

This Provisional PDF corresponds to the article as it appeared upon acceptance. Fully formatted PDF and full text (HTML) versions will be made available soon.

Genome-wide DNA methylation analysis for diabetic nephropathy in type 1 diabetes mellitus

BMC Medical Genomics 2010, **3**:33 doi:10.1186/1755-8794-3-33

Christopher G Bell (christopher.bell@ucl.ac.uk)
Andrew E Teschendorff (a.teschendorff@ucl.ac.uk)
Vardhman K Rakyan (v.rakyan@qmul.ac.uk)
Alexander P Maxwell (A.P.Maxwell@qub.ac.uk)
Stephan Beck (s.beck@ucl.ac.uk)
David A Savage (d.savage@qub.ac.uk)

ISSN 1755-8794

Article type Research article

Submission date 4 February 2010

Acceptance date 5 August 2010

Publication date 5 August 2010

Article URL <http://www.biomedcentral.com/1755-8794/3/33>

Like all articles in BMC journals, this peer-reviewed article was published immediately upon acceptance. It can be downloaded, printed and distributed freely for any purposes (see copyright notice below).

Articles in BMC journals are listed in PubMed and archived at PubMed Central.

For information about publishing your research in BMC journals or any BioMed Central journal, go to

<http://www.biomedcentral.com/info/authors/>

Genome-wide DNA methylation analysis for diabetic nephropathy in type 1 diabetes mellitus

Christopher G. Bell¹, Andrew E. Teschendorff¹, Vardhman K. Rakyan², Alexander P. Maxwell³, Stephan Beck¹, David A. Savage³

¹Medical Genomics, UCL Cancer Institute, University College London, London, United Kingdom

²Institute of Cell and Molecular Science, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom

³Nephrology Research Group, Centre for Public Health, Queen's University Belfast, Belfast, Northern Ireland, United Kingdom

CGB: christopher.bell@ucl.ac.uk

AT: a.teschendorff@ucl.ac.uk

VKR: v.rakyan@qmul.ac.uk

APM: a.p.maxwell@qub.ac.uk

SB: s.beck@ucl.ac.uk

DAS: david.savage@belfasttrust.hscni.net

Corresponding authors: Dr. Christopher G. Bell (christopher.bell@ucl.ac.uk) & Dr. David A. Savage (david.savage@belfasttrust.hscni.net)

Abstract

Background: Diabetic nephropathy is a serious complication of diabetes mellitus and is associated with considerable morbidity and high mortality. There is increasing evidence to suggest that dysregulation of the epigenome is involved in diabetic nephropathy. We assessed whether epigenetic modification of DNA methylation is associated with diabetic nephropathy in a case-control study of 192 Irish patients with type 1 diabetes mellitus (T1D). Cases had T1D and nephropathy whereas controls had T1D but no evidence of renal disease.

Methods: We performed DNA methylation profiling in bisulphite converted DNA from cases and controls using the recently developed Illumina Infinium® HumanMethylation27 BeadChip, that enables the direct investigation of 27,578 individual cytosines at CpG loci throughout the genome, which are focused on the promoter regions of 14,495 genes.

Results: Singular Value Decomposition (SVD) analysis indicated that significant components of DNA methylation variation correlated with patient age, time to onset of diabetic nephropathy, and sex. Adjusting for confounding factors using multivariate Cox-regression analyses, and with a false discovery rate (FDR) of 0.05, we observed 19 CpG sites that demonstrated correlations with time to development of diabetic nephropathy. Of note, this included one CpG site located 18 bp upstream of the transcription start site of *UNC13B*, a gene in which the first intronic SNP rs13293564 has recently been reported to be associated with diabetic nephropathy.

Conclusion: This high throughput platform was able to successfully interrogate the methylation state of individual cytosines and identified 19 prospective CpG sites associated with risk of diabetic nephropathy. These differences in DNA methylation are worthy of further follow-up in replication studies using larger cohorts of diabetic patients with and without nephropathy.

Background

Diabetic nephropathy is a serious microvascular complication of diabetes and has become the most common cause of end-stage renal disease (ESRD) in many national ESRD registries [1, 2]. Approximately one third of diabetic individuals will develop clinically apparent nephropathy characterised by persistent proteinuria, hypertension and eventual progressive decline in glomerular filtration rate [3, 4]. Whilst chronic hyperglycaemia is integral to the pathogenesis of diabetic nephropathy there is also strong evidence for a genetic susceptibility to this common complication of diabetes [5-7]. In genetically susceptible individuals, prolonged hyperglycaemia leads to chronic metabolic and haemodynamic changes [8, 9] whose effects, including those driven by transforming growth factor beta 1 (TGF β 1), promote structural abnormalities in the kidney such as glomerular basement membrane thickening, podocyte injury, and mesangial matrix expansion, with the later development of irreversible glomerular sclerosis and tubulointerstitial fibrosis [10].

Genetic studies have found evidence for linkage to nephropathy at a number of chromosome loci including 2q, 3q22 and 19q [11, 12]. While no consistently replicated genetic associations have been identified, various candidate gene associations in diabetic nephropathy in type 1 diabetes (T1D) have been proposed, such as the SNP rs13293564 G/T substitution in intron 1 of *UNC13B* [13], and *SOD1* SNPs [14]. An association with the *ELMO1* gene and diabetic nephropathy was initially identified in patients with type 2 diabetes mellitus [15] and this association has now also been reported for diabetic nephropathy in T1D [16]. A recent genome-wide association study identified *FRMD3* and *CARS* gene variants as possible genetic susceptibility factors for T1D diabetic nephropathy [17].

