27,372	-0,17498	ELA3B
145	cg26848126	0,48696	0,312	-33,393	-0,17496	CYSLTR1
146	cg00078867	0,72867	0,55399	-33,393	-0,17468	WAS
147	cg10925082	0,59551	0,42107	-22,601	-0,17444	ARHGDIB
148	cg23815000	0,63681	0,46243	-22,601	-0,17438	LCN1
149	cg21307628	0,49794	0,32367	-20,605	-0,17427	URB
150	cg24939733	0,54862	0,37458	-22,601	-0,17405	FLJ38159
151	cg05186188	0,50671	0,33272	-20,605	-0,17398	AQP8
152	cg21880328	0,35241	0,17878	-24,942	-0,17363	CTTNBP2
153	cg15746445	0,56673	0,39314	-30,382	-0,17359	TRPM8
154	cg19103609	0,65512	0,48159	-22,601	-0,17353	PKN1
155	cg14324200	0,2402	0,06693	-20,605	-0,17327	SAMD11
156	cg15662251	0,46529	0,29211	-20,605	-0,17318	PAQR7
157	cg16334519	0,62956	0,4565	-27,372	-0,17306	ARL14
158	cg25463779	0,66996	0,4969	-33,393	-0,17306	FAM101A
159	cg16272420	0,53203	0,3592	-22,601	-0,17283	PNLIPRP2
160	cg10321196	0,48322	0,31048	-20,605	-0,17274	SLC25A5
161	cg15746187	0,5458	0,37322	-20,605	-0,17258	FBXO44
162	cg04802221	0,67159	0,49927	-33,393	-0,17232	LOC283849
163	cg12940073	0,66151	0,48992	-24,942	-0,1716	ABCC12
164	cg14672994	0,25101	0,07968	-24,942	-0,17132	FLJ20920
165	cg14401837	0,57374	0,40259	-20,605	-0,17115	GPR154
166	cg12113819	0,38805	0,2172	-30,382	-0,17086	THRAP2
167	cg21686987	0,68513	0,51457	-27,372	-0,17056	CTRB1
168	cg23082877	0,5774	0,40687	-33,393	-0,17054	RASIP1
169	cg04049033	0,51186	0,34257	-24,942	-0,1693	RILP
170	cg03770548	0,6871	0,51814	-22,601	-0,16895	CDH24
171	cg00403724	0,44007	0,27119	-27,372	-0,16888	TNMD
172	cg15479752	0,73716	0,56894	-33,393	-0,16822	FFAR2
173	cg16796951	0,71167	0,54357	-22,601	-0,1681	C20orf152
174	cg05441133	0,64997	0,48194	-20,605	-0,16804	GDF2
175	cg22340747	0,54182	0,37387	-22,601	-0,16795	GATM
176	cg03245641	0,53117	0,36345	-27,372	-0,16772	GPHA2
177	cg08816023	0,69959	0,53192	-24,942	-0,16767	FGF1

WO 2012/097903

PCT/EP2011/070796

178	cg05828624	0,63651	0,46887	-24,942	-0,16764	REG1A
179	cg16101800	0,65212	0,4846	-27,372	-0,16751	PCK1
180	cg12958813	0,6332	0,46582	-22,601	-0,16738	ATP6V1G3
181	cg04086834	0,50839	0,34123	-33,393	-0,16716	SMAP1
182	cg16545105	0,77694	0,61005	-27,372	-0,16689	CRHBP
183	cg15154229	0,47631	0,30967	-24,942	-0,16664	CPA1
184	cg25809905	0,75705	0,5906	-33,393	-0,16645	ITGA2B
185	cg05379350	0,48283	0,31683	-24,942	-0,166	GIT1
186	cg02131995	0,5007	0,33481	-24,942	-0,16588	HAMP
187	cg01515887	0,67205	0,50666	-30,382	-0,16539	SAA2
188	cg12932195	0,69631	0,53125	-20,605	-0,16506	ALG12
189	cg21660392	0,74294	0,57795	-22,601	-0,16499	ABCA8
190	cg23320649	0,32424	0,15933	-22,601	-0,16491	C3orf18
191	cg11812202	0,58843	0,42406	-24,942	-0,16437	PNLIP
192	cg23276695	0,49954	0,33541	-22,601	-0,16414	CNR1
193	cg26066361	0,5455	0,38154	-30,382	-0,16395	CLEC7A
194	cg01197831	0,74458	0,58071	-27,372	-0,16387	FBP2
195	cg14123992	0,46171	0,29784	-27,372	-0,16387	APOE
196	cg21049762	0,71941	0,55553	-27,372	-0,16387	TCIRG1
197	cg07572435	0,73412	0,57048	-27,372	-0,16365	LY6D
198	cg03156547	0,50299	0,33944	-22,601	-0,16355	CDH24
199	cg27160701	0,61567	0,45215	-27,372	-0,16352	SBEM
200	cg02763671	0,59935	0,436	-24,942	-0,16335	RANBP1
201	cg16792160	0,5345	0,37194	-24,942	-0,16255	ASAH2
202	cg22686523	0,69283	0,53036	-20,605	-0,16246	FLJ25006
203	cg05517711	0,58187	0,41958	-22,601	-0,16229	ST6GALNAC6
204	cg00259755	0,20497	0,04299	-27,372	-0,16198	PWWP2
205	cg14481339	0,66172	0,50038	-22,601	-0,16134	KCNJ1
206	cg16752583	0,36106	0,2002	-22,601	-0,16087	TRPV6
207	cg20011134	0,38458	0,22426	-22,601	-0,16032	DDO
208	cg09676390	0,37676	0,21652	-22,601	-0,16024	ADARB1
209	cg02497428	0,69676	0,53671	-20,605	-0,16005	IGSF6
210	cg22586527	0,5303	0,37027	-27,372	-0,16004	SOX6
211	cg24607535	0,75289	0,59303	-22,601	-0,15986	CDH26
212	cg05744229	0,52241	0,36268	-22,601	-0,15973	MYH7
213	cg19856444	0,56152	0,40199	-27,372	-0,15953	SLC39A12
214	cg05535113	0,54181	0,38254	-22,601	-0,15926	CHST4
215	cg01593385	0,47541	0,31639	-30,382	-0,15902	FGG
216	cg15013019	0,40922	0,25043	-30,382	-0,15879	LYL1
217	cg20392764	0,74737	0,58867	-33,393	-0,1587	ASCL2
218	cg13379236	0,50373	0,34528	-22,601	-0,15844	EGF
219	cg05663262	0,55965	0,40135	-20,605	-0,1583	HYI
220	cg23756272	0,81269	0,65444	-24,942	-0,15825	BCL2
221	cg07908874	0,58667	0,42849	-20,605	-0,15818	TUBGCP2
222	cg15770654	0,77312	0,61529	-27,372	-0,15783	IIP45
223	cg25915982	0,47213	0,31467	-27,372	-0,15746	GRB10
224	cg27655855	0,63662	0,47915	-20,605	-0,15746	CST9L
225	cg06186808	0,67558	0,51826	-20,605	-0,15732	LOC161247
226	cg04151683	0,48781	0,33057	-20,605	-0,15723	LRRN6C
227	cg02552945	0,46844	0,31151	-22,601	-0,15693	C5
228	cg09069593	0,61571	0,45921	-24,942	-0,1565	GPR114
229	cg07959477	0,56274	0,40634	-22,601	-0,15639	CLEC1A
230	cg07294734	0,3251	0,1688	-20,605	-0,1563	ATP5D
231	cg13587552	0,55762	0,40135	-22,601	-0,15628	SCNN1D
232	cg22598744	0,61064	0,4544	-22,601	-0,15623	MLANA

233	cg18034859	0,84475	0,6887	-33,393	-0,15605	MYLK2
234	cg08886154	0,60656	0,45066	-27,372	-0,1559	PAX4
235	cg17260954	0,67254	0,51666	-20,605	-0,15588	ATP10A
236	cg24363955	0,67324	0,51767	-24,942	-0,15556	FLJ14054
237	cg01260219	0,74052	0,5852	-24,942	-0,15532	ADAMTS8
238	cg02882813	0,61333	0,45804	-33,393	-0,1553	CST5
239	cg02347487	0,70923	0,55405	-27,372	-0,15519	NALP14
240	cg17886959	0,31953	0,16445	-24,942	-0,15508	MT2A
241	cg06434428	0,73339	0,57837	-27,372	-0,15502	HAPLN1
242	cg12288726	0,51054	0,35584	-24,942	-0,1547	ARF5
243	cg02121427	0,44954	0,29497	-33,393	-0,15456	LRRC15
244	cg11750883	0,59256	0,43806	-22,601	-0,1545	C1orf42
245	cg10978355	0,22295	0,06865	-33,393	-0,15431	CKMT2
246	cg12619162	0,8215	0,66737	-24,942	-0,15414	FXYD4
247	cg22194129	0,42427	0,27023	-22,601	-0,15404	CLEC4C
248	cg26624914	0,45809	0,30406	-20,605	-0,15403	AQP3
249	cg08996986	0,77349	0,61971	-22,601	-0,15377	EPS8L1
250	cg22415472	0,68373	0,53014	-24,942	-0,15359	SLC5A7
251	cg26203861	0,42393	0,2708	-33,393	-0,15313	TMEM58
252	cg05275605	0,64618	0,49312	-27,372	-0,15306	C21orf123
253	cg25107791	0,48191	0,32904	-20,605	-0,15287	CLPS
254	cg10003443	0,50097	0,34831	-33,393	-0,15266	FOXA2
255	cg03116740	0,6439	0,49125	-30,382	-0,15265	TSPAN4
256	cg04578090	0,67607	0,52378	-22,601	-0,1523	PROCA1
257	cg09440243	0,64619	0,49389	-22,601	-0,1523	PTPRD
258	cg11479877	0,32623	0,17416	-20,605	-0,15207	B3GALT5
259	cg12658552	0,63117	0,47961	-22,601	-0,15156	SFRS2IP
260	cg26711820	0,33734	0,18621	-22,601	-0,15113	MYF6
261	cg06323290	0,70643	0,55533	-20,605	-0,1511	HK1
262	cg17327492	0,57214	0,4215	-20,605	-0,15064	FSHR
263	cg26065841	0,73882	0,58819	-30,382	-0,15064	CHAC1
264	cg13944141	0,46884	0,31871	-30,382	-0,15013	PRSS2
265	cg19782598	0,69015	0,54007	-30,382	-0,15008	FLJ30046
266	cg21518947	0,69864	0,54862	-27,372	-0,15002	CABP4
267	cg16872071	0,39948	0,54964	20,605	0,15016	RALGDS
268	cg24815853	0,33013	0,48094	22,601	0,15081	KLK12
269	cg15780361	0,38725	0,55509	27,372	0,16785	ALS2CR11
270	cg07644368	0,30195	0,47843	33,393	0,17648	CDO1
271	cg10805676	0,68341	0,86024	20,605	0,17683	MRPL28
272	cg14948436	0,68126	0,86352	20,605	0,18226	KCNK16
273	cg14204735	0,33629	0,52427	20,605	0,18799	CYB561
274	cg26365553	0,31438	0,50655	20,605	0,19217	MADD
275	cg05664072	0,2089	0,4055	30,382	0,1966	PER2
276	cg04660410	0,38579	0,59548	27,372	0,20969	VILL

Table 4B

Marker/ SEQ ID No.	CHR	MAPINFO	GENE_ID	ACCESSION	SYNONYM
1	22	18091327	GeneID:2812	NM_000407.3	CD42c;
2	3	161878114	GeneID:80117	NM_025047.1	ARF7; FLJ22595;
3	12	54910819	GeneID:283375	NM_173596.1	ZIP5; LZT-Hs7; MGC34778;
4	3	147452269	GeneID:57088	NM_020353.1	
5	11	67534528	GeneID:221	NM_000694.2	ALDH4; ALDH7;
6	11	2122001	GeneID:51214	NT_009237.17	PEG8;
7	22	41158069	GeneID:150372	NM_145912.4	CNAIP; FLJ40652;
8	2	201755704	GeneID:843	NM_001230.3	MCH4; ALPS2; FLICE2;
9	20	57015516	GeneID:1522	NT_011362.9	CTSX;
10	9	138763172	GeneID:158062	NM_198946.2	LCN5; hLcn5; UNQ643;
11	1	153428408	GeneID:4582	NM_002456.4	EMA; PEM; PUM; MAM6; PEMT; CD227; H23AG; mucin;
12	9	33786088	GeneID:5646	NM_002771.2	MTG; TRY3; TRY4; PRSS4;
13	2	85748559	GeneID:6439	NM_000542.2	SP-B; PSP-B; SFTB3; SFTP3;
14	6	25862198	GeneID:10050	NM_005495.1	KAIA2138; MGC129623;
15	17	27837258	GeneID:8851	NM_003885.2	p23; p25; p35; CDK5R; NCK5A; CDK5P35; MGC33831; p35nck5a;
16	8	6825434	GeneID:1667	NM_004084.2	MRS; DEF1; HP-1; DEFA2; HNP-1;
17	9	130911826	GeneID:5524	NM_021131.3	PP2A; PR53; PTPA; MGC2184;
18	4	139382258	GeneID:23657	NM_014331.2	xCT; CCBP1;
19	17	33078559	GeneID:11056	NM_152300.2	ROK1;
20	3	52839699	GeneID:3700	NM_002218.3	H4P; IHRP; PK120; ITIHL1; DKFZp686G21125;
21	15	73728631	GeneID:257364	NM_153271.1	MGC32065; SH3PXD3C;
22	5	179267849	GeneID:23061	NM_198868.2	
23	1	183210911	GeneID:116496	NM_022083.1	C1orf24;
24	16	66523368	GeneID:1506	NM_001907.1	CTRL1; MGC70821;
25	20	60559844	GeneID:253868	NM_152757.1	
26	15	23660518	GeneID:57194	NM_024490.2	ATPVA; ATPVC; ATP10C; KIAA0566;
27	11	67108617	GeneID:2950	NM_000852.2	PI; DFN7; GST3; FAEES3;
28	12	50723566	GeneID:3164	NM_002135.3	HMR; N10; TR3; NP10; GFRP1; NAK-1; NGFIB; NUR77; MGC9485;
29	17	73637896	GeneID:147138	NM_152468.3	EV2; EVER2; EVIN2; MGC40121;
30	1	32933378	GeneID:81493	NM_030786.1	SYNC; SYNCOILIN;
31	1	15656100	GeneID:63036	NM_033440.1	ELA1; PE-1;
32	3	48388773	GeneID:285231	NM_207102.1	Fbw12; FBXO35; MGC120385; MGC120386; MGC120387;
33	14	69107481	GeneID:56936	NM_020181.1	PLPL;
34	1	54392291	GeneID:200008	NM_201546.2	
35	2	192420171	GeneID:8436	NM_004657.4	SDR; PS-p68;
36	2	11728082	GeneID:23620	NM_012344.2	NTR2;
37	1	15655777	GeneID:63036	NM_033440.1	ELA1; PE-1;
38	7	129718965	GeneID:51200	NM_016352.2	CPA3;

39	9	136912707	GenelD:2220	NM_004108.2	P35; FCNL; EBP-37; ficolin-2;
40	14	102045158	GenelD:122416	NM_152326.2	MGC21990;
41	9	98186585	GenelD:11046	NM_007001.1	hfrc; HFRC1; SQV7L; UGTrel8; MGC117215;
42	14	20341153	GenelD:6035	NM_198232.1	RIB1; RNS1; MGC12408;
43	3	48388617	GenelD:285231	NM_207102.1	Fbw12; FBXO35; MGC120385; MGC120386; MGC120387;
44	1	22200613	GenelD:10136	NM_005747.3	ELA3;
45	21	40160387	GenelD:5121	NM_006198.2	PEP-19;
46	11	64459768	GenelD:170589	NM_130769.2	GPA2; ZSIG51;
47	16	28515789	GenelD:6799	NM_001054.2	STP2; HAST4; P-PST; ST1A2; TSPST2;
48	21	43722518	GenelD:114038	NM_153752.1	MGC129789;
49	3	52839856	GenelD:3700	NM_002218.3	H4P; IHRP; PK120; ITIHL1; DKFZp686G21125;
50	11	2122232	GenelD:51214	NT_009237.17	PEG8;
51	11	66581226	GenelD:29984	NM_014578.2	Rho; ARHD; RHOM; RHOHP1;
52	1	42390900	GenelD:2981	NM_007102.1	UGN; GCAP-II;
53	3	9861232	GenelD:285367	NM_173659.2	FLJ34707; MGC29784;
54	4	111054039	GenelD:1950	NM_001963.2	URG;
55	16	1369016	GenelD:65259	NM_023076.2	FLJ23360;
56	20	30062162	GenelD:140706	NM_080625.2	dJ310O13.5;
57	1	42402385	GenelD:2980	NM_033553.2	GUCA2; STARA; GUANYLIN;
58	7	97679527	GenelD:168620	NM_177455.2	MIST1;
59	11	66976619	GenelD:390212	NM_206997.1	PGR5;
60	9	36159175	GenelD:881	NM_005893.1	
61	1	23566513	GenelD:80818	NM_030634.1	ZNF; KIAA1710;
62	1	22201054	GenelD:10136	NM_005747.3	ELA3;
63	8	86537833	GenelD:761	NM_005181.2	CAIII;
64	19	47638973	GenelD:284340	NM_198477.1	VCC1;
65	19	4132854	GenelD:51548	NM_016539.1	SIR2L6;
66	3	32254265	GenelD:152189	NM_178868.3	CKLFSF8;
67	14	20340764	GenelD:6035	NM_198232.1	RIB1; RNS1; MGC12408;
68	19	15644791	GenelD:66002	NM_023944.1	F22329_1;
69	8	28229651	GenelD:5368	NM_006228.2	PPNOC;
70	15	88150515	GenelD:290	NM_001150.1	CD13; LAP1; PEPN; gp150;
71	12	50723838	GenelD:3164	NM_002135.3	HMR; N10; TR3; NP10; GFRP1; NAK-1; NGFIB; NUR77; MGC9485;
72	6	35811413	GenelD:221481	NM_145028.3	FLJ25390;
73	11	113671846	GenelD:4837	NM_006169.2	
74	7	129693863	GenelD:1358	NM_001869.1	
75	21	34658253	GenelD:9992	NM_172201.1	LQT5; LQT6; MIRP1; MGC138292;
76	6	33879909	GenelD:4295	NM_002418.1	
77	11	45784339	GenelD:55343	NM_018389.3	FUCT1; FLJ11320; FLJ14841;
78	16	67072	GenelD:4350	NM_00101505 2.1	AAG; MDG; APNG; Mid1; anpg; PIG11; PIG16; CRA36.1;
79	11	67108362	GenelD:2950	NT_033903.7	PI; DFN7; GST3; FAES3;
80	9	96395913	GenelD:8789	NM_003837.2	
81	12	42440120	GenelD:83448	NM_031292.2	DKFZP434G1415; DKFZp781B2386;

82	17	28343567	GenelD:124912	NM_173847.2	LYC3; LYZL3; SLLP1; ALLP17; MGC119058; 1700025M08Rik;
83	17	39438719	GenelD:5697	NM_004160.3	PYY1;
84	11	60858741	GenelD:26007	NM_015533.2	DKFZP586B1621;
85	1	15675181	GenelD:51032	NM_015849.2	MGC97052;
86	12	15005660	GenelD:397	NM_001175.4	D4; GDIA2; GDID4; LYGDI; Ly-GDI; RAP1GN1;
87	7	30758396	GenelD:11185	NM_006774.4	MGC125940; MGC125941;
88	21	44598210	GenelD:7226	NM_00100118.3	KNP3; EREG1; TRPC7; LTRPC2; NUDT9H; NUDT9L1;
89	10	88418237	GenelD:11155	NM_007078.1	ZASP; CYPHER; ORACLE; PDLIM6; ldb3z4; KIAA0613; KIAA01613;
90	1	43056449	GenelD:374969	NM_199342.1	MGC45441; RP5-994D16.10;
91	20	44313670	GenelD:64405	NM_021248.1	C20orf25; MGC39564; dJ998H6.1;
92	15	24570449	GenelD:2562	NM_021912.2	MGC9051;
93	5	147191374	GenelD:6690	NM_003122.2	PCTT; PSTI; TATI; Spink3;
94	21	41802424	GenelD:7113	NM_005656.2	PRSS10;
95	1	13903198	GenelD:7799	NM_012231.3	RIZ; RIZ1; RIZ2; MTB-ZF; HUMHOXY1;
96	3	150064527	GenelD:1359	NM_001870.1	
97	16	73810196	GenelD:1504	NM_001906.3	CTRB; FLJ42412; MGC88037;
98	20	25329464	GenelD:9837	NM_021067.2	KIAA0186; RP4-691N24.2;
99	17	61656565	GenelD:350	NM_000042.1	BG; B2G1;
100	1	16411675	GenelD:128272	NM_153213.2	WGEF; FLJ33962; RP4-733M16.1;
101	11	72606702	GenelD:5029	NM_002564.2	P2U; HP2U; P2U1; P2UR; P2Y2; P2RU1; P2Y2R; MGC20088; MGC40010;
102	9	138866186	GenelD:158056	NM_206920.1	AEGP; DKFZp434M1411;
103	21	42644424	GenelD:7032	NM_005423.3	SP; SML1;
104	10	118370165	GenelD:5408	NM_005396.3	PLRP2;
105	1	152431575	GenelD:7170	NM_152263.1	TRK; NEM1; MGC72094;
106	1	205328813	GenelD:725	NM_000716.3	C4BP;
107	11	112851598	GenelD:1813	NM_016574.2	D2R; D2DR;
108	19	7459324	GenelD:92960	NM_080662.2	MGC4281;
109	X	48570887	GenelD:3266	NM_181532.2	HRAS2; HRASP; MGC126691; MGC126693;
110	13	108046048	GenelD:23026	NM_015011.1	KIAA0865;
111	1	22176158	GenelD:23436	NM_007352.2	
112	19	44081059	GenelD:4793	NM_002503.3	IKBB; TRIP9;
113	11	6298933	GenelD:112464	NM_145040.2	SRBC; HSRBC; MGC20400;
114	11	67534258	GenelD:221	NM_000694.2	ALDH4; ALDH7;
115	17	23721408	GenelD:7448	NM_000638.2	V75; VNT;
116	20	1257722	GenelD:27111	NM_015685.3	ST-2; SITAC18;
117	3	197428110	GenelD:200931	NM_152672.3	MGC39807;
118	19	48076917	GenelD:5669	NM_006905.2	SP1; B1G1; PBG1; CD66f; PSBG1; PSGGA; DHFRP2; FLJ90598; FLJ90654; PSGIIA-a; PSGIIA-b; PSGIIA-c; PSG IIA-d;
119	22	20552597	GenelD:5594	NM_002745.4	ERK; p38; p40; p41; ERK2; ERT1; MAPK2; PRKM1;

