


RESEARCH

Open Access



# Assessment of differentially methylated loci in individuals with end-stage kidney disease attributed to diabetic kidney disease: an exploratory study

L. J. Smyth<sup>1\*</sup> , J. Kilner<sup>1</sup>, V. Nair<sup>2</sup>, H. Liu<sup>3</sup>, E. Brennan<sup>4</sup>, K. Kerr<sup>1</sup>, N. Sandholm<sup>5,6,7</sup>, J. Cole<sup>8,9,10</sup>, E. Dahlström<sup>5,6,7</sup>, A. Syreeni<sup>5,6,7</sup>, R. M. Salem<sup>11</sup>, R. G. Nelson<sup>12</sup>, H. C. Looker<sup>12</sup>, C. Wooster<sup>1</sup>, K. Anderson<sup>1</sup>, G. J. McKay<sup>1</sup>, F. Kee<sup>1</sup>, I. Young<sup>1</sup>, D. Andrews<sup>4</sup>, C. Forsblom<sup>5,6,7</sup>, J. N. Hirschhorn<sup>8,9,13</sup>, C. Godson<sup>4</sup>, P. H. Groop<sup>5,6,7,14</sup>, A. P. Maxwell<sup>1,15</sup>, K. Susztak<sup>3</sup>, M. Kretzler<sup>2</sup>, J. C. Florez<sup>8,9,16</sup> and A. J. McKnight<sup>1</sup>

## Abstract

**Background:** A subset of individuals with type 1 diabetes mellitus (T1DM) are predisposed to developing diabetic kidney disease (DKD), the most common cause globally of end-stage kidney disease (ESKD). Emerging evidence suggests epigenetic changes in DNA methylation may have a causal role in both T1DM and DKD. The aim of this exploratory investigation was to assess differences in blood-derived DNA methylation patterns between individuals with T1DM-ESKD and individuals with long-duration T1DM but no evidence of kidney disease upon repeated testing to identify potential blood-based biomarkers. Blood-derived DNA from individuals (107 cases, 253 controls and 14 experimental controls) were bisulphite treated before DNA methylation patterns from both groups were generated and analysed using Illumina's Infinium MethylationEPIC BeadChip arrays ( $n = 862,927$  sites). Differentially methylated CpG sites (dmCpGs) were identified (false discovery rate adjusted  $p \leq \times 10^{-8}$  and fold change  $\pm 2$ ) by comparing methylation levels between ESKD cases and T1DM controls at single site resolution. Gene annotation and functionality was investigated to enrich and rank methylated regions associated with ESKD in T1DM.

**Results:** Top-ranked genes within which several dmCpGs were located and supported by functional data with methylation look-ups in other cohorts include: *AFF3*, *ARID5B*, *CUX1*, *ELMO1*, *FKBP5*, *HDAC4*, *ITGAL*, *LY9*, *PIM1*, *RUNX3*, *SEPTIN9* and *UPF3A*. Top-ranked enrichment pathways included pathways in cancer, TGF- $\beta$  signalling and Th17 cell differentiation.

**Conclusions:** Epigenetic alterations provide a dynamic link between an individual's genetic background and their environmental exposures. This robust evaluation of DNA methylation in carefully phenotyped individuals has identified biomarkers associated with ESKD, revealing several genes and implicated key pathways associated with ESKD in individuals with T1DM.

**Keywords:** Association, Diabetes, EPIC, End-stage, Kidney, Methylation, Nephropathy

## Background

Type 1 diabetes mellitus (T1DM) is a polygenic disease characterised by autoimmune destruction of the insulin producing beta cells in the pancreas which subsequently

\*Correspondence: laura.smyth@qub.ac.uk

<sup>1</sup> Molecular Epidemiology Research Group, Centre for Public Health, Queen's University Belfast, Belfast, UK  
Full list of author information is available at the end of the article



© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

leads to hyperglycaemia [1]. Up to 40% of individuals with T1DM are predisposed to diabetic kidney disease (DKD), a microvascular diabetic complication [2, 3]. Diabetes remains the primary disease causing end-stage kidney disease (ESKD) [4, 5]. Individuals with DKD are also at increased risk of developing cardiovascular disease and premature mortality [6].

Emerging evidence indicates that autoimmune diseases such as T1DM are influenced by the interaction of genetic and epigenetic factors [7]. Genome-wide association studies (GWAS) have identified >50 loci associated with T1DM risk [8, 9], including rs3802604 in *GATA3*, rs1770 in *MHC* [10], rs2292239 in *ERBB3* [11] and rs2476601 in *PTPN22* [12].

GWAS have been completed in populations with T1DM assessing the genetic predisposition to DKD and ESKD. One recent GWAS encompassing almost 20,000 individuals with T1DM [13] identified 16 risk loci including the missense mutation rs55703767 in *COL4A3*. Additional GWAS have identified nucleotide variants in rs7583877 in *AFF3* [14], rs1888747 and rs10868025 in *FRMD3*, [15] and rs4972593 in *CDCA7-SP3* [16].

Epigenetic changes in DNA methylation have been associated with T1DM [17] and DKD [18] using peripheral whole blood samples. Bell et al. [19] used Illumina's Infinium 27K array in 2010 to assess epigenetic profiles associated with DKD in T1DM within 192 individuals. Nineteen differently methylated CpG sites (dmCpGs) were reported and correlated with time to the development of DKD. Six genes of interest with at least two dmCpGs including *CUX1*, *ELMO1*, *FKBP5*, *PRKAG2* and *PTPRN2* were identified from an epigenome-wide association study (EWAS) published in 2014 [20]. This investigation included individuals with DKD, caused by either T1DM or type 2 diabetes mellitus (T2DM).

In 2014, Gu et al. [21] employed bisulphite pyrosequencing of 778 individuals with T1DM, with and without DKD, and reported a decrease in the DNA methylation levels within *IGFBP1*. In 2015, Swan et al. [22] assessed DNA methylation variation in genes which encode mitochondrial proteins using Illumina's 450K and 27K methylation arrays in 442 individuals with T1DM and DKD. In all, 46 dmCpGs were identified in individuals with both DKD and ESKD. The largest change in methylation was evident for cg03169527 within *TAMM41*. Previous EWAS have identified differential methylation features with associations to chronic kidney disease (CKD) [23–26]. Improvements in sequencing and profiling technologies have prompted a rise in the number of studies assessing diseases using these techniques for biomarker discovery [23, 27].

The aim of the present exploratory analysis was to assess differences in blood-derived DNA methylation patterns (using a high-density array) between individuals with T1DM-ESKD (including individuals requiring dialysis, or those having received a kidney transplant) and individuals with at least 15 years of T1DM and no evidence of kidney disease to identify blood-based biomarkers for T1D-ESKD. Careful and precise phenotyping has provided the ability to minimise differences caused by dialysis treatment; we also evaluated differences in methylation between individuals with T1DM-ESKD who had received a kidney transplant and thus were receiving immunosuppressive medication, versus those with long duration of T1DM and no evidence of kidney disease.

## Results

The Illumina Infinium MethylationEPIC BeadChip array examines 862,927 CpG sites. Each resulting .idat file generated from the iScan was assessed using Illumina's BACR software. This software assessed the data in connection with a pre-set standard set of thresholds [28] which are indicated alongside the BACR data for each analysis. QC was completed with no significant difference in intensity levels detected. On average, fewer than 30 probes failed per sample. Houseman estimates [29] were calculated for the proportional WCCs of each sample, and the results were included as Additional file 1 for each of the four discovery analyses; Additional file 1: Table S2, Table S10, Table S19 and Table S27. Proportional WCCs for each group were compared with a general population control group, NICOLA (Northern Ireland Cohort for the Longitudinal Study of Ageing).

Blood-derived DNA from 360 individuals was included in this analysis. Of these participants, 253 were control individuals with a long duration of T1DM and no evidence of kidney disease upon repeated testing. The mean age of these participants was 41 years and with an average duration of diabetes of 26 years and without anti-hypertensive medication usage. The remaining 107 individuals were diagnosed with T1DM-ESKD, 73 of whom had received a kidney transplant. The average age of these participants was 41 years. The participant characteristics are provided in Table 1.

This study was conducted using four complementary analyses; matching of the participants was required for *Analyses 1* and *3*. Individuals were matched for age at sample collection (differing by <5 years), sex and duration of diabetes (differing by ≤10 years). Concordance plots were drawn for seven duplicate samples; average  $r^2 = 0.99$ ; Additional file 2: Figure S1.

**Table 1** Participant group characteristics, per comparison

	Analysis 1*	Analysis 2	Analysis 3*	Analysis 4
<i>Cases</i>				
Number of participants	107	107	73	73
Diagnosis	T1DM with ESKD (transplant recipients or dialysis)	T1DM with ESKD (transplant recipients or dialysis)	T1DM with ESKD (transplant recipients only)	T1DM with ESKD (transplant recipients only)
Number of F:M	43:64	43:64	26:47	26:47
Average Age (St. Dev.)	41.7 years (9.7)	41.7 years (9.7)	40.5 years (9.3)	40.5 years (9.3)
Average duration of T1DM (St. Dev.)	29.1 years (6.5)	29.1 years (6.5)	30.1 years (5.7)	30.1 years (5.7)
<i>Controls</i>				
Number of participants	107	253	73	253
Diagnosis	Long-duration of T1DM with no evidence of kidney disease	Long-duration of T1DM with no evidence of kidney disease	Long-duration of T1DM with no evidence of kidney disease	Long-duration of T1DM with no evidence of kidney disease
Number of F:M	43:64	121:132	26:47	121:132
Average Age (St. Dev.)	41.2 years (10.2)	41.2 years (9.6)	40.1 years (9.1)	41.2 years (9.6)
Average duration of T1DM (St. Dev.)	26.8 years (7.9)	26.3 years (7.4)	25.5 years (7.8)	26.3 years (7.4)

ESKD end-stage kidney disease, F females, M males, St. Dev. standard deviation, T1DM type 1 diabetes

\*Analyses 1 and 3: Matched cases and controls for sex, average age ( $\leq 5$  years) and duration of diabetes ( $\leq 10$  years)

### Analysis 1: individuals with T1DM and ESKD compared to matched control individuals with T1DM and no evidence of kidney disease

This first analysis included 107 matched pairs, where both the case and control individuals were matched on sex and age at diagnosis (Table 1). The BACR QC report is included within Additional file 1: Table S1. Proportional WCCs were estimated for all included participant blood samples. Additional file 1: Table S2 shows the average WCCs for *Analysis 1* and the results of a *t* test comparing these between the two sample groups. Two cells, granulocytes and monocytes showed significant differences at the 0.05 level, the most significant of which was 0.04; but after correction for multiple testing of the WCC analyses, these results are not significant ( $p > 0.002$ ).

Comparison of the methylation patterns between cases and controls identified 4383 top-ranked dmCpGs (FDR  $p \leq \times 10^{-8}$ , Additional file 1: Table S3). Two genes, *ARID5B* and *SEPTIN9*, contained  $\geq 10$  top-ranked dmCpGs. When the stringency was increased to include a larger change in methylation between cases and controls (FDR  $p \leq \times 10^{-8}$  and  $FC \geq \pm 2$ ), 489 CpG sites remained significant (Additional file 1: Table S4) of which 331 were gene centric. Of these, 22 genes contained at least two dmCpGs including *HDAC4*, *ITGAL*, *LY9*, *PBX1*, *RPTOR*, *RUNX3* and *SEPTIN9* (Additional file 1: Table S4). A Manhattan plot was drawn using the qqman R package for all CpG sites from this analysis (Additional file 2: Figure S2).

To further assess the functional significance of these changes in DNA methylation between case and control groups, a GO enrichment analysis was undertaken. This analysis assessed the biological processes, cellular components and molecular functions of the genes within which the 489 top-ranked CpG sites were located (FDR  $p \leq \times 10^{-8}$  and  $FC \geq \pm 2$ ). A total of 325 GO functions had an enrichment score  $\geq 4$  ( $p < 0.01$ , Additional file 1: Table S5 and Additional file 2: Figure S3). The processes with the top enrichment scores include cell surface receptor signalling pathway, immune system process and positive regulation of immune system processes, lymphocyte activation and T cell activation. The KEGG and REACTOME pathway databases were additionally searched to identify key pathways linked to the genes where the top-ranked dmCpGs were located. Eleven KEGG pathways were identified with an enrichment score of  $\geq 2$  and  $p \leq 0.01$  (Additional file 1: Table S6). This analysis of differentially methylated genes returned pathways including Th17 cell differentiation, human T-cell leukaemia virus 1 infection and natural killer cell-mediated cytotoxicity. Ten REACTOME pathways were identified including six related to *RUNX1* or *RUNX3*; *RUNX3* regulates *RUNX1*-mediated transcription, transcriptional regulation by *RUNX3*, regulation of *RUNX1* expression and activity, *RUNX3* regulates immune response and cell migration, *RUNX1* regulates transcription of genes involved in differentiation of myeloid cells and *RUNX3* regulates p14-ARF (Additional file 1: Table S7).

The eFORGE-TF database was searched for transcription factor (TF) motifs associated with the top-ranked dmCpGs (FDR  $p \leq \times 10^{-8}$  and  $FC \geq \pm 2$ ) from within the most biologically plausible genes (*ARID5B*, *HDAC4*, *ITGAL*, *LY9*, *PBX1*, *PIM1*, *RUNX3*, *SEPT9* and *UPF3A*, available in Table 2). Of the 17 dmCpGs examined, eight reported TF motifs where  $q < 0.05$ . These are detailed in Additional file 1: Table S8.

#### Analysis 2: individuals with T1DM and ESKD compared to a larger cohort of unmatched control individuals with T1DM and no evidence of kidney disease

This second analysis included the same 107 individual cases from *Analysis 1* with a larger sample size of (unmatched) control individuals with T1DM and no evidence of kidney disease (Table 1). The BACR QC report and proportional WCCs for this analysis are included within Additional file 1: Table S9 and Additional file 1: Table S10, respectively.