Epigenetic mechanisms allow alteration of genome function without mutating the underlying sequence. They involve the interacting actions of DNA methylation (the addition of a methyl group to the 5th carbon position of cytosine), histone modifications and non-coding RNAs [18]. A number of indirect lines of evidence point to the involvement of epigenetic changes in diabetic nephropathy. Murine models of disease progression

displaying temporal variation in gene expression have indicated these supra-sequence devices may be involved in the pathogenesis [19]. Gene expression changes reflect dynamic alterations in gene transcription and also messenger RNA stability, which may be influenced by the epigenetic modification of the genome in response to chronic hyperglycaemic stress. Altered DNA methylation has been additionally implicated in vascular disease [20, 21]. Furthermore, characteristics observed in diabetic nephropathy such as hyperhomocysteinaemia, dyslipidaemia, inflammation and oxidative stress can promote aberrant DNA methylation [22-24].

DNA methylation in mammalian species, except for pluripotent stem cells [25, 26], appears to occur almost exclusively at consecutive CpG dinucleotides, of which there are approximately 28 million located within the human genome [27]. The majority of cytosines in these dinucleotides remain methylated (~80% in *Homo sapiens*) [27], which contrasts with CpG islands (CGIs) in which cytosines are generally unmethylated. CGIs are regions that exceed genome average levels of CpG density and co-localise with 72% of gene promoters [28], notably in those of highly expressed genes.

In order to assess possible changes in CpG methylation in a genome-wide manner we utilised the high throughput Illumina Infinium® Human Methylation27 BeadChip DNA methylation array to investigate association with nephropathy in a case-control study of 192 T1D patients. These probes interrogate the methylation state of 27,578 individual CpGs located predominately in CGIs within proximal promoter regions, between 1.5 kb upstream and 1 kb downstream of the transcription start sites of 14,475 consensus coding sequence genes throughout the genome. Furthermore 110 miRNA promoters are targeted with 254 CpG loci probes [29]. This platform enables the high-throughput array based investigation of individual CpGs at a magnitude higher level than the previous array based Golden-Gate assay [30].

The development, as well as the progression, of diabetic nephropathy has been linked to the same inflammatory cell processes that are implicated in the progression of T1D itself, such as the activated T cells involved in the destruction of islet β cells [7]. Thus changes in the expression of cytokines and other inflammatory markers from distinct blood

cell types make the investigation of methylation changes in DNA derived from blood a plausible option for examining epigenetic alterations in this disease. DNA methylation changes in peripheral blood have additionally been shown to indicate or be surrogate markers of active disease processes [31].

This study has investigated the possibility to identify DNA methylation biomarkers in peripheral blood cell-derived DNA that may be associated with the pathogenesis of T1D nephropathy. The study was designed to detect associations between the disease and differences in DNA methylation but not their functional relationship. Undertaken in a case control design, those with and without nephropathy were matched by age of onset and duration of diabetes.

Methods

Participants

Cases (n= 96) and controls (n = 96) were recruited from nephrology and diabetic clinics in Belfast and Dublin (Table 1). Genomic DNA was extracted from whole blood samples from cases and controls using a salting out method. All participants were white, with parents and grandparents born in Northern Ireland or the Republic of Ireland. Both cases and controls were diagnosed with T1D before the age of 31 years and required insulin from time of diagnosis. The definition of a case was based on development of persistent proteinuria (>0.5g protein/24h) at least 10 years after diagnosis of diabetes, hypertension (BP >135/85 mmHg or treatment with antihypertensive agents) and presence of diabetic retinopathy. In contrast, a control was defined as a patient with T1D duration of at least 15 years with urinary albumin excretion in the normal range and not receiving antihypertensive treatment. Patients with microalbuminuria were excluded from both groups. Cases and controls were matched for confounders that may affect DNA methylation namely age, gender and duration of diabetes within five year windows. Study approval for each recruitment site was obtained from respective research ethics committees (Queens University Belfast and Mater Misericordiae Hospital, Dublin) and written informed consent was obtained from all participants.

Bisulphite conversion of DNA

Bisulphite conversion of 1ug genomic DNA was performed using the EZ-96 DNA Methylation-Gold™ Kit (©Zymo Research Corp., USA) in two 96 well plates according to the manufacturer's instructions. Cases and controls were randomly located between the two plates. Effective bisulphite conversion was checked by successful PCR amplification with a pair of primers specific for converted DNA and unsuccessful amplification with a pair of

primers for unconverted DNA, of 24 random samples across both plates (details available on request from authors).

Illumina Infinium methylation assay

DNA samples from cases and controls were interrogated utilising the Illumina Infinium® Human Methylation27 BeadChip. This platform detects the methylation status of 27,578 CpG sites by sequencing-based genotyping of bisulphite treated DNA. The chromosomal distribution and other statistics relevant to these CpG sites are provided in Additional File 1: Supplementary Figure S1. This chemical modification of DNA converts only unmethylated cytosines to uracils, thereafter allowing for highly multiplexed genotyping with single site resolution. Bisulphite converted and unconverted (i.e. methylated) sites are simultaneously evaluated by hybridisation of DNA to site-specific probes attached to beads, one for unmethylated and the other for methylated sites, followed by allele specific base extension that includes a fluorescent label. Two different labels are used, and fluorescent signals are specific for either the T (unmethylated) or C (methylated) alleles. Methylation scores represented as β values are generated for each site using BeadStudio 3.2 software (© Illumina, Inc. 2003-2008) and are computed based on the ratio of methylated to methylated plus unmethylated signal outputs. Thus the β values range from 0 (unmethylated) to 1 (fully methylated) on a continuous scale.

This platform for quantitative methylation, essentially an adaptation of the highly successful Illumina Infinium I Whole Genome Genotype SNP genotyping assay used extensively for GWAS [32, 33], is extremely accurate and powerful. Reproducibility estimates are extremely high with an average r^2 correlation for β -values of 0.992 for 24 technical replicates [29]. Due to the high density of probes up to 12 separate samples can be assayed on a single BeadChip, thus allowing for high-throughput processing.