					PRKM2; P42MAPK; p41mapk;
120	10	48010942	GeneID:5949	NM_002900.2	IRBP; RBPI; D10S64; D10S65; D10S66;
121	6	151602579	GeneID:9590	NM_005100.2	AKAP250; DKFZp686M0430; DKFZp686O0331;
122	5	179853188	GeneID:57472	NM_015455.3	CCR4; KIAA1194;
123	3	152530425	GeneID:53829	NM_023914.2	GPCR1; GPR86; GPR94; P2Y13; SP174; FKSG77;
124	7	21952409	GeneID:55536	NM_018719.2	R1; RAM2; DKFZp762L0311;
125	13	37071340	GeneID:10631	NM_006475.1	PN; OSF-2; MGC119510; MGC119511; periostin; RP11-412K4.1;
126	12	112345493	GeneID:113675	NM_138432.2	SDS-RS1;
127	4	8491500	GeneID:152992	NM_152544.1	FLJ12891;
128	10	96820092	GeneID:1558	NM_030878.1	CPC8; P450 MP-12/MP-20;
129	19	61998443	GeneID:23619	NT_011109.15	
130	8	121894651	GeneID:6641	NM_021021.2	A1B; SNT2; BSYN2; 59-DAP; DAPA1B; SNT2B1; TIP-43; FLJ22442; MGC111389;
131	7	142178834	GeneID:5645	NM_002770.2	TRY2; TRY8; TRYP2; MGC111183; MGC120174;
132	19	53936486	GeneID:54922	NM_017805.2	RAIN; FLJ20401;
133	12	56453750	GeneID:4234	NM_023032.1	TRM8; C12orf1; YDL201w;
134	X	152413272	GeneID:633	NM_001711.3	PGI; DSPG1; PG-S1; SLRR1A;
135	19	2003564	GeneID:2872	NM_199054.1	MNK2; GPRK7;
136	9	136911327	GeneID:2220	NM_004108.2	P35; FCNL; EBP-37; ficolin-2;
137	3	109958816	GeneID:84666	NM_032579.1	XCP2; FIZZ1; FIZZ2; HXCP2; RELMb; RELMbeta; RELM- beta;
138	22	36305887	GeneID:3957	NM_006498.2	HL14; MGC75071;
139	13	50537954	GeneID:2974	NM_004129.2	GC-SB2;
140	3	144378065	GeneID:5088		
141	15	46287532	GeneID:6557	NM_000338.1	BSC1; NKCC2; MGC48843;
142	11	11988084	GeneID:27122	NM_015881.5	REIC;
143	1	103961780	GeneID:279	NM_000699.2	PA; AMY2; AMY2B; amylase 2A; amylase 2B;
144	1	22175799	GeneID:23436	NM_007352.2	
145	X	77469569	GeneID:10800	NM_006639.2	HG55; CYSLT1; CYSLTR; CYSLT1R; HMTMF81; MGC46139;
146	X	48427342	GeneID:7454	NM_000377.1	THC; IMD2; WASP;
147	12	15005970	GeneID:397	NM_001175.4	D4; GDIA2; GDID4; LYGDI; Ly-GDI; RAP1GN1;
148	9	137552833	GeneID:3933	NM_002297.2	TP; PMFA; VEGP; MGC71975;
149	3	113843642	GeneID:151887	NM_199511.1	DRO1; SSG1;
150	11	85075332	GeneID:220388	NM_152723.1	
151	16	25136070	GeneID:343	NM_001169.2	
152	7	117301253	GeneID:83992	NM_033427.1	Orf4; C7orf8; CORTBP2; KIAA1758; MGC104579;
153	2	234489672	GeneID:79054	NM_024080.3	TRPP8; LTRPC6; MGC2849;
154	19	14404524	GeneID:5585	NM_002741.3	DBK; PKN; PAK1; PRK1; PRKCL1; MGC46204;
155	1	851180	GeneID:148398	NM_152486.2	MGC45873;
156	1	26070442	GeneID:164091	NM_178422.4	mSR; MPRA;

157	3	161878388	GeneID:80117	NM_025047.1	ARF7; FLJ22595;
158	12	123339189	GeneID:144347	NM_181709.2	FLJ44614;
159	10	118371034	GeneID:5408	NM_005396.3	PLRP2;
160	X	118485919	GeneID:292	NM_001152.1	T2; T3; 2F1; ANT2;
161	1	11635710	GeneID:93611	NM_00101476 5.1	FBG3; FBX30; Fbx44; Fbxo6a; MGC14140; DKFZp781J0852;
162	16	65781456	GeneID:283849	NM_178516.2	MGC88052;
163	16	46738165	GeneID:94160	NM_033226.1	MRP9; MGC27071;
164	17	45858056	GeneID:80221	NM_025149.3	
165	7	34664018	GeneID:387129	NM_207173.1	GPRA; VRR1; NPSR1; PGR14;
166	12	115200537	GeneID:23389	NM_015335.2	MED13L; FLJ21627; KIAA1025; TRAP240L; PROSIT240; DKFZp781D0112;
167	16	73810474	GeneID:1504	NM_001906.3	CTRB; FLJ42412; MGC88037;
168	19	53935239	GeneID:54922	NM_017805.2	RAIN; FLJ20401;
169	17	1500411	GeneID:83547	NM_031430.1	PP10141; FLJ31193;
170	14	22594880	GeneID:64403	NM_144985.2	CDH11L; FLJ25193;
171	X	99725945	GeneID:64102	NM_022144.1	TEM; CHM1L; BRICD4; tendin; myodulin; CHM1- LIKE;
172	19	40632702	GeneID:2867	NM_005306.1	FFA2R; GPR43;
173	20	34018487	GeneID:140894	NM_080834.1	dJ954P9.1;
174	10	48036897	GeneID:2658	NM_016204.1	BMP9; BMP-9;
175	15	43470502	GeneID:2628	NM_001482.1	AGAT;
176	11	64459991	GeneID:170589	NM_130769.2	GPA2; ZSIG51;
177	5	142045723	GeneID:2246	NM_033136.1	AFGF; ECGF; FGFA; ECGFA; ECGFB; HBGF1; GLIO703; ECGF-beta; FGF- alpha;
178	2	79201207	GeneID:5967	NM_002909.3	P19; PSP; PTP; REG; ICRF; PSPS; PSPS1; MGC12447;
179	20	55569975	GeneID:5105	NM_002591.2	PEPCK1; PEPCKC; MGC22652;
180	1	196776292	GeneID:127124	NM_133262.2	Vma10; ATP6G3; MGC119810; MGC119813;
181	6	71433785	GeneID:60682	NM_021940.2	SMAP-1; FLJ13159;
182	5	76284505	GeneID:1393	NM_001882.3	CRFBP; CRF-BP;
183	7	129807247	GeneID:1357	NM_001868.1	CPA;
184	17	39823254	GeneID:3674	NM_000419.2	GTA; CD41; GP2B; HPA3; CD41B; GPIIb;
185	17	24941283	GeneID:28964	NM_014030.2	
186	19	40464757	GeneID:57817	NM_021175.2	HEPC; HFE2B; LEAP1; LEAP-1;
187	11	18226210	GeneID:6289	NM_030754.2	
188	22	48699347	GeneID:79087	NM_024105.3	ECM39; hALG12; MGC3136; PP14673; MGC111358;
189	17	64462140	GeneID:10351	NM_007168.2	KIAA0822;
190	3	50579617	GeneID:51161	NM_016210.2	G20;
191	10	118295399	GeneID:5406	NM_000936.2	PL;
192	6	88912435	GeneID:1268	NM_016083.3	CB1; CNR; CB-R; CB1A; CANN6; CB1K5;
193	12	10173888	GeneID:64581	NM_022570.3	BGR; DECTIN1; CLECSF12;
194	9	96395762	GeneID:8789	NM_003837.2	

195	19	50099708	GenelD:348	NM_000041.2	AD2; MGC1571; apoprotein;
196	11	67562032	GenelD:10312	NM_006019.2	a3; Stv1; Vph1; Atp6i; OC116; OPTB1; TIRC7; ATP6N1C; ATP6V0A3; OC- 116kDa;
197	8	143865015	GenelD:8581	NM_003695.2	E48;
198	14	22594843	GenelD:64403	NM_144985.2	CDH11L; FLJ25193;
199	12	53534691	GenelD:118430	NM_058173.1	
200	22	18483541	GenelD:5902	NM_002882.2	MGC88701;
201	10	51678599	GenelD:56624	XM_936277.1	
202	17	23965053	GenelD:124923	NM_144610.1	
203	9	129702272	GenelD:30815	NM_013443.3	SIAT7F; ST6GALNACVI; RP11-203J24.3;
204	10	134059753	GenelD:170394	NM_138499.2	pp8607; FLJ46823; bA432J24.1; RP11-273H7.1;
205	11	128242529	GenelD:3758	NM_000220.2	ROMK; ROMK1; KIR1.1;
206	7	142293853	GenelD:55503	NM_018646.2	CAT1; CATL; ZFAB; ECAC2; ABP/ZF; LP6728; HSA277909;
207	6	110843115	GenelD:8528	NM_003649.2	DASOX; DDO-1; DDO-2; FLJ45203;
208	21	45317773	GenelD:104	NM_001112.2	RED1; ADAR2; ADAR2d; ADAR2g; DRABA2; DRADA2; ADAR2a-L1; ADAR2a-L2; ADAR2a-L3;
209	16	21572639	GenelD:10261	NM_005849.1	DORA;
210	11	16455987	GenelD:55553	NM_017508.1	HSSOX6;
211	20	57966890	GenelD:60437	NM_177980.1	VR20;
212	14	22974518	GenelD:4625	NM_000257.1	CMH1; MPD1; CMD1S; MYHCB; MGC138376; MGC138378;
213	10	18281122	GenelD:221074	NM_152725.1	FLJ30499; MGC43205; MGC51099; bA570F3.1;
214	16	70117080	GenelD:10164	NM_005769.1	LSST;
215	4	155753599	GenelD:2266	NM_000509.4	
216	19	13074451	GenelD:4066	NM_005583.3	
217	11	2249693	GenelD:430	NM_005170.2	ASH2; HASH2; MASH2;
218	4	111053257	GenelD:1950	NM_001963.2	URG;
219	1	43702866	GenelD:81888	NM_031207.2	HT036; MGC20767; RP11- 506B15.5;
220	18	59055398	GenelD:596	NT_025028.13	Bcl-2;
221	10	134972996	GenelD:10844	NM_006659.1	GCP2; SPBC97; Spc97p;
222	1	12001408	GenelD:60672	NM_00102537 4.1	FLJ12438; FLJ38609;
223	7	50816909	GenelD:2887	NT_033968.5	RSS; IRBP; MEG1; GRB-IR; KIAA0207;
224	20	23497455	GenelD:128821	NM_080610.1	bA218C14.1;
225	14	23670741	GenelD:161247	NM_203402.1	MGC46490;
226	9	28708459	GenelD:158038	NM_152570.1	LERN3; LINGO2; FLJ31810;
227	9	122851435	GenelD:727	NM_001735.2	CPAMD4;
228	16	56133786	GenelD:221188	NM_153837.1	PGR27;
229	12	10143146	GenelD:51267	NM_016511.2	CLEC1; MGC34328;
230	19	1193323	GenelD:513	NM_001687.4	
231	1	1206304	GenelD:6339	NM_002978.2	ENaCd; SCNED; dNaCh; ENaCdelta;
232	9	5880866	GenelD:2315	NM_005511.1	MART1; MART-1;
233	20	29870349	GenelD:85366	NM_033118.2	KMLC; MLCK; skMLCK;

234	7	127043912	GeneID:5078	NM_006193.1	
235	15	23577440	GeneID:57194	NT_026446.13	ATPVA; ATPVC; ATP10C; KIAA0566;
236	5	32824224	GeneID:79614	NM_024563.2	
237	11	129804543	GeneID:11095	NM_007037.3	METH2; ADAM-TS8;
238	20	23808078	GeneID:1473	NM_001900.4	MGC71922;
239	11	7015679	GeneID:338323	NM_176822.2	NOD5; GC-LRR;
240	16	55199525	GeneID:4502	NM_005953.2	MT2;
241	5	83052386	GeneID:1404	NM_001884.2	CRTL1;
242	7	127015093	GeneID:381	NM_001662.2	
243	3	195572481	GeneID:131578	NM_130830.2	LIB;
244	1	150753574	GeneID:54544	NM_019060.1	NICE-1;
245	5	80565096	GeneID:1160	NM_001825.1	SMTCK;
246	10	43187161	GeneID:53828	NM_173160.2	CHIF;
247	12	7792928	GeneID:170482	NM_130441.2	DLEC; HECL; BDCA2; CD303; CLECSF7; CLECSF11; PRO34150; MGC125791; MGC125792; MGC125793;
248	9	33438382	GeneID:360	NM_004925.3	GIL;
249	19	60275375	GeneID:54869	NM_017729.2	DRC3; EPS8R1; MGC4642; PP10566; FLJ20258; MGC23164;
250	2	107968793	GeneID:60482	NM_021815.2	CHT; CHT1; hCHT;
251	1	200125284	GeneID:149345	NM_198149.1	C1orf40; MGC129813;
252	21	45669311	GeneID:378832	NM_199175.1	PRED80;
253	6	35872936	GeneID:1208	NM_001832.2	
254	20	22515170	GeneID:3170	NM_021784.3	HNF3B; TCF3B; MGC19807;
255	11	831335	GeneID:7106	NM_003271.4	NAG2; NAG-2; TM4SF7; TSPAN-4; TETRASPAN;
256	17	24063650	GeneID:147011	NM_152465.1	MGC39650;
257	9	9009757	GeneID:5789	NM_130391.1	HPTP; PTPD; HPTPD; MGC119750; MGC119751; MGC119752; MGC119753; HPTP-DELTA; R-PTP- DELTA;
258	21	39951370	GeneID:10317	NM_033170.1	B3T5; GLCT5; B3GalTx; B3GalT-V; beta3Gal-T5;
259	12	44609923	GeneID:9169	NM_004719.1	SIP1; CASP11; SRRP129;
260	12	79625820	GeneID:4618	NM_002469.1	MRF4; HERCULIN;
261	10	70699885	GeneID:3098	NM_033496.1	HKI; HXK1; HK1-ta; HK1-tb;
262	2	49235663	GeneID:2492	NM_181446.1	LGR1; ODG1; FSHRO;
263	15	39032145	GeneID:79094	NM_024111.2	MGC4504;
264	7	142178732	GeneID:5645	NM_002770.2	TRY2; TRY8; TRYP2; MGC111183; MGC120174;
265	13	77213286	GeneID:122060	NM_144595.2	
266	11	66979224	GeneID:57010	NM_145200.2	
267	9	134986976	GeneID:5900	NM_006266.2	RGF; RaiGEF;
268	19	56230228	GeneID:43849	NM_019598.2	KLK-L5; MGC42603;
269	2	202192343	GeneID:151254	NM_152525.3	FLJ25351; FLJ40332;
270	5	115180684	GeneID:1036	NM_001801.2	
271	16	361991	GeneID:10573	NM_006428.3	p15; MAAT1; MGC8499;
272	6	39398137	GeneID:83795	NM_032115.2	TALK1; TALK-1;
273	17	58878036	GeneID:1534	NM_001915.3	FRRS2;
274	11	47247189	GeneID:8567	NM_130470.1	DENN; IG20; RAB3GEP; KIAA0358;

275	2	238863033	GeneID:8864	NM_003894.3	FASPS; KIAA0347;
276	3	38011038	GeneID:50853	NM_015873.2	

Table 4C

Marker/SEQ ID No.	CpGIsland	CpGIsland	Class	GCContent	CpGRatio	MaxGC Content	MaxCpG Ratio
1	CGI	true	HCP	0,691	0,767	0,67	0,767
2	other_CpG	false	LCP	0,386	0,203	0,414	0,31
3	other_CpG	false	ICP	0,565	0,147	0,606	0,198
4	shore	true	HCP	0,61	0,824	0,668	0,885
5	other_CpG	false	ICP	0,587	0,28	0,672	0,379
6	CGI	true	No coord.	nd	nd	nd	nd
7	other_CpG	false	ICP	0,604	0,339	0,658	0,473
8	other_CpG	false	ICP	0,429	0,346	0,488	0,498
9	CGI	true	No coord.	nd	nd	nd	nd
10	other_CpG	false	ICP	0,624	0,172	0,66	0,309
11	other_CpG	true	ICP	0,609	0,284	0,646	0,406
12	other_CpG	false	ICP	0,504	0,137	0,556	0,22
13	other_CpG	false	ICP	0,588	0,077	0,646	0,144
14	other_CpG	false	LCP	0,398	0,044	0,432	0,093
15	other_CpG	true	HCP	0,807	1,033	0,802	1,055
16	other_CpG	false	ICP	0,474	0,209	0,51	0,282
17	shore	false	HCP	0,683	0,894	0,722	0,878
18	other_CpG	true	ICP	0,451	0,382	0,518	0,448
19	shore	false	HCP	0,478	0,733	0,554	0,857
20	other_CpG	false	ICP	0,537	0,291	0,578	0,391
21	shore	false	ICP	0,583	0,366	0,588	0,578
22	shore	true	HCP	0,711	0,939	0,802	1,031
23	shore	false	HCP	0,59	0,729	0,726	0,896
24	other_CpG	false	ICP	0,575	0,214	0,616	0,325
25	shore	true	ICP	0,663	0,62	0,72	0,707
26	shore	true	HCP	0,605	0,746	0,688	0,982
27	shore	true	HCP	0,62	0,815	0,554	1,044
28	other_CpG	false	ICP	0,606	0,228	0,628	0,329
29	other_CpG	false	ICP	0,609	0,181	0,648	0,24
30	other_CpG	false	ICP	0,559	0,261	0,58	0,368
31	other_CpG	false	ICP	0,489	0,253	0,54	0,312
32	other_CpG	false	ICP	0,513	0,241	0,552	0,289
33	other_CpG	false	HCP	0,586	0,62	0,62	0,779
34	other_CpG	false	ICP	0,496	0,326	0,552	0,482
35	other_CpG	false	ICP	0,46	0,335	0,572	0,396
36	shore	false	HCP	0,568	0,713	0,718	0,839
37	other_CpG	false	ICP	0,489	0,253	0,54	0,312
38	other_CpG	false	LCP	0,429	0,146	0,462	0,238
39	other_CpG	true	ICP	0,578	0,269	0,622	0,4
40	CGI	true	HCP	0,752	0,975	0,828	1,05
41	shore	true	HCP	0,585	0,801	0,682	0,775
42	other_CpG	false	ICP	0,525	0,254	0,558	0,369
43	other_CpG	false	ICP	0,513	0,241	0,552	0,289
44	other_CpG	false	ICP	0,525	0,182	0,578	0,311
45	other_CpG	false	ICP	0,482	0,381	0,51	0,57

46	other_CpG	false	ICP	0,55	0,209	0,588	0,337
47	other_CpG	false	ICP	0,555	0,156	0,594	0,325
48	other_CpG	false	ICP	0,643	0,291	0,662	0,444
49	other_CpG	false	ICP	0,537	0,291	0,578	0,391
50	CGI	true	No coord.	nd	nd	nd	nd
51	shore	true	HCP	0,656	0,62	0,708	0,755
52	other_CpG	false	ICP	0,595	0,189	0,646	0,258
53	shore	true	HCP	0,565	0,658	0,648	0,751
54	other_CpG	false	LCP	0,429	0,183	0,472	0,3
55	shore	true	HCP	0,689	0,837	0,748	0,849
56	other_CpG	false	ICP	0,543	0,17	0,592	0,255
57	other_CpG	false	ICP	0,588	0,261	0,638	0,357
58	shore	true	ICP	0,673	0,523	0,712	0,729
59	other_CpG	false	ICP	0,567	0,353	0,67	0,418
60	other_CpG	false	ICP	0,533	0,259	0,558	0,488
61	other_CpG	false	HCP	0,639	0,668	0,674	0,765
62	other_CpG	false	ICP	0,525	0,182	0,578	0,311
63	shore	true	ICP	0,461	0,565	0,584	0,651
64	other_CpG	false	ICP	0,501	0,239	0,528	0,367
65	shore	true	HCP	0,61	0,705	0,648	0,776
66	shore	true	HCP	0,729	0,819	0,672	0,762
67	other_CpG	false	ICP	0,525	0,254	0,558	0,369
68	other_CpG	false	ICP	0,592	0,152	0,616	0,18
69	other_CpG	false	ICP	0,56	0,304	0,612	0,466
70	other_CpG	true	ICP	0,565	0,385	0,646	0,555
71	other_CpG	false	ICP	0,606	0,228	0,628	0,329
72	other_CpG	false	ICP	0,555	0,21	0,61	0,272
73	other_CpG	false	ICP	0,48	0,275	0,534	0,407
74	other_CpG	false	LCP	0,409	0,12	0,43	0,162
75	other_CpG	false	ICP	0,467	0,2	0,552	0,328
76	other_CpG	false	ICP	0,545	0,325	0,618	0,414
77	shore	true	ICP	0,65	0,518	0,706	0,66
78	shore	false	ICP	0,605	0,282	0,644	0,37
79	CGI	true	No coord.	nd	nd	nd	nd
80	other_CpG	true	ICP	0,5	0,293	0,596	0,405
81	shore	false	HCP	0,58	0,696	0,588	0,751
82	other_CpG	false	ICP	0,529	0,119	0,586	0,214
83	CGI	true	HCP	0,625	0,644	0,708	0,75
84	shore	false	ICP	0,522	0,257	0,568	0,399
85	other_CpG	false	ICP	0,475	0,223	0,53	0,282
86	other_CpG	false	ICP	0,435	0,372	0,508	0,54
87	other_CpG	false	ICP	0,583	0,129	0,598	0,171
88	other_CpG	false	ICP	0,563	0,342	0,612	0,453
89	other_CpG	false	ICP	0,606	0,245	0,648	0,393
90	shore	false	HCP	0,586	0,756	0,74	0,752
91	CGI	true	HCP	0,645	0,558	0,752	0,836
92	shore	false	HCP	0,608	0,764	0,768	0,895
93	other_CpG	false	LCP	0,4	0,104	0,476	0,225
94	CGI	true	HCP	0,714	0,662	0,756	0,75
95	other_CpG	false	LCP	0,355	0,189	0,384	0,316
96	other_CpG	false	LCP	0,416	0,25	0,448	0,497
97	other_CpG	false	ICP	0,58	0,228	0,67	0,359

98	other_CpG	true	ICP	0,465	0,478	0,514	0,646
99	other_CpG	false	ICP	0,418	0,405	0,482	0,595
100	other_CpG	false	ICP	0,58	0,179	0,648	0,263
101	CGI	true	HCP	0,677	0,737	0,726	0,752
102	shore	false	ICP	0,641	0,292	0,682	0,399
103	other_CpG	false	ICP	0,553	0,283	0,59	0,423
104	other_CpG	false	ICP	0,543	0,271	0,628	0,298
105	other_CpG	false	ICP	0,487	0,084	0,51	0,182
106	other_CpG	false	ICP	0,422	0,131	0,484	0,154
107	shore	true	HCP	0,712	0,776	0,732	0,763
108	shore	true	HCP	0,554	0,794	0,61	0,755
109	CGI	true	ICP	0,495	0,34	0,564	0,364
110	other_CpG	false	LCP	0,403	0,287	0,456	0,418
111	other_CpG	false	ICP	0,525	0,217	0,568	0,342
112	shore	false	HCP	0,6	0,753	0,632	0,831
113	shore	true	HCP	0,636	0,66	0,712	0,759
114	other_CpG	false	ICP	0,587	0,28	0,672	0,379
115	other_CpG	false	ICP	0,493	0,22	0,528	0,339
116	other_CpG	false	ICP	0,502	0,175	0,584	0,24
117	other_CpG	true	ICP	0,617	0,411	0,678	0,493
118	other_CpG	false	ICP	0,545	0,113	0,57	0,148
119	shore	true	HCP	0,675	1,036	0,778	1,072
120	shore	false	ICP	0,511	0,179	0,62	0,234
121	CGI	true	HCP	0,657	0,752	0,626	0,765
122	shore	true	HCP	0,69	0,947	0,606	0,761
123	other_CpG	false	LCP	0,321	0,065	0,346	0,149
124	shore	true	HCP	0,642	0,752	0,688	0,802
125	other_CpG	false	LCP	0,326	0,252	0,388	0,416
126	other_CpG	false	ICP	0,57	0,213	0,628	0,337
127	other_CpG	true	LCP	0,41	0,321	0,47	0,529
128	other_CpG	false	LCP	0,351	0,082	0,438	0,13
129	shore	true	No coord.	nd	nd	nd	nd
130	shore	false	HCP	0,56	0,557	0,688	0,753
131	other_CpG	false	ICP	0,499	0,141	0,534	0,201
132	shore	true	ICP	0,6	0,361	0,658	0,456
133	shore	true	HCP	0,637	0,704	0,666	0,767
134	other_CpG	false	ICP	0,646	0,431	0,706	0,609
135	shore	true	HCP	0,727	0,782	0,788	0,89
136	other_CpG	false	ICP	0,578	0,269	0,622	0,4
137	other_CpG	false	ICP	0,455	0,177	0,504	0,249
138	other_CpG	false	ICP	0,53	0,293	0,608	0,483
139	other_CpG	false	ICP	0,5	0,254	0,568	0,332
140	other_CpG	true	ICP	0,539	0,399	0,712	0,502
141	other_CpG	false	LCP	0,352	0,243	0,44	0,372
142	shore	false	HCP	0,656	0,635	0,72	0,763
143	other_CpG	false	LCP	0,383	0,252	0,434	0,326
144	other_CpG	false	ICP	0,525	0,217	0,568	0,342
145	other_CpG	false	LCP	0,419	0,152	0,442	0,294
146	other_CpG	true	ICP	0,573	0,257	0,602	0,377
147	other_CpG	false	ICP	0,435	0,372	0,508	0,54
148	other_CpG	false	ICP	0,591	0,23	0,662	0,294
149	other_CpG	false	LCP	0,435	0,228	0,45	0,384
150	other_CpG	false	ICP	0,459	0,239	0,546	0,366