Following the same analysis path as previously described, comparison of methylation between case individuals and control individuals identified 13,926 top-ranked dmCpGs (FDR  $p \leq \times 10^{-8}$ ; Additional file 1: Table S11). Two genes *ETS1* and *UBAC2* contained over 20 dmCpGs. When the stringency levels were increased (FDR  $p \leq \times 10^{-8}$  and  $FC \pm 2$ ), 1106 CpG sites remained, of which 764 were gene centric (Additional file 1: Table S12). *SEPTIN9* contained the largest number of dmCpGs ( $n=5$ ) with the criteria set to include  $FC \pm 2$ . Comparison of results from *Analyses 1* and 2 (FDR  $p \leq \times 10^{-8}$  and  $FC \pm 2$ ) identified 325 dmCpGs (223 within genes) which overlapped. Each dmCpG demonstrated the same direction of effect (Additional file 1: Table S13). A Manhattan plot (Additional file 2: Figure S4) was drawn using the qqman R package for all CpG sites from this analysis. A QQ plot for combined  $p$  values from *Analyses 1* and 2 is included in Additional file 2: Figure S5.

GO enrichment analysis was similarly undertaken to assess the functional significance of the 1,106 significant DNA methylation alterations between the case and control groups. A total of 505 GO functions had an enrichment score  $\geq 4$ , alongside  $p < 0.01$  (Additional file 1: Table S14 and Additional file 2: Figure S6). The processes with the top enrichment scores include several linked to immune responses including regulation of immune system processes and lymphocyte activation. The KEGG and REACTOME pathway analyses assessed the same list of genes in which the 1,112 dmCpGs were located. KEGG pathway analysis revealed 16 pathways (enrichment score of  $\geq 2$ , and  $p \leq 0.01$ ; Additional file 1: Table S15), including cancer, acute myeloid leukaemia and natural killer cell-mediated cytotoxicity. REACTOME pathway analysis identified 18 pathways ( $p < 0.01$ ) including

transcriptional regulation of granulopoiesis and transcriptional activation of cell cycle inhibitor p21 (Additional file 1: Table S16).

The eFORGE-TF database was searched for TF motifs associated with the top-ranked dmCpGs (FDR  $p \leq \times 10^{-8}$  and  $FC \geq \pm 2$ ) from within the most biologically plausible genes (*ARID5B*, *CUX1*, *FKBP5*, *HDAC4*, *ITGAL*, *LY9*, *PBX1*, *PIM1*, *RUNX3* and *SEPT9*, available in Table 2). Of the 21 dmCpGs examined, 11 reported TF motifs where  $q < 0.05$ . These are detailed in Additional file 1: Table S17.

#### Analysis 3: individuals with T1DM and ESKD who have received a kidney transplant compared to matched control individuals with T1DM and no evidence of kidney disease

The inclusion criterion for the case subjects was restricted to individuals with a functioning kidney transplant for the third analysis. Only blood-derived DNA samples from individuals who had received a kidney transplant were included to minimise potential confounding due to differences in medication and renal replacement therapy modalities ( $n=73$ ). The methylation status of the CpG sites for these individuals was compared to matched controls with T1DM and no evidence of kidney disease ( $n=73$ ). The BACR QC report and proportional WCCs for this analysis are included within Additional file 1: Tables S18 and S19, respectively.

In total, 1,515 top-ranked dmCpGs were different between cases and controls (FDR  $p \leq \times 10^{-8}$ ; Additional file 1: Table S20). Of those, 132 of the top-ranked dmCpGs remained when the stringency levels were increased (FDR  $p \leq \times 10^{-8}$  and  $FC \pm 2$ ), 81 of which were gene centric including two within *MTURN* and *UPF3A* (Additional file 1: Table S21). A Manhattan plot (Additional file 2: Figure S7) was drawn using the qqman R package for all CpG sites from this analysis.

Additional GO enrichment and pathway analyses were undertaken for the 132 top-ranked genes in which the dmCpGs were located (FDR  $p \leq \times 10^{-8}$  and  $FC \pm 2$ ) to assess their functional significance. In total, 75 GO functions were enriched with a score  $\geq 4$ , alongside  $p < 0.01$  (Additional file 1: Table S22, Additional file 2: Figure S8). The KEGG pathway analysis showed one result—the TGF- $\beta$  signalling pathway with an enrichment score of  $\geq 2$  and  $p \leq 0.01$ , Additional file 1: Table S23) and four REACTOME pathways were identified (entities  $p < 0.01$ ) including the nuclear receptor transcription pathway and signalling by TGF- $\beta$  family members (Additional file 1: Table S24).

The eFORGE-TF database was searched for TF motifs associated with the top-ranked dmCpGs (FDR  $p \leq \times 10^{-8}$  and  $FC \geq \pm 2$ ) from within the most biologically plausible genes (*ARID5B*, *FKBP5* and *UPF3A*, available in Table 2).

**Table 2** Summary results, highlighting top-ranked dmCpGs, GO enrichment functions and pathways per analysis

	<b>Analysis 1*</b> 107 T1DM-ESKD (transplant and dialysis) vs 107 T1DM	<b>Analysis 2</b> 107 T1DM-ESKD (transplant and dialysis) vs 253 T1DM	<b>Analysis 3*</b> 73 T1DM-ESKD (transplant only) vs 73 T1DM	<b>Analysis 4</b> 73 T1DM-ESKD (transplant only) vs 253 T1DM
<i>Genes with dmCpGs (FDR <math>p \leq 10^{-5}</math>)</i>				
Number of dmCpGs	4391	13,983	1518	13,739
Biological candidates linked to T1DM-ESKD (most significant dmCpG, <i>p</i> value, FC and direction in individuals with ESKD)	<p><b>AFF3</b> cg27315109 <math>p = 3.61 \times 10^{-11}</math> FC + 1.26 No additional dmCpGs</p> <p><b>ARID5B</b> cg22846816 <math>p = 8.14 \times 10^{-16}</math> FC + 1.47 Additional dmCpGs FC↑</p> <p><b>FKBP5<sup>b,g</sup></b> cg03546163 <math>p = 1.17 \times 10^{-24}</math> FC + 1.83 Additional dmCpGs FC↑</p> <p><b>HDAC4<sup>a</sup></b> cg25569341 <math>p = 1.38 \times 10^{-12}</math> FC - 1.60 Additional dmCpGs FC↑</p> <p><b>ITGAL<sup>b,c</sup></b> cg14889891 <math>p = 7.06 \times 10^{-14}</math> FC + 2.42 Additional dmCpGs FC↑</p> <p><b>LY9<sup>b,c</sup></b> cg18920397 <math>p = 3.57 \times 10^{-14}</math> FC + 2.59 Additional dmCpG FC↑</p> <p><b>PBX1</b> cg20178893 <math>p = 2.40 \times 10^{-12}</math> FC - 2.41 Additional dmCpG FC↓</p> <p><b>PIM1<sup>a</sup></b> cg25325512 <math>p = 2.79 \times 10^{-12}</math> FC + 2.61 Additional dmCpG FC↑</p> <p><b>RUNX3<sup>a,b,c,g</sup></b> cg25616056 <math>p = 3.69 \times 10^{-13}</math> FC + 2.27 Additional dmCpGs FC↑</p> <p><b>SEPTIN<sup>a,b,c,d,e,g</sup></b> cg11328665 <math>p = 8.07 \times 10^{-14}</math> FC + 1.43 Additional dmCpGs FC↑ + ↓</p> <p><b>UPF3A<sup>b</sup></b> cg12615906 <math>p = 8.86 \times 10^{-16}</math> FC + 2.32 Additional dmCpGs FC↑</p>	<p><b>AFF3</b> cg16812723 <math>p = 4.37 \times 10^{-13}</math> FC + 1.20 Additional dmCpGs FC↑</p> <p><b>ARID5B</b> cg06456847 <math>p = 1.54 \times 10^{-18}</math> FC + 1.42 Additional dmCpGs FC↑</p> <p><b>CUX1<sup>a,f</sup></b> cg25553730 <math>p = 1.04 \times 10^{-12}</math> FC + 1.40 Additional dmCpGs FC↑</p> <p><b>ELMO1<sup>a,c</sup></b> cg18707120 <math>p = 9.41 \times 10^{-14}</math> FC + 1.34 Additional dmCpGs FC↑</p> <p><b>FKBP5<sup>b,g</sup></b> cg03546163 <math>p = 6.70 \times 10^{-43}</math> FC + 1.66 Additional dmCpGs FC↑</p> <p><b>HDAC4<sup>a</sup></b> cg25569341 <math>p = 1.33 \times 10^{-17}</math> FC - 1.59 Additional dmCpGs FC↑</p> <p><b>ITGAL<sup>b,c</sup></b> cg16391678 <math>p = 3.77 \times 10^{-14}</math> FC + 1.27 Additional dmCpGs FC↑</p> <p><b>LY9<sup>b,c</sup></b> cg18920397 <math>p = 3.03 \times 10^{-13}</math> FC + 1.99 Additional dmCpGs FC↑</p> <p><b>PBX1</b> cg20178893 <math>p = 2.56 \times 10^{-16}</math> FC - 2.31 Additional dmCpGs FC↑ + ↓</p> <p><b>PIM1<sup>a</sup></b> cg19799178 <math>p = 1.62 \times 10^{-15}</math> FC + 3.40 Additional dmCpGs FC↑</p> <p><b>PRKAG2<sup>a</sup></b> cg07461953 <math>p = 1.17 \times 10^{-13}</math> FC + 1.68 Additional dmCpGs FC↑</p> <p><b>PTPRN2<sup>a,b,c,f</sup></b> cg23361114 <math>p = 1.66 \times 10^{-12}</math> FC + 1.18 Additional dmCpGs FC↑</p> <p><b>RUNX3<sup>a,b,c,g</sup></b> cg05656688 <math>p = 1.62 \times 10^{-14}</math> FC + 1.36 Additional dmCpGs FC↑</p> <p><math>p = 1.74 \times 10^{-17}</math> FC + 1.87 Additional dmCpGs FC↑</p>	<p><b>ARID5B</b> cg06456847 <math>p = 3.29 \times 10^{-15}</math> FC + 1.65 Additional dmCpGs FC↑</p> <p><b>ELMO1<sup>a,c</sup></b> cg00260664 <math>p = 1.67 \times 10^{-11}</math> FC + 1.26 No additional dmCpGs</p> <p><b>FKBP5<sup>b,g</sup></b> cg03546163 <math>p = 1.95 \times 10^{-15}</math> FC + 2.20 Additional dmCpGs FC↑</p> <p><b>RUNX3<sup>a,b,c,g</sup></b> cg08544331 <math>p = 9.98 \times 10^{-12}</math> FC + 1.23 No additional dmCpGs</p> <p><b>UPF3A<sup>b</sup></b> cg12615906 <math>p = 2.29 \times 10^{-12}</math> FC + 2.31 Additional dmCpGs FC↑</p>	<p><b>AFF3</b> cg16812723 <math>p = 8.80 \times 10^{-14}</math> FC + 1.23 Additional dmCpGs FC↑</p> <p><b>ARID5B</b> cg06456847 <math>p = 3.13 \times 10^{-20}</math> FC + 1.50 Additional dmCpGs FC↑</p> <p><b>CUX1<sup>a,f</sup></b> cg25553730 <math>p = 8.44 \times 10^{-13}</math> FC + 1.45 Additional dmCpGs FC↑</p> <p><b>FKBP5<sup>b,g</sup></b> cg03546163 <math>p = 3.65 \times 10^{-35}</math> FC + 1.68 Additional dmCpGs FC↑</p> <p><b>HDAC4<sup>a</sup></b> cg07315452 <math>p = 8.79 \times 10^{-16}</math> FC + 1.64 Additional dmCpGs FC↑ + ↓</p> <p><b>ITGAL<sup>b,c</sup></b> cg16391678 <math>p = 6.96 \times 10^{-14}</math> FC + 1.30 Additional dmCpGs FC↑</p> <p><b>LY9<sup>b,c</sup></b> cg18920397 <math>p = 2.92 \times 10^{-12}</math> FC + 2.10 Additional dmCpGs FC↑</p> <p><b>PBX1</b> cg21895155 <math>p = 1.53 \times 10^{-13}</math> FC - 2.60 Additional dmCpG FC↓</p> <p><b>PIM1<sup>a</sup></b> cg19799178 <math>p = 6.77 \times 10^{-12}</math> FC + 2.88 No additional dmCpGs</p> <p><b>RUNX3<sup>a,b,c,g</sup></b> cg05656688 <math>p = 7.38 \times 10^{-14}</math> FC + 1.39 Additional dmCpGs FC↑</p> <p><b>SEPTIN<sup>a,b,c,d,e,g</sup></b> cg00871371 <math>p = 4.73 \times 10^{-14}</math> FC - 3.05 Additional dmCpGs FC↑ + ↓</p> <p><b>TAMM41</b> cg21194040 <math>p = 4.75 \times 10^{-11}</math> FC = + 1.23 No additional dmCpGs</p> <p><b>UPF3A<sup>b</sup></b> cg12615906 <math>p = 2.66 \times 10^{-17}</math> FC + 1.97 Additional dmCpGs FC↑</p>