The Illumina Infinium methylation assay was performed according to the manufacturer's instructions, and this along with the technical validation of the assay is detailed in Bibikova *et al.* [29]. Bisulphite-converted DNA was denatured, amplified,

fragmented and subsequently hybridised, with use of a specific hybridisation buffer, to the chip arrays (Sentrix positions A-L). Cases and controls were arranged to randomly approximate a 50:50 proportion on each chip. Primer extension was performed, then staining and coating, followed by imaging with the Illumina BeadArray reader. Internal quality controls included assessment of staining, hybridisation, target removal, extension, bisulphite conversion efficiency, dye specificity and additionally negative controls.

Statistical analysis

Initial array results were visualised using Illumina® BeadStudio 3.2 (© Illumina Inc 2003-2008). All computations and statistical analyses were performed using the R package (R 2.8.1) (<http://www.r-project.org>) [34] and Bioconductor [35].

Quality control and data normalisation

Quality control and normalisation of methylation data was undertaken using the analysis pipeline reported elsewhere [31]. Briefly, samples were monitored for coverage (fraction of CpGs with detectable intensity values above background) and bisulphite conversion efficiency (BSCE) using the controls provided on the Illumina Beadchip. Of the 192 samples, one sample had failed BSCE and an additional four samples had relatively low coverage and BSCE. These samples were removed, leaving a total of 187 samples with approximately 81% global coverage (22,486 CpG sites). Intra-array normalisation was performed by the Illumina BeadStudio 3.2. software. The resulting β -valued data matrix was normalised further using a quantile-normalisation strategy, designed to reduce unwanted inter-array variation. The normalised and raw data are available from GEO (Gene Expression Omnibus, NCBI) under the accession number GSE20067.

Singular Value Decomposition and Significance of Singular Values

Singular Value Decomposition (SVD) provides a linear representation of the data in terms of a relatively small number of components, which capture the most salient patterns of variation [36]. We verified that the first component of a SVD analysis on the β -valued data matrix captures over 90% of the variation in the data, representing the inherent bimodality of the methylation value distribution (see Figures 1a and 1b). Thus, to correctly evaluate the statistical significance of the remaining components of variation, the normalised data matrix needs to be adjusted by removing this background variation, by subtracting out the top principal component. This is essentially equivalent to normalising the data matrix by the mean profile. SVD was subsequently reapplied to the normalised adjusted data and the spectrum of singular values compared to the null distribution obtained by considering random matrices [37, 38]. Specifically, the normalised adjusted data was randomised by permuting the CpGs, using a distinct permutation for each sample. Subsequently, SVD was performed on the randomised data matrix and the fraction of variation of the inferred singular values compared to the fractions of variation of the unpermuted data (Figure 2). Using multiple randomisations we verified that the null distribution of singular values was very tight (Figure 2), as expected since null singular values reflect a global property of the randomised data, which should be robust to further randomisations of the data. Thus, significant components of variation were selected as those whose variation was larger than the expected variation under the null-hypothesis. This gave five principal components in addition to the top component, making a total of six. The biological significance of the five components of variation (top principal component did not correlate with any phenotype) was tested by correlation to phenotypes of interest and experimental factors (Figure 3). For continuous and ordinal variables (e.g. BSCE, age) associations were evaluated using linear regression, for categorical variables (e.g. Beadchip) we used a Kruskal-Wallis test. For the time to onset of nephropathy (CC-case/control) we used a Cox-regression with time to event = time to onset of nephropathy (NP), and event = 1 (NP), 0 (no NP).

Supervised analysis

Association between CpG-valued methylation profiles and diabetic nephropathy was performed using a Cox-regression model where time to event equals the time between diagnosis of T1D to the onset of nephropathy. Multivariate regressions were performed for each CpG separately and included factors for chip, BSCE, sex, cohort and age at sample draw. To correct for multiple testing we estimated the false discovery rate (FDR) using the q-value framework [39]. We have shown previously that the analytical q-value estimates are similar to those obtained using permutations of sample labels preserving the potential correlative structure between CpGs [31].

Results

Methylation profile of arrays

The methylation values as defined by β scores ranging between 1 and 0, or fully to non-methylated, are displayed for all 27,578 CpGs queried as histograms in Figures 1a and 1b, for controls and cases respectively. Both show a similar pattern with a high peak of hypomethylated loci and a low peak in the hypermethylated loci. This is the expected U shape of methylation pattern of the genome but due to the array design which is skewed towards promoter region CpGs in CGIs shows the reverse pattern with a larger peak at the low end of the distribution. The minimum and maximum β values results for a representative chromosome, chromosome 1, are shown in Figure 4. This displays the individual results for the 2,903 probed CpGs that reside across this chromosome, for cases and controls separately. Scatter plots showed highly equivalent intensity scores ($r^2 = 0.9990$) and average β scores ($r^2 = 0.9996$) between cases and controls (data not shown).

DNA methylation analysis

SVD of the normalised data revealed significant components of variation that correlated with age, time to onset of nephropathy, sex, but also with experimental factors including chip, cohort and bisulphite conversion efficiency (Figures 2 and 3). As age and experimental factors could confound the analysis, we identified CpGs correlating with time to onset of nephropathy by applying a multivariate Cox-regression model to each CpG site, including the confounding factors as covariates. We observed that the resulting p-value distribution was significantly different from a uniform distribution with an excess of p-values close to 0, suggesting that a significant number of CpGs are correlated with time to onset of nephropathy (Figure 5). This involves the contribution of directional methylation increases or decreases observed in cases, with respect to the period of time that nephropathy took to develop, and controls, whom had had a variable period of T1D without

developing nephropathy. At an FDR = 0.15, we observed 263 CpGs as conferring increased risk and 162 CpGs conferring decreased risk for nephropathy, respectively (Additional File 2: Supplementary Table S1). However, whether these changes are related to cause or effect of early/late onset of nephropathy cannot be answered by this study. Gene set enrichment analysis searches for genes that share common biological function, chromosomal location, or regulation, and this analysis run on these data using the Molecular Signatures Database (MSigDB v2.5) [40] did not reveal any significant associations, after correction for multiple testing. Using a more stringent FDR of 0.05 resulted in a set of 19 CpGs and these are listed in Table 2. The CpG cg07341907, located 5' of the *UNC13B* gene, was the only CpG near a previously identified T1D nephropathy-associated candidate gene within this group. The box plot for adjusted methylation Z-score differences for this CpG is shown in Figure 6, with higher levels of methylation in the case group (0.00557) versus controls (-0.00563) (Uncorrected average β values were 0.147 and 0.157, for controls and cases respectively). The putative association SNP rs13293564 resides within the first intron of this gene [13]. MUNC13, the protein coded for by *UNC13B*, has been shown to be up-regulated in the renal cortex of rats with streptozotocin-induced diabetes and its expression is induced by hyperglycaemia [41].