151	other_CpG	true	ICP	0,545	0,36	0,586	0,43
152	shore	false	HCP	0,65	0,966	0,758	0,961
153	other_CpG	false	LCP	0,341	0,258	0,394	0,343
154	shore	true	HCP	0,689	0,856	0,556	0,813
155	CGI	true	HCP	0,723	0,84	0,76	0,959
156	other_CpG	false	ICP	0,488	0,112	0,574	0,16
157	other_CpG	false	LCP	0,386	0,203	0,414	0,31
158	other_CpG	false	ICP	0,474	0,433	0,538	0,609
159	other_CpG	false	ICP	0,543	0,271	0,628	0,298
160	shore	false	HCP	0,595	0,705	0,614	0,751
161	other_CpG	false	HCP	0,703	0,662	0,676	0,755
162	other_CpG	false	ICP	0,57	0,186	0,616	0,253
163	other_CpG	false	ICP	0,505	0,15	0,53	0,216
164	CGI	true	HCP	0,67	0,602	0,694	0,751
165	other_CpG	false	ICP	0,484	0,229	0,59	0,308
166	other_CpG	true	HCP	0,777	0,968	0,742	1,022
167	other_CpG	false	ICP	0,58	0,228	0,67	0,359
168	shore	false	ICP	0,6	0,361	0,658	0,456
169	shore	false	ICP	0,608	0,487	0,69	0,657
170	other_CpG	false	ICP	0,594	0,181	0,63	0,213
171	other_CpG	false	LCP	0,376	0,095	0,424	0,168
172	other_CpG	true	ICP	0,572	0,448	0,632	0,559
173	other_CpG	false	ICP	0,482	0,074	0,522	0,12
174	other_CpG	true	ICP	0,572	0,387	0,636	0,525
175	other_CpG	false	LCP	0,399	0,21	0,44	0,343
176	other_CpG	false	ICP	0,55	0,209	0,588	0,337
177	other_CpG	false	ICP	0,537	0,254	0,598	0,346
178	other_CpG	false	ICP	0,43	0,109	0,488	0,229
179	other_CpG	false	ICP	0,499	0,322	0,522	0,492
180	other_CpG	false	LCP	0,366	0,176	0,42	0,305
181	shore	false	HCP	0,617	0,863	0,554	0,863
182	shore	false	ICP	0,467	0,306	0,562	0,448
183	shore	false	ICP	0,614	0,243	0,646	0,359
184	other_CpG	false	ICP	0,53	0,12	0,574	0,211
185	shore	false	HCP	0,739	0,789	0,834	0,88
186	other_CpG	false	ICP	0,575	0,322	0,63	0,478
187	other_CpG	false	ICP	0,514	0,164	0,558	0,249
188	shore	true	HCP	0,734	0,749	0,716	0,765
189	other_CpG	false	LCP	0,316	0,133	0,392	0,266
190	shore	false	ICP	0,626	0,586	0,72	0,719
191	other_CpG	false	LCP	0,371	0,291	0,43	0,512
192	other_CpG	false	ICP	0,481	0,249	0,514	0,308
193	other_CpG	false	LCP	0,369	0,245	0,438	0,395
194	other_CpG	true	ICP	0,5	0,293	0,596	0,405
195	other_CpG	false	ICP	0,594	0,35	0,664	0,475
196	shore	true	ICP	0,676	0,452	0,724	0,661
197	other_CpG	false	ICP	0,639	0,302	0,69	0,429
198	other_CpG	false	ICP	0,594	0,181	0,63	0,213
199	other_CpG	false	LCP	0,409	0,18	0,476	0,243
200	shore	false	HCP	0,731	0,912	0,684	0,821
201	other_CpG	false	LCP	0,42	0,189	0,436	0,286
202	other_CpG	false	ICP	0,505	0,108	0,566	0,184
203	shore	true	HCP	0,673	0,723	0,756	0,776

204	CGI	true	HCP	0,791	0,969	0,83	0,801
205	other_CpG	false	ICP	0,4	0,249	0,506	0,412
206	other_CpG	false	ICP	0,568	0,207	0,638	0,24
207	other_CpG	false	ICP	0,504	0,184	0,558	0,25
208	shore	true	HCP	0,806	1,026	0,774	0,915
209	other_CpG	false	LCP	0,381	0,233	0,448	0,567
210	other_CpG	false	LCP	0,319	0,133	0,334	0,218
211	other_CpG	false	ICP	0,44	0,172	0,552	0,283
212	other_CpG	false	ICP	0,56	0,14	0,616	0,219
213	other_CpG	false	LCP	0,387	0,135	0,404	0,187
214	other_CpG	false	ICP	0,461	0,266	0,528	0,343
215	other_CpG	false	ICP	0,44	0,138	0,506	0,318
216	CGI	true	ICP	0,666	0,406	0,748	0,644
217	shore	true	HCP	0,708	0,829	0,688	0,93
218	other_CpG	false	LCP	0,429	0,183	0,472	0,3
219	other_CpG	false	LCP	0,391	0,22	0,438	0,28
220	other_CpG	true	No coord.	nd	nd	nd	nd
221	CGI	true	HCP	0,714	0,856	0,686	0,799
222	shore	true	HCP	0,61	0,814	0,552	0,848
223	shore	true	No coord.	nd	nd	nd	nd
224	other_CpG	false	ICP	0,51	0,181	0,576	0,264
225	shore	false	ICP	0,58	0,208	0,658	0,335
226	other_CpG	false	LCP	0,375	0,115	0,43	0,238
227	other_CpG	false	LCP	0,373	0,168	0,45	0,317
228	other_CpG	false	ICP	0,552	0,165	0,612	0,267
229	other_CpG	false	ICP	0,426	0,239	0,52	0,371
230	CGI	true	HCP	0,656	0,82	0,578	0,759
231	other_CpG	false	ICP	0,641	0,276	0,688	0,34
232	other_CpG	false	ICP	0,417	0,249	0,484	0,5
233	other_CpG	false	ICP	0,602	0,175	0,632	0,356
234	other_CpG	false	ICP	0,544	0,158	0,608	0,274
235	other_CpG	true	No coord.	nd	nd	nd	nd
236	other_CpG	false	LCP	0,356	0,16	0,374	0,199
237	shore	false	HCP	0,545	0,452	0,6	0,766
238	other_CpG	false	ICP	0,551	0,188	0,61	0,225
239	other_CpG	false	LCP	0,409	0,13	0,462	0,195
240	shore	true	HCP	0,577	0,709	0,566	0,757
241	shore	false	ICP	0,453	0,292	0,556	0,377
242	CGI	true	HCP	0,689	0,802	0,674	0,757
243	other_CpG	false	ICP	0,527	0,324	0,64	0,444
244	shore	false	ICP	0,505	0,173	0,54	0,255
245	other_CpG	true	ICP	0,557	0,419	0,642	0,613
246	other_CpG	false	ICP	0,546	0,268	0,604	0,414
247	other_CpG	false	ICP	0,507	0,261	0,55	0,369
248	shore	false	ICP	0,639	0,522	0,722	0,692
249	other_CpG	false	ICP	0,565	0,181	0,596	0,247
250	shore	false	HCP	0,621	0,669	0,632	0,756
251	other_CpG	true	ICP	0,693	0,475	0,752	0,665
252	other_CpG	true	ICP	0,645	0,393	0,666	0,613
253	other_CpG	false	ICP	0,516	0,19	0,596	0,303
254	shore	true	HCP	0,578	0,609	0,66	0,856

255	other_CpG	false	HCP	0,738	0,889	0,702	0,797
256	shore	true	HCP	0,653	0,745	0,706	0,895
257	other_CpG	false	LCP	0,301	0,33	0,326	0,578
258	other_CpG	false	ICP	0,499	0,187	0,538	0,268
259	other_CpG	false	LCP	0,332	0,151	0,384	0,193
260	shore	true	ICP	0,488	0,337	0,574	0,494
261	other_CpG	false	ICP	0,458	0,176	0,506	0,238
262	other_CpG	false	LCP	0,441	0,172	0,47	0,226
263	shore	true	HCP	0,658	0,708	0,686	0,757
264	other_CpG	false	ICP	0,499	0,141	0,534	0,201
265	other_CpG	false	LCP	0,382	0,229	0,422	0,431
266	other_CpG	false	ICP	0,622	0,26	0,656	0,408
267	shore	false	HCP	0,697	0,771	0,748	0,76
268	other_CpG	false	ICP	0,549	0,188	0,596	0,351
269	shore	true	ICP	0,505	0,557	0,644	0,696
270	shore	true	HCP	0,548	0,844	0,636	0,755
271	shore	true	HCP	0,624	0,61	0,738	0,755
272	other_CpG	false	ICP	0,551	0,186	0,638	0,266
273	CGI	true	HCP	0,744	0,796	0,744	0,81
274	shore	false	HCP	0,595	0,594	0,658	0,758
275	shore	true	HCP	0,697	0,975	0,686	0,762
276	shore	true	ICP	0,59	0,527	0,67	0,704

The methylation pattern of any one of the markers above is important on the diagnostic method according to the present invention. More particularly, it was shown that using the methylation pattern of a panel of the following genes is sufficient for obtaining good prognostic results: markers 1, 3, 5, 6, 8, 15, 17, 23, 26, 28, 50, 65, 71, 92, 114, 129, 142, 178, 210, 223, 235, 263, 270, and 275. Said CpG sites are defined by SEQ ID NOs 1, 3, 5, 6, 8, 15, 17, 23, 26, 28, 50, 65, 71, 92, 114, 129, 142, 178, 210, 223, 235, 263, 270, and 275 respectively

Example 6: Functional analysis of differentially methylated genes reveals their involvement in specific biological processes in islets and beta-cells

This experiment sought to provide experimental proof that at least some of the differentially methylated genes could indeed affect beta-cell function and viability. For this, the biological role of two of the differentially methylated genes with previously unknown function in beta-cells, namely NIBAN and CHAC1, was further studied.

ER stress response can initiate apoptosis and has been implicated in beta-cell demise in diabetes. To test the role of NIBAN and CHAC1 in ER stress, human islets were treated with the physiological ER stressors oleate and palmitate or the synthetic ER stressors thapsigargin (THA), tunicamycin (TUN) or brefeldin (BRE). Expression of NIBAN was induced about two-fold by the saturated fatty acid palmitate (Figures 8A, line 3) but not with oleate (Figure 8A, line 2), which is a less potent inducer of ER stress. A similar induction of expression was observed for CHAC1 when islets were treated with palmitate (Figure 8D, column 3) but not with oleate (Figure 8D, column 2). The synthetic ER stressors thapsigargin (THA), tunicamycin (TUN) and brefeldin

(BRE) also increased NIBAN and CHAC1 mRNA expression in human islets (Figure 8A and 7D, right panels). Thapsigargin, which causes ER calcium depletion by blocking the sarcoendoplasmic reticulum Ca²⁺ ATPase (SERCA), tunicamycin, which blocks glycosylation of nascent proteins (Hickman Set al., 1977 J Biol Chem 252: 4402-4408), and especially brefeldin, which inhibits ER-to-Golgi transport, induced NIBAN and CHAC1 gene expression. The magnitude of ER stress induced by these three chemicals and palmitate, as measured by ATF3, CHOP and BiP mRNA expression, was closely correlated with the NIBAN and CHAC1 induction.

Next, the functional role of NIBAN and CHAC1 expression on the outcome of ER stress in beta-cells was determined. For this purpose, the rat beta-cell line INS-1E was exposed to palmitate or cyclopiazonic acid (CPA), a specific inhibitor of the SERCA pump. ER stress induced by these agents increased Niban mRNA expression (columns 1, 3, 5 in Figure 8B). As shown in Figure 8B, expression of Niban is efficiently diminished by a specific siRNA (siNiban; compare columns 1 and 2, 3 and 4, 5 and 6). RNAi-mediated knockdown of Niban increased apoptosis induced by palmitate (Figure 8C; columns 3, 4) as well as CPA (columns 5, 6 in Figure 8C). This is mirrored by an elevated caspase 3 activation (Figure 8C, lower panel). Taken together, the increase of Niban expression during ER stress and augmented apoptosis after its knockdown indicates an anti-apoptotic role for Niban during the beta-cell ER stress response.

Chac1 mRNA expression was also induced by palmitate and CPA in INS-1E cells (Figure 8E). Chac1 knockdown by siRNA (siChac) significantly reduced its expression upon incubation with palmitate (Figure 8E, columns 3, 4) or CPA (columns 5, 6 in Figure 8E). Next it was analysed whether the knockdown of Chac1 affected ER stress-induced apoptosis stimulated by palmitate and CPA (Figure 8F). We found that Chac1 knockdown protected beta-cells from apoptosis (cf. columns 3, 4 and 5, 6 in Figure 8F), which was confirmed by a lessened activation of caspase 3 (Figure 8F, lower panel). The induction of Chac1 expression by ER stressors and the fact that its knockdown protected beta-cells from ER stress-induced apoptosis suggest that Chac1 is involved in apoptosis triggered after the different branches of ER stress response fail to re-establish ER homeostasis in beta-cells.

These experiments investigated two novel genes with opposite effects on ER stress-induced apoptosis in human islets, CHAC1 and NIBAN. Expression of both genes is strongly induced by ER stress, elicited by different agents. While NIBAN protects beta-cells against apoptosis, the activation of CHAC1 indicates failure of the processes restoring beta-cell functionality and heralds ER stress-induced apoptosis.

These experiments provide the proof of concept that the genes identified by the inventors to be differentially methylated in type-II diabetes mellitus patients versus healthy subjects, are indeed functionally implicated in beta-cell physiology.

Methods:

INS-1E Cell and Human Islet Culture. The rat insulin-producing INS-1E cell line (a kind gift from Prof. C. Wollheim, Centre Medical Universitaire, Geneva, Switzerland) was cultured in RPMI 1640 (with 2mM GlutaMAX-I) containing 5% FBS and used at passages 59-73. Human islets were isolated from 11 organ donors (age 69±6 years; body mass index 26±1kg/m²) in Pisa, Italy, as described above. The islets were cultured in Ham's F-10 medium containing 6.1 or 28mM glucose as previously described (Cunha DA, et al., 2008, J Cell Sci 121: 2308-2318). The percentage of beta-cells, assessed in dispersed islet preparations following staining with mouse monoclonal anti-insulin antibody (1:1000, Sigma) and donkey anti-mouse IgG Rhodamine (1:200, Jackson ImmunoResearch Europe, Soham, Cambridgeshire, UK), was 53±3%. Palmitate and oleate (Sigma-Aldrich, Schnelldorf, Germany) were dissolved in 90% ethanol, and used at a final concentration of 0.5mM in the presence of 1% BSA. The chemical ER stressors thapsigargin (diluted in DMSO and used at a final concentration of 1µM), cyclopiazonic acid (CPA, diluted in DMSO and used at final concentration of 25µM), tunicamycin (diluted in PBS and used at a final concentration of 5µg/ml) and brefeldin (diluted in ethanol and used at a final concentration of 0.1µg/ml) were obtained from Sigma-Aldrich. The control condition contained similar dilutions of vehicle.

Assessment of Beta-Cell Apoptosis. Quantitative evaluation of INS-1E cell apoptosis was done by fluorescence microscopy following staining with the DNA-binding dyes propidium iodide (5µg/ml) and Hoechst 33342 (5µg/ml). Caspase 3 activation was assessed by Western blot, as previously described (Gurzov et al., 2009, Cell Death Differ 16: 1539-1550), using anti-cleaved caspase 3 antibody (1:1000; from Cell Signaling, Beverly, MA, USA).

RNA Interference. NIBAN and CHAC1 were knocked down using small interfering RNA (siRNA). The Niban siRNA was SMARTpool (L-080179-01 from Dharmacon, Chicago, IL, USA) and CHAC1 was Stealth RNAiTM (RSS324745 from Invitrogen, Carlsbad, CA, USA). A negative control of 21 nucleotide duplex RNA with no known sequence homology was obtained from Qiagen (Hilden, Germany). Lipid-RNA complexes were formed in OptiMem1 with 1.5µl Lipofectamine 2000 (Invitrogen) to 150nM siRNA and added at a final concentration of 30nM siRNA for transfection as described. Transfected cells were cultured for 2 days and subsequently treated.

Real-time PCR. Poly(A)⁺ RNA was isolated and reverse transcribed as previously described (Chen MC, et al., 2001, Diabetologia 44: 325-332.). The PCR was done in 3mM MgCl₂, 0.5µM forward and reverse primers, 2µl SYBR Green PCR master mix (Qiagen, Hilden, Germany) and 2µl cDNA. Standards for each gene were prepared using appropriate primers in a conventional PCR. The samples were assayed on a LightCycler instrument (Roche Diagnostics, Mannheim, Germany) and their concentration was calculated as copies per µl using the standard curve. The

expression level of the gene of interest was corrected for the expression of the housekeeping gene glyceraldehyde-3-phosphate dehydrogenase (Gapdh, for INS-1E cells) or beta-actin (for human islets). The different treatments utilised in the study did not change expression of the housekeeping gene (data not shown).

CLAIMS

1. A method for the prognosis, diagnosis or prediction of Type 2 Diabetes (T2D) and/or follow up of T2D intervention therapies comprising the steps of:
 - 5 a) measuring the methylation status of one or more of the CpG site(s) defined in Table 4 in a sample of the subject, defined by any one of SEQ ID Nos 1-276, and
 - b) comparing the methylation status of said one or more CpG site(s) obtained from step a) with the methylation status of said CpG site(s) in a control sample, wherein a difference in methylation status as detected in step b) indicates the subject has
 - 10 or is at risk of developing T2D, optionally comprising the step of:
 - c) comparing the methylation status of said one or more CpG site(s) obtained from step a) with the methylation status of said CpG site(s) in a sample obtained after an intervention therapy aimed to prevent or treat T2D.
- 15 2. The method of claim 1, wherein the difference in methylation status is due to hypermethylation or hypomethylation.
3. The method according to claim 1 or 2, wherein the sample is pancreatic islet tissue, a blood, or serum sample, adipose tissue, muscle, or any other biological sample that serve
- 20 as surrogate material for the pancreatic islet tissue.
4. The method according to any one of claims 1 to 3, wherein the methylation status of up to or more than: 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190,
- 25 195, 200, 205, 210, 215, 220, 225, 230, 235, 240, 245, 250, 255, 260, 265, 270, 275 or of all 276 CpG sites as defined by SEQ ID Nos 1 to 276 in Table 4, is measured.
5. The method according to any one of claims 1 to 4, wherein the methylation status of the CpG sites defined by SEQ ID Nos 1, 3, 5, 6, 8, 15, 17, 23, 26, 28, 50, 65, 71, 92, 114,
- 30 129, 142, 178, 210, 223, 235, 263, 270, and/or 275, taken from Table 4, is measured.
6. A method of treating T2D by targeting one or more genes having aberrant methylation in T2D in any one or more of the CpG sites defined by SEQ ID Nos 1 to 276, taken from Table 4.
- 35 7. The method according to claim 6, wherein said one or more CpG sites are defined by SEQ ID Nos 1, 3, 5, 6, 8, 15, 17, 23, 26, 28, 50, 65, 71, 92, 114, 129, 142, 178, 210, 223, 235, 263, 270, and/or 275, taken from Table 4.

8. The method according to claim 6 or 7, wherein said targeting implies changing the methylation status by using demethylating or methylating agents, by changing the expression level, or by changing the protein activity of the protein encoded by said one or more genes.
- 5
9. The method according to claim 8, wherein said methylating agents are methyl donors such as folic acid, methionine, choline or any other chemicals capable of elevating DNA methylation.
- 10
10. The method according to claim 1, wherein the methylation status is analysed by one or more techniques selected from the group consisting of nucleic acid amplification, polymerase chain reaction (PCR), methylation specific PCR (MCP), methylated-CpG island recovery assay (MIRA), combined bisulfite-restriction analysis (COBRA), bisulfite pyrosequencing, single-strand conformation polymorphism (SSCP) analysis, restriction analysis, microarray analysis, or bead-chip technology.
- 15
11. The method according to anyone of claims 1 to 10, wherein the patient is in a high risk group for developing diabetes or suffering from any beta-cell related disorder such as: type 1 diabetes mellitus, type 2 diabetes mellitus, hyperinsulinemia, obesity, neuroendocrine tumors or occurrence of insulinoma.
- 20
12. A method for identifying an agent that modulates one or more of the genes having aberrant methylation in T2D in any one or more of the CpG sites defined by SEQ ID Nos 1 to 276, taken from Table 4, comprising the steps of:
- 25
- a) contacting the candidate agent with said one or more genes, and
- c) analysing the modulation of said one or more gene by the candidate agent.
13. The method according to claim 12, wherein said CpG sites are defined by SEQ ID Nos 1, 3, 5, 6, 8, 15, 17, 23, 26, 28, 50, 65, 71, 92, 114, 129, 142, 178, 210, 223, 235, 263, 270, and/or 275, taken from Table 4.
- 30
14. The method according to claim 12 or 13, wherein said agent modulates the methylation status, the expression level or the activity of said one or more gene.
- 35
15. A method for establishing a reference methylation status profile comprising the steps of: measuring the methylation status of one or more genes corresponding to CpG sites defined by SEQ ID Nos 1-276 as defined in Table 4 having aberrant methylation in T2D, in a sample of subject.

16. The method according to claim 15, wherein said subject is healthy, thereby producing a reference profile of a healthy subject, or wherein said subject is suffering from T2D, thereby producing a T2D reference profile.
- 5 17. The method according to claim 15 or 16, wherein said reference profile concerns the methylation status profile of up to, or more than: 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, 200, 205, 210, 215, 220, 225, 230, 235, 240, 245, 250, 255, 260, 265, 270, or 275 of the CpG sites defined by SEQ ID Nos 1 to 276 taken
10 from Table 4.
18. The method according to any one of claims 15 to 17, wherein said reference profile concerns the methylation status profile of the CpG sites defined by SEQ ID Nos 1, 3, 5, 6, 8, 15, 17, 23, 26, 28, 50, 65, 71, 92, 114, 129, 142, 178, 210, 223, 235, 263, 270, and/or
15 275, taken from Table 4.
19. A microarray or chip comprising one or more T2D-specific CpG regions as defined by SEQ ID Nos 1 to 276 taken from Table 4.
- 20 20. The microarray or chip according to claim 19, comprising up to, or more than: 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, 200, 205, 210, 215, 220, 225, 230, 235, 240, 245, 250, 255, 260, 265, 270, 275 or all of the CpG sites defined by SEQ ID Nos 1 to 276 taken from Table 4.
- 25 21. The microarray or chip according to claim 19 or 20, comprising the CpG sites defined by SEQ ID Nos 1, 3, 5, 6, 8, 15, 17, 23, 26, 28, 50, 65, 71, 92, 114, 129, 142, 178, 210, 223, 235, 263, 270, and/or 275, taken from Table 4.
- 30 22. Use of the methylation status of one or more of the CpG sites defined by SEQ ID Nos 1-276 taken from Table 4, in the prognosis, diagnosis or prediction of Type 2 Diabetes (T2D).
23. The use according to claim 22, wherein said one or more of the CpG sites defined by
35 SEQ ID Nos 1, 3, 5, 6, 8, 15, 17, 23, 26, 28, 50, 65, 71, 92, 114, 129, 142, 178, 210, 223, 235, 263, 270, and/or 275, taken from Table 4.
24. A method for identifying T2D specific SNPs in comprising the step of comparing the sequence of one or more of the differentially methylated genes corresponding to CpG

sites defined by SEQ ID Nos 1-276 taken from Table 4, in a sample from a healthy versus a T2D subject, wherein a difference or polymorphism in said one or more gene regions between the healthy and T2D subject sample is identified as a T2D-specific SNP.