**Table 2** (continued)

	<b>Analysis 1*</b> 107 T1DM-ESKD (transplant and dialysis) vs 107 T1DM	<b>Analysis 2</b> 107 T1DM-ESKD (transplant and dialysis) vs 253 T1DM	<b>Analysis 3*</b> 73 T1DM-ESKD (transplant only) vs 73 T1DM	<b>Analysis 4</b> 73 T1DM-ESKD (transplant only) vs 253 T1DM
		<b>SEPTIN9<sup>a,b,c,d,e,g</sup></b> cg20772590 $p = 2.82 \times 10^{-15}$ FC + 1.50 Additional dmCpGs FC ↑ <b>UPF3A<sup>b</sup></b> cg12615906		
<i>Genes with dmCpGs (FDR <math>p \leq \times 10^{-8}</math> and FC <math>\pm 2</math>)</i>				
Number of dmCpGs	490	1112	132	1082
Biological candidates linked to T1DM-ESKD (most significant $p$ value, FC and direction in individuals with ESKD)	<b>ARID5B</b> cg02479789 $p = 3.89 \times 10^{-13}$ FC + 2.15 No additional dmCpGs <b>HDAC4<sup>a</sup></b> cg25487819 $p = 1.42 \times 10^{-11}$ FC + 2.54 Additional dmCpG FC ↑ <b>ITGAL<sup>b,c</sup></b> cg14889891 $p = 7.06 \times 10^{-14}$ FC + 2.42 Additional dmCpG FC ↑ <b>LY9<sup>b,c</sup></b> cg18920397 $p = 3.57 \times 10^{-14}$ FC + 2.59 Additional dmCpG FC ↑ <b>PBX1</b> cg20178893 $p = 2.4 \times 10^{-12}$ FC - 2.41 Additional dmCpG FC ↓ <b>PIM1<sup>a</sup></b> cg25325512 $p = 2.79 \times 10^{-12}$ FC 2.61 Additional dmCpG FC ↑ <b>RUNX3<sup>a,b,c,g</sup></b> cg25616056 $p = 4.7 \times 10^{-13}$ FC + 2.27 Additional dmCpG FC ↑ <b>SEPTIN9<sup>a,b,c,d,e,g</sup></b> cg04661929 $p = 1.35 \times 10^{-12}$ FC - 6.10 Additional dmCpGs FC ↑ + ↓ <b>UPF3A<sup>b</sup></b> cg12615906 $p = 8.86 \times 10^{-16}$ FC + 2.32 No additional dmCpGs	<b>CUX1<sup>a,f</sup></b> cg00243880 $p = 9.07 \times 10^{-12}$ FC + 2.10 Additional dmCpGs FC ↑ <b>FKBP5<sup>b,g</sup></b> cg03591753 $p = 5.72 \times 10^{-12}$ FC + 14.75 No additional dmCpGs <b>HDAC4<sup>a</sup></b> cg25487819 $p = 4.2 \times 10^{-13}$ FC + 2.06 Additional dmCpGs FC ↑ <b>ITGAL<sup>b,c</sup></b> cg23954865 $p = 4.39 \times 10^{-13}$ FC + 8.04 No additional dmCpGs <b>LY9<sup>b,c</sup></b> cg06056332 $p = 4.61 \times 10^{-11}$ FC + 2.17 No additional dmCpGs <b>PBX1</b> cg20178893 $p = 2.56 \times 10^{-16}$ FC - 2.31 Additional dmCpGs FC ↓ <b>PIM1<sup>a</sup></b> cg19799178 $p = 1.62 \times 10^{-15}$ FC + 3.40 Additional dmCpGs FC ↑ <b>RUNX3<sup>a,b,c,g</sup></b> cg03961551 $p = 2.84 \times 10^{-13}$ FC + 2.04 Additional dmCpGs FC ↑ <b>SEPTIN9<sup>a,b,c,d,e,g</sup></b> cg04661929 $p = 1.09 \times 10^{-13}$ FC - 5.33 Additional dmCpGs FC ↑ + ↓	<b>ARID5B</b> cg02479789 $p = 1.27 \times 10^{-12}$ FC + 2.28 No additional dmCpGs <b>FKBP5<sup>b,g</sup></b> cg03546163 $p = 1.95 \times 10^{-15}$ FC + 2.20 No additional dmCpGs <b>UPF3A<sup>b</sup></b> cg12615906 $p = 2.29 \times 10^{-12}$ FC + 2.31 Additional dmCpG FC ↑	<b>CUX1<sup>a,f</sup></b> cg00243880 $p = 3.18 \times 10^{-11}$ FC + 2.23 Additional dmCpGs FC ↑ <b>FKBP5<sup>b,g</sup></b> cg03591753 $p = 1.72 \times 10^{-11}$ FC + 16.14 Additional dmCpGs FC ↑ <b>HDAC4<sup>a</sup></b> cg25487819 $p = 6.43 \times 10^{-13}$ FC + 2.21 No additional dmCpGs <b>ITGAL<sup>b,c</sup></b> cg23954865 $p = 9.15 \times 10^{-13}$ FC + 9.03 No additional dmCpGs <b>LY9<sup>b,c</sup></b> cg18920397 $p = 2.92 \times 10^{-12}$ FC + 2.10 Additional dmCpG FC ↑ <b>PBX1</b> cg21895155 $p = 1.53 \times 10^{-13}$ FC - 2.59 Additional dmCpG FC ↓ <b>PIM1<sup>a</sup></b> cg19799178 $p = 6.77 \times 10^{-12}$ FC + 2.88 No additional dmCpGs <b>RUNX3<sup>a,b,c,g</sup></b> cg03961551 $p = 4.36 \times 10^{-12}$ FC + 2.14 Additional dmCpG FC ↑ <b>SEPTIN9<sup>a,b,c,d,e,g</sup></b> cg00871371 $p = 4.73 \times 10^{-14}$ FC - 3.05 Additional dmCpGs FC ↑ + ↓
<i>GO enrichment functions (<math>\geq 4</math> and <math>p &lt; 0.01</math>)</i>				
Number of GO functions	325	505	75	679
Top-ranked GO functions linked to T1DM-ESKD	Positive regulation of: Immune response immune system process leukocyte activation lymphocyte activation Ras protein signal transduction T cell activation	Positive regulation of: immune response immune system process leukocyte activation lymphocyte activation Ras protein signal transduction T cell activation	Positive regulation of: immune response immune system process	Positive regulation of: immune response immune system process leukocyte activation lymphocyte activation melanocyte differentiation T cell activation
<i>KEGG pathways (<math>\geq 2</math>, and <math>p \leq 0.01</math>)</i>				

**Table 2** (continued)

	<b>Analysis 1*</b> 107 T1DM-ESKD (transplant and dialysis) vs 107 T1DM	<b>Analysis 2</b> 107 T1DM-ESKD (transplant and dialysis) vs 253 T1DM	<b>Analysis 3*</b> 73 T1DM-ESKD (transplant only) vs 73 T1DM	<b>Analysis 4</b> 73 T1DM-ESKD (transplant only) vs 253 T1DM
Number of KEGG pathways	11	16	1	14
Top-ranked pathways linked to T1DM-ESKD	Human T-cell leukaemia virus 1 infection Th17 cell differentiation	Human T-cell leukaemia virus 1 infection Pathways in cancer Th17 cell differentiation Transcriptional misregulation in cancer	TGF-beta signalling pathway	Human T-cell leukaemia virus 1 infection Th17 cell differentiation Transcriptional misregulation in cancer

All genes included in Table 2 are reported in at least two of the categories except *PRKAG2*, *PTPRN2* and *TAMM41*. Where additional dmCpGs were located within a candidate gene listed, their FCs are indicated. Approved gene symbols are included in Table 2, the gene symbols as included in the manifest file from Illumina are included in the Additional file 1

CpGs cytosine-phosphate-guanine sites, dmCpGs differentially methylated CpG sites, eGFR estimated glomerular filtration rate, eQTL expression quantitative trait loci, ESKD end-stage kidney disease, FC fold change, FDR false discovery rate, GO gene ontology, KEGG Kyoto encyclopedia of genes and genomes, T1DM type 1 diabetes mellitus, T2DM type 2 diabetes mellitus

<sup>a</sup> eQTL support ( $p < \times 10^{-5}$ ) in American Indians (T2DM); <sup>b</sup>glomerular kidney tissue support ( $p < 0.05$ ); <sup>c</sup>tubular kidney tissue support ( $p < 0.05$ ); <sup>d</sup>kidney tubule support for eGFR slope ( $p < 0.05$ ) using 450K data; <sup>e</sup>kidney tubule support for fibrosis ( $p < 0.05$ ) using 450K data; <sup>f</sup>blood support for eGFR ( $p < 0.05$ ) using 450K data; <sup>g</sup>blood support for eGFR slope ( $p < 0.05$ ) using 450K data

Of the four dmCpGs examined, one reported a TF motif ( $q < 0.05$ ) which is detailed in Additional file 1: Table S25.

#### **Analysis 4: Individuals with T1DM who had received a kidney transplant compared to unmatched control individuals with T1DM and no evidence of kidney disease**

The methylation status of the CpG sites for individuals who had received a kidney transplant was compared to unmatched controls with T1DM and no evidence of kidney disease ( $n = 253$ ). The BACR and proportional WCCs QC report are included within Additional file 1: Table S26 and Additional file 1: Table S27, respectively. In total, 13,654 top-ranked dmCpGs were identified between both groups (FDR  $p \leq \times 10^{-8}$ , Additional file 1: Table S28). Of these, 9,064 dmCpGs were gene centric. When the stringency levels were increased (FDR  $p \leq \times 10^{-8}$  and  $FC \pm 2$ ), 1,070 CpG sites remained of which 714 were located within genes (Additional file 1: Table S29). A Manhattan plot (Additional file 2: Figure S9) was drawn using the qqman R package for all CpG sites from this analysis. Combined  $p$  values from Analyses 3 and 4 are presented in a QQ plot (Additional file 2: Figure S10) drawn using qqman. Seven genes including *ACAD8*, *LIME1*, *RPTOR* and *SEPTIN9* each contained three dmCpGs. In total, 78 dmCpGs (48 were within genes) overlapped between Analyses 3 and 4 (FDR  $p \leq \times 10^{-8}$  and  $FC \pm 2$ , Additional file 1: Table S30).

GO enrichment and pathway analyses were completed for the genes in which the 1,070 top-ranked dmCpGs were located (FDR  $p \leq \times 10^{-8}$  and  $FC \pm 2$ ). A total of 679 GO functions had an enrichment score  $\geq 4$  and  $p < 0.01$  (Additional file 1: Table S31 and Additional file 2: Figure S11). The processes with the top enrichment scores include regulation of cell activation, enzyme binding and

immune system processes. Fourteen KEGG pathways were identified (enrichment score of  $\geq 2$ , and  $p \leq 0.01$ ; Additional file 1: Table S32) including cancer and platelet activation, and 16 REACTOME pathways (entities  $p < 0.01$ ) including the SUMOylation of intracellular receptors and nuclear receptor transcription pathway (Additional file 1: Table S33).

The eFORGE-TF database was searched for TF motifs associated with the top-ranked dmCpGs (FDR  $p \leq \times 10^{-8}$  and  $FC \geq \pm 2$ ) from within the most biologically plausible genes (*CUX1*, *FKBP5*, *HDAC4*, *ITGAL*, *LY9*, *PBX1*, *PIMI1*, *RUNX3* and *SEPT9*, available in Table 2). Of the 14 dmCpGs examined, seven reported TF motifs ( $q < 0.05$ ), detailed in Additional file 1: Table S34.

#### **Significantly associated dmCpGs from analyses 1–4**

Thirty-six dmCpGs were identified as significantly different between cases and controls in each of the four analyses (FDR  $p \leq \times 10^{-8}$  and  $FC \pm 2$ ; Additional file 1: Table S35 and Additional file 2: Figure S12). The direction of fold change was consistent for each *Analysis*.

#### **Search for overlapping SNPs**

Previous studies have reported that SNPs could potentially impact on methylation status [30]. To assess this, we searched the Infinium HD Methylation SNP List [31] for any SNPs that could potentially impact the methylation array results if present in the test population. Of the top-ranked dmCpGs from this analysis (FDR  $p \leq \times 10^{-8}$ ,  $FC \pm 2$ ), five single CpG sites have the potential to be affected by SNPs (Additional file 1: Table S36). All but two of these SNPs, rs4788986 (*SEPTIN9*) and rs742232 (*RUNX3*) are very rare in European populations and would therefore be unlikely to impact in this study. Of

note, multiple dmCpGs were identified in both *SEPTIN9* and *RUNX3* genes with functional support provided for both these genes, making it improbable that these genes are falsely identified.

Additional searches were conducted for the presence of SNP-CpGs within the dmCpGs identified from within the most biologically plausible genes which are included in Table 2 using the BBMRI-NL atlas platform. Of the 28 dmCpGs examined, eight were shown to have been potentially affected by SNPs. These are included within Additional file 1: Table S37.

### STRING functional analyses

Functional network analyses were undertaken using STRING v11 [32] for the list of genes in which the top-ranked dmCpGs were located. Those which showed an increase in FC in the individuals with T1DM-ESKD compared to individuals with T1DM were analysed separately to those genes which showed a decrease in FC. All pathway interactions are shown in Additional file 2: Figures S13–S20.

### Functional data results

Functional support was sought from existing glomerular and tubular expression data alongside eGFR and fibrosis in both blood and kidney biopsy tissue using data ascertained from the 450K array (all  $p < 0.05$ ) [24, 33]. Seven differentially methylated genes were replicated in the glomerular database and six in the tubular (Additional file 1: Table S38; Table 2). The genes in which top-ranked dmCpGs were located, which also had the most biological plausibility for ESKD, were also assessed for eQTLs in an American Indian population with T2DM and known renal status, in the absence of a more suitable cohort with the same phenotype as our discovery population [34]. eQTLs demonstrated support for eight of the top-ranked dmCpG genes; *CUX1*, *ELMO1*, *HDAC4*, *PIMI*, *PRKAG2*, *PTPRN2*, *RUNX3*, *SEPTIN9* ( $p \leq 10^{-5}$ , Additional file 1: Table S38) [24, 33].