To investigate whether there was a significant co-directional change of CpGs within 1 kb of each other, the normalised results for each of those proximal pairs within this distance were combined and compared in a case versus control analysis, with two of the genes in Table 2, *PPAPR3* (cg02590345 and cg16898420 – separated by 119 bp) and *TRPS1* (cg12569516 and cg21312090 – separated by 534 bp) also being in the top ten results of this analysis (data not shown).

Discussion

This study investigated the methylation state of approximately 14,000 CGIs in promoter regions of genes spread throughout the genome in a cohort of 192 type 1 diabetic patients, half of whom had developed diabetic nephropathy. A Cox-regression model was

utilised whereby the effect of a genetic susceptibility has a multiplicative effect over time on the risk of a subject developing nephropathy. This analysis, after accounting for confounding effects, identified an excess of CpG sites which were observed to have a change in methylation status that could be correlated with duration of diabetes prior to onset of nephropathy with a FDR of 0.15. Of these observations 263 CpGs were identified as conferring increased risk and 162 CpGs conferring decreased risk for nephropathy, respectively and co-localised to 421 unique genic regions. However this dataset did not reveal any significant enrichment of particular physiological gene set pathways.

Using a more stringent FDR cut-off of 0.05, a set of 19 CpGs was identified (in 19 unique gene CGIs). One CpG in this list, cg07341907 is located 18 bp upstream of the transcription start site of a previously identified type 1 diabetic nephropathy gene, *UNC13B* [13]. Levels of methylation for this CpG were shown to be slightly higher in the case group compared with controls. Whilst expression of this gene has been shown to be increased in models of nephropathy and hyperglycaemia [41] the expression effect of methylation changes can often be difficult to predict [42]. The other CpG assayed within this genic region, cg02096633, is located ~0.5 kb further downstream, 197 bp into the first intron and shows consistently low methylation levels in both cases and controls (β -value averages 0.051 and 0.052 respectively). The susceptibility SNP, rs13293564 resides within the first intron and in Caucasians (from CEU HapMap data) is within a linkage disequilibrium block of 23 kb that includes the CGI. Tregouet *et al.* could determine no plausible function for the intronic SNP, however did identify five SNPs in strong linkage disequilibrium with this SNP that reside within the plausible promoter region [13]. Two of which (rs10081672 and rs10972333) were found to affect potential transcription factor binding sites. Additionally, the former SNP (rs10081672) creates or abrogates a CpG site depending on which allele is present. Therefore, modification of the promoter region of this gene whether genetically or epigenetically or in combination could influence its expression by affecting transcription factor binding. Tregouet *et al.* proposed that *UNC13B* mediates apoptosis in glomerular cells due to hyperglycemia, and they therefore suggested that this association could indicate initiation of nephropathy [13].

Whilst the 27K Infinium assay enables high-throughput investigation of individual CpG sites with high resolution, and is a considerable improvement on the previous Golden-Gate methylation assay, this methodology still has some limitations. Only approximately 0.1% of the total number of CpGs within the genome are assessed, however the placement of these CpGs are predominately within the CpG Islands with high likelihood of critical effects on promoter activity. As these CpGs can significantly affect or silence expression, changes can be more deleterious than gene sequence variants, therefore increasing the potential power of the assay. Additionally not all CpG sites need to be assessed as correlation of CpG status between sites can extend up to 1kb [43]. Whilst dramatic switching on or off facilitated by DNA methylation may occur within these promoter regions, more subtle and perhaps dynamic effects may however be missed that are occurring in the surrounding CGI shores (regions within 2 kb of the islands) [44]. Changes in gene body methylation that may influence the expression of certain isoforms, by effecting the inclusion or exclusion of certain exons, are also not assayed. Additionally this array does have a bias towards cancer and imprinted genes; with ~200 of these having 3-20 CpG probes per promoter. Increasing knowledge with regards to the most biologically informative CpG sites and future up scaling will undoubtedly improve targeted platforms, although a more in depth investigation of all of the CpGs of the genome would require a differing approach. An enrichment technique such as MeDIP-Seq would gain a representative genome wide picture but this lacks the base pair resolution of this array [45]. Isolated studies utilising full bisulphite sequencing of a very small number of genomes have recently been published [26], however scaling up to large numbers of cases and controls currently remains prohibitive due to cost.

Like all genome-wide association studies this study was designed to find statistically significant associations, in this case with DNA methylation variation not with genotype, but not to identify the underlying mechanism(s) for the cause or function of the observed variation. Although DNA methylation is tissue-specific, examination of peripheral whole blood was informative in this case in order to determine whether any genome-wide methylation signal changes could be detected in this easily accessible surrogate tissue.

Whilst only a small number of significant sites were identified, due to the stringent False Discovery Rate cut-off implemented, this experiment did identify possible biomarkers in this pathogenic process with more easily implementable clinical utility potential. Therefore, this indicates that active disease processes can be identified in the DNA methylation pattern of peripheral blood and be a possible marker of these [31]. The development, as well as the progression, of diabetic nephropathy has been linked to the same inflammatory cell changes that are implicated in the progression of T1D itself, such as the activated T cells involved in the destruction of islet β cells [7]. Furthermore, additional inflammatory-related or other signals may be present in peripheral blood DNA that require isolation of specific cell types to detect and other methylation change signals may only be able to be identified in the affected renal cells themselves.