- 5 25. The method according to claim 24, wherein said differentially methylated gene regions are selected from the group consisting of those corresponding to CpG sites defined by SEQ ID Nos 1, 3, 5, 6, 8, 15, 17, 23, 26, 28, 50, 65, 71, 92, 114, 129, 142, 178, 210, 223, 235, 263, 270, and/or 275, taken from Table 4.

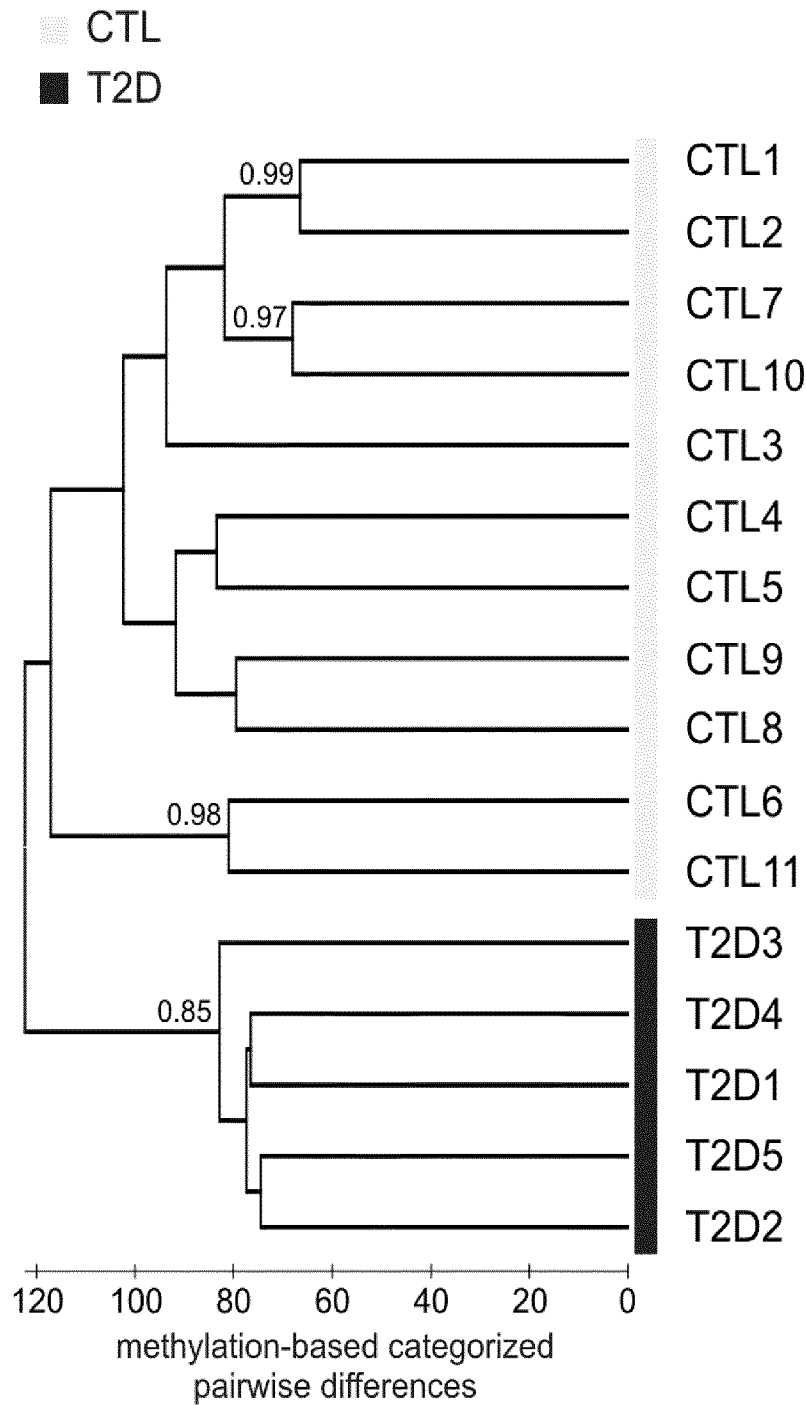


Figure 1A

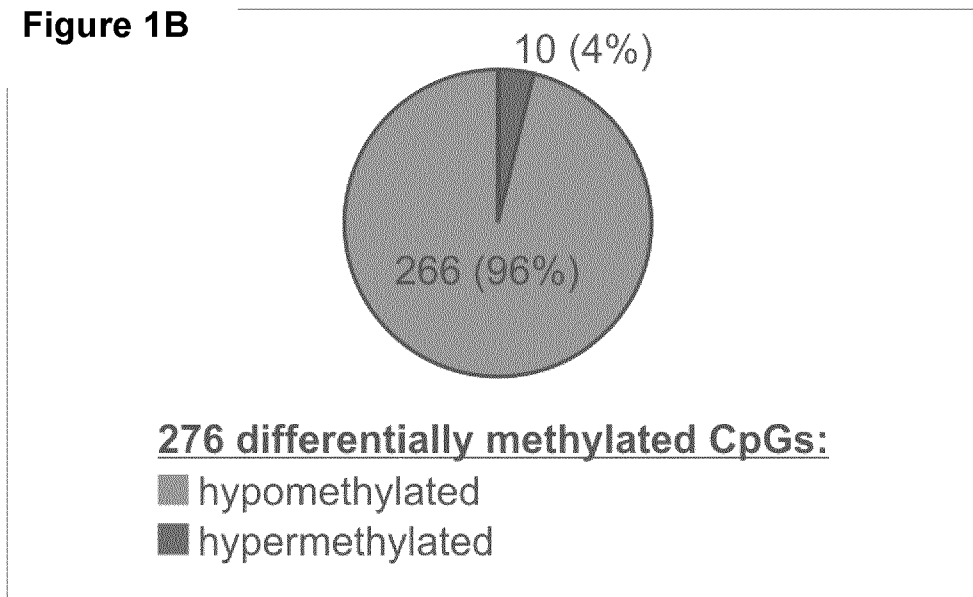
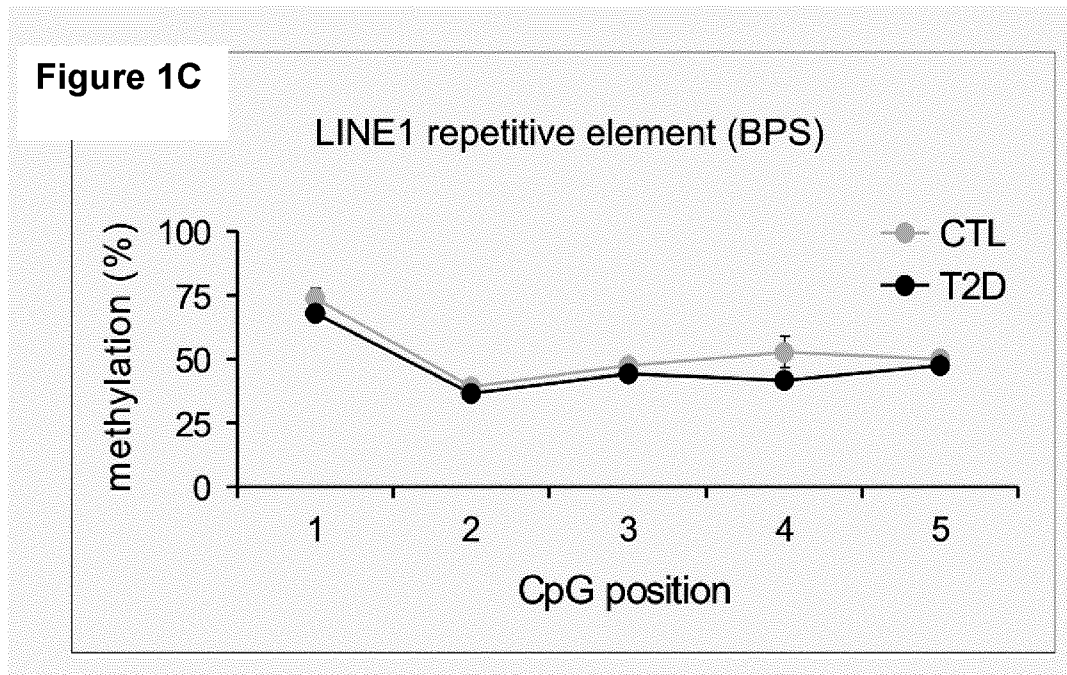
Figure 1B**Figure 1C**