### Look-ups of top-ranked methylation markers

In the absence of an adequately powered, independent replication cohort using an EWAS array, we sought alternatives to provide supporting data to the top-ranked methylation markers reported from the discovery analysis. We have examined the association between the methylation of top-ranked CpGs and kidney function from this study and previous studies which had been performed on subjects with T2DM and known renal status, including cohorts with blood sample methylome analysis and others where methylation was studied in micro-dissected human kidney tubule samples. We confirmed

nominally significant association with kidney function at the *CUX1*, *PTPRN2* and *SEPTIN9* locus and with kidney function decline at *AFF3*, *HDAC4* and *SEPTIN9* regions. Furthermore, the methylation level at the *AFF3*, *CUX1* and *PIMI* regions correlated with the degree of fibrosis in micro-dissected human kidney tissue samples.

### Summary

A summary of results is included within Table 2, Additional file 2: Figures S21 and S22, including specific details regarding the strongest biologically plausible candidates linked to T1DM-ESKD, the GO functions and pathways. Table 2 highlights the top-ranked genes and pathways of interest and states the number of significant results from each section of the analysis.

### Discussion

Epigenetic alterations provide a dynamic link between genetic background and environmental exposures. These alterations have been proposed to play an important role in kidney disease [20, 23] and T1DM [35]. Previous research assessing T1DM-DKD [19–22, 36] has identified dmCpGs using bisulphite pyrosequencing and either the Illumina Infinium 450K or 27K methylation arrays. This manuscript describes differences in DNA methylation patterns between individuals with T1DM and those with T1DM-ESKD, utilising the MethylationEPIC array technology.

Several top-ranked dmCpGs identified through this study showed an increase in FC related to the more severe phenotype of T1DM-ESKD compared to the T1DM control population.

The genes from each analysis with the most biological plausibility having assessed the literature are included in Table 2 and Additional file 2: Figures S21 and S22. Of the 16 genes in which dmCpGs were located with a significance level of  $FDR p \leq \times 10^{-8}$ , (Additional file 2: Figure S21), 11 retained this status when the stringency level was increased ( $FDR p \leq \times 10^{-8}$  and  $FC \pm 2$ ; Additional file 2: Figure S22). DmCpGs were located within *FKBP5* and *RUNX3* in each of the analyses where  $FDR p \leq \times 10^{-8}$ . When the stringency was increased ( $FDR p \leq \times 10^{-8}$  and  $FC \pm 2$ ), they remained top-ranked genes in all but *Analyses 1* and *3*, respectively.

*FKBP5*, also known as FK506 binding protein 51, acts as a co-chaperone of hsp90 to aid the modulation of glucocorticoid receptor sensitivity in response to stress [37, 38]. Polymorphisms in this gene lead to an extended stress hormone response following exposure [37]. Recently, genome-wide analyses of human blood found associations between *FKBP5* mRNA and a pro-inflammatory profile [39]. Moreover, aberrant *FKBP5* methylation has previously been implicated in the pathology



of numerous diseases, particularly in diseases common in older populations. This hypermethylation is exemplified in myocardial infarction [39] and conditions such as T2DM [40] and CKD [20]. All top-ranked dmCpGs within this gene were located in either the gene body or 5' UTR and showed increased methylation for individuals with T1DM-ESKD. As a hallmark feature of CKD is persistent, low-to-moderate levels of circulating inflammatory markers [41], with distinguishing features such as nephron loss with subsequent acceleration of organ fibrosis, further study is required to determine if *FKBP5* plays a mechanistic role in CKD development.

The runt-related transcription factor 3 (*RUNX3*) plays a downstream role in the TGF- $\beta$  signalling pathway [42]. Its suppression has been implicated in tumour growth, migration and invasion [43]. In 2019, Cen et al. reported that higher methylation levels in *RUNX3* were associated with a shorter renal cell carcinoma survival time [44]. Additionally, they suggested independent predictors of heightened methylation levels of this gene included the presence of intra-tumour vascularity. We identified increased FC in dmCpGs within the *RUNX3* gene body and the UTRs in association with T1DKD-ESKD.

As indicated in Additional file 2: Figure S22, a further six genes, *HDAC4*, *ITGAL*, *LY9*, *PBX1*, *PIMI* and *SEP-TIN9* had top-ranked dmCpGs from three of the four analyses. These genes did not reach statistical significance in *Analysis 3*, which had the smallest population which compared DNA methylation patterns in 73 individuals with T1DM-ESKD (transplant only) to 73 individuals, matched, with T1DM.

Histone deacetylases (HDACs) are a group of enzymes that are characterised into three defined classes, known as I, II and III [45, 46]. They each have roles in removal of acetyl groups from histone and non-histone proteins, chromatin condensation and transcriptional repression [45, 46]. They have the ability to impact cellular function via both epigenetic and non-epigenetic mechanisms [46]. *HDAC4* (a member of class II) was reported to reduce kidney injury during in vivo animal studies [45–47]. Additionally, *HDAC4* was found to be over-expressed in kidney epithelial cells of a murine kidney fibrosis model [48]. Subsequent treatment with HDAC inhibitors demonstrated that the development and progression of kidney fibrosis can be inhibited by suppressing the activation and expression of numerous pro-fibrotic molecules, such as fibronectin and collagen 1 [48]. In the present study, *HDAC4* showed an increased FC for all dmCpGs except cg25569341 in individuals with ESKD undergoing dialysis or who had undergone a transplant.

Integrins are integral membrane proteins which are heterodimeric in nature. They are comprised of alpha and

beta chains which together form the integrin lymphocyte function-associated antigen-1, expressed in leukocytes. *ITGAL* encodes the alpha L chain and its expression has been previously linked to renal cell carcinoma [49]. A second report has demonstrated that DNA methylation of the *ITGAL* gene is heavily methylated in fibroblasts and demethylated in T lymphocytes [50]. The top-ranked dmCpGs present within the gene body or north shelf of *ITGAL* showed an increase in FC in individuals with T1DM-ESKD who were immunosuppressed following transplantation, compared to those with T1DM with no evidence of kidney disease.

Lymphocyte antigen 9 (*LY9*) is a member of the signalling lymphocyte activation molecule family receptor and is involved in immune responses. In 2019, Parikova et al. [51] reported expression of *LY9* to be significantly increased in individuals receiving long-term dialysis compared to those who had received dialysis over a shorter time period. Our findings support these observations showing an increased FC in *LY9* dmCpGs located within either the gene body or 5' UTR in individuals with T1DM-ESKD compared to those with no kidney disease.

*PBX1* encodes a nuclear protein within the PBX homeobox family of transcription factors. Through this, it can affect the expression of several genes including those which regulate insulin action and glucose metabolism [52]. In the current study, dmCpGs within the *PBX1* gene body showed a consistent decrease in FC at the highest stringency level (FDR  $p \leq \times 10^{-8}$ , FC  $\pm 2$ ) in individuals with T1DM-ESKD across each of the four analyses. Previous analyses of this gene have linked differential methylation patterns to higher birth weight-for-gestational age [53] and a translocational rearrangement of this gene and *TCF3/E2A* has been associated with B-cell acute lymphoblastic leukaemia [54]. Additionally, *PBX1* haploinsufficiency has been linked to congenital anomalies of the kidney and urinary tract [55], reported to be involved in the proliferation of cells in renal cell carcinoma [56] and has been recognised as a candidate gene for T2DM [52, 57].

*PIMI* belongs to the serine/threonine kinase family [58] and its overexpression has previously been implicated in diseases such as ovarian cancer [59] and breast cancer [60]. Overexpression of *PIMI* appears to influence cancer development in three ways, preventing apoptosis, enhancing cellular proliferation and through promoting genomic instability [61]. More specific to kidney research, *PIMI* is aberrantly overexpressed in renal cell carcinoma [62] and lupus nephritis [63]. In this analysis, we identified top-ranked dmCpGs present within the 3' UTR and north shore of *PIMI* which resulted in an increase in the FC in individuals with T1D-ESKD.

Septins are a group of GTPase proteins that play a role in cytoskeleton organisation through their links with microtubules and actin filaments [64]. Although some septins are associated with the advancement of kidney fibrosis [64], *SEPTIN9* specifically has been implicated in a host of disease pathologies, such as prostate cancer [65], colorectal cancer [66], liver fibrosis [67] and T2DM [68]. Moreover, *SEPTIN9* overexpression has been shown to promote kidney epithelial cell migration [69]. Additionally Dayeh et al. reported *SEPTIN9*, alongside *PTPRN2*, as one of the top-ranked differentially methylated genes when comparing pancreatic islets from individuals with and without T2DM [70]. In this analysis, 20 dmCpGs were located in *SEPTIN9*, 17 of which displayed an increase in FC in individuals with T1DM-ESKD.

An additional three top-ranked genes which contained dmCpGs were common in two analyses; *ARID5B* and *UPF3A*, which were identified from *Analyses 1* and *3*, and *CUX1* which resulted from *Analyses 2* and *4* (FDR  $p \leq \times 10^{-8}$  and  $FC \pm 2$ ; Additional file 2: Figure S22).

Each of the dmCpGs within *ARID5B* were located within the gene body and all were consistently increased in cases with ESKD (both chronic dialysis and transplant recipients) compared with controls with long duration of T1DM and no evidence of kidney disease. It has been previously demonstrated that *ARID5B* has a regulatory role in the phenotypic change of smooth muscle cells and SNPs within *ARID5B* have been linked to T2DM and coronary artery disease [71]. Differential methylation of a six-probe region spanning 99 base pairs within *ARID5B* gene has also been reported in Alzheimer's disease [72]. *ARID5B* is also known to contribute to cell growth, the differentiation of B-lymphocyte progenitors and has additionally been linked to acute lymphoblastic leukaemia [73–75].

*UPF3A* encodes a protein involved in mRNA nuclear export and mRNA surveillance, it has a crucial role in downregulating aberrant mRNAs [76]. Gotoh et al. in 2014 reported that this gene alongside 14 others had significantly reduced mRNA expression levels in renal cell carcinoma compared to non-cancerous kidney cortex tissue [77]. Top-ranked dmCpGs present within *UPF3A* consistently displayed an increase in FC within individuals with T1DM and ESKD.

*CUX1* is a protein coding member of the homeodomain family of DNA binding proteins and is involved in cell cycle regulation and kidney development through the inhibition of p27 which promotes cell proliferation in the nephrogenic zone [20, 78, 79]. Previously, significant DNA methylation alterations have been reported in CKD where more than one dmCpG site was located within *CUX1* compared to individuals with no evidence of kidney disease [20]. Additionally, genetic abnormalities

within *CUX1* have been linked to polycystic kidney disease [80] and myelodysplastic syndrome [81] in mouse models. Significant dmCpGs from *Analyses 2* and *4* were located within the body of this gene which additionally showed an increase in FC in individuals with T1DM-ESKD, but this gene has not previously been explored in individuals post-kidney transplant.

Lastly, an additional five genes with biological plausibility for ESKD had statistical significance for this phenotype but did not show a large change in methylation ( $FC \pm 2$ ); *AFF3*, *ELMO1*, *PRKAG2*, *PTPRN2* and *TAMM41* (Additional file 2: Figure S21, FDR  $p \leq \times 10^{-8}$ ). These genes contained dmCpGs from either one (*PRKAG2*, *PTPRN2* and *TAMM41*), two (*ELMO1*) or three (*AFF3*) analyses.

In 2012, the GENIE consortium conducted a meta-analysis of GWAS in T1D-DKD, which revealed an intronic SNP (rs7583877) located in the AF4/FMR2 family member 3 (*AFF3*) gene as significantly associated with ESKD [14, 82]. Functional studies have indicated that *AFF3* influences kidney tubule fibrosis through the TGF- $\beta$ 1 pathway [14]. The findings from our study show increased methylation levels in the dmCpGs within the body of *AFF3* in the individuals with T1D-ESKD, which could result in decreased gene expression. A link between DNA methylation of this gene and T1DM has previously been considered as a mediator of the genetic risk [83]. Each of the top-ranked dmCpGs identified from this analysis were present within the body of *AFF3* with increased methylation in individuals with T1DM-ESKD.

*ELMO1* encodes a member of the engulfment and cell motility protein family and has been previously linked to T2DM [84–86], hepatocellular carcinoma [87] and inflammatory arthritis [88]. Previous methylation analyses of this gene have shown associations with gastric cancer [89] and CKD [20]. The eight top-ranked dmCpGs within *ELMO1* were present within various regions of the gene and each dmCpG site reported an increase in the FC in individuals with T1DM-ESKD. One hypermethylated dmCpG site, cg01119452, had been previously reported in 2014, where it also showed hypermethylation in its association with CKD [20].

*PRKAG2* is an important regulator of cellular energy status and has previously been associated with eGFR [90]. SNP rs7805747 located in *PRKAG2* has been reported in association with both CKD and serum creatinine at genome-wide significance level [91]. Differential methylation within this gene has also previously been reported in association with CKD [20]. The top-ranked dmCpGs were present within the 5' UTR or body of *PRKAG2* and each showed an increase in FC in individuals with T1DM-ESKD compared to those with no evidence of kidney disease.

Identified as an auto-antigen in diabetes, *PTPRN2* has previously been linked to CKD [20, 92], fasting plasma glucose and obesity [93]. *PTPRN2* encodes islet antigen (IA)-2 $\beta$  and together with IA-2, these are integral membrane proteins of dense core vesicles which are expressed throughout the body in neuroendocrine cells [94]. *PTPRN2* was reported by Dayeh et al. as second of the top-ranked differentially methylated genes in a comparison of pancreatic islets from individuals with and without T2DM [70]. Each of the dmCpGs present within *PTPRN2* showed a consistent increase in FC in association with T1DM-ESKD.

The function of *TAMM41* in higher vertebrates still remains largely undetermined, yet it is known to play a critical role in yeast cell mitochondrial membrane maintenance [95]. Using zebrafish models of human CVD, it was determined that the developing heart overexpressed *tamm41* [95, 96]. Furthermore, CRISPR/Cas9-mediated knockout of the *tamm41* gene resulted in immature heart valve formation [95]. Differential methylation in *TAMM41* has previously been reported in both DKD and ESKD [22].