Conclusion

This study is one of the first to investigate a complex disease trait utilising this high-throughput DNA methylation 27K array based assay. The effect of length of T1D status before developing or not developing nephropathy was observed to show a correlation with respect to methylation scores of individual CpG loci. By the examination of a cohort of 192 T1D patients this study has confirmed the utility of this approach to genome wide DNA methylation analysis and the potential prospects for larger and subsequent replication studies of epigenomic factors in common diseases.

List of Abbreviations

BSCE – Bisulphite Conversion Efficiency, CGI – CpG Island, ESRD – End-Stage Renal Disease, FDR – False Discovery Rate, NP – Nephropathy, T1D – Type 1 Diabetes Mellitus, SVD – Singular Value Decomposition.

Competing interests

The Authors declare that they have no competing interests.

Authors' contributions

CGB participated in the design of the study, Bisulphite conversion, analysis and drafting of the manuscript. AET performed the statistical analysis. VRK participated in the design of the study and coordination. DAS, APM and SB conceived of the study, participated in its design and coordination, and drafting of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank the Renal Unit Fund, Belfast Health and Social Care Trust for funding this project. CGB and SB were funded by the Wellcome Trust (084071), and AET was supported by a Heller Research Fellowship. SB is a Royal Society-Wolfson Research Merit Award Holder. We also thank Dr Denise Sadlier, University College Dublin, for providing DNA samples from cases and controls from the Republic of Ireland.

We would further like to thank UCL Genomics for excellent technical support.

Table 1: Subjects

| | Controls | Cases |
|------------------|----------|-------|
| Number | 96 | 96 |
| Age | 43.2 | 42.8 |
| Sex (M/F) | 47/49 | 48/48 |
| Age at Diagnosis | 16.3 | 16.2 |
| Duration T1D | 27.8 | 27.5 |

Subjects: Number, Average Age, Sex proportion (Male/Female), Duration of T1D

Table 2: Methylation Associations

| Ilmn CpG ID | Gene | Gene Function | Location | coef | exp(coef) | se(coef) | z | Pr(> z) | Q value |
|-------------|-----------------|---|--------------|--------|-----------|----------|--------|---------------------------|---------|
| cg01172656 | <i>NRBF2</i> | Transcription regulation | 10:64563126 | 0.631 | 1.880 | 0.136 | 4.653 | 3.266 x 10 ⁻⁰⁶ | 0.031 |
| cg18346038 | <i>PIGU</i> | GPI-anchor biosynthesis | 20:32728489 | 0.624 | 1.867 | 0.136 | 4.603 | 4.164 x 10 ⁻⁰⁶ | 0.031 |
| cg00117172 | <i>RUNX3</i> | Transcription regulation | 1:25128425 | 0.682 | 1.977 | 0.150 | 4.548 | 5.427 x 10 ⁻⁰⁶ | 0.031 |
| cg02068676 | <i>COBRA1</i> | Negative regulation of transcription | 9:139269304 | 0.662 | 1.938 | 0.149 | 4.447 | 8.703 x 10 ⁻⁰⁶ | 0.037 |
| cg00025138 | <i>MAP3K9</i> | Mixed-lineage kinase | 14:70345670 | 0.543 | 1.721 | 0.125 | 4.335 | 1.457 x 10 ⁻⁰⁵ | 0.046 |
| cg03221776 | <i>MFSD3</i> | Transmembrane transport | 8:145704779 | 0.520 | 1.682 | 0.121 | 4.308 | 1.644 x 10 ⁻⁰⁵ | 0.046 |
| cg12181621 | <i>HIST1H3I</i> | Nucleosome assembly | 6:27948047 | 0.558 | 1.748 | 0.132 | 4.216 | 2.486 x 10 ⁻⁰⁵ | 0.048 |
| cg03752885 | <i>DAPK3</i> | Induction of apoptosis | 19:3920736 | 0.492 | 1.636 | 0.118 | 4.189 | 2.798 x 10 ⁻⁰⁵ | 0.048 |
| cg12688215 | <i>KTI12</i> | ATP binding | 1:52271816 | 0.446 | 1.563 | 0.107 | 4.175 | 2.976 x 10 ⁻⁰⁵ | 0.048 |
| cg07341907 | <i>UNC13B</i> | Exocytosis | 9:35151971 | 0.606 | 1.833 | 0.145 | 4.169 | 3.056 x 10 ⁻⁰⁵ | 0.048 |
| cg26850145 | <i>ZBTB5</i> | Transcription regulation | 9:37455023 | -0.485 | 0.616 | 0.117 | -4.142 | 3.441 x 10 ⁻⁰⁵ | 0.048 |
| cg21312090 | <i>TRPS1</i> | NLS-bearing substrate import into nucleus | 8:116750901 | 0.441 | 1.555 | 0.107 | 4.135 | 3.543 x 10 ⁻⁰⁵ | 0.048 |
| cg23266266 | <i>DCUN1D4</i> | - | 4:52403466 | 0.560 | 1.751 | 0.136 | 4.117 | 3.847 x 10 ⁻⁰⁵ | 0.048 |
| cg04466870 | <i>SFXN4</i> | Transmembrane transport | 10:120915170 | 0.477 | 1.611 | 0.116 | 4.102 | 4.092 x 10 ⁻⁰⁵ | 0.048 |

| | | | | | | | | | |
|------------|---------------|--|-------------|--------|-------|-------|--------|-------------------------|-------|
| cg16898420 | <i>PPAPR3</i> | Catalytic activity | 9:102830899 | -0.711 | 0.491 | 0.175 | -4.066 | 4.786×10^{-05} | 0.048 |
| cg17260725 | <i>CCNB2</i> | Cell division | 15:57184580 | 0.523 | 1.687 | 0.129 | 4.0596 | 4.915×10^{-05} | 0.048 |
| cg17428423 | <i>DOC2A</i> | Regulation of calcium ion-dependent exocytosis | 16:29929633 | -0.728 | 0.483 | 0.180 | -4.053 | 5.049×10^{-05} | 0.048 |
| cg18731014 | <i>ZNF639</i> | Transcription regulation | 3:180524284 | 0.556 | 1.744 | 0.137 | 4.050 | 5.122×10^{-05} | 0.048 |
| cg21830413 | <i>VPS26</i> | Protein transport | 10:70554257 | 0.472 | 1.604 | 0.117 | 4.036 | 5.430×10^{-05} | 0.048 |