Figure 2 A

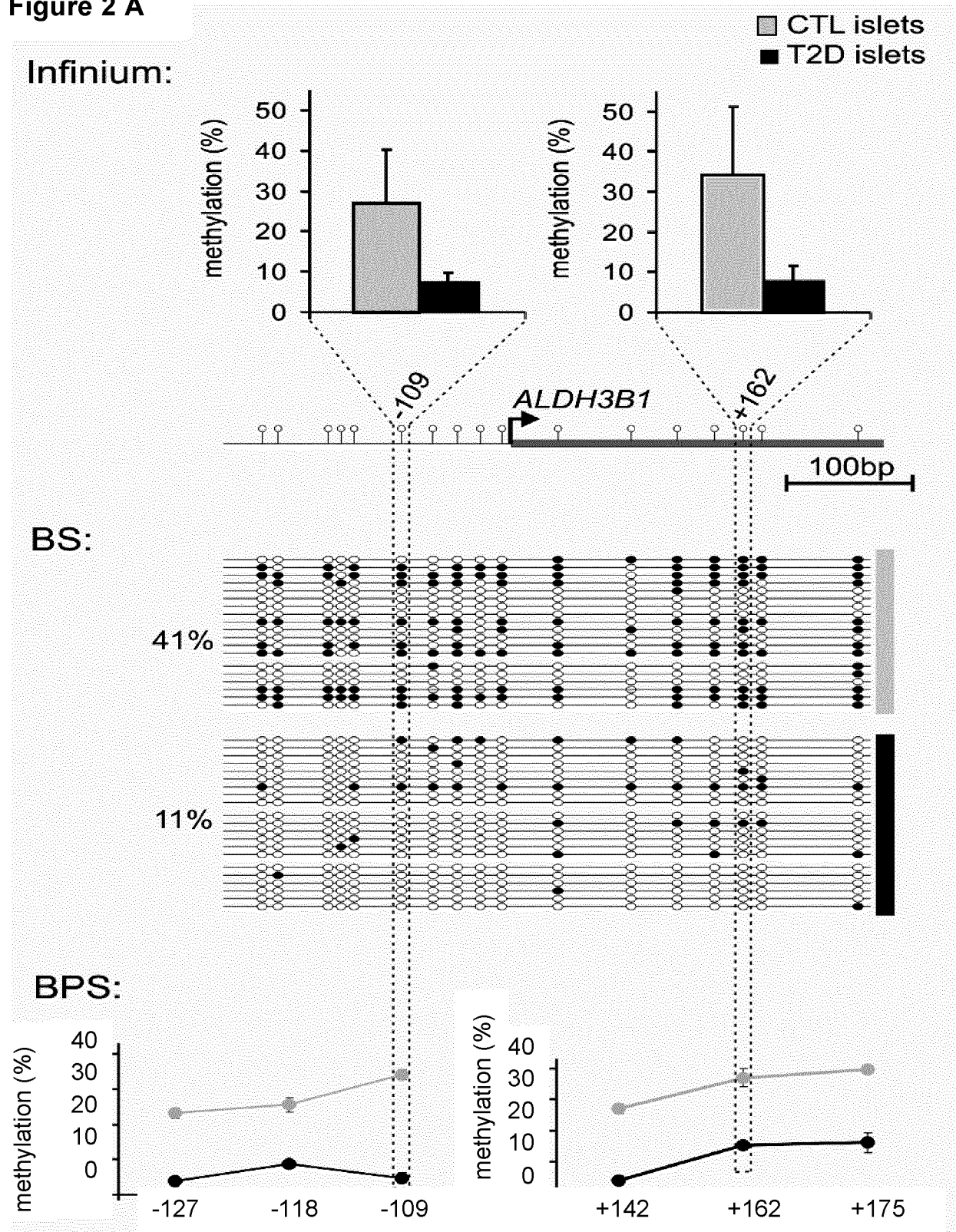


Figure 2 B

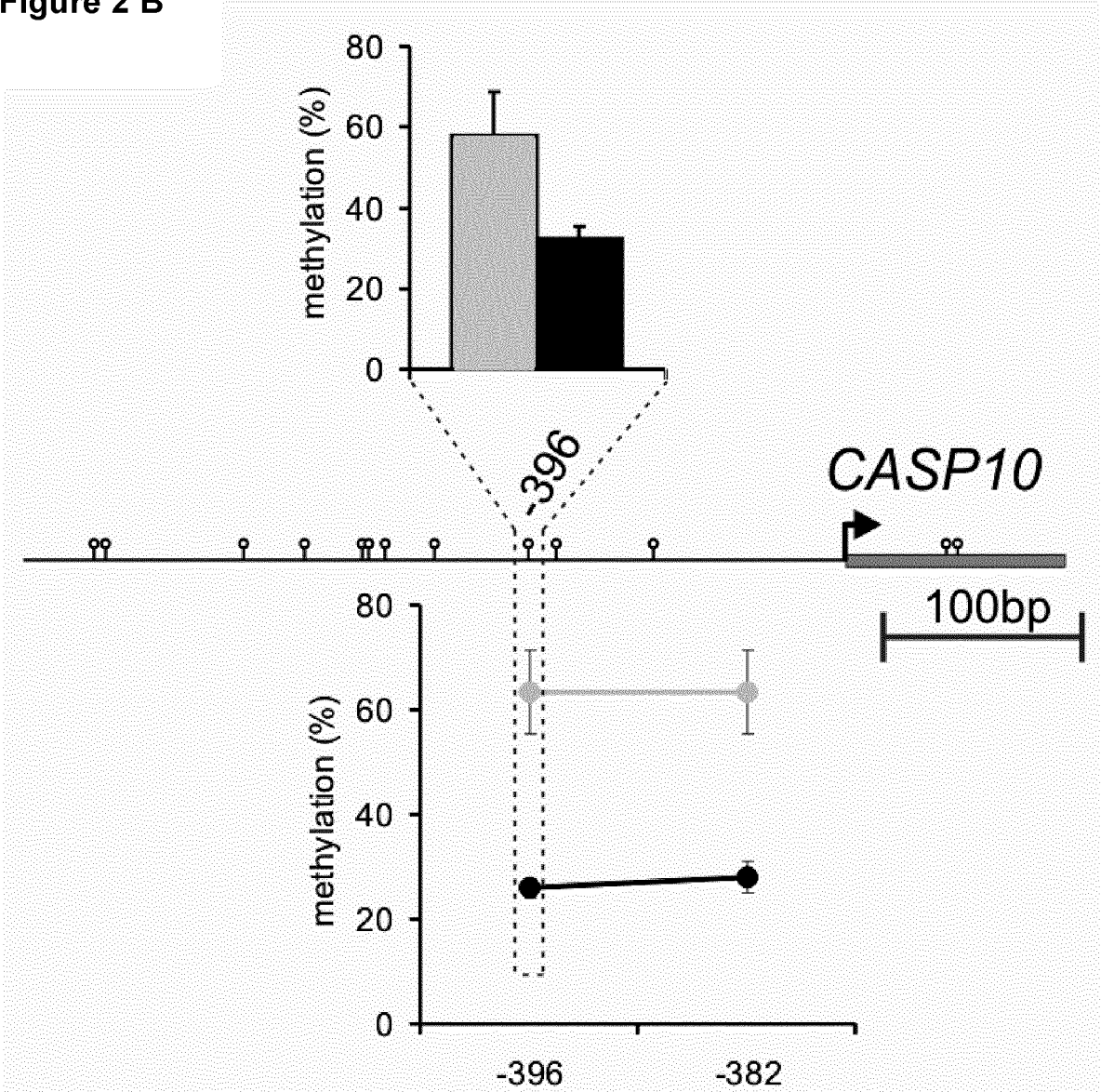


Figure
2 C

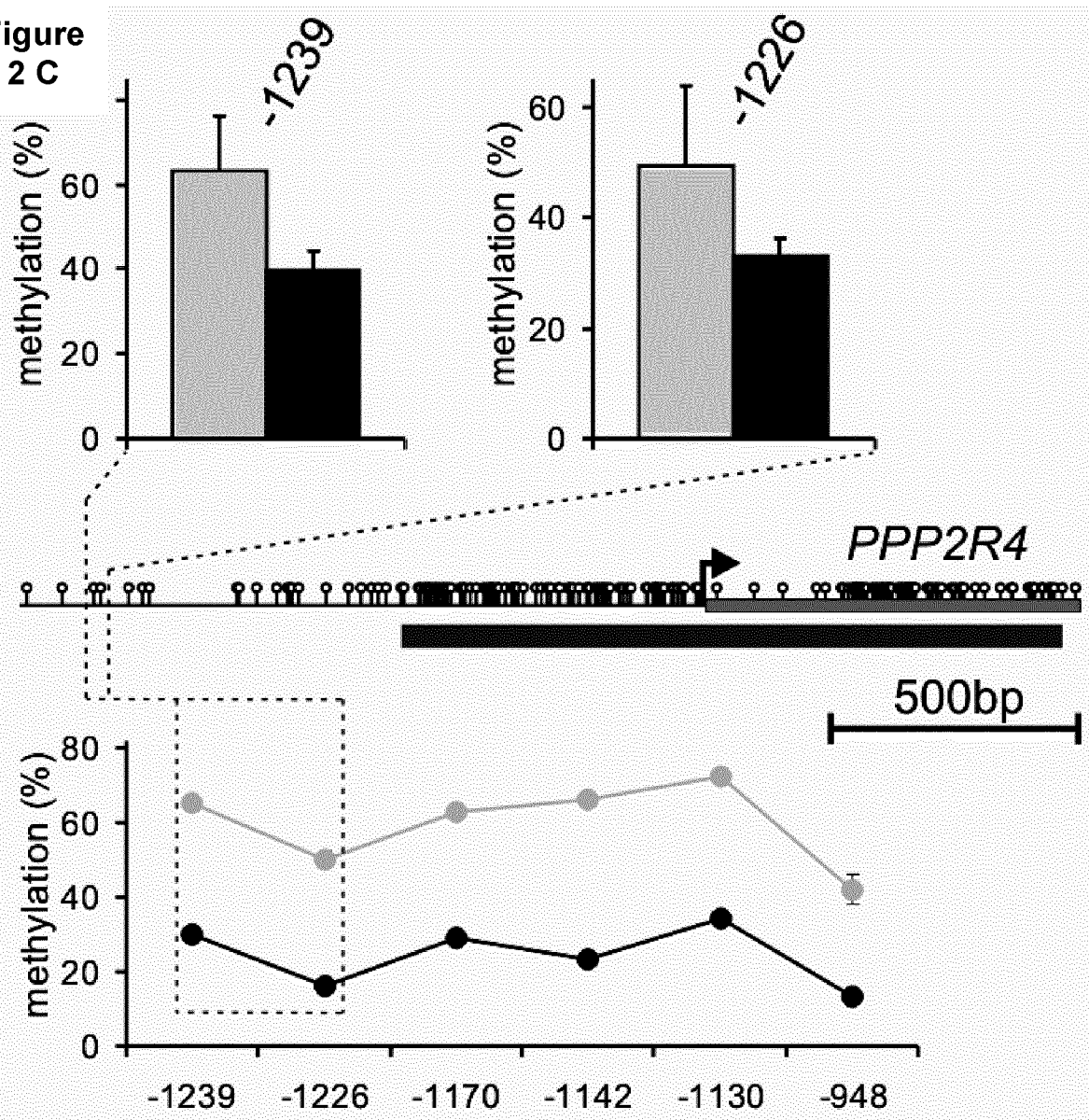


Figure 2 D

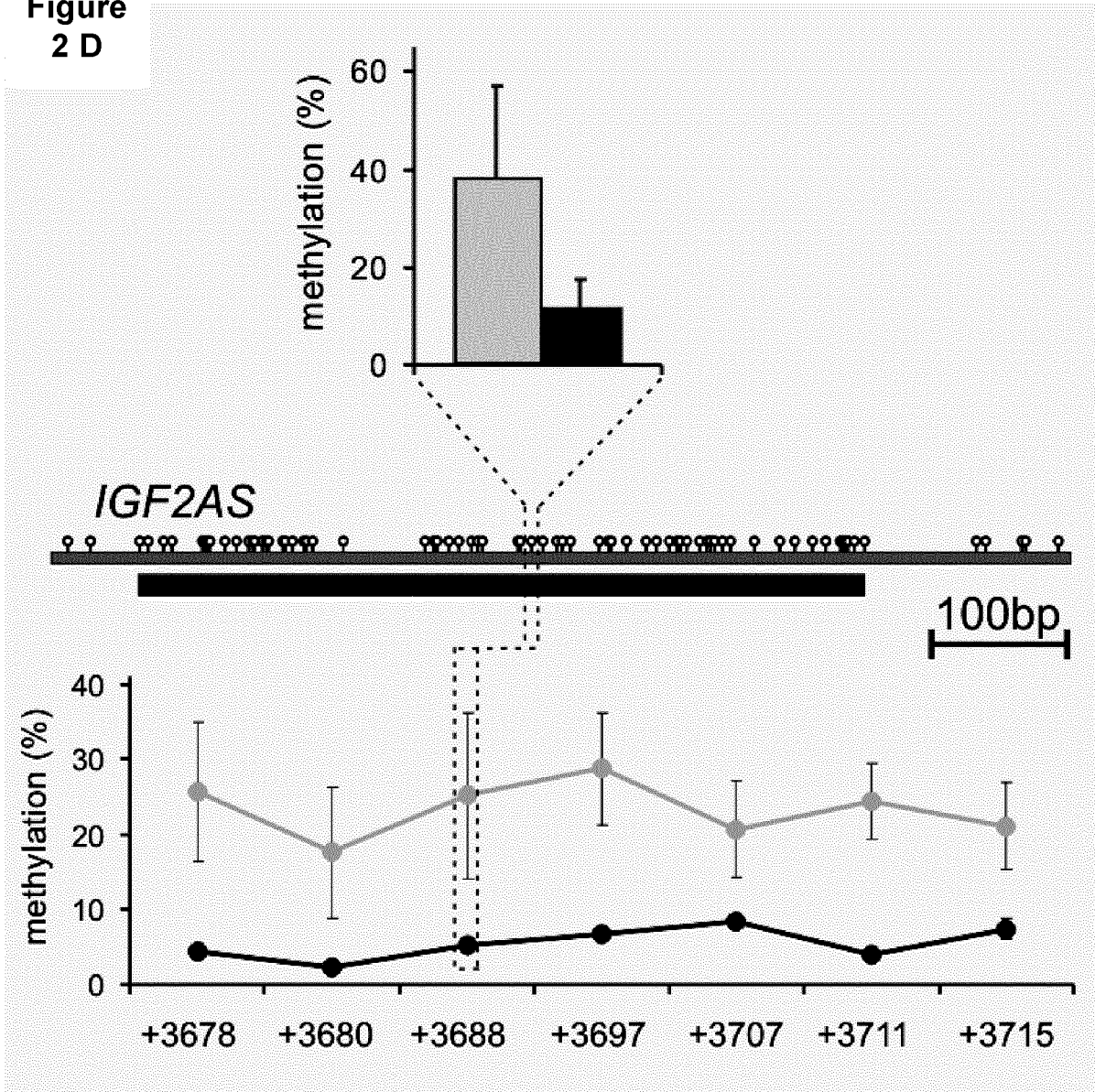


Figure
2 E

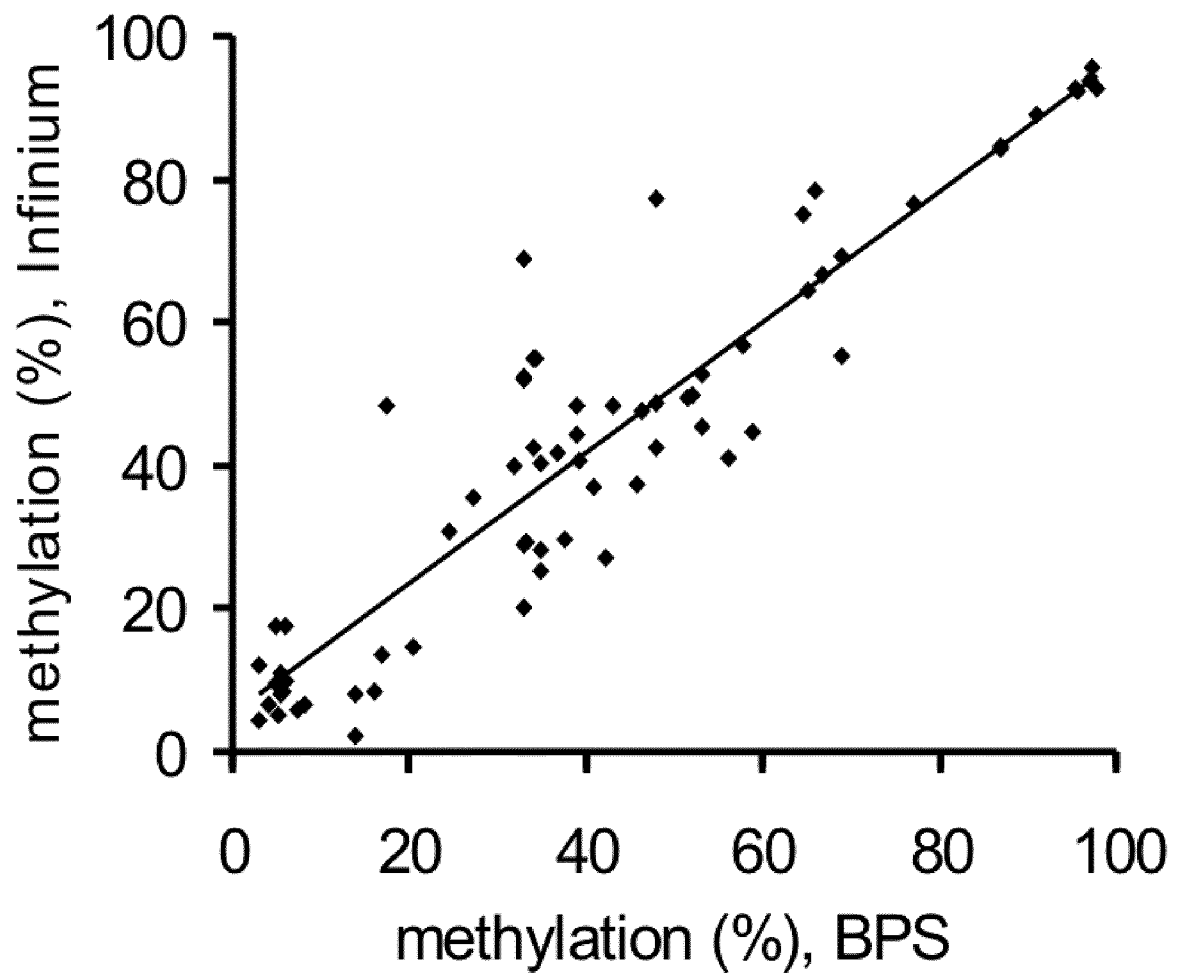
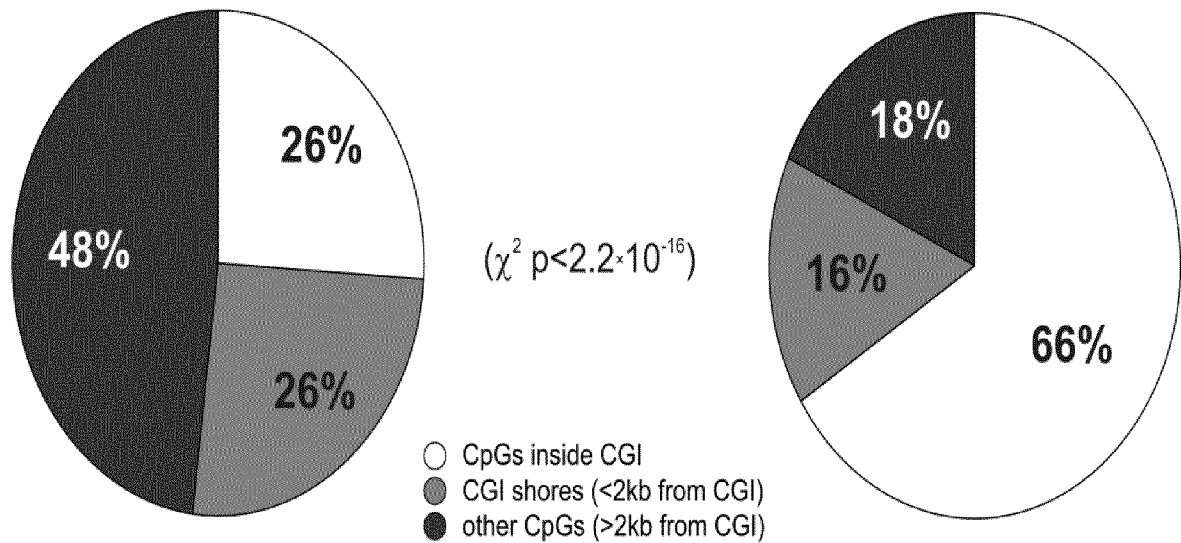


Figure 3

A**CpG location versus CGI**

Differentially methylated in T2D islets

Infinium Array

**B****CpG promoter class**

Differentially methylated in T2D islets

Infinium Array

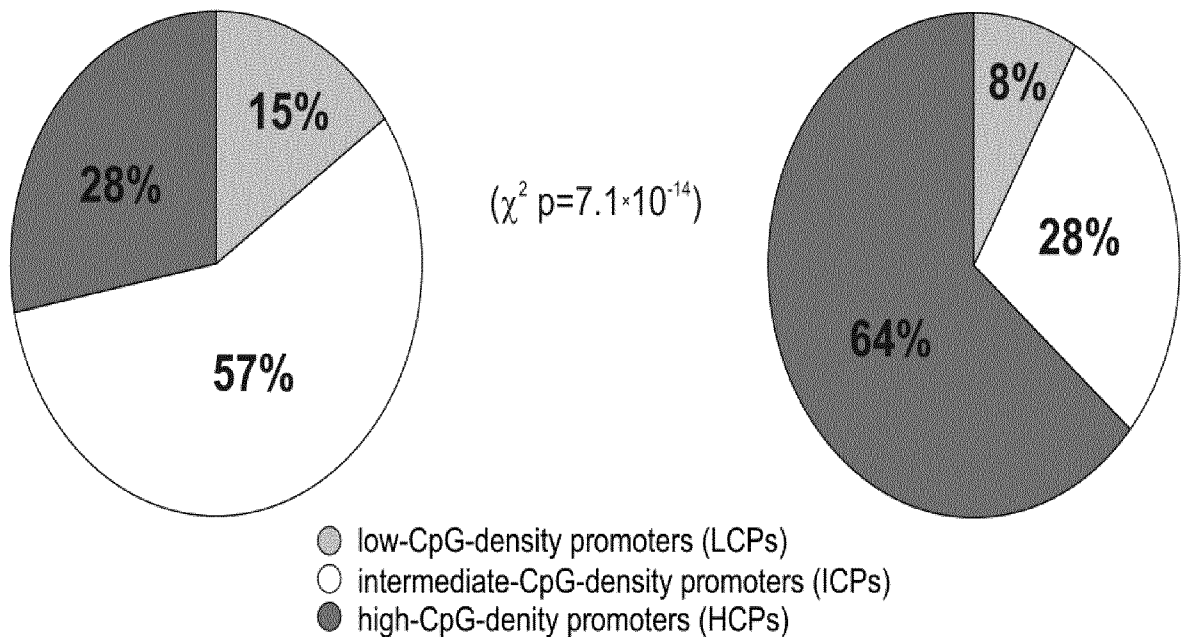


Figure 3

C

Transcriptional regulatory elements prediction

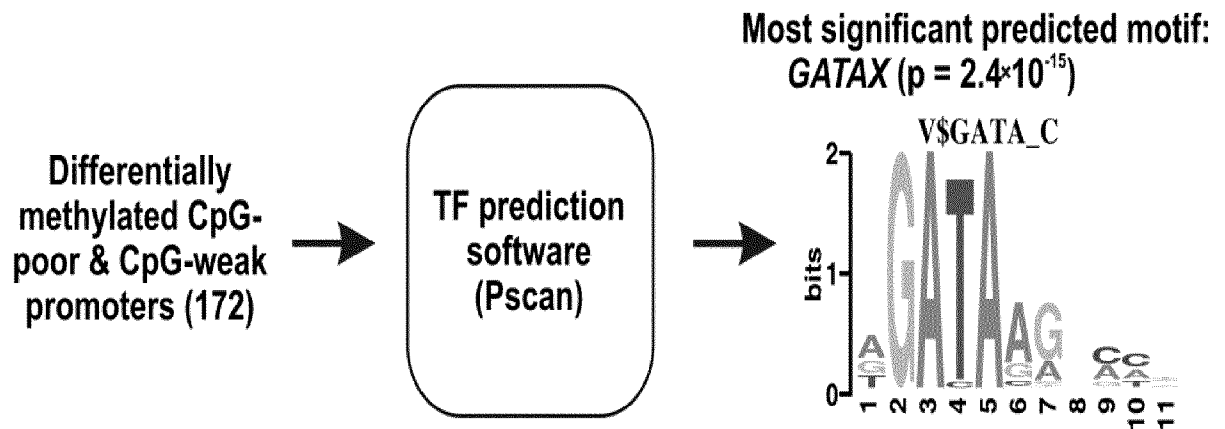


Figure 4

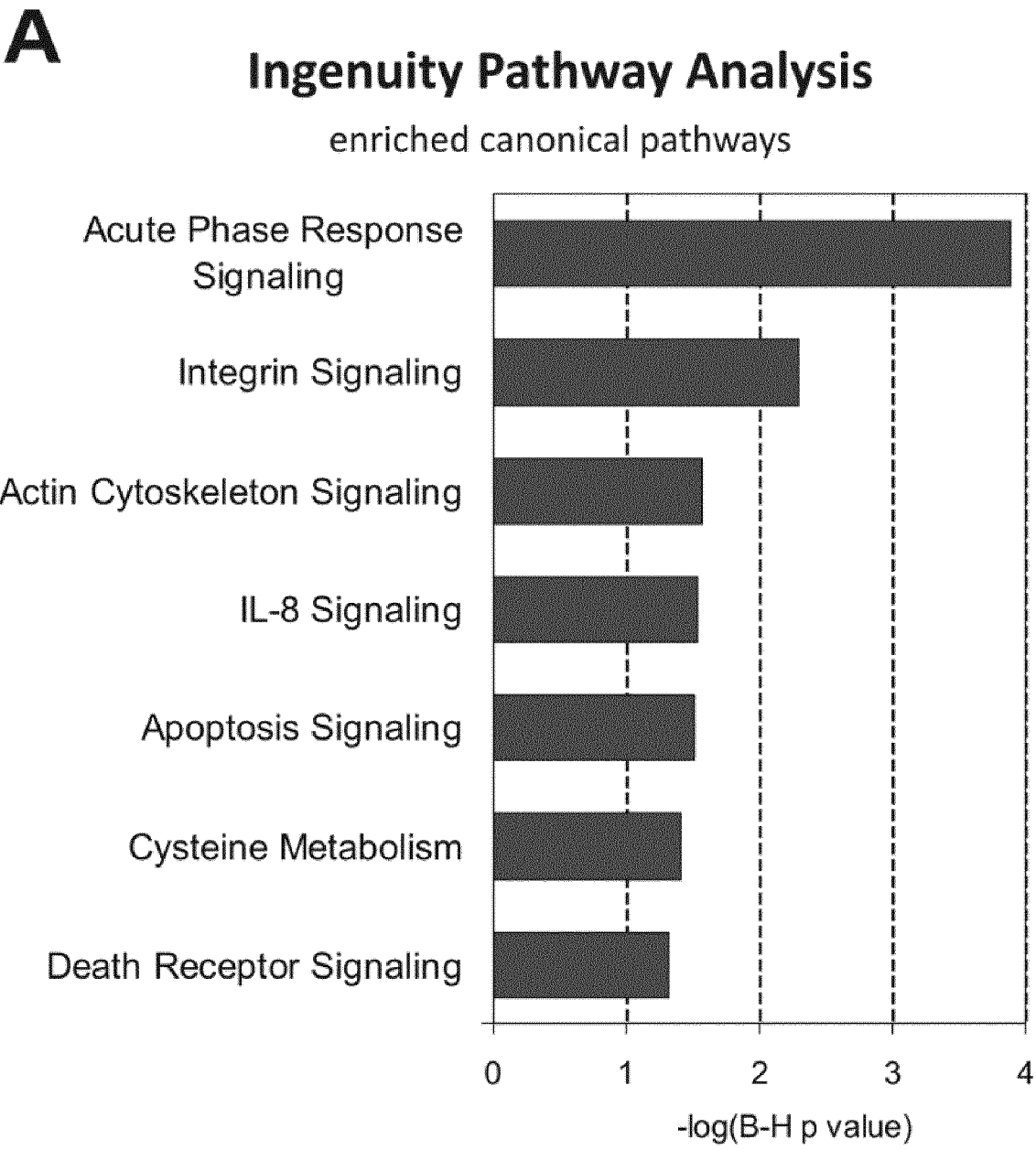
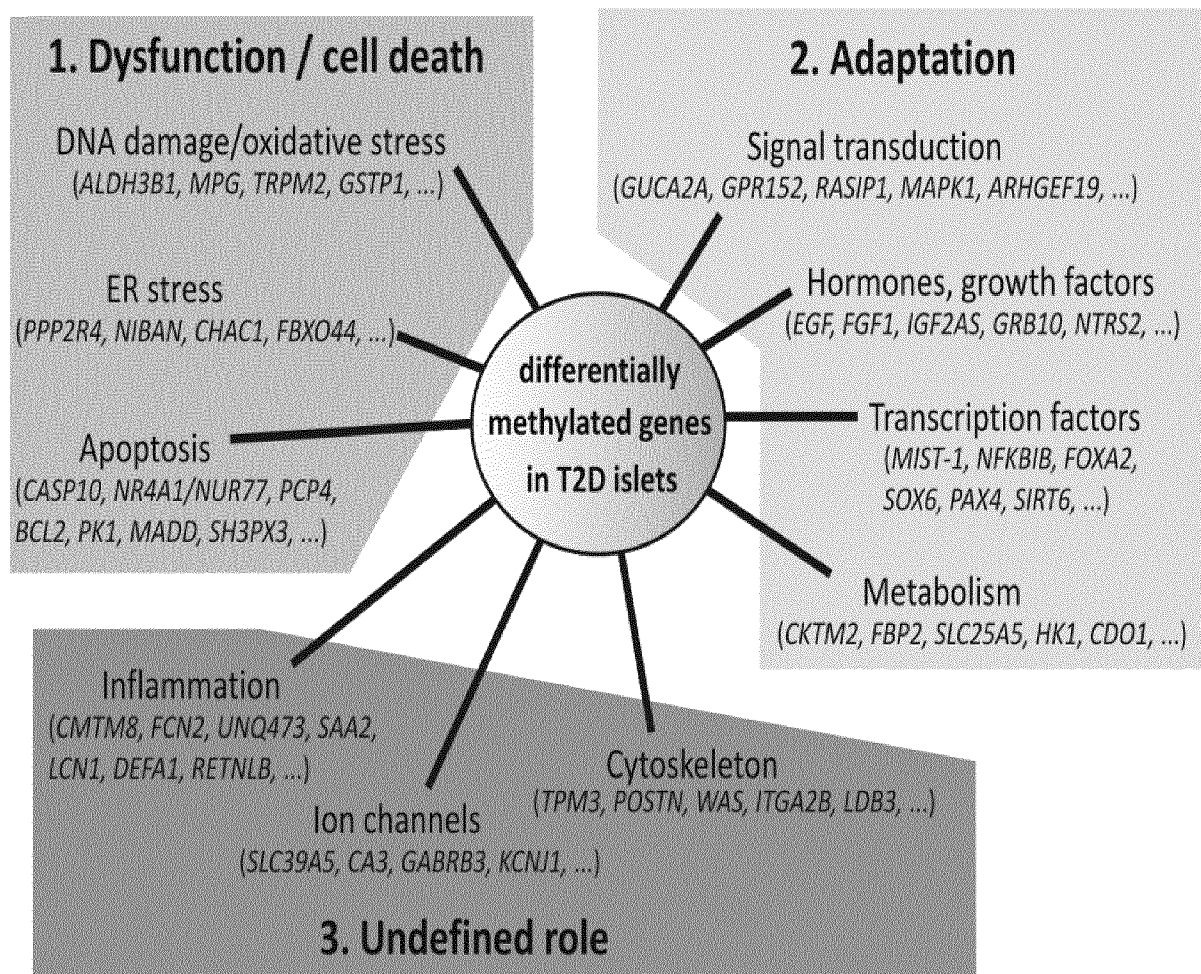