Several of the top-ranked pathways and genes defined by dmCpGs have been previously linked to leukaemia. This is not unexpected as an elevated risk of leukaemia during dialysis and after transplant failure has previously been reported in an Australian and New Zealand population [97, 98]. Although rare, Alfano et al. suggested that leukaemia can occur during the post-transplant period [99].

Five of the genes in which top-ranked gene-centric dmCpGs were located from this analysis have been previously identified as strong biological candidates for chronic kidney disease in a previous investigation published by Smyth et al. [20]. The *CUX1*, *ELMO1*, *FKBP5*, *PRKAG2* and *PTPRN2* genes included CpGs with statistically significant differential methylation levels between cases and controls in both studies. Furthermore, four dmCpGs, two within *ELMO1* cg08044454 and cg01119452, and two within *FKBP5*, cg03546163 and cg14284211 show evidence of differential methylation between cases and controls in both investigations. Different Infinium BeadChip arrays were used for the analyses. The Infinium 450K array utilised in 2014 does not cover all CpGs sites which are included in the analysis using the Infinium MethylationEPIC BeadChip.

T1DM is a polygenic disease with several candidate genes previously reported [100]. Of these candidate genes, 23 are found to include significantly dmCpGs in at least one of the four analyses. *BACH2*, a gene previously linked to immunodeficiency, is a candidate gene for T1DM [101, 102] which showed increased methylation levels at dmCpGs resulting from Analyses 1 and 2.

Additionally, of the 10 dmCpGs within *BACH2* from Analysis 4, nine presented with increased methylated levels in cases compared to control individuals. DmCpGs were also located within *CIQTNF6*, *CCR5*, *CLEC16A*, *IL2RA*, *IL7R*, *PGM1*, *PRKCQ*, *TNFAIP3* and *UBASH3A* (Analyses 1, 2 and 4), alongside *TAGAP* (Analyses 2, 3 and 4). Each of these candidate genes were first reported to have links with T1DM in 2007–8 [100, 101, 103–107].

Additional previously reported candidate genes including *CD226*, *CTLA4*, *IL10*, *IL18RAP* and *IL27* [100, 103, 104, 108] were shown to include dmCpGs in two analyses undertaken in this study. *COBL*, *GLIS3*, *IFIH1*, *ORMDL3*, *PTPN2*, *RGS1* included dmCpG sites in one of the four Analyses [100, 103, 104, 109].

*SH2B3* encodes a member of the SH2B adaptor family of proteins, involved in a range of signalling activities and is involved in haematopoiesis [110, 111]. Mutations within this gene have been shown to increase susceptibility for T1DM [104, 112]. DmCpGs within this gene were identified within each of the four analyses undertaken, with each dmCpG site showing an increased level of methylation in cases compared to controls.

STRING v11: protein–protein association networks were searched for all genes in which top-ranked dmCpGs from each analysis (FDR  $p \leq 10^{-8}$  and  $FC \pm 2$ ) were located. Those with an increased FC were searched separately to those with a decreased FC when cases and controls were compared. A strong interaction network was identified between LCK and three proteins including MAPK8IP3, SLA2 and PRKCE, each gaining a score of  $\geq 0.9$  when the genes with a decreased FC from Analysis 1 were searched. Scores  $\geq 0.95$  were obtained for 20 interactions resulting from the genes with an increased FC from Analysis 1 including UPF1-UPF3A, ITGAL-ITGB2 and LCK-CD247. *ITGAL* and *UPF3A* were selected as two of the most biologically plausible genes linked to T1DM-ESKD, as discussed previously. CD247 [113] and LCK [114] have both been previously linked to diabetes and which could warrant further investigation. Furthermore, two interactions scored 0.999 from both Analyses 2 and 4: ITGAL-ITGB2 and BRE-UIMC1. Neither *BRE* nor *UIMC1* have previously been linked to T1DM or ESKD.

In the absence of a complementary replication cohort with available data, further support for top-ranked genes was sought from eQTL analysis from an American Indian cohort where the individuals had T2DM and known renal status [34]. eQTL analysis supports associations for eight of the top-ranked genes: *CUX1*, *ELMO1*, *HDAC4*, *PIM1*, *PRKAG2*, *PTPRN2*, *RUNX3*, *SEPTIN9* ( $p \leq 10^{-5}$ , Additional file 1: Table S33). Functional support was also generated for these top-ranked markers in kidney tissue (glomerular tissue and tubule tissue from two cohorts

[33, 34]) alongside supporting data from additional blood-derived DNA from individuals with kidney function assessed using the 450K array (unpublished data) [33]. Unfortunately, the 450K array for DNA methylation includes <25% of the top-ranked dmCpGs identified by our study which used the more comprehensive EPIC array. It is not unexpected that the dmCpGs did not correlate across all top-ranked results due to the differences in material and technologies employed.

Through this analysis, we have demonstrated that similar association results are reported for individuals who have received a kidney transplant or persons who are receiving dialysis. This shows that transplant recipients can be analysed alongside individuals receiving dialysis to increase the power of future EWAS for ESKD. The results, genes,  $p$  values and FC directions are consistent across each of the four analyses. *RPTOR* is the only top-ranked gene linked to immunosuppressant medication for transplants [115].

We have reported associations with several biologically plausible genes. Employing a considered approach, including four analyses (two of which were matched and two were unmatched), we show broad overlap in results from the analyses. This indicates that for future large-scale studies, it is not essential to stratify this analysis by age and sex matching the participants, which will facilitate a larger, more powerful unmatched case–control design.

### Strengths and limitations

Overall, this study has several strengths. We have carefully defined phenotypes for each analysis, ensuring individuals were well matched for paired *Analyses 1* and *3*, with a larger cohort of individuals employed for unmatched *Analyses 2* and *4*. *Analyses 3* and *4* represented an extension of the initial analyses to evaluate the differences between individuals with T1DM and DKD compared to those with a more extreme phenotype who had progressed to T1DM-ESKD and had received a kidney transplant.

We have extended previous investigations through assessing T1DM-ESKD using the most cost-effective high-density methylation array available, the Infinium MethylationEPIC [116]. Methylation signatures were assessed using peripheral blood samples, while considering estimates of proportional WCCs and cell heterogeneity. We also assessed both the beta and M values which were derived from the methylationEPIC arrays. Furthermore, only top-ranked dmCpGs that met a significance level of  $FDR p \leq \times 10^{-8}$  were reported, a threshold previously reported to reduce the rate of false-positives in

studies which use the Infinium MethylationEPIC array [117].

This study is limited in not having a complementary replication cohort; however, we took the approach that a larger discovery cohort using the most comprehensive array commercially available was the optimal study design given similar kidney transplant cohorts do not have EPIC array data available for replication. This is similar to the approach taken for GWAS, in that the largest possible discovery cohort is selected whilst supporting data are sought for top-ranked results. While we have provided details of all dmCpGs with  $FDR p \leq \times 10^{-8}$  and  $FC \pm 2$  in supplementary material, this manuscript focuses mainly on genes that have previously been linked to similar phenotypes in the literature. A larger-scale multi-omic analysis incorporating genetic variation, epigenetic alterations and gene expression would be required to further determine the markers of interest for this phenotype and improve understanding of the biological mechanisms involved. Additional investigations could be undertaken to assess this phenotype in different ethnicities, sex-specific effects and analyses could be conducted to compare these results with those derived from kidney biopsy samples.

### Conclusion

Epigenetic alterations, unlike genetic changes, are potentially reversible, offering opportunistic therapeutic interventions. Through this exploratory investigation, we have reported associations between dmCpGs and genes with T1DM-ESKD, several of which suggest complementary genetic and epigenetic influences to alter gene expression. Eight top-ranked genes also showed eQTL support in a T2DM American Indian cohort and 13 were supported by gene expression and / or methylation data from kidney tubule or glomerular tissues. Additional prospective studies may help identify whether the underlying methylation influences are causal or consequential.

The identification of unique epigenetic profiles associated with developing ESKD could highlight additional biological mechanisms to study kidney disease. Epigenetic profiles may also help to identify patients at greater risk for progression to ESKD. Targeting healthcare resources to individuals at highest risk for ESKD remains an important clinical goal.

### Methods

#### Samples

Each participant was recruited as part of the All Ireland-Warren 3-Genetics of Kidneys in Diabetes (GoKinD) United Kingdom Collection. All participants were previously recruited, had White ancestry and provided written informed consent for research. DNA was frozen in

multiple aliquots following extraction from whole blood using the salting out method [118] and normalised using PicoGreen quantitation [119] using the CytoFluor® Series 4000 (Applied Biosystems, Thermo Fisher Scientific, CA, USA).

Individuals with both T1DM and ESKD were defined as cases ( $n = 107$ ). These individuals had  $\geq 10$  years duration of T1DM alongside a diagnosis of DKD defined as persistent macroalbuminuria ( $\geq 500$  mg/24 hr), estimated glomerular filtration rate (eGFR)  $< 60$  mL/min/m<sup>2</sup> calculated using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) creatinine equation, hypertension (systolic/diastolic blood pressure  $\geq 135/85$  mmHg) and ESKD. They were not taking any anti-hypertensive medication. The control individuals had  $\geq 15$  years duration of T1DM and no evidence of kidney disease on repeat testing, i.e. they all had normal urinary albumin excretion and eGFR  $> 60$  mL/min/m<sup>2</sup> ( $n = 253$ ).

#### MethylationEPIC array

Blood-derived DNA from each participant was accurately quantitated using PicoGreen® prior to normalisation. In total, 800 ng of DNA from each participant was bisulphite treated using the EZ Zymo Methylation Kit (D5002, Zymo Research, CA, USA) following the alternative overnight incubation conditions for use prior to the Illumina® Infinium MethylationEPIC Kit provided in the published protocol. All samples were prepared and analysed using the Infinium MethylationEPIC Kit and BeadChips (Illumina, CA, USA) with no protocol deviations. All samples were processed in a consistent laboratory workstream by the same members of trained staff and methylation arrays were scanned using a dedicated iScan machine with regular monitoring of laser intensity levels. Case and control samples were randomly distributed across the BeadChip arrays. This array is a high-throughput platform which provides quantitative evaluation of methylation levels ( $\beta$  values) with single nucleotide resolution. In total, 862,927 sites were examined by the Infinium MethylationEPIC array. The Infinium HD Methylation SNP List was searched for any SNPs that may have impacted upon the methylation array results [31]. Top-ranked dmCpGs located within biologically plausible genes were further examined using the web-based Biobanking and Biomolecular Resources Research Infrastructure, Netherlands (BBMRI-NL) database tool to identify potential SNP-CpGs [120].

#### Quality control

Each resulting.idat file generated from the iScan was assessed using BeadArray Controls Reporter (BACR) Software (Illumina) for quality control (QC). This

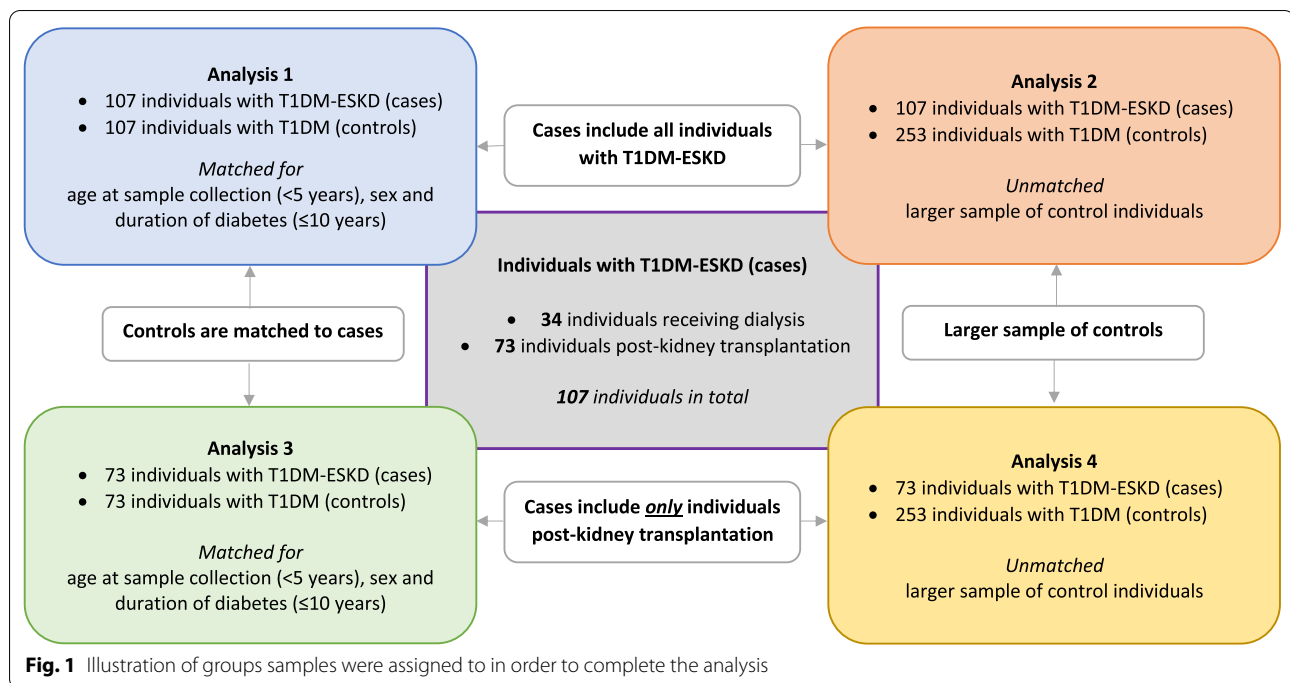
software assessed the data in connection with a pre-set standard set of controls and evaluated the hybridisation, extension, dye specificity and bisulphite conversion process. An additional QC measure to determine the concordance of average beta values generated for seven duplicate samples was completed using GenomeStudio (Illumina) v1.8, methylation module including a sex check of all included individuals.