Cox-Regression analysis results of CpGs, with FDR (q value) <0.05, for time to develop nephropathy. (Location given as Chromosome:Base Pair (Build Hs 36.1))

Figure Legends

Figure 1 – Average β values.

Figure 1a (top): Controls, Figure 1b (bottom): Cases - Average β values across array.

Figure 2 – Fraction of variation associated with principal components.

Fraction of variation associated with principal components after adjustment for the top component which captures the inherent bi-modality of the methylation distribution. Red denotes the observed values, black denotes those obtained by randomly scrambling up the data matrix. Each black dot represents the value of 10 distinct randomisations, which all yield the same value as singular values represent measures of global variation which themselves are invariant under a global randomisation of the data.

Figure 3 – Correlation of principle components to phenotype/experimental factors.

Correlation of principal components (top component removed as it did not correlate with any phenotype) to phenotypes of interest and experimental factors: BS conversion (bisulphite conversion, C1(green), C2(blue)), Sex, Cohort, Chip, Batch, Age at Draw (Age when sample taken), Duration T1D, Age at Diagnosis, CC (time to onset of NP)

Colour key: dark red = p value $< 10^{-10}$, red = p value $< 10^{-5}$, orange = p value $< 10^{-2}$, pink = p value < 0.05 , white = p value > 0.05 .

Figure 4: Maximum and minimum β values for Chromosome 1.

Maximum and minimum β values across the 2903 probed CpGs located on Chromosome 1 for Controls and Cases.

Figure 5 – Nephropathy and duration type 1 diabetes p values.

Histogram of p -values after explicit correction for all potentially confounding factors reveals an excess of significant p values for CpGs correlated with Nephropathy and duration of Type 1 Diabetes.

Figure 6 – Z score values for cg07341907.

Box plot for adjusted comparison of methylation Z-score values for the *UNC13B* promoter CpG:cg07341907 on controls (0) and cases (1).

Additional File 1

Supplementary Figure S1

Distribution of CpG sites on Infinium platform.

(Top, Left) Chromosome Distribution of CpG sites. (Top, Right) Distance to Transcription Start Site of CpG Locus. (Bottom, Left) Interrogated CpG Islands: near genes, near miRNA. (Bottom, Right) Number of CpG sites per CpG Island.

Additional File 2

Supplementary Table S1.

List of CpGs with q value < 0.15.

Column headings are IlmnID = Illumina CpG ID, Gene_ID = Entrez Gene ID, Symbol = Gene Symbol, coef = coefficient, exp(coef) = exponential of coefficient, se(coef) = standard error of coefficient, z = z score, Pr(>|z|) = probability greater than absolute z score, q value

References

1. Ansell D, Feehally J, Fogarty D, Tomsom CRV, Williams AJ, Warwick G: *UK Renal Registry Annual Report*. 2009/06/30 edn. Bistol; 2008.
2. U.S. Renal Data System: **US Renal Data System 2008 Annual Data Report**. In *Atlas of Chronic Kidney Disease and End Stage Renal Disease in the United States* (National Institutes of Health NIDaDaKD ed. Bethesda, MD. : National Institutes of Health, National Institutes of Diabetes and Digestive and Kidney Diseases 2008.
3. Ibrahim HN, Hostetter TH: **Diabetic nephropathy**. *J Am Soc Nephrol* 1997, **8**:487-493.
4. O'Connor AS, Schelling JR: **Diabetes and the kidney**. *Am J Kidney Dis* 2005, **46**:766-773.
5. Seaquist ER, Goetz FC, Rich S, Barbosa J: **Familial clustering of diabetic kidney disease. Evidence for genetic susceptibility to diabetic nephropathy**. *N Engl J Med* 1989, **320**:1161-1165.