Figure 4**B**

12/31

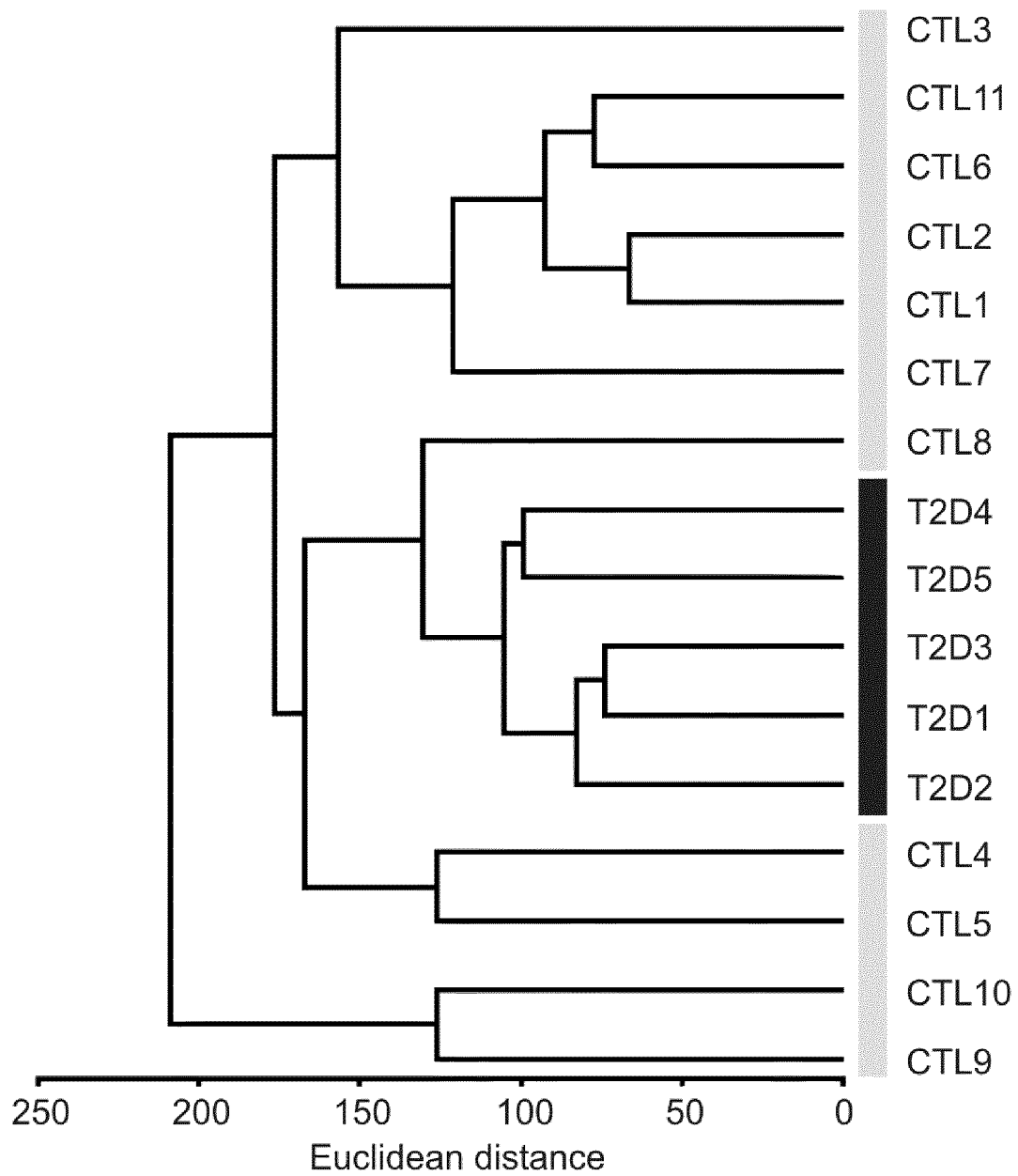
**Figure 5**

Figure 6

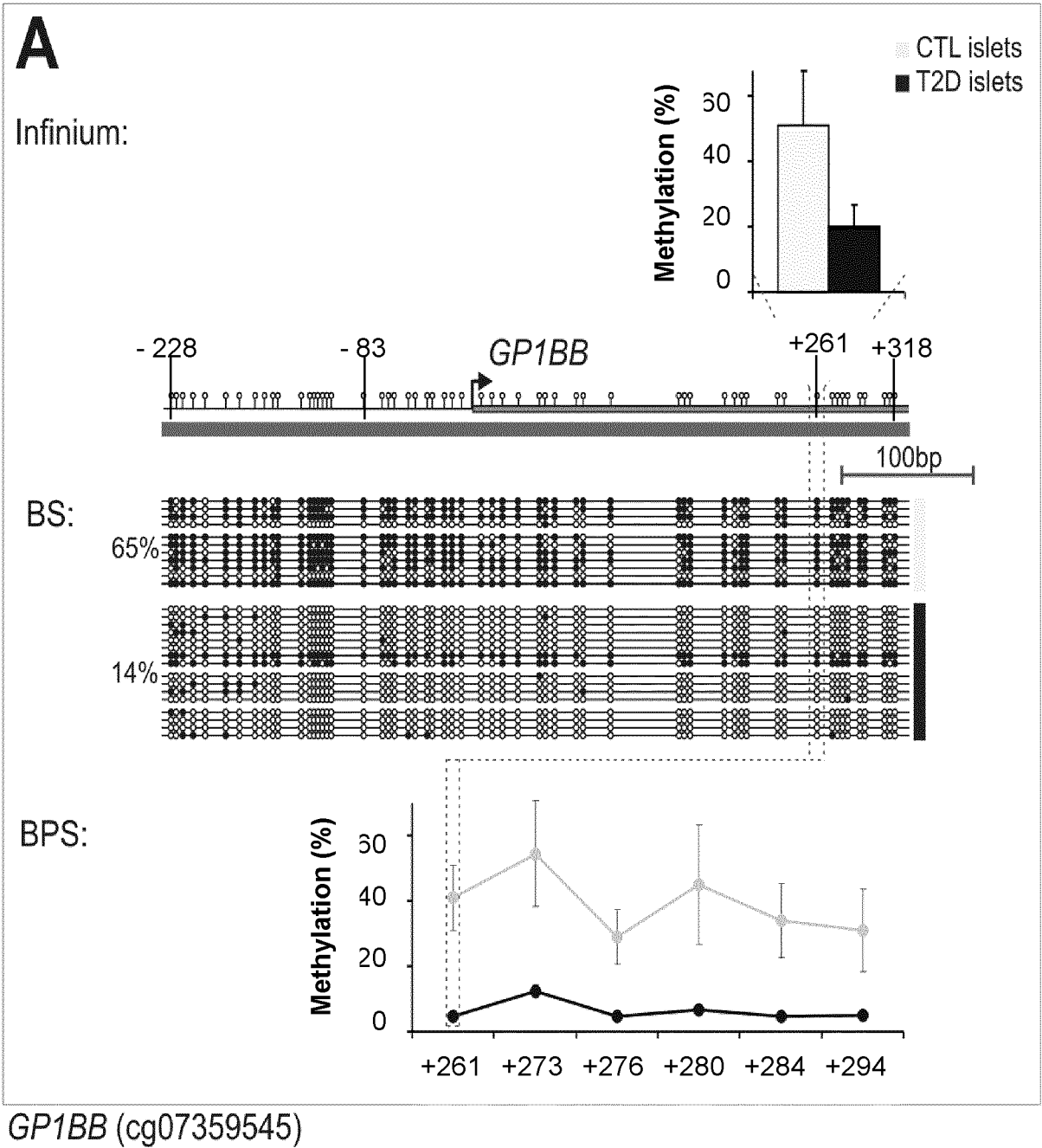


Figure 6

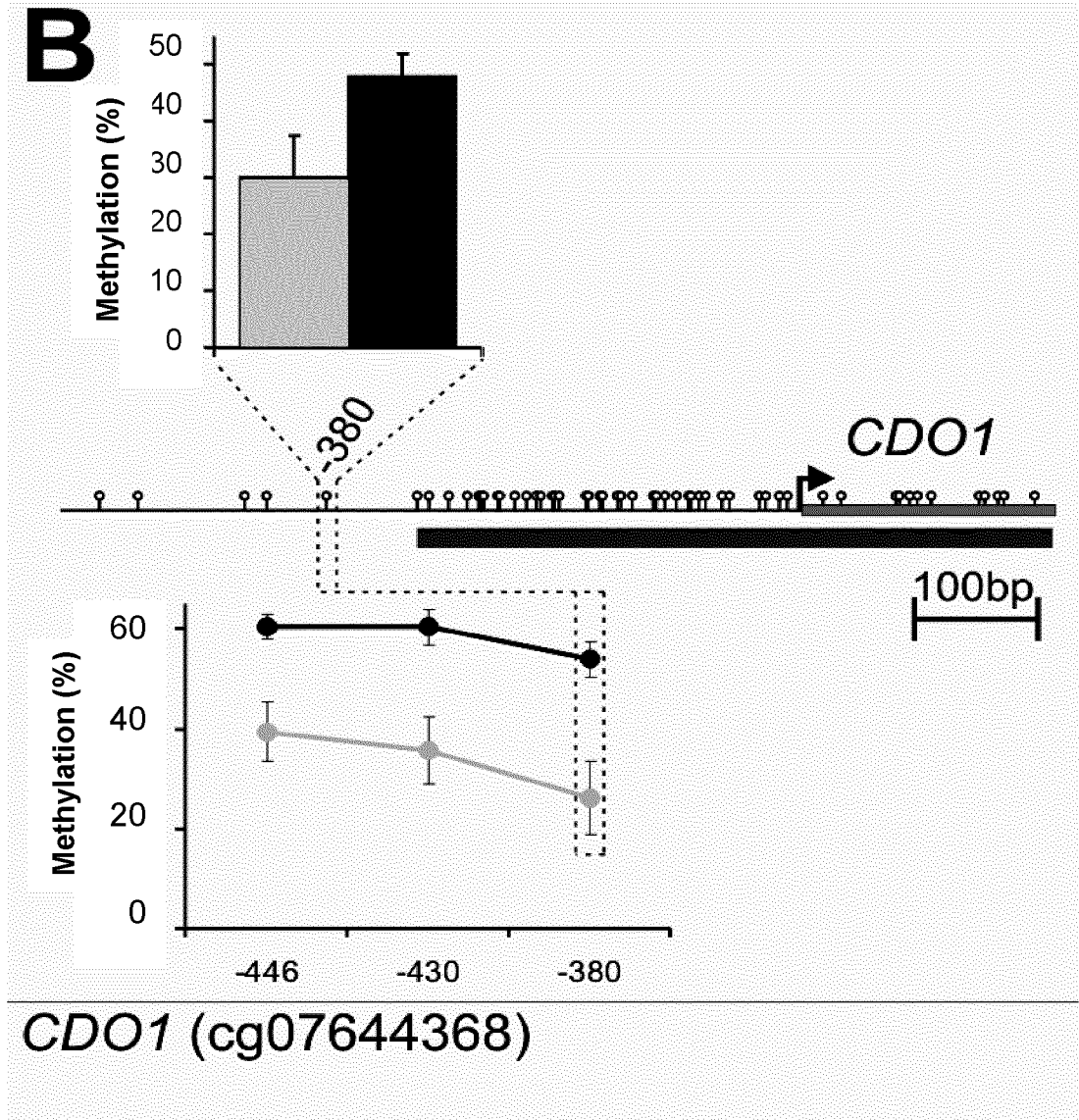


Figure 6 C

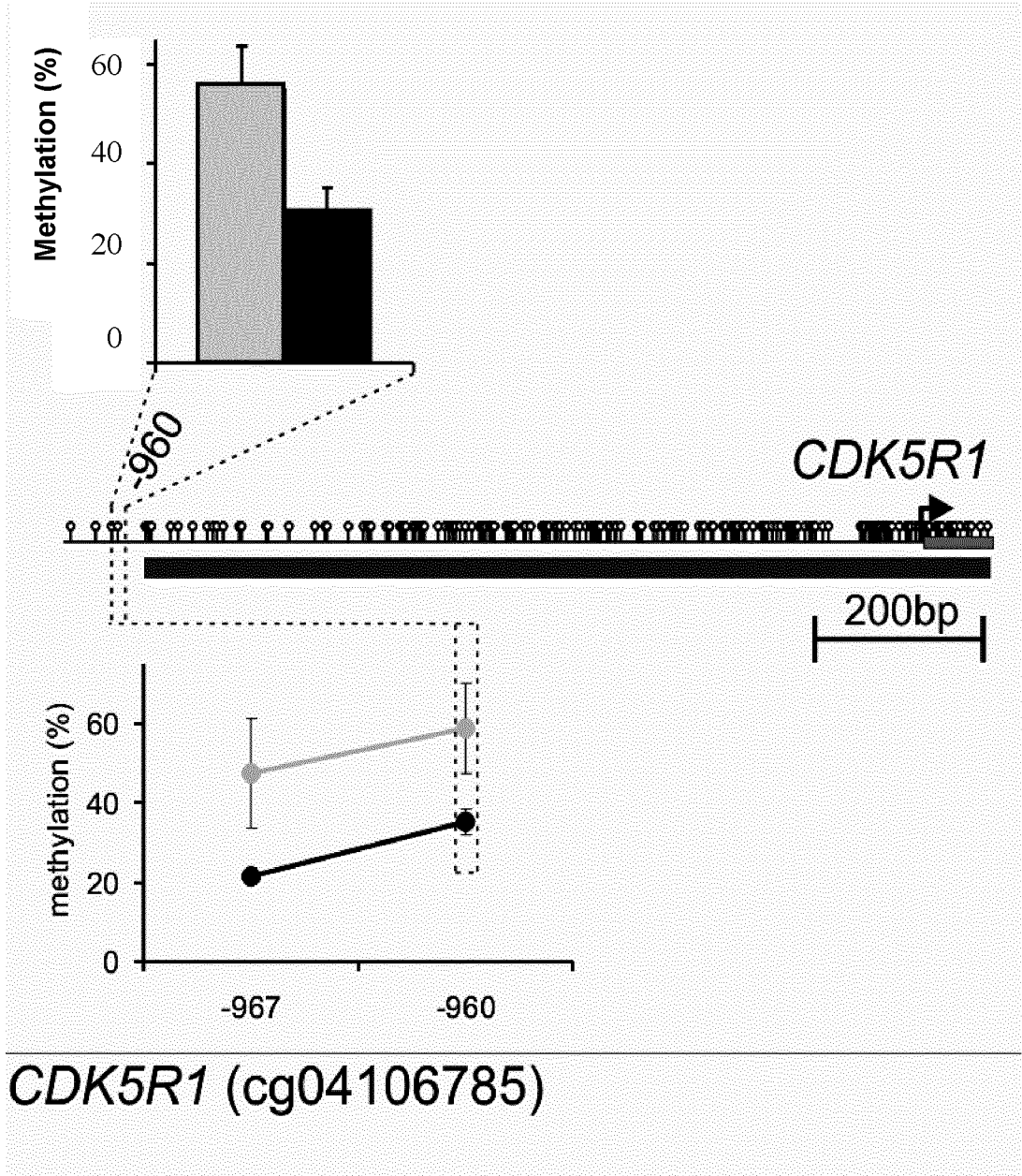
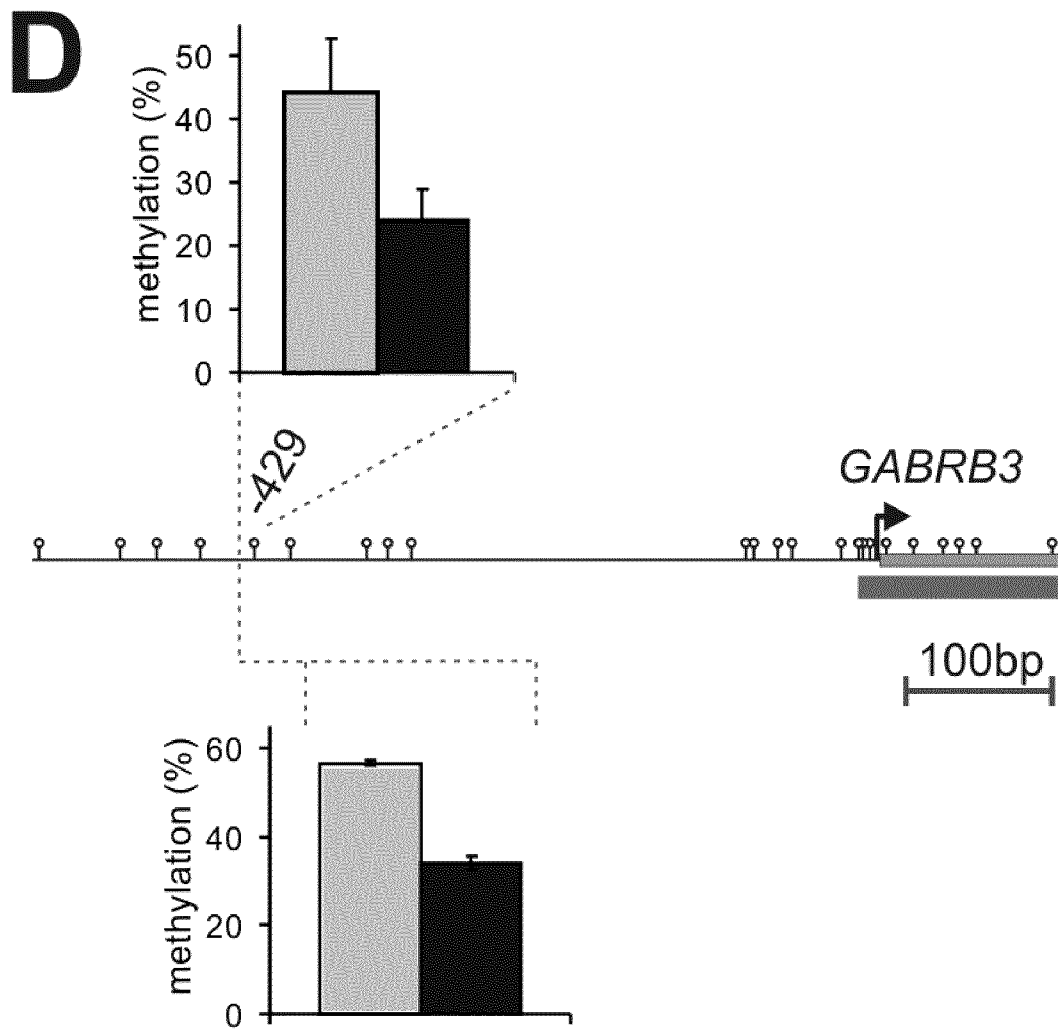


Figure 6



GABRB3 (cg15940569)

Figure 6

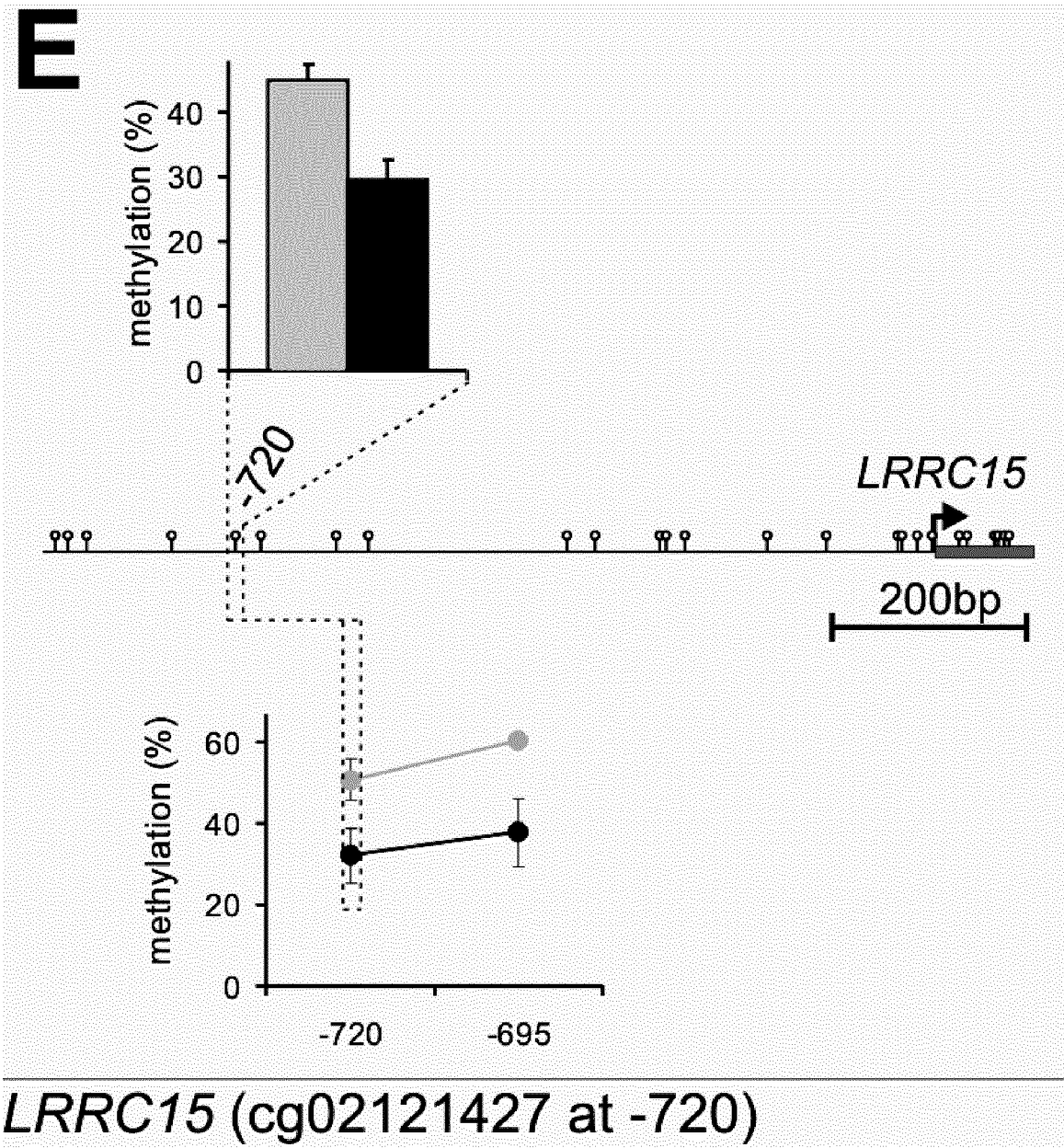


Figure 6

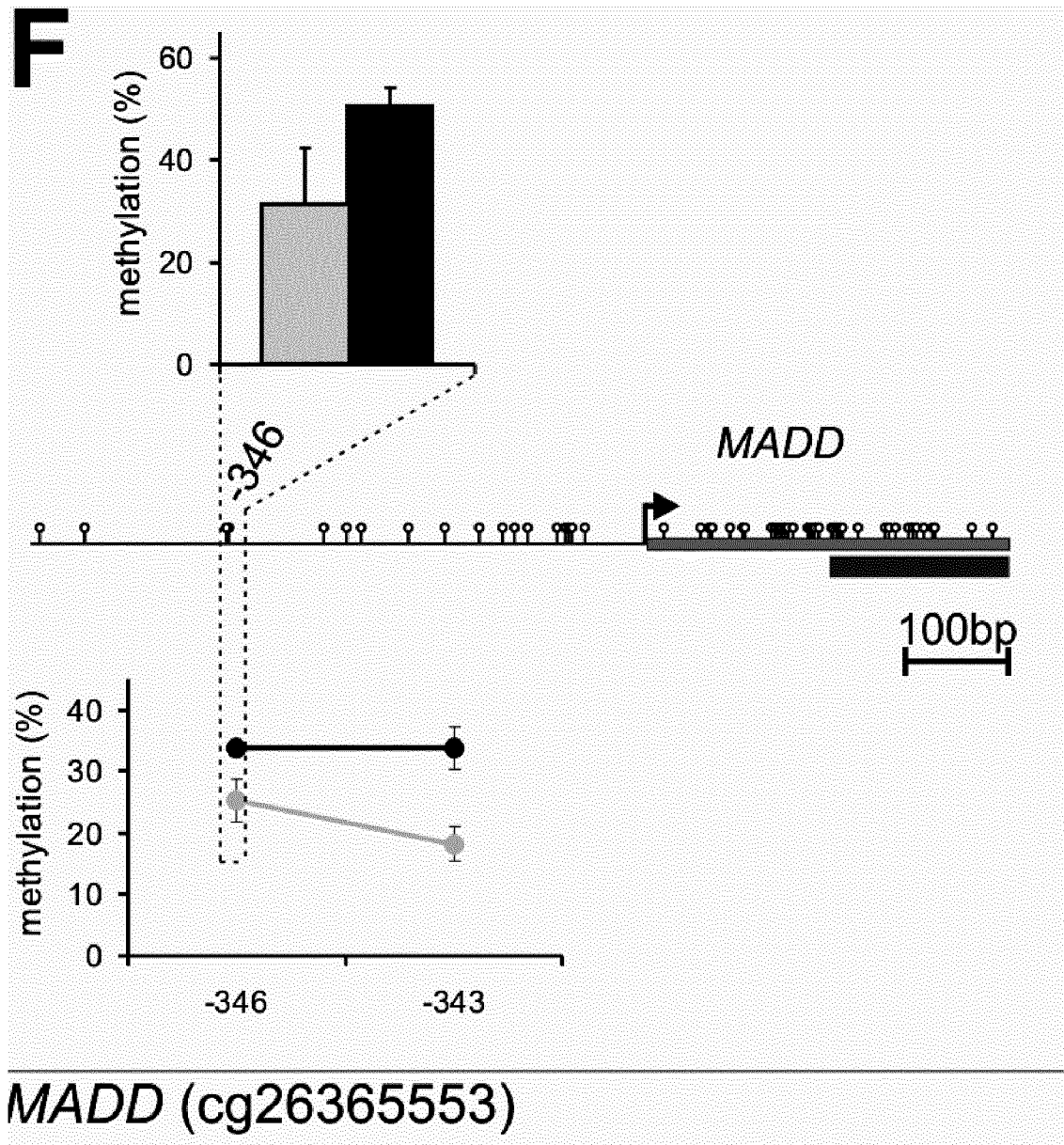
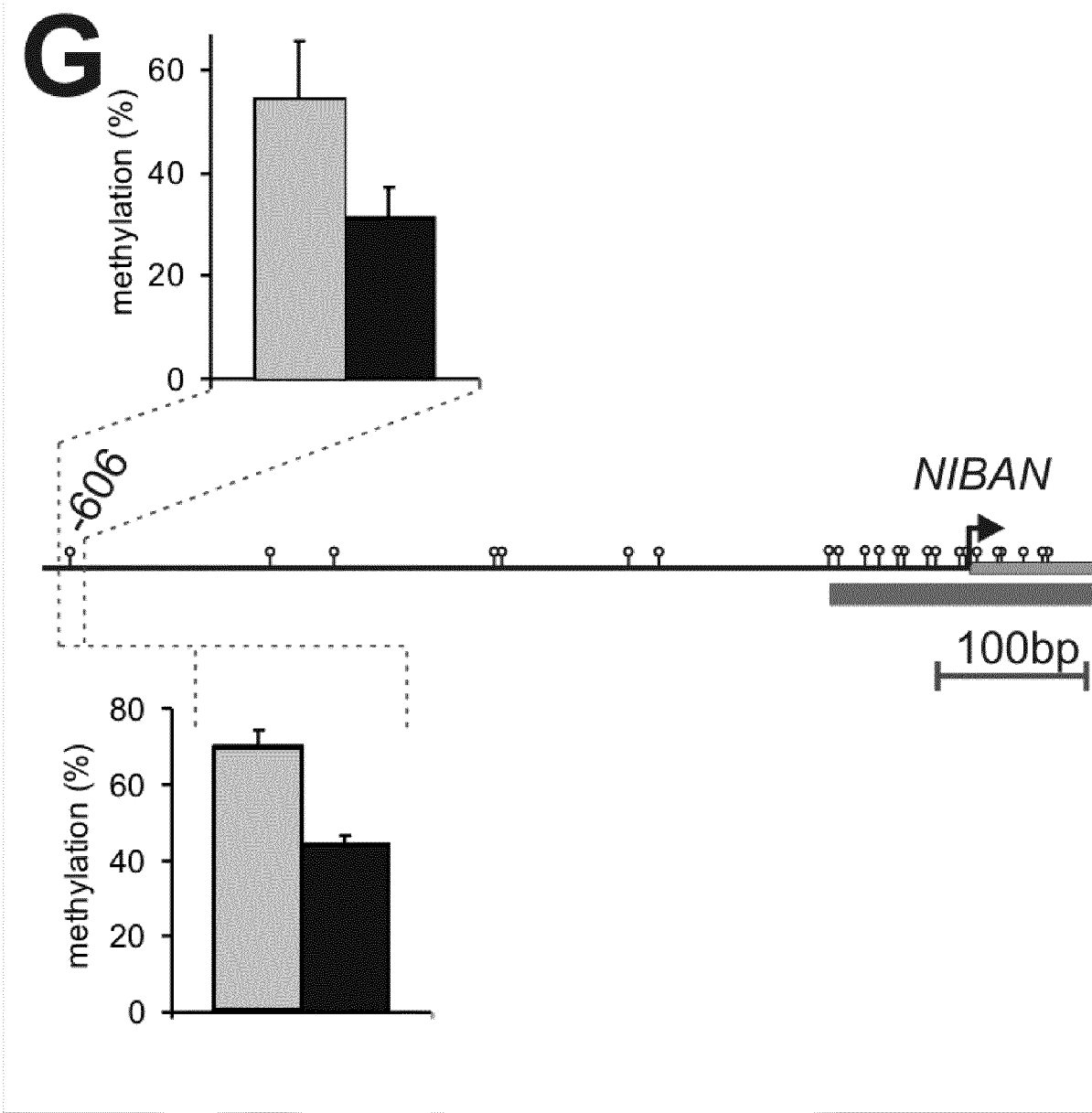


Figure 6



NIBAN / C1orf24 (cg25182523)