Proportional white cell counts (WCCs) were estimated following the Houseman method [29] using the raw.idat files output from the iScan machine. The *minfi* Bioconductor (v3.10) package was utilised to estimate six WCCs, CD8+T, CD4+T and CD19+B lymphocytes, CD56+natural killer cells, CD14+monocytes and CD15+granulocytes using the *estimateCellCounts* function for both the case and control groups, and a Northern Irish general population control group [121].

#### Differentially methylated loci analysis

Case and control groups were investigated for dmCpGs using Partek Genomics Suite (PGS) v7.19.1125 (Partek, MO, USA) following functional normalisation. All software was used following the developer's instructions. Beta values were generated before M values were calculated. DmCpGs were determined using the M values for individuals with T1DM and ESKD compared to controls with long duration of T1DM and no evidence of kidney disease on repeat testing. Parameters were set at false discovery rate (FDR) adjusted  $p$  value threshold of  $\leq \times 10^{-8}$  alongside a fold change (FC)  $\geq \pm 2$ , both calculated using PGS. Sex chromosomes were removed during the analysis (846,232 CpG sites were examined) alongside any probes with a poor detection  $p$  value ( $p > 0.01$ ). Related genes for each of the CpG sites were annotated based on Homo sapiens hg19 genome build using PGS and the Infinium MethylationEPIC v1.0 B4 Manifest File. Four analyses were performed (Table 1), each assessing differential methylation patterns in individuals with T1DM-ESKD to those with T1DM and no evidence of renal disease (Fig. 1). Briefly,

1. 107 matched pairs, where both the case and control individuals were matched for age at sample collection (differing by  $< 5$  years), sex and duration of diabetes (differing by  $\leq 10$  years). Case individuals had a functioning kidney transplant or were receiving dialysis. Controls individuals had T1DM and no evidence of kidney disease.
2. The same 107 individual cases from *Analysis 1* were compared to a larger sample size of (unmatched) control individuals with T1DM and no evidence of kidney disease ( $n = 253$ ).



- 73 matched pairs for age at sample collection (differing by <5 years), sex and duration of diabetes (differing by ≤10 years). The case subjects were restricted to individuals with a functioning kidney transplant to minimise potential confounding due to differences in medication and renal replacement therapy modalities. These individuals were compared to matched controls with T1DM and no evidence of kidney disease.
- The same 73 individuals from *Analysis 3* were compared to a larger sample size of (unmatched) control individuals with T1DM and no evidence of kidney disease ( $n = 253$ ).

Resulting dmCpGs which overlapped each of the four analyses were extracted and displayed in a heatmap which highlighted their degree of hyper- and hypo-methylation. Manhattan plots and QQ plots were drawn using the qqman R package [122].

#### Functional gene ontology and network analyses

Gene functionality was examined by gene ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis using PGS and the Reactome pathway database [123]. Solely genes in which top-ranked dmCpGs were located ( $FDR p \leq \times 10^{-8}$  and  $FC \pm 2$ ) were examined. Enrichment score thresholds were set at an enrichment score  $\geq 4$  and  $p < 0.01$  for GO, an enrichment score of  $\geq 2$ , and  $p \leq 0.01$  for KEGG pathway analysis and  $p < 0.01$  for the Reactome pathway

database. An additional functional analysis was undertaken using STRING v11: protein–protein association networks [32]. For each of the four analyses, genes with top-ranked dmCpGs were categorised according to increased or decreased FC.

#### Tertiary analysis of dmCpGs and potential overlaps with transcription factors (TFs)

Top-ranked dmCpGs ( $FDR p \leq \times 10^{-8}$  and  $FC \pm 2$ ) from biologically plausible genes (Table 2) were examined using eFORGE-TF, a software suite that assesses CpG sites for TF motif enrichment [124]. The CpGs were searched against previously acquired kidney data available in the online database. All significant motifs ( $q < 0.05$ ) have been reported.

#### Kidney gene expression data

Participants for this study were American Indians with type-2 diabetes [125]. Protocol kidney biopsy samples were obtained from this cohort ( $n = 97$ ). The study was approved by the Institutional Review Board of the National Institute of Diabetes and Digestive and Kidney Diseases and each participant signed an informed consent document. Expression profiling of the kidney biopsies was previously carried out using Affymetrix GeneChips (HumanGenome U133 Array and U133Plus2 Array and Affymetrix Human Gene ST GeneChip 2.1) as reported [126, 127], and RNA-seq (Illumina)[13]. Expression quantitative trait (eQTL) mapping was performed using an EPACTS (<https://genome.sph.umich.edu/>)

wiki/EPACTS) software tool using linear mixed model accounting for hidden familial relatedness, after inverse Gaussian transformation of expression levels, adjusting for age and sex. Significance threshold was set at  $p \leq 10^{-5}$  for eQTL and  $p \leq 0.05$  for association studies.

For morphometry measures, kidney biopsy tissue was prepared for light and electron microscopy studies according to standard procedures [128–130]. The following glomerular structural parameters were measured by unbiased morphometry on electron microscopy images as described elsewhere [128, 129, 131]: glomerular basement membrane (GBM) width [132, 133], mesangial fractional volume (VVMES) [132, 133], glomerular filtration surface area per glomerular volume (SV) [132, 133], percentage of endothelial fenestrations (P\_FEN) [134] and the fractional podocyte volume per glomerulus (VVPC) [135]. Cortical interstitial fractional volume (VVINT) [130] was estimated using light microscopy with all specimens embedded in paraffin.

#### Analysis of top-ranked genes using the 450K array

In the absence of a complementary replication cohort, which is unavailable for a population with ESKD and T1DM, we sought Infinium 450K data from collaborators. In order to further assess the top-ranked methylation markers from the discovery cohorts, the top-ranked genes showing alterations in DNA methylation were assessed in 327 peripheral blood samples obtained from Pima Indians [24]. Blood samples for DNA isolation and cytosine methylation analysis were collected at the baseline examination from subjects. EGFR measures were available from clinical records before and after the nephrectomy. Among these subjects, 318 had longitudinal eGFR measurements with mean follow-up time of 9.8 years with a standard deviation (SD) of 3.9 years. The best linear unbiased predictor (BLUP) model was used to calculate eGFR slope for each subject.

Human kidney samples ( $n=91$ ) were obtained from surgical full/partial nephrectomies. Kidney samples were ~0.5 cm in diameter and were surrounded by at least 2 cm of normal tissue margins. Clinical data were obtained at the time of the sample collection and for a subset of samples eGFR measures were available from clinical records before and after the nephrectomy ( $n=69$ ). Only subjects with longitudinal eGFR measurements for at least 3 months post-nephrectomy were included for the analysis. The mean timespan of follow-up was 2.4 years (SD = 1.5 years).

DNA methylation was detected using Illumina 450K arrays and values were extracted using the minfi package. Several quality control measures were carried out and after the removal of CpG probes that (1) were in the

proximity of regions with genetic variations, (2) were located on the sex chromosomes, (3) were known to cross-hybridise to other locations, (4) had poor detection  $p$  values ( $p > 0.01$ ) or (5) were control probes, 321,473 CpG probes remained in the blood cohort used for ‘look-ups’ cohort. The same pre-processing was performed for this kidney cohort which resulted in 355,141 CpG probes remaining.

In blood-derived DNA cohort, linear regression models were used to determine the association between DNA methylation, eGFR and eGFR slope, with covariates including sex, age, duration, mean arterial blood pressure, HbA1c, batch, converse and estimated cell fraction. Association between cytosine methylation level, eGFR and interstitial fibrosis was determined by linear regression models adjusted by age, sex, race, diabetes, hypertension, batch effect, bisulphite conversion efficiency and degree of lymphocytic infiltrate on histology. A  $p$  value of  $< 0.05$  was used to determine significance in this population.

#### Abbreviations

BACR: Bead Array Controls Reporter; CKD: Chronic kidney disease; DKD: Diabetic kidney disease; dmCpGs: Differentially methylated; eGFR: Estimated glomerular filtration rate; eQTL: Expression quantitative trait loci; ESKD: End-stage kidney disease; EWAS: Epigenome-wide association study; FC: Fold change; FDR: False discovery rate; GBM: Glomerular basement membrane width; GENIE: Genetics of Nephropathy an International Effort; GO: Gene ontology; GoKinD: Genetics of Kidneys in Diabetes; GWAS: Genome wide association studies; HbA1c: Haemoglobin A1c; HDACs: Histone deacetylases; KEGG: Kyoto Encyclopedia of Genes and Genomes; NICOLA: Northern Ireland Cohort for the Longitudinal Study of Ageing; P\_FEN: Percent of endothelial fenestration falling on the peripheral glomerular basement membrane; PGS: Partek Genomics Suite; QC: Quality control; QQ: Quantile–quantile; SF: Supplementary figure; ST: Supplementary table; SV: Surface volume of peripheral glomerular basement membrane per glomerulus; T1DM: Type 1 diabetes mellitus; T2DM: Type 2 diabetes mellitus; TF: Transcription factor; VVINT: Cortical interstitial fractional volume; VVMES: Mesangial fractional volume; VVPC: Volume fraction of podocyte cell per glomerulus; WCCs: White cell counts.

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13148-021-01081-x>.

**Additional file 1.** Supplementary Tables.

**Additional file 2.** Supplementary Figures.

#### Acknowledgements

We thank all individuals who participated in this study and acknowledge all physicians and nurses who were involved in their recruitment. We would like to acknowledge the GENIE consortium and Warren 3 and Genetics of Kidneys in Diabetes (GoKinD) Study Group. The GENIE\_UK data include samples recruited as part of the Warren3/U.K. GoKinD Study Group, jointly funded by Diabetes UK and the Juvenile Diabetes Research Foundation and includes the following individuals: Professor A.P. Maxwell, Professor A.J. McKnight, Dr. D.A. Savage (Belfast); Dr. J. Walker (Edinburgh); Dr. S. Thomas, Professor G.C. Viberti (London); Professor A.J.M. Boulton (Manchester); Professor S. Marshall (Newcastle); Professor A.G. Demaine and Dr. B.A. Millward (Plymouth); and Professor S.C. Bain (Swansea). We are grateful to all the participants of the NICOLA Study, and the whole NICOLA team, which includes nursing staff, research

scientists, clerical staff, computer and laboratory technicians, managers and receptionists. The Atlantic Philanthropies, the Economic and Social Research Council, the UKCRC Centre of Excellence for Public Health Northern Ireland, the Centre for Ageing Research and Development in Ireland, the Office of the First Minister and Deputy First Minister, the Health and Social Care Research and Development Division of the Public Health Agency, the Wellcome Trust/Wolfson Foundation and Queen's University Belfast provide core financial support for NICOLA. The authors alone are responsible for the interpretation of the data and any views or opinions presented are solely those of the authors and do not necessarily represent those of the NICOLA Study team.

#### Authors' contributions

LJS, AJM and APM designed the study. LJS, JK, KK, CW and KA generated the methylationEPIC data. LJS analysed the data. VN, RGN, HCL and MK provided the kidney gene expression and kidney morphometry data. HL and KS provided the Illumina 450K look-up data for top-ranked genes using blood-derived and kidney derived samples. LJS drafted the article. All authors read, provided important intellectual input, critically revised and approved the final manuscript.

#### Funding

LJS is the recipient of a Northern Ireland Kidney Research Fund Fellowship (NIKRF). The Genetics of Nephropathy an International Effort (GENIE) through the Medical Research Council (MC\_PC\_15025), Public Health Agency R&D Division (STL/4760/13), Science Foundation Ireland (SFI15/US/B3130), NIH R01\_DK081923 and R01\_DK105154. LJS, KA, APM and AJM are also supported by Science Foundation Ireland and the Department for the Economy, Northern Ireland US partnership award 15/IA/3152. The data referenced for the NICOLA cohort were funded by Economic and Social Research Council—Award Reference ES/L008459/1. CG, EB and DA are supported by SFI Award15/IA/3152, the US Ireland R&D partnership and a strategic grant from the JDRF. This study was supported, in part, by the George M. O'Brien Michigan Kidney Translational Core Center, funded by NIH/NIDDK grant 2P30-DK-081943 and by the Intramural Research Program of NIDDK. None of the funding bodies had a role in the study design or analysis.

#### Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

#### Declarations

##### Ethics approval and consent to participate

Research ethics approval was obtained from the South and West Multicentre Research Ethics Committee (MREC/98/6/71). All participants provided written informed consent for research. No individual level data is being reported in this manuscript. Ethical approval for the NICOLA study was obtained from the School of Medicine, Dentistry and Biomedical Sciences Ethics Committee, Queen's University Belfast (Ref 12/23).

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare that they have no competing interests.

##### Author details

<sup>1</sup>Molecular Epidemiology Research Group, Centre for Public Health, Queen's University Belfast, Belfast, UK. <sup>2</sup>Internal Medicine, Department of Nephrology, University of Michigan, Ann Arbor, MI, USA. <sup>3</sup>Department of Department of Medicine/ Nephrology, Department of Genetics, Institute of Diabetes, Obesity and Metabolism, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. <sup>4</sup>Diabetes Complications Research Centre, Conway Institute of Biomolecular and Biomedical Research, School of Medicine, University College Dublin, Dublin 4, Ireland. <sup>5</sup>Folkhälsan Institute of Genetics, Folkhälsan Research Center, Helsinki, Finland. <sup>6</sup>Abdominal Center, Nephrology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. <sup>7</sup>Research Program for Clinical and Molecular Metabolism, Faculty of Medicine, University of Helsinki, Helsinki, Finland. <sup>8</sup>Programs in Metabolism and Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge,

MA, USA. <sup>9</sup>Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA, USA. <sup>10</sup>Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. <sup>11</sup>Department of Family Medicine and Public Health, UC San Diego, San Diego, CA, USA. <sup>12</sup>Chronic Kidney Disease Section, National Institute of Diabetes and Digestive and Kidney Diseases, Phoenix, AZ, USA. <sup>13</sup>Department of Genetics, Harvard Medical School, Boston, MA, USA. <sup>14</sup>Department of Diabetes, Central Clinical School, Monash University, Melbourne, VIC, Australia. <sup>15</sup>Regional Nephrology Unit, Belfast City Hospital, Belfast, Northern Ireland, UK. <sup>16</sup>Department of Medicine, Harvard Medical School, Boston, MA, USA.