6. Quinn M, Angelico MC, Warram JH, Krolewski AS: **Familial factors determine the development of diabetic nephropathy in patients with IDDM.** *Diabetologia* 1996, **39**:940-945.
7. Ichinose K, Kawasaki E, Eguchi K: **Recent advancement of understanding pathogenesis of type 1 diabetes and potential relevance to diabetic nephropathy.** *Am J Nephrol* 2007, **27**:554-564.
8. Weigert C, Sauer U, Brodbeck K, Pfeiffer A, Haring HU, Schleicher ED: **AP-1 proteins mediate hyperglycemia-induced activation of the human TGF-beta1 promoter in mesangial cells.** *J Am Soc Nephrol* 2000, **11**:2007-2016.
9. Katavetin P, Miyata T, Inagi R, Tanaka T, Sassa R, Ingelfinger JR, Fujita T, Nangaku M: **High glucose blunts vascular endothelial growth factor response to hypoxia via the oxidative stress-regulated hypoxia-inducible factor/hypoxia-responsible element pathway.** *J Am Soc Nephrol* 2006, **17**:1405-1413.
10. Navarro-Gonzalez JF, Mora-Fernandez C: **The role of inflammatory cytokines in diabetic nephropathy.** *J Am Soc Nephrol* 2008, **19**:433-442.
11. He B, Osterholm AM, Hoverfalt A, Forsblom C, Hjorleifsdottir EE, Nilsson AS, Parkkonen M, Pitkanieni J, Hreidarsson A, Sarti C, et al: **Association of genetic variants at 3q22 with nephropathy in patients with type 1 diabetes mellitus.** *Am J Hum Genet* 2009, **84**:5-13.
12. Rogus JJ, Poznik GD, Pezzolesi MG, Smiles AM, Dunn J, Walker W, Wanic K, Moczulski D, Canani L, Araki S, et al: **High-density single nucleotide polymorphism genome-wide linkage scan for susceptibility genes for diabetic nephropathy in type 1 diabetes: discordant sibpair approach.** *Diabetes* 2008, **57**:2519-2526.
13. Tregouet DA, Groop PH, McGinn S, Forsblom C, Hadjadj S, Marre M, Parving HH, Tarnow L, Telgmann R, Godefroy T, et al: **G/T substitution in intron 1 of the UNC13B gene is associated with increased risk of nephropathy in patients with type 1 diabetes.** *Diabetes* 2008, **57**:2843-2850.
14. Al-Kateb H, Boright AP, Mirea L, Xie X, Sutradhar R, Mowjoodi A, Bharaj B, Liu M, Bucksa JM, Arends VL, et al: **Multiple superoxide dismutase 1/splicing factor serine alanine 15 variants are associated with the development and progression of diabetic nephropathy: the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Genetics study.** *Diabetes* 2008, **57**:218-228.
15. Leak TS, Perlegas PS, Smith SG, Keene KL, Hicks PJ, Langefeld CD, Mychaleckyj JC, Rich SS, Kirk JK, Freedman BI, et al: **Variants in intron 13 of the ELMO1 gene are associated with diabetic nephropathy in African Americans.** *Ann Hum Genet* 2009, **73**:152-159.
16. Pezzolesi MG, Katavetin P, Kure M, Poznik GD, Skupien J, Mychaleckyj JC, Rich SS, Warram JH, Krolewski AS: **confirmation of genetic associations at ELMO1 in the GoKinD collection supports its role as a susceptibility gene in diabetic nephropathy.** *Diabetes* 2009, **58**:2698-2702.
17. Pezzolesi MG, Poznik GD, Mychaleckyj JC, Paterson AD, Barati MT, Klein JB, Ng DP, Placha G, Canani LH, Bochenski J, et al: **Genome-wide association scan for diabetic nephropathy susceptibility genes in type 1 diabetes.** *Diabetes* 2009, **58**:1403-1410.
18. Richards EJ: **Inherited epigenetic variation--revisiting soft inheritance.** *Nat Rev Genet* 2006, **7**:395-401.

19. Makino H, Miyamoto Y, Sawai K, Mori K, Mukoyama M, Nakao K, Yoshimasa Y, Suga S: **Altered gene expression related to glomerulogenesis and podocyte structure in early diabetic nephropathy of db/db mice and its restoration by pioglitazone.** *Diabetes* 2006, **55**:2747-2756.
20. Krause B, Sobrevia L, Casanello P: **Epigenetics: new concepts of old phenomena in vascular physiology.** *Curr Vasc Pharmacol* 2009, **7**:513-520.
21. Turunen MP, Aavik E, Yla-Herttuala S: **Epigenetics and atherosclerosis.** *Biochim Biophys Acta* 2009, **1790**:886-891.
22. Ingrosso D, Perna AF: **Epigenetics in hyperhomocysteinemic states. A special focus on uremia.** *Biochim Biophys Acta* 2009, **1790**:892-899.
23. Ekstrom TJ, Stenvinkel P: **The epigenetic conductor: a genomic orchestrator in chronic kidney disease complications?** *J Nephrol* 2009, **22**:442-449.
24. Lund G, Zaina S: **Atherosclerosis risk factors can impose aberrant DNA methylation patterns: a tale of traffic and homocysteine.** *Curr Opin Lipidol* 2009, **20**:448-449.
25. Ramsahoye BH, Biniszkiwicz D, Lyko F, Clark V, Bird AP, Jaenisch R: **Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a.** *Proc Natl Acad Sci U S A* 2000, **97**:5237-5242.
26. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, et al: **Human DNA methylomes at base resolution show widespread epigenomic differences.** *Nature* 2009, **462**:315-322.
27. Beck S, Rakyen VK: **The methylome: approaches for global DNA methylation profiling.** *Trends Genet* 2008, **24**:231-237.
28. Saxonov S, Berg P, Brutlag DL: **A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters.** *Proc Natl Acad Sci U S A* 2006, **103**:1412-1417.
29. Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, Gunderson K: **Genome-wide DNA methylation profiling using Infinium assay.** *Epigenomics* 2009, **1**:177-200.
30. Bibikova M, Fan JB: **GoldenGate assay for DNA methylation profiling.** *Methods Mol Biol* 2009, **507**:149-163.
31. Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Gayther SA, Apostolidou S, Jones A, Lechner M, Beck S, Jacobs IJ, Widschwendter M: **An epigenetic signature in peripheral blood predicts active ovarian cancer.** *PLoS One* 2009, **4**:e8274.
32. Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, Jonasdottir A, Sigurdsson A, Baker A, Palsson A, Masson G, et al: **A common variant on chromosome 9p21 affects the risk of myocardial infarction.** *Science* 2007, **316**:1491-1493.
33. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, et al: **A genome-wide association study identifies novel risk loci for type 2 diabetes.** *Nature* 2007, **445**:881-885.
34. Ihaka R, Gentleman R: **R: A Language for Data Analysis and Graphics.** *Journal of Computational and Graphical Statistics* 1996, **Vol 5**:299-314.
35. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, et al: **Bioconductor: open software development for computational biology and bioinformatics.** *Genome Biol* 2004, **5**:R80.