Figure 6

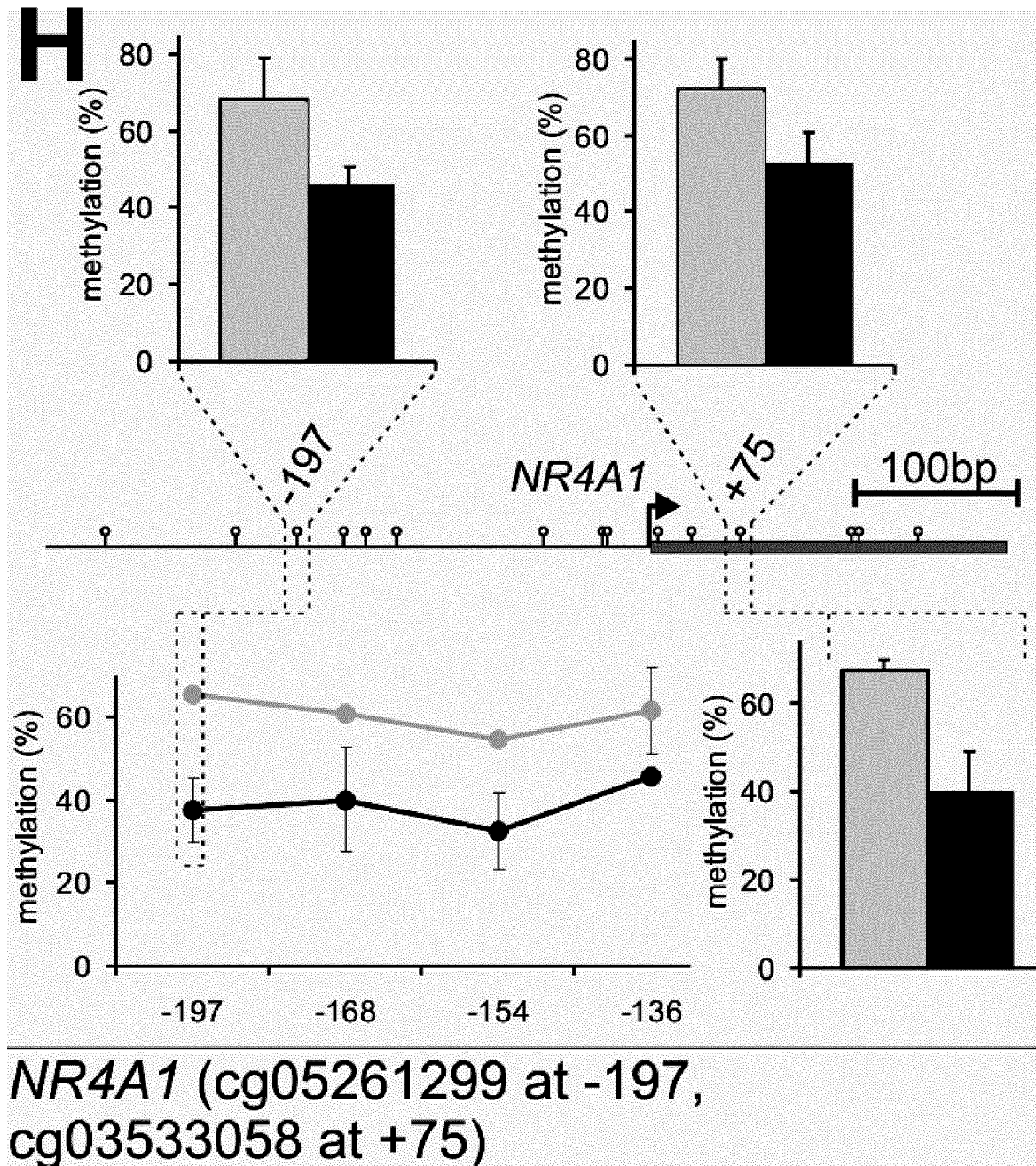


Figure 6

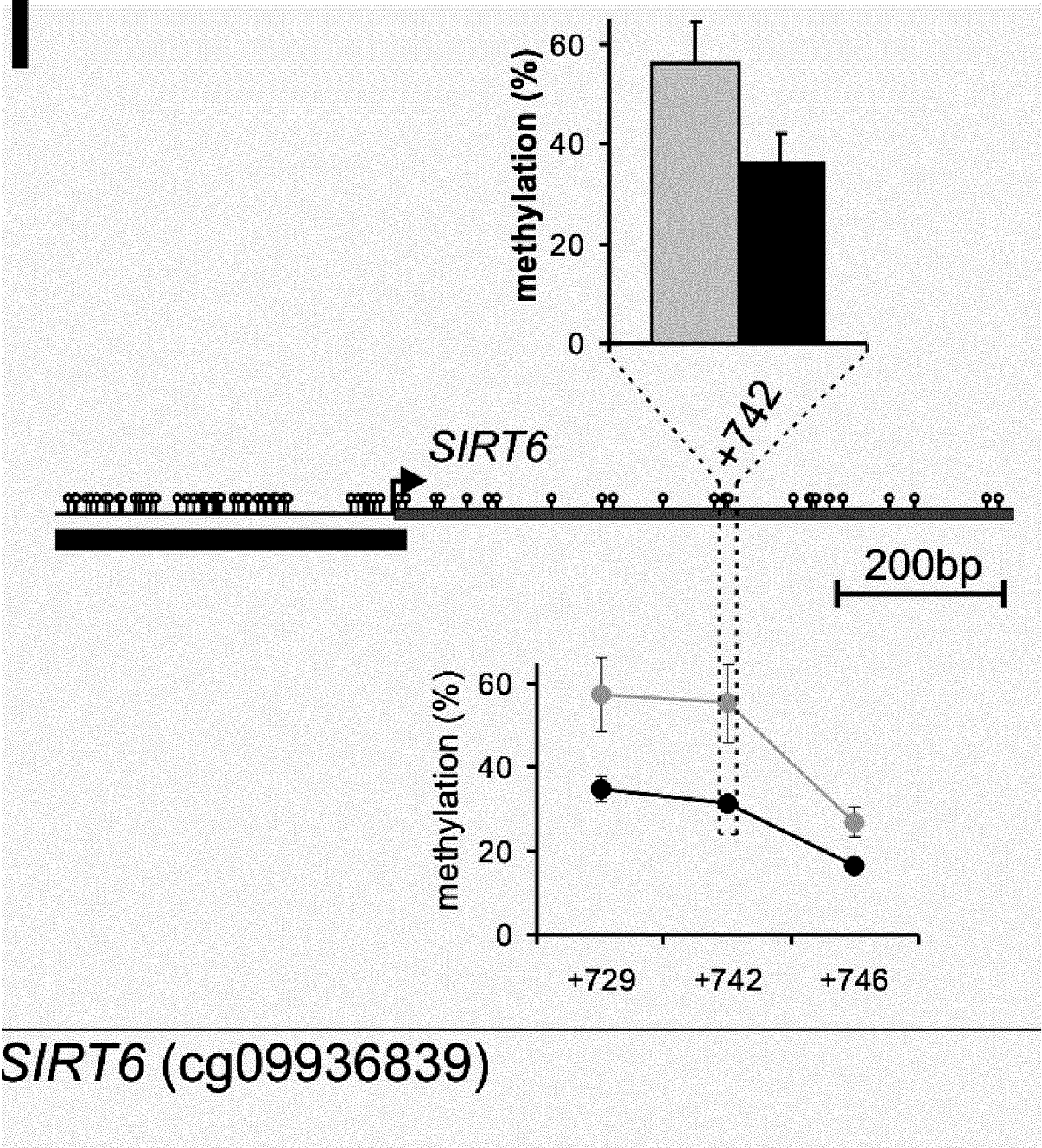


Figure 6

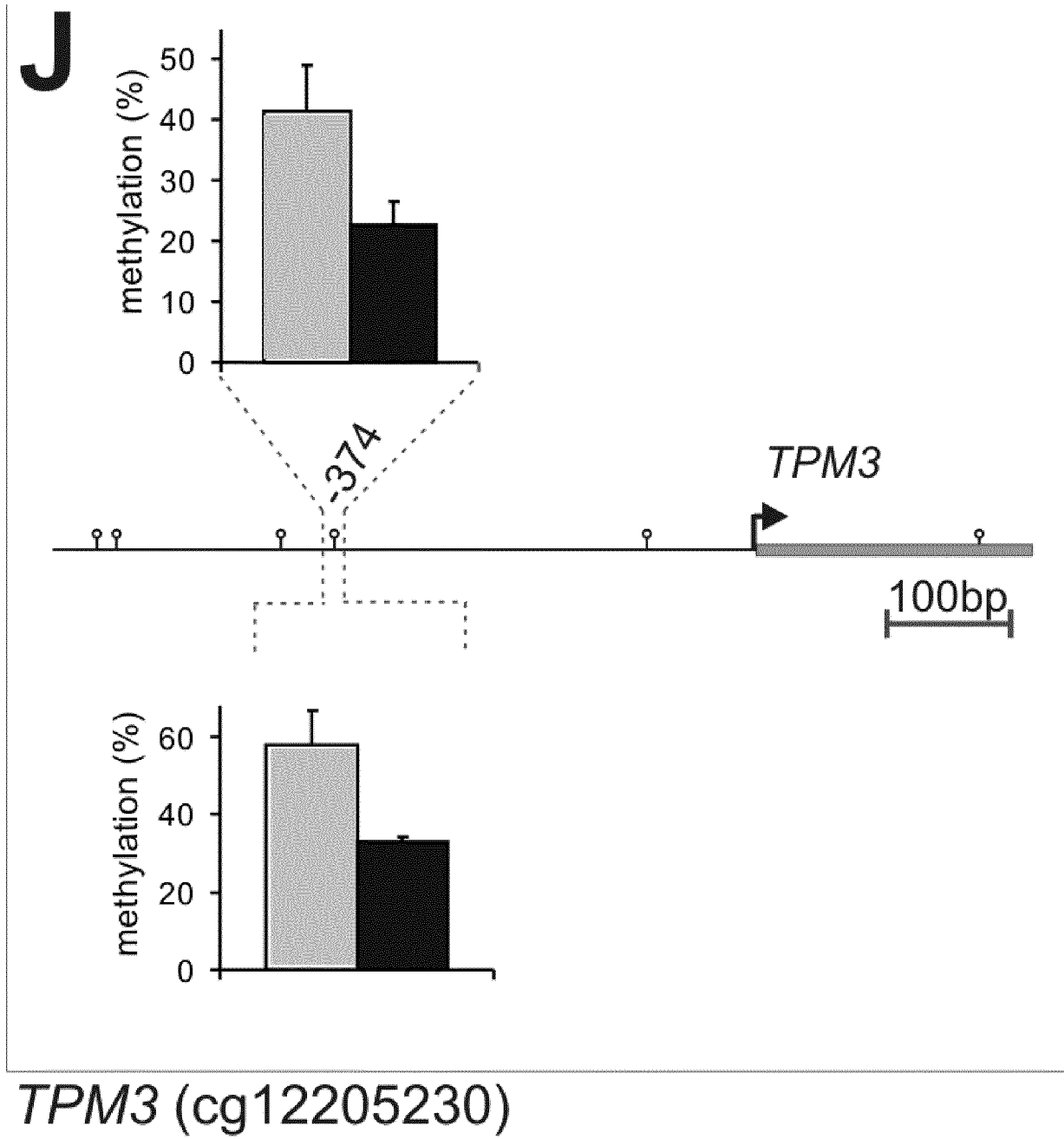


Figure 6

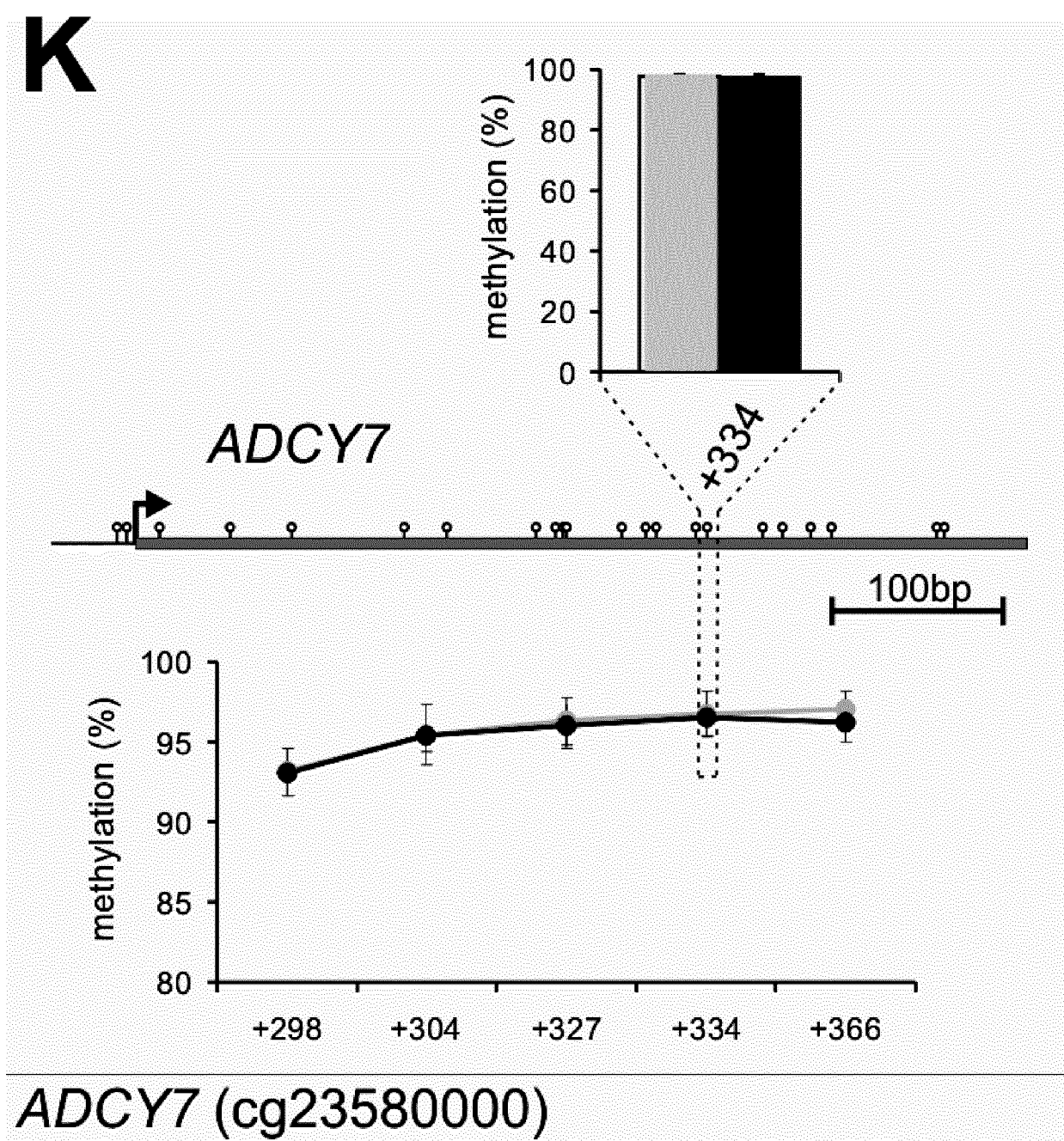


Figure 6

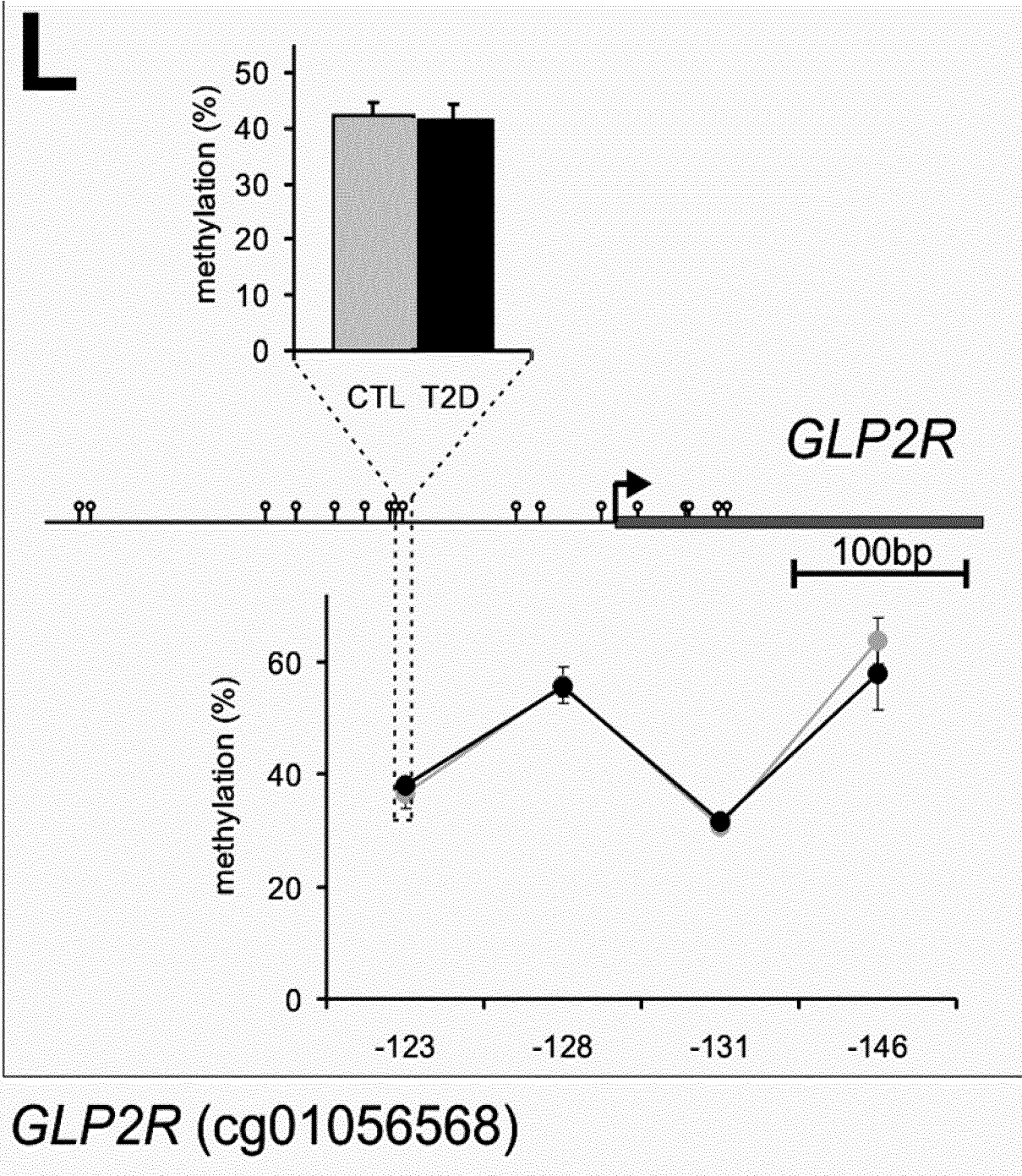
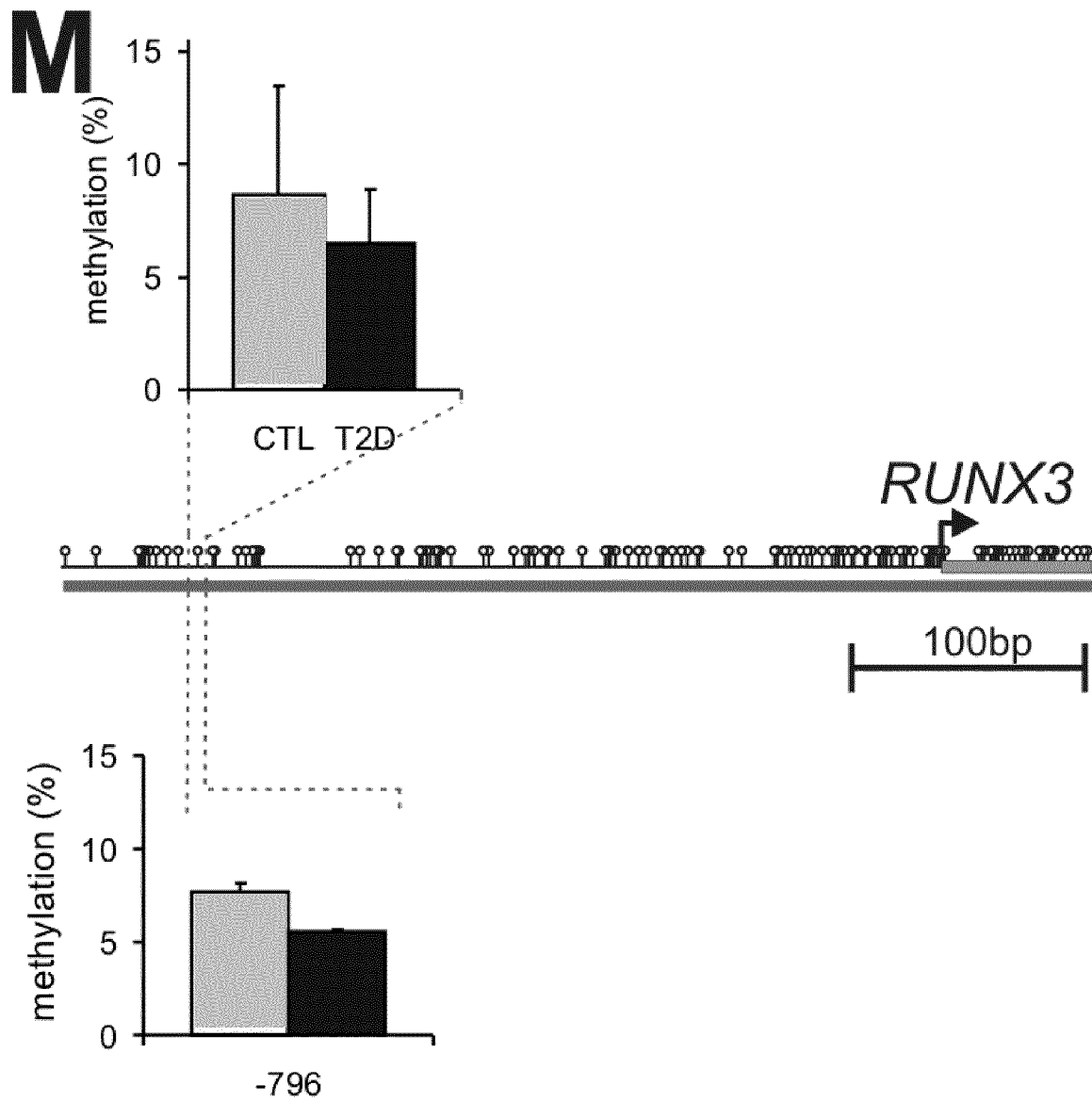


Figure 6



RUNX3 (cg24019564)

Figure 7 A

Most significant TFs predicted by Pscan in differentially methylated set

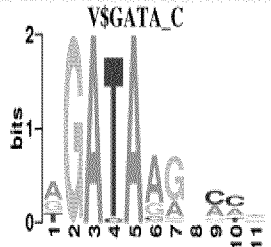
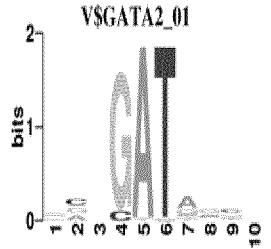
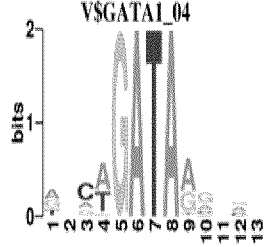
Rank	TRANSFAC Motif ID	TF	P-value	Motif Logo
1	V\$GATA_C	<i>GATA1</i>	2.38E-15	 V\$GATA_C bits 1 2 3 4 5 6 7 8 9 10 11
2	V\$GATA2_02	<i>GATA2</i>	1.33E-13	 V\$GATA2_01 bits 1 2 3 4 5 6 7 8 9 10
3	V\$GATA1_04	<i>GATA1</i>	3.28E-13	 V\$GATA1_04 bits 1 2 3 4 5 6 7 8 9 10 11 12 13

Figure 7 B **Negative control: most significant TFs predicted in differentially methylated set vs distribution of the best p value on random data sets**