Received: 8 December 2020 Accepted: 15 April 2021

Published online: 01 May 2021

#### References

- DiMeglio LA, Evans-Molina C, Oram RA. Type 1 diabetes. *Lancet*. 2018;391(10138):2449–62.
- Wang G, Ouyang J, Li S, Wang H, Lian B, Liu Z, et al. The analysis of risk factors for diabetic nephropathy progression and the construction of a prognostic database for chronic kidney diseases. *J Transl Med*. 2019;17(1):1–12.
- Sandholm N, Van Zuydam N, Ahlqvist E, Juliusdottir T, Deshmukh HA, Rayner NW, et al. The genetic landscape of renal complications in type 1 diabetes. *J Am Soc Nephrol*. 2017;28(2):557–74.
- UK Renal Registry (2019) UK Renal Registry 21st Annual Report. Data to 31/12/2017, Bristol, UK. <https://www.renalreg.org/publications-reports/>.
- United States Renal Data System. 2018 USRDS annual data report: epidemiology of kidney disease in the United States. Bethesda: National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases; 2018.
- Sabanayagam C, Chee ML, Banu R, Cheng CY, Lim SC, Tai ES, et al. Association of diabetic retinopathy and diabetic kidney disease with all-cause and cardiovascular mortality in a multiethnic Asian population. *JAMA Netw open*. 2019;2(3):e1915.
- Jerram ST, Dang MN, Leslie RD. The role of epigenetics in type 1 diabetes. *Curr Diab Rep*. 2017;17(10):89.
- Pociot F. Type 1 diabetes genome-wide association studies: not to be lost in translation. *Clin Transl Immunol*. 2017;6(12):e162.
- Cole JB, Florez JC. Genetics of diabetes mellitus and diabetes complications. *Nat Rev Nephrol*. 2020. <https://doi.org/10.1038/s41581-020-0278-5>.
- Zhu M, Xu K, Chen Y, Gu Y, Zhang M, Luo F, et al. Identification of novel T1D risk loci and their association with age and islet function at diagnosis in autoantibody-positive T1D individuals: based on a two-stage genome-wide association study. *Diabetes Care*. 2019;42(8):1414–21.
- Lemos NE, Dieter C, Dorfman LE, Assmann TS, Duarte GCK, Canani LH, et al. The rs2292239 polymorphism in ERBB3 gene is associated with risk for type 1 diabetes mellitus in a Brazilian population. *Gene*. 2018;644:122–8.
- Bottini N, Musumeci L, Alonso A, Rahmouni S, Nika K, Rostamkhani M, et al. A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. *Nat Genet*. 2004;36(4):337–8.
- Salem RM, Todd JN, Sandholm N, Cole JB, Chen WM, Andrews D, et al. Genome-Wide association study of diabetic kidney disease highlights biology involved in glomerular basement membrane collagen. *J Am Soc Nephrol*. 2019;30(10):2000–16.
- Sandholm N, Salem RM, McKnight AJ, Brennan EP, Forsblom C, Isakova T, et al. New susceptibility loci associated with kidney disease in type 1 diabetes. *PLoS Genet*. 2012;8(9):e1002921.
- Pezzolesi MG, Poznik GD, Mychaleckyj JC, Paterson AD, Barati MT, Klein JB, et al. Genome-wide association scan for diabetic nephropathy susceptibility genes in type 1 diabetes. *Diabetes*. 2009;58(6):1403–10.
- Sandholm N, McKnight AJ, Salem RM, Brennan EP, Forsblom C, Harjutsalo V, et al. Chromosome 2q31.1 associates with ESRD in women with type 1 diabetes. *J Am Soc Nephrol*. 2013;24(10):1537–43.
- Bansal A, Pinney SE. DNA methylation and its role in the pathogenesis of diabetes. *Pediatr Diabetes*. 2017;18(3):167–77.



18. Kato M, Natarajan R. Epigenetics and epigenomics in diabetic kidney disease and metabolic memory. *Nat Rev Nephrol*. 2019;15(6):327–45.
19. Bell CG, Teschendorff AE, Rakan VK, Maxwell AP, Beck S, Savage DA. Genome-wide DNA methylation analysis for diabetic nephropathy in type 1 diabetes mellitus. *BMC Med Genomics*. 2010;3(33):1–11.
20. Smyth LJ, McKay GJ, Maxwell AP, McKnight AJ. DNA hypermethylation and DNA hypomethylation is present at different loci in chronic kidney disease. *Epigenetics*. 2014;9(3):366–76.
21. Gu T, Falhammar H, Gu HF, Brismar K. Epigenetic analyses of the insulin-like growth factor binding protein 1 gene in type 1 diabetes and diabetic nephropathy. *Clin Epigenet*. 2014;6(1):1–6.
22. Swan EJ, Maxwell AP, McKnight AJ. Distinct methylation patterns in genes that affect mitochondrial function are associated with kidney disease in blood-derived DNA from individuals with Type 1 diabetes. *Diabet Med*. 2015;32(8):1110–5.
23. Chu AY, Tin A, Schlosser P, Ko YA, Qiu C, Yao C, et al. Epigenome-wide association studies identify DNA methylation associated with kidney function. *Nat Commun*. 2017;8(1286):1–12.
24. Qiu C, Hanson RL, Fufaa G, Kobes S, Gluck C, Huang J, et al. Cytosine methylation predicts renal function decline in American Indians. *Kidney Int*. 2018;93(6):1417–31.
25. Smyth LJ, Duffy S, Maxwell AP, McKnight AJ. Genetic and epigenetic factors influencing chronic kidney disease. *Am J Physiol Renal Physiol*. 2014;307(7):F757–76.
26. Smyth LJ, Maxwell AP, Benson KA, Kilner J, McKay GJ, McKnight AJ. Validation of differentially methylated microRNAs identified from an epigenome-wide association study; Sanger and next generation sequencing approaches. *BMC Res Notes*. 2018;11:767.
27. Dirks RAM, Stunnenberg HG, Marks H. Genome-wide epigenomic profiling for biomarker discovery. *Clin Epigenet*. 2016;8(122):1–17.
28. Illumina. BeadArray Controls Reporter Software Guide. 2015. [https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry\\_documentation/infinium\\_assays/infinium\\_hd\\_methylation/beadarray-controls-reporter-user-guide-1000000004-009-00.pdf](https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/infinium_assays/infinium_hd_methylation/beadarray-controls-reporter-user-guide-1000000004-009-00.pdf).
29. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinform*. 2012;8(13):86.
30. Daca-Roszak P, Pfeifer A, Zembracka-Gala J, Rusinek D, Szybińska A, Jarzab B, et al. Impact of SNPs on methylation readouts by Illumina Infinium HumanMethylation450 BeadChip array: implications for comparative population studies. *BMC Genomics*. 2015;16(1003):1–13.
31. Illumina. Infinium HD Methylation SNP List. 2013. [https://support.illumina.com/downloads/infinium\\_hd\\_methylation\\_snp\\_list.html](https://support.illumina.com/downloads/infinium_hd_methylation_snp_list.html).
32. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607–13.
33. Gluck C, Qiu C, Han SY, Palmer M, Park J, Ko YA, et al. Kidney cytosine methylation changes improve renal function decline estimation in patients with diabetic kidney disease. *Nat Commun*. 2019;10(2461):1–12.
34. Nair V, Komorowsky CV, Weil EJ, Yee B, Hodgin J, Harder JL, et al. A molecular morphometric approach to diabetic kidney disease can link structure to function and outcome. *Kidney Int*. 2018;93(2):439–49.
35. Dang MN, Buzzetti R, Pozzilli P. Epigenetics in autoimmune diseases with focus on type 1 diabetes. *Diabetes Metab Res Rev*. 2013;29(1):8–18.
36. Smyth LJ, Patterson CC, Swan EJ, Maxwell AP, McKnight AJ. DNA methylation associated with diabetic kidney disease in Blood-Derived DNA. *Front Cell Dev Biol*. 2020;8:561907.
37. Binder EB. The role of FKBP5, a co-chaperone of the glucocorticoid receptor in the pathogenesis and therapy of affective and anxiety disorders. *Psychoneuroendocrinology*. 2009;34(1):S186–95.
38. Zannas AS, Wiechmann T, Gassen NC, Binder EB. Gene-stress-epigenetic regulation of FKBP5: clinical and translational implications. *Neuropsychopharmacology*. 2016;41:261–74.
39. Zannas AS, Jia M, Hafner K, Baumert J, Wiechmann T, Pape JC, et al. Epigenetic upregulation of FKBP5 by aging and stress contributes to NF- $\kappa$ B-driven inflammation and cardiovascular risk. *Proc Natl Acad Sci U S A*. 2019;116(23):11370–9.
40. Ortiz R, Joseph JJ, Lee R, Wand GS, Golden SH. Type 2 diabetes and cardiometabolic risk may be associated with increase in DNA methylation of FKBP5. *Clin Epigenet*. 2018;10(82):1–8.
41. Mihai S, Codrici E, Popescu ID, Enciu AM, Albulescu L, Necula LG, et al. Inflammation-related mechanisms in chronic kidney disease prediction, progression, and outcome. *J Immunol Res*. 2018. <https://doi.org/10.1155/2018/2180373>.
42. Chen F, Liu X, Bai J, Pei D, Zheng J. The emerging role of RUNX3 in cancer metastasis (review). *Oncol Rep*. 2016;35(3):1227–36.
43. Mei PJ, Bai J, Liu H, Li C, Wu YP, Yu ZQ, et al. RUNX3 expression is lost in glioma and its restoration causes drastic suppression of tumor invasion and migration. *J Cancer Res Clin Oncol*. 2011;137(12):1823–30.
44. Cen D, Xu L, Zhang S, Chen Z, Huang Y, Li Z, et al. Renal cell carcinoma: predicting RUNX3 methylation level and its consequences on survival with CT features. *Eur Radiol*. 2019;29(10):5415–22.
45. Wang X, Qin G, Zhao TC. HDAC4: Mechanism of regulations and biological functions. *Epigenomics*. 2014;6(1):139–50.
46. Hadden MJ, Advani A. Histone deacetylase inhibitors and diabetic kidney disease. *Int J Mol Sci*. 2018;19(9):2630.
47. Liu N, Zhuang S. Treatment of chronic kidney diseases with histone deacetylase inhibitors. *Front Physiol*. 2015;6:121.
48. Xiong C, Guan Y, Zhou X, Liu L, Zhuang MA, Zhang W, et al. Selective inhibition of class IIa histone deacetylases alleviates renal fibrosis. *FASEB J*. 2019;33(7):8249–62.
49. Boguslawska J, Kedzierska H, Poplawski P, Rybicka B, Tanski Z, Piekliko-Witkowska A. Expression of genes involved in cellular adhesion and extracellular matrix remodeling correlates with poor survival of patients with renal cancer. *J Urol*. 2016;195(6):1982–2902.
50. Lu Q, Ray D, Gutsch D, Richardson B. Effect of DNA methylation and chromatin structure on ITGAL expression. *Blood*. 2002;99(12):4503–8.
51. Parikova A, Hrupa P, Krediet RT, Krejci Z, Stranecky V, Striz I, et al. Long-term peritoneal dialysis treatment provokes activation of genes related to adaptive immunity. *Physiol Res*. 2019;68(5):775–83.
52. Thameem F, Wolford JK, Bogardus C, Prochazka M. Analysis of PBX1 as a candidate gene for type 2 diabetes mellitus in Pima Indians. *Biochim Biophys Acta Gene Struct Expr*. 2001;1518(1–2):215–20.
53. Agha G, Hajj H, Rifas-Shiman SL, Just AC, Hivert MF, Burris HH, et al. Birth weight-for-gestational age is associated with DNA methylation at birth and in childhood. *Clin Epigenet*. 2016;8(118):1–12.
54. Deucher AM, Qi Z, Yu J, George TI, Etezzel JE. BCL6 expression correlates with the t(1;19) translocation in B-lymphoblastic leukemia. *Am J Clin Pathol*. 2015;143(4):547–57.
55. Le TP, Breton J, Bidart M, Satre V, Harbuz R, Ray PF, et al. PBX1 haploinsufficiency leads to syndromic congenital anomalies of the kidney and urinary tract (CAKUT) in humans. *J Med Genet*. 2017;54(7):502–10.
56. Wei X, Yu L, Li Y. PBX1 promotes the cell proliferation via JAK2/STAT3 signaling in clear cell renal carcinoma. *Biochem Biophys Res Commun*. 2018;500(3):650–7.
57. Duesing K, Charpentier G, Marre M, Tichet J, Hercberg S, Balkau B, et al. Evaluating the association of common PBX1 variants with type 2 diabetes. *BMC Med Genet*. 2008;9(14):1–6.
58. Merkel AL, Meggers E, Ocker M. PIM1 kinase as a target for cancer therapy. *Expert Opin Investig Drugs*. 2012;21(4):425–36.
59. Wu Y, Deng Y, Zhu J, Duan Y, Weng WW, Wu X. Pim1 promotes cell proliferation and regulates glycolysis via interaction with MYC in ovarian cancer. *Onco Targets Ther*. 2018;11:6647–56.
60. Gao X, Liu X, Lu Y, Wang Y, Cao W, Liu X, et al. PIM1 is responsible for IL-6-induced breast cancer cell EMT and stemness via c-myc activation. *Breast Cancer*. 2019;26(5):663–71.
61. Magnuson NS, Wang Z, Ding G, Reeves R. Why target PIM1 for cancer diagnosis and treatment? *Futur Oncol*. 2010;9:1461–78.
62. Zhao B, Liu L, Mao J, Zhang Z, Wang Q, Li Q. PIM1 mediates epithelial-mesenchymal transition by targeting Smads and c-Myc in the nucleus and potentiates clear-cell renal-cell carcinoma oncogenesis article. *Cell Death Dis*. 2018;9(3):307.
63. Fu R, Xia Y, Li M, Mao R, Guo C, Zhou M, et al. Pim-1 as a therapeutic target in lupus nephritis. *Arthritis Rheumatol*. 2019;71(8):1308–18.
64. Neubauer K, Neubauer B, Seidl M, Zieger B. Characterization of septin expression in normal and fibrotic kidneys. *Cytoskeleton*. 2019;76(1):143–53.