36. Hastie T, Tibshirani R, Friedman JH: *The elements of statistical learning : data mining, inference, and prediction : with 200 full-color illustrations*. New York: Springer; 2001.
37. Leek JT, Storey JD: **Capturing heterogeneity in gene expression studies by surrogate variable analysis**. *PLoS Genet* 2007, **3**:1724-1735.
38. Buja A, Eyuboglu N: **Remarks on Parallel Analysis**. *Multivariate Behavioral Research* 1992, **27**:509-540.
39. Storey JD, Tibshirani R: **Statistical significance for genomewide studies**. *Proc Natl Acad Sci U S A* 2003, **100**:9440-9445.
40. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles**. *Proc Natl Acad Sci U S A* 2005, **102**:15545-15550.
41. Goldenberg NM, Silverman M: **Rab34 and its effector munc13-2 constitute a new pathway modulating protein secretion in the cellular response to hyperglycemia**. *Am J Physiol Cell Physiol* 2009, **297**:C1053-1058.
42. Suzuki MM, Bird A: **DNA methylation landscapes: provocative insights from epigenomics**. *Nat Rev Genet* 2008, **9**:465-476.
43. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, Burton J, Cox TV, Davies R, Down TA, et al: **DNA methylation profiling of human chromosomes 6, 20 and 22**. *Nat Genet* 2006, **38**:1378-1385.
44. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M, et al: **The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores**. *Nat Genet* 2009, **41**:178-186.
45. Down TA, Rakyan VK, Turner DJ, Flicek P, Li H, Kulesha E, Graf S, Johnson N, Herrero J, Tomazou EM, et al: **A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis**. *Nat Biotechnol* 2008, **26**:779-785.

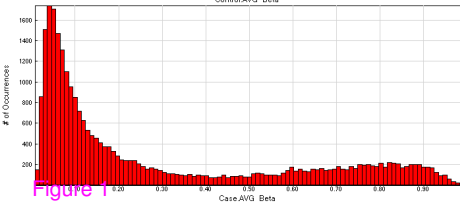
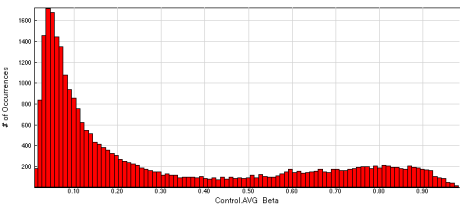
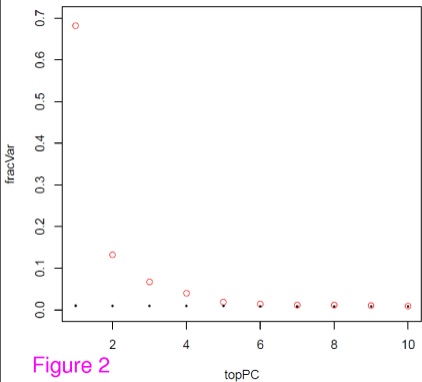


Figure 1



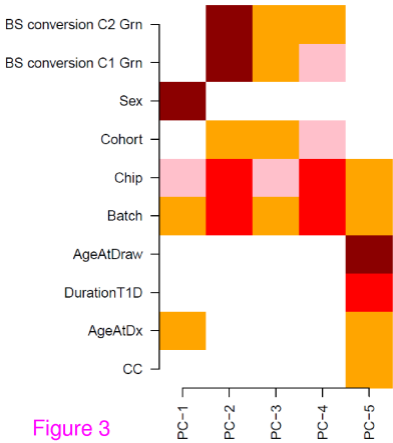


Figure 3

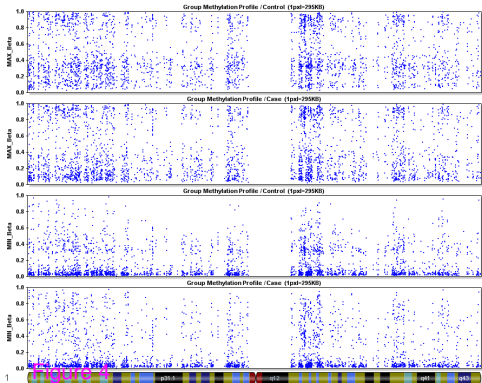


Figure 4

Histogram of p-values: NP-durT1D

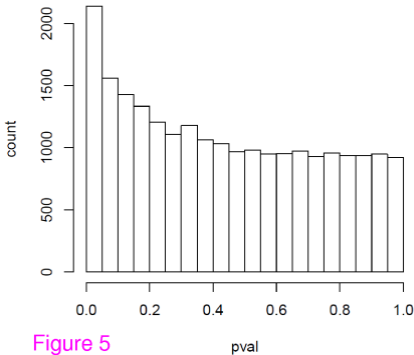


Figure 5

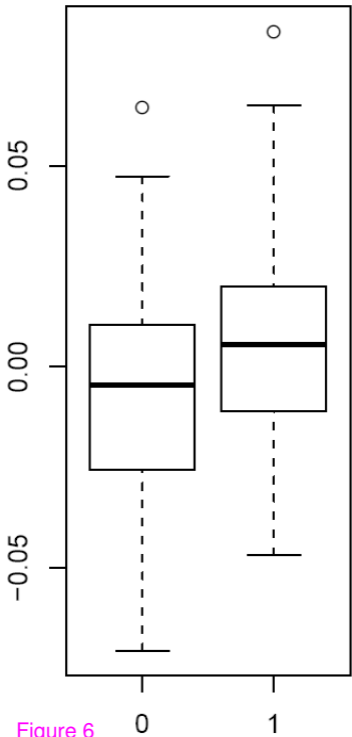


Figure 6

Additional files provided with this submission:

Additional file 1: Supplementary-Figure-S1.png, 130K

<http://www.biomedcentral.com/imedia/4854591274308819/supp1.png>

Additional file 2: Supplementary_table_S1.xls, 106K

<http://www.biomedcentral.com/imedia/5302899444308821/supp2.xls>