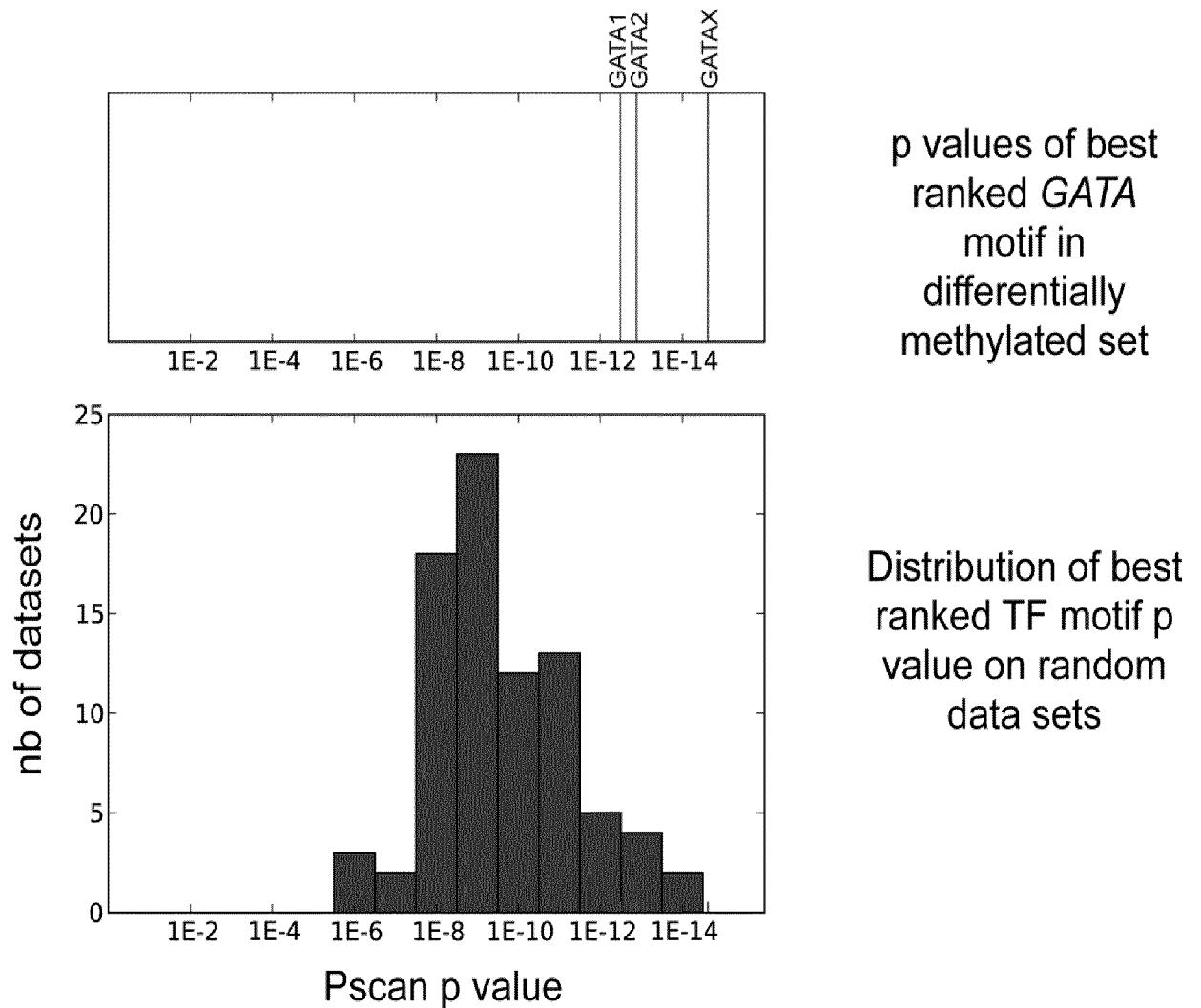


Figure 8

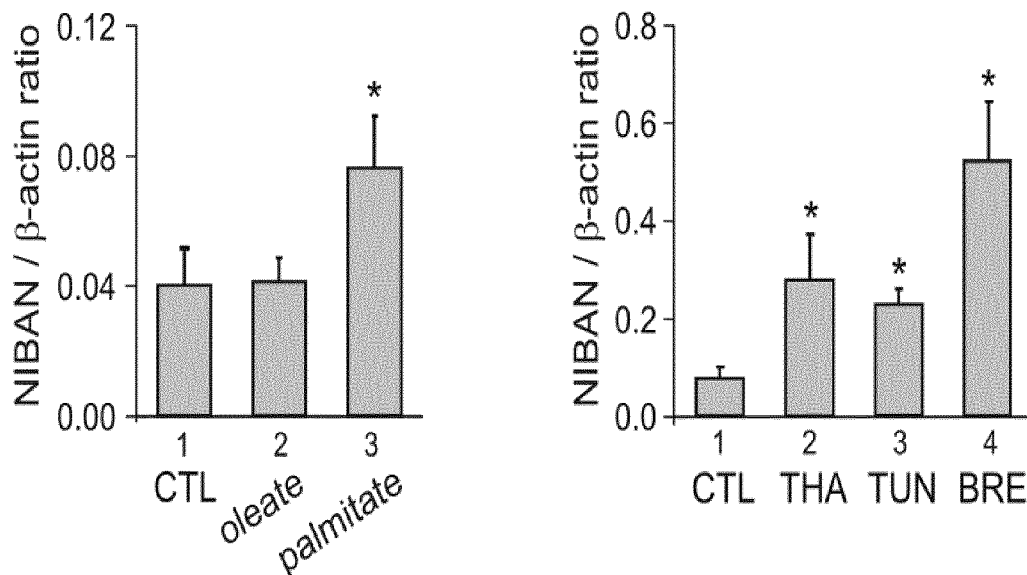
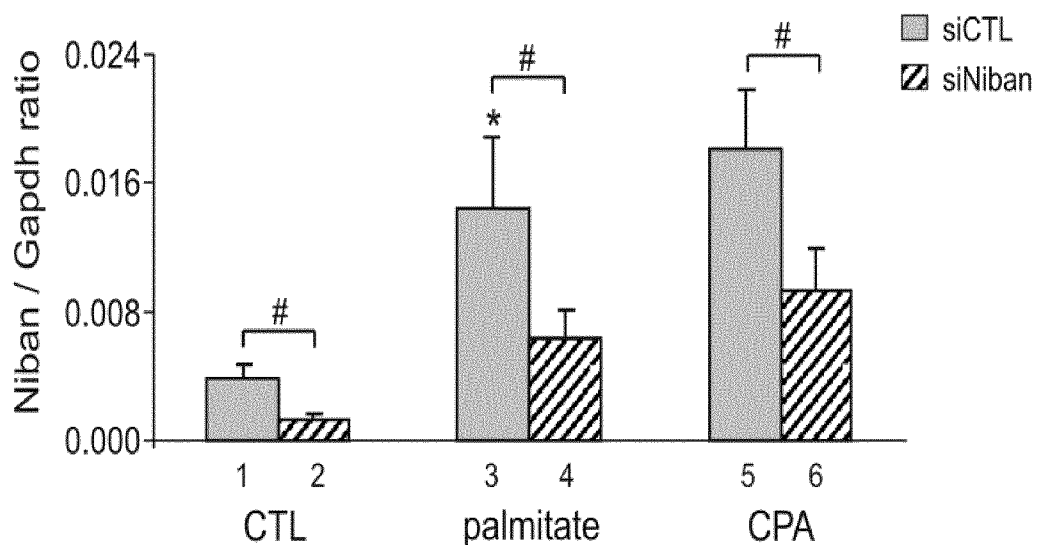
A *NIBAN* expression in human islets upon ER stress**B** Effect of *Niban* knockdown on ER stress in rat INS-1E cells

Figure 8

C ER stress-induced apoptosis in the presence of
Niban knockdown

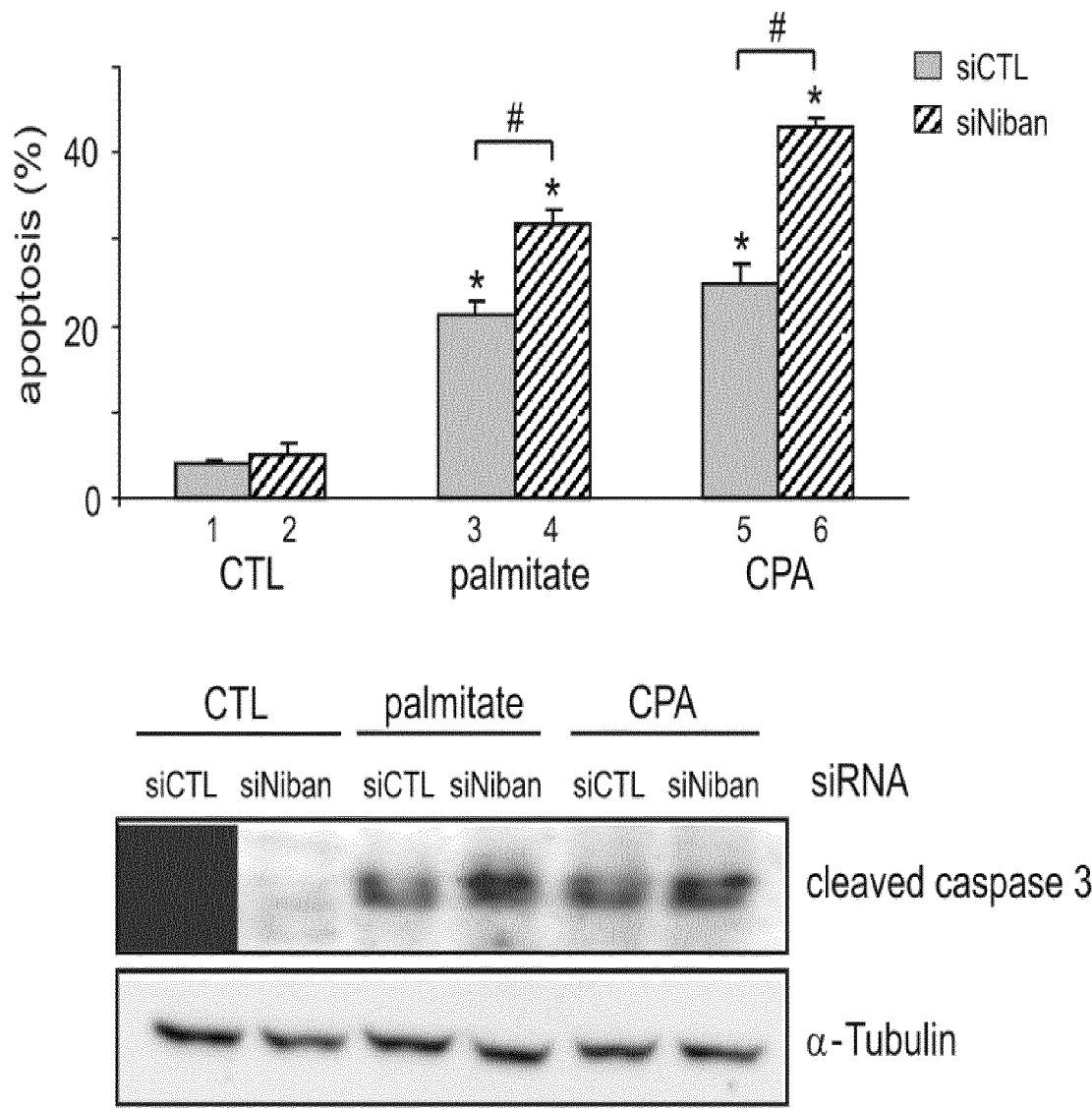


Figure 8

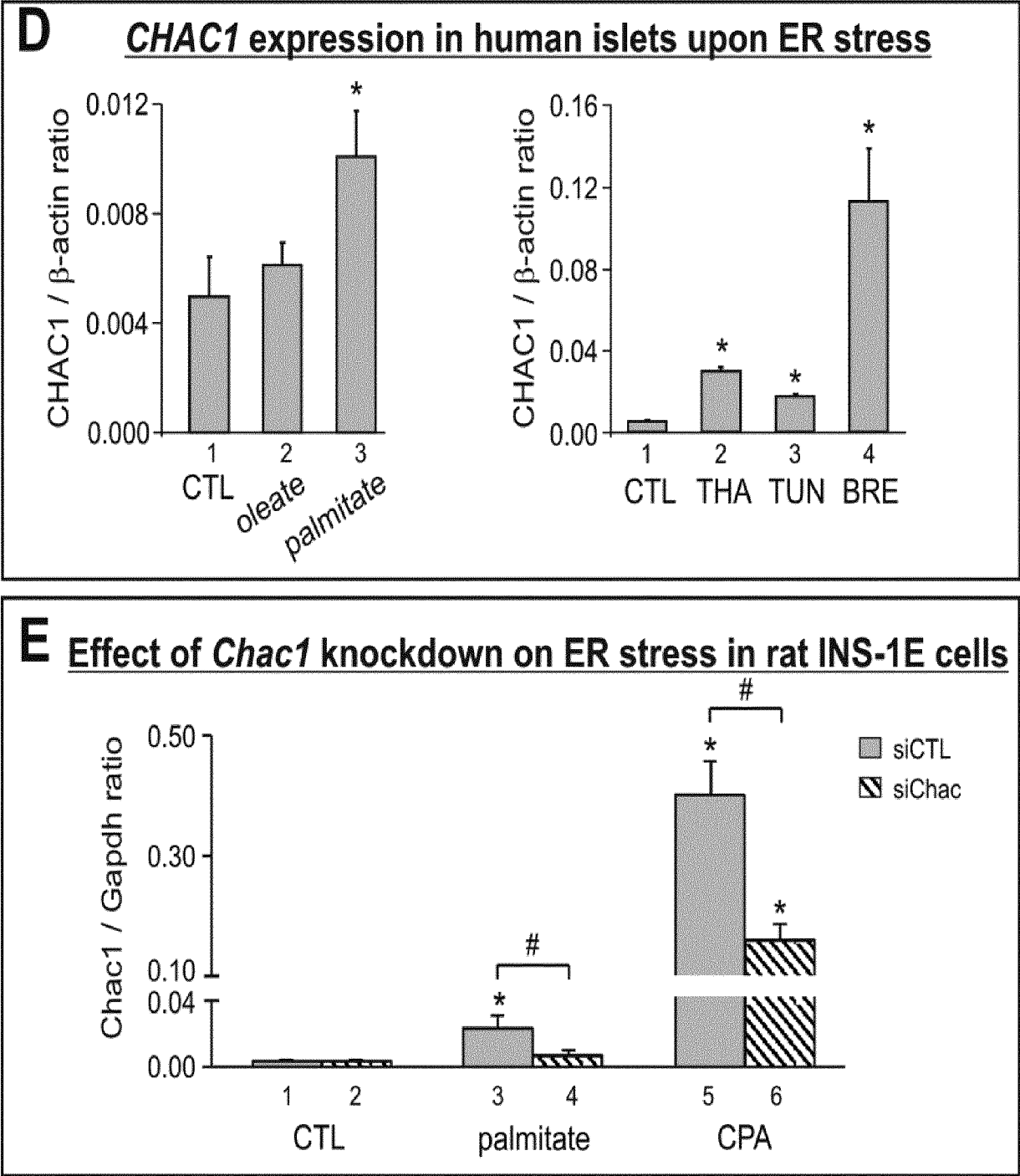
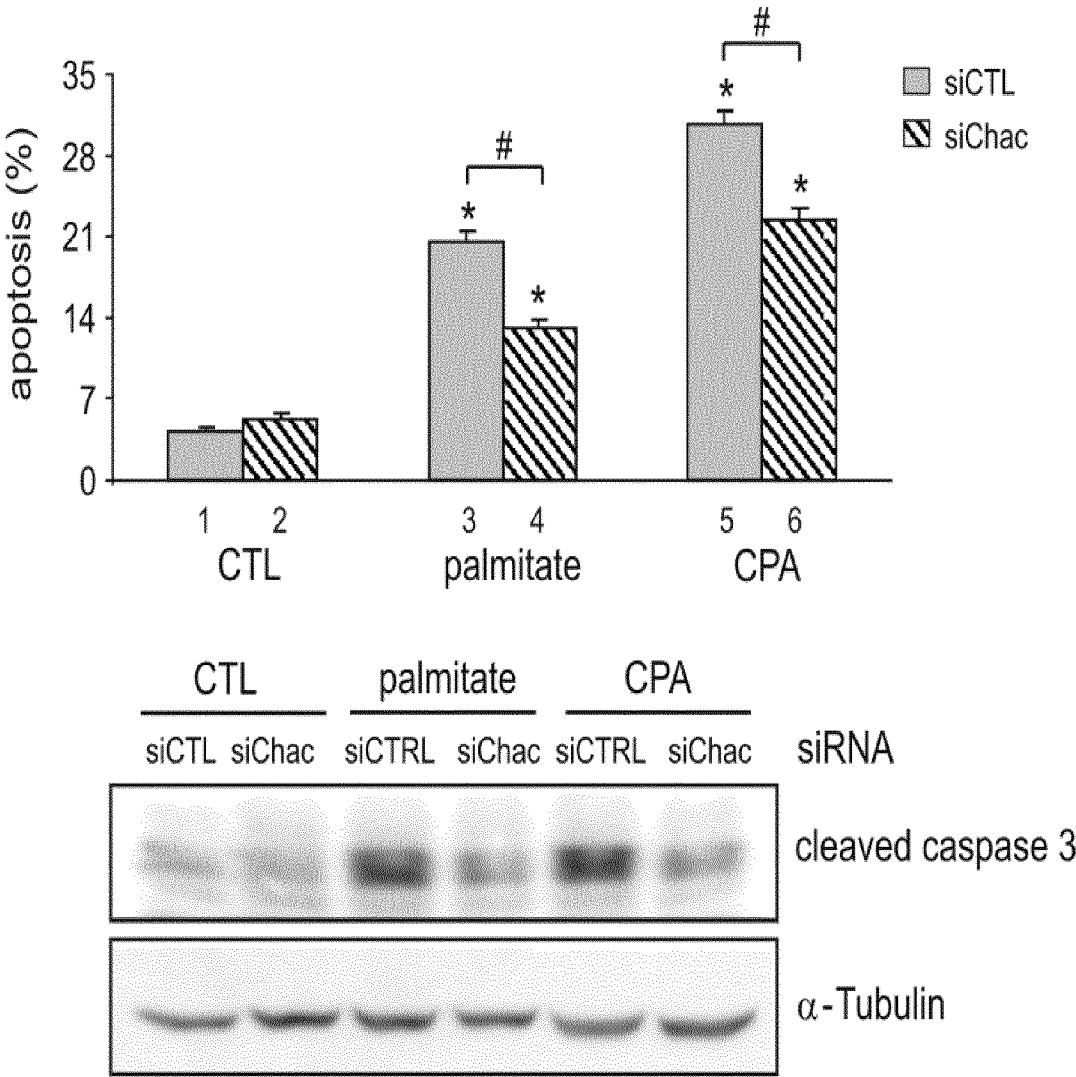


Figure 8

F ER stress-induced apoptosis in the presence of
Chac1 knockdown



INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2011/070796

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EP0-Internal, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	BIBIKOVA MARINA ET AL: "Genome-wide DNA methylation profiling using Infinium (R) assay", EPIGENOMICS, vol. 1, no. 1, October 2009 (2009-10), pages 177-200, XP009158623, abstract page 179, column 2, paragraph 2 - page 181, column 2, paragraph 2 page 182, column 2, paragraph 3 - page 185, column 1 ----- -/-	15-21



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

12 June 2012

Date of mailing of the international search report

18/06/2012

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Tilkorn, A

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2011/070796

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	PARK JUN H ET AL: "Development of type 2 diabetes following intrauterine growth retardation in rats is associated with progressive epigenetic silencing of Pdx1", JOURNAL OF CLINICAL INVESTIGATION, vol. 118, no. 6, June 2008 (2008-06), pages 2316-2324, XP002674723, ISSN: 0021-9738 cited in the application abstract; figures 1-7 -----	1-25
X	LING C ET AL: "Epigenetic regulation of PPARGC1A in human type 2 diabetic islets and effect on insulin secretion", DIABETOLOGIA ; CLINICAL AND EXPERIMENTAL DIABETES AND METABOLISM, SPRINGER, BERLIN, DE, vol. 51, no. 4, 13 February 2008 (2008-02-13), pages 615-622, XP019582981, ISSN: 1432-0428 cited in the application abstract page 619, column 2, paragraph 3 - page 621, column 2, paragraph 3 -----	1-25
X	PINNEY S E ET AL: "Epigenetic mechanisms in the development of type 2 diabetes", TRENDS IN ENDOCRINOLOGY AND METABOLISM, ELSEVIER SCIENCE PUBLISHING, NEW YORK, NY, US, vol. 21, no. 4, 1 April 2010 (2010-04-01), pages 223-229, XP026987124, ISSN: 1043-2760 [retrieved on 2009-10-26] abstract -----	1-25
X	YANG B T ET AL: "Insulin promoter DNA methylation correlates negatively with insulin gene expression and positively with HbA1c levels in human pancreatic islets", DIABETOLOGIA ; CLINICAL AND EXPERIMENTAL DIABETES AND METABOLISM, SPRINGER, BERLIN, DE, vol. 54, no. 2, 23 November 2010 (2010-11-23), pages 360-367, XP019871269, ISSN: 1432-0428, DOI: 10.1007/S00125-010-1967-6 abstract page 363, column 2, paragraph 2 - page 366, column 1, paragraph 3; figure 2 ----- -/--	1-25

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2011/070796

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CHRISTOPHER G. BELL ET AL: "Integrated Genetic and Epigenetic Analysis Identifies Haplotype-Specific Methylation in the FTO Type 2 Diabetes and Obesity Susceptibility Locus", PLOS ONE, vol. 5, no. 11, 1 January 2010 (2010-01-01), page E14040, XP055024141, ISSN: 1932-6203, DOI: 10.1371/journal.pone.0014040 the whole document	1-25
A	KELLY MICHAEL D ET AL: "Complementary DNA cloning of the alternatively expressed endothelial cell glycoprotein Ib-beta (GPIb-beta) and localization of the GPIb-beta gene to chromosome 22", JOURNAL OF CLINICAL INVESTIGATION, vol. 93, no. 6, 1994, pages 2417-2424, XP002674724, ISSN: 0021-9738 abstract; figure 4	1-25
T	MICHAEL VOLKMAR ET AL: "DNA methylation profiling identifies epigenetic dysregulation in pancreatic islets from type 2 diabetic patients", THE EMBO JOURNAL, vol. 31, no. 6, 1 January 2012 (2012-01-01), pages 1405-1426, XP055023960, ISSN: 0261-4189, DOI: 10.1038/emboj.2011.503 the whole document	1-25
A	GEMMA C ET AL: "Methylation of TFAM gene promoter in peripheral white blood cells is associated with insulin resistance in adolescents", MOLECULAR GENETICS AND METABOLISM, ACADEMIC PRESS, SAN DIEGO, CA, US, vol. 100, no. 1, 1 May 2010 (2010-05-01), pages 83-87, XP027212134, ISSN: 1096-7192 [retrieved on 2010-02-12] the whole document	1-25
	----- -/--	

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2011/070796

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	MICHAEL L. STITZEL ET AL: "Global Epigenomic Analysis of Primary Human Pancreatic Islets Provides Insights into Type 2 Diabetes Susceptibility Loci", CELL METABOLISM, vol. 12, no. 5, 1 November 2010 (2010-11-01), pages 443-455, XP055024096, ISSN: 1550-4131, DOI: 10.1016/j.cmet.2010.09.012 the whole document -----	1-25

INTERNATIONAL SEARCH REPORT

International application No.

PCT/EP2011/070796

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application and necessary to the claimed invention, the international search was carried out on the basis of:
 - a. (means)
☐ on paper
☒ in electronic form
 - b. (time)
☒ in the international application as filed
☐ together with the international application in electronic form
☐ subsequently to this Authority for the purpose of search
2. ☐ In addition, in the case that more than one version or copy of a sequence listing and/or table relating thereto has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.
3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2011/070796

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-25(partially)

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-25(partially)

A method for diagnosis, prognosis, therapy or prediction of T2D or follow up of T2D intervention therapies on the basis of the methylation status of the CpG site as defined in SEQ DI NO: 1. A method of identifying a modulating agent for aberrantly methylated GP1BB, a method for establishing a reference methylation status profile and a method for identifying T2D specific SNPs on the basis of SEQ Id NO: 1, use of the methylation status and a microarray comprising SEQ ID NO: 1.

2. claims: 1-25(partially)

A method for diagnosis, prognosis, therapy or prediction of T2D or follow up of T2D intervention therapies on the basis of the methylation status of the CpG site as defined in SEQ DI NO: 2. A method of identifying a modulating agent for aberrantly methylated ARL14 a method for establishing a reference methylation status profile and a method for identifying T2D specific SNPs on the basis of SEQ Id NO: 2, use of the methylation status and a microarray comprising SEQ ID NO: 2.

3-254. claims: 1-25(partially)

A method for diagnosis, prognosis, therapy or prediction of T2D or follow up of T2D intervention therapies on the basis of the methylation status of a gene promoter wherein the respective CpG sites are defined in SEQ DI NOs: 3-276. A method of identifying a modulating agent for aberrantly methylated respective marker gene, a method for establishing a reference methylation status profile and a method for identifying T2D specific SNPs on the basis of one of SEQ Id NOs: 3-276, use of the methylation status and a microarray comprising one of SEQ ID NOs: 3-276