65. Angulo JC, Andrés G, Ashour N, Sánchez-Chapado M, López JL, Ropero S. Development of castration resistant prostate cancer can be predicted by a DNA hypermethylation profile. *J Urol*. 2016;195(3):619–26.
66. Tóth K, Galamb O, Spisák S, Wichmann B, Sipos F, Valcz G, et al. The influence of methylated septin 9 gene on RNA and protein level in colorectal cancer. *Pathol Oncol Res*. 2011;17(3):503–9.
67. Wu Y, Bu F, Yu H, Li W, Huang C, Meng X, et al. Methylation of Septin9 mediated by DNMT3a enhances hepatic stellate cells activation and liver fibrogenesis. *Toxicol Appl Pharmacol*. 2017;315:35–49.
68. Guo J, Sun C, Wang B, Ma K, Li F, Wang Y, et al. Associations between Vitamin D and  $\beta$ -cell function and colorectal cancer-associated tumor markers in Chinese type 2 diabetic patients with albuminuria. *Clin Lab*. 2019;65(4):30969095.
69. Dolat L, Hunyara JL, Bowen JR, Karasmanis EP, Elgawly M, Galkin VE, et al. Septins promote stress fiber-mediated maturation of focal adhesions and renal epithelial motility. *J Cell Biol*. 2014;207(2):225–35.
70. Dayeh T, Volkov P, Saló S, Hall E, Nilsson E, Olsson AH, et al. Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion. *PLoS Genet*. 2014;10(3):e1004160.
71. Wang G, Watanabe M, Imai Y, Hara K, Manabe I, Maemura K, et al. Associations of variations in the MRF2/ARID5B gene with susceptibility to type 2 diabetes in the Japanese population. *J Hum Genet*. 2012;57:727–33.
72. Smith AR, Smith RG, Pishva E, Hannon E, Roubroeks JAY, Burrage J, et al. Parallel profiling of DNA methylation and hydroxymethylation highlights neuropathology-associated epigenetic variation in Alzheimer's disease. *Clin Epigenet*. 2019;11(52):1–13.
73. Hiwatari M, Seki M, Akahoshi S, Yoshida K, Miyano S, Shiraiishi Y, et al. Molecular studies reveal MLL-MLLT10/AF10 and ARID5B-MLL gene fusions displaced in a case of infantile acute lymphoblastic leukemia with complex karyotype. *Oncol Lett*. 2017;14(2):2295–9.
74. Tan SH, Leong WZ, Ngoc PCT, Tan TK, Bertulfo FC, Lim MC, et al. The enhancer RNA ARIEL activates the oncogenic transcriptional program in T-cell acute lymphoblastic leukemia. *Blood*. 2019;134(3):239–51.
75. Xu H, Zhao X, Bhojwani D, Shuyu E, Goodings C, Zhang H, et al. ARID5B influences antimetabolite drug sensitivity and prognosis of acute lymphoblastic leukemia. *Clin Cancer Res*. 2020;26(1):256–64.
76. Chan WK, Bhalla AD, Le Hir H, Nguyen LS, Huang L, Gécz J, et al. A UPF3-mediated regulatory switch that maintains RNA surveillance. *Nat Struct Mol Biol*. 2009;16:747–53.
77. Gotoh M, Ichikawa H, Arai E, Chiku S, Sakamoto H, Fujimoto H, et al. Comprehensive exploration of novel chimeric transcripts in clear cell renal cell carcinomas using whole transcriptome analysis. *Genes Chromosom Cancer*. 2014;53(12):1018–32.
78. Sharma M, Brantley JG, Vassmer D, Chaturvedi G, Baas J, Vanden Heuvel GB. The homeodomain protein Cux1 interacts with Grg4 to repress p27kip1 expression during kidney development. *Gene*. 2009;439(1–2):87–94.
79. Livingston S, Carlton C, Sharma M, Kearns D, Baybutt R, Vanden Heuvel GB. Cux1 regulation of the cyclin kinase inhibitor p27 kip1 in polycystic kidney disease is attenuated by HDAC inhibitors. *Gene X*. 2019;2:100007.
80. Porath B, Livingston S, Andres EL, Petrie AM, Wright JC, Woo AE, et al. Cux1 promotes cell proliferation and polycystic kidney disease progression in an ADPKD mouse model. *Am J Physiol Ren Physiol*. 2017;313(4):F1050–9.
81. An N, Khan S, Imgruet MK, Gurbuxani SK, Konecki SN, Burgess MR, et al. Gene dosage effect of CUX1 in amurinemodel disruptsHSC homeostasis and controls the severity and mortality of MDS. *Blood*. 2018;131(24):2682–97.
82. Reidy K, Kang HM, Hostetter T, Susztak K. Molecular mechanisms of Diabetic kidney disease. *J Clin Invest*. 2014;124(6):2333–40.
83. Ye J, Richardson TG, McArdle WL, Relton CL, Gillespie KM, Suderman M, et al. Identification of loci where DNA methylation potentially mediates genetic risk of type 1 diabetes. *J Autoimmun*. 2018;93(June):66–75.
84. Wu YH, Wang Y, Chen M, Zhang X, Wang D, Pan Y, et al. Association of ELMO1 gene polymorphisms with diabetic nephropathy in Chinese population. *J Endocrinol Invest*. 2013;36(5):298–302.
85. Turki A, Mzoughi S, Mtitaoui N, Khairallah M, Marmouch H, Hammami S, et al. Gender differences in the association of ELMO1 genetic variants with type 2 diabetes in Tunisian Arabs. *J Endocrinol Invest*. 2018;41(3):285–91.
86. Yahya MJ, Ismail P, Nordin N, Akim AM, Yusuf WSM, Adam NL, et al. Association of CCL2, CCR5, ELMO1, and IL8 polymorphism with diabetic nephropathy in Malaysian type 2 diabetic patients. *Int J Chronic Dis*. 2019;2019:1–13.
87. Peng H, Zhang Y, Zhou Z, Guo Y, Huang X, Westover KD, et al. Integrated analysis of ELMO1, serves as a link between tumour mutation burden and epithelial-mesenchymal transition in hepatocellular carcinoma. *EBioMedicine*. 2019;46:105–18.
88. Arandjelovic S, Perry JSA, Lucas CD, Penberthy KK, Kim TH, Zhou M, et al. A noncanonical role for the engulfment gene ELMO1 in neutrophils that promotes inflammatory arthritis. *Nat Immunol*. 2019;20(2):141–51.
89. Pirini F, Noazin S, Jahuira-Arias MH, Rodriguez-Torres S, Friess L, Michailidi C, et al. Early detection of gastric cancer using global, genome-wide and IRF4, ELMO1, CLIP4 and MSC DNA methylation in endoscopic biopsies. *Oncotarget*. 2017;8(24):38501–16.
90. van Zuydam NR, Ahlqvist E, Sandholm N, Deshmukh H, Rayner NW, Abdalla M, et al. A genome-wide association study of diabetic kidney disease in subjects with type 2 diabetes. *Diabetes*. 2018;67(7):1414–27.
91. Köttgen A, Pattaro C, Böger CA, Fuchsberger C, Olden M, Glazer NL, et al. Multiple new loci associated with kidney function and chronic kidney disease: the CKDGen consortium. *Nat Genet*. 2010;42(5):376–84.
92. Yoshida T, Kato K, Yokoi K, Oguri M, Watanabe S, Metoki N, et al. Association of genetic variants with chronic kidney disease in Japanese individuals with or without hypertension or diabetes mellitus. *Exp Ther Med*. 2010;1(1):137–45.
93. Lee S. The association of genetically controlled CpG methylation (cg158269415) of protein tyrosine phosphatase, receptor type N2 (PTPRN2) with childhood obesity. *Sci Rep*. 2019;9(1):4855.
94. Abuhatzira L, Xu H, Tahhan G, Boulougoura A, Schäffer AA, Notkins AL. Multiple microRNAs within the 14q32 cluster target the mRNAs of major type 1 diabetes autoantigens IA-2, IA-2b, and GAD65. *FASEB J*. 2015;29(10):4374–83.
95. Yang RM, Tao J, Zhan M, Yuan H, Wang HH, Chen SJ, et al. TAMM41 is required for heart valve differentiation via regulation of PINK-PARK2 dependent mitophagy. *Cell Death Differ*. 2019;26:2430–46.
96. Staudt D, Stainier D. Uncovering the molecular and cellular mechanisms of heart development using the zebrafish. *Annu Rev Genet*. 2012;46:397–418.
97. Vajdic CM, McDonald SP, McCredie MRE, Van Leeuwen MT, Stewart JH, Law M, et al. Cancer incidence before and after kidney transplantation. *J Am Med Assoc*. 2006;296(23):2823–31.
98. Van Leeuwen MT, Webster AC, McCredie MRE, Stewart JH, McDonald SP, Amin J, et al. Effect of reduced immunosuppression after kidney transplant failure on risk of cancer: population based retrospective cohort study. *BMJ*. 2010;340:c570.
99. Alfano G, Fontana F, Colaci E, Mori G, Cerami C, Messerotti A, et al. T-cell large granular lymphocyte leukemia in solid organ transplant recipients: case series and review of the literature. *Int J Hematol*. 2019;110(3):313–21.
100. Barrett JC, Clayton D, Concannon P, Akolkar B, Jason D, Erlich HA, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet*. 2010;41(6):703–7.
101. Cooper JD, Smyth DJ, Smiles AM, Plagnol V, Walker NM, Allen JE, et al. Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci. *Nat Genet*. 2008;40(12):1399–401.
102. Afzali B, Grönholm J, Vandrovicova J, Brien CO, Vanderleyden I, Davis FP, et al. BACH2 immunodeficiency illustrates an association between super-enhancers and haploinsufficiency. *Nat Immunol*. 2017;18(7):813–23.
103. Smyth DJ, Plagnol V, Walker NM, Cooper JD, Downes K, Yang JHM, et al. Shared and distinct genetic variants in type 1 diabetes and celiac disease. *N Engl J Med*. 2008;359(26):2767–77.
104. Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V, et al. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet*. 2007;39(7):857–64.
105. Concannon P, Onengut-Gumuscu S, Todd JA, Smyth DJ, Pociot F, Bergholdt R, et al. A human type 1 diabetes susceptibility locus maps to chromosome 21q22.3. *Diabetes*. 2008;57(10):2858–61.

106. Fung EYMG, Smyth DJ, Howson JMM, Cooper JD, Walker NM, Stevens H, et al. Analysis of 17 autoimmune disease-associated variants in type 1 diabetes identifies 6q23/TNFAIP3 as a susceptibility locus. *Genes Immun*. 2009;10(2):188–91.
107. Hakonarson H, Grant SFA, Bradfield JP, Marchand L, Kim CE, Glessner JT, et al. A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. *Nature*. 2007;448(7153):591–4.
108. Nisticò L, Buzzetti R, Pritchard LE, Van Der Auwera B, Giovannini C, Bosi E, et al. The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. *Hum Mol Genet*. 1996;5(7):1075–80.
109. Smyth DJ, Cooper JD, Bailey R, Field S, Burren O, Smink LJ, et al. A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region. *Nat Genet*. 2006;38(6):617–9.
110. Wang TC, Chiu H, Chang YJ, Hsu TY, Chiu IM, Chen L. The adaptor protein SH2B3 (LNK) negatively regulates neurite outgrowth of PC12 cells and cortical neurons. *PLoS ONE*. 2011;6(10):e26433.
111. Maslah N, Cassinat B, Verger E, Kiladjian JJ, Velazquez L. The role of LNK/SH2B3 genetic alterations in myeloproliferative neoplasms and other hematological disorders. *Leukemia*. 2017;31(8):1661–70.
112. Auburger G, Gispert S, Lahut S, Omur O, Damrath E, Heck M, et al. 12q24 locus association with type 1 diabetes: SH2B3 or ATXN2? *World J Diabetes*. 2014;5(3):316.
113. Holmberg D, Ruikka K, Lindgren P, Eliasson M, Mayans S. Association of CD247 (CD3ζ) gene polymorphisms with T1D and AITD in the population of northern Sweden. *BMC Med Genet*. 2016;17(1):1–7.
114. Zhu Q, Wang J, Zhang L, Bian W, Lin M, Xu X, et al. LCK rs10914542-G allele associates with type 1 diabetes in children via T cell hyporesponsiveness. *Pediatr Res*. 2019;86(3):311–5.
115. Woillard JB, Kamar N, Rousseau A, Rostaing L, Marquet P, Picard N. Association of sirolimus adverse effects with m-TOR, p70S6K or rapTOR polymorphisms in kidney transplant recipients. *Pharmacogenet Genomics*. 2012;22(10):725–32.
116. Yong WS, Hsu FM, Chen PY. Profiling genome-wide DNA methylation. *Epigenet Chromatin*. 2016. <https://doi.org/10.1186/s13072-016-0075-3>.
117. Mansell G, Gorrie-Stone TJ, Bao Y, Kumari M, Schalkwyk LS, Mill J, et al. Guidance for DNA methylation studies: statistical insights from the Illumina EPIC array. *BMC Genomics*. 2019;20(1):366.
118. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res*. 1988;16(3):1215.
119. Ahn SJ, Costa J, Emanuel JR. PicoGreen quantitation of DNA: effective evaluation of samples pre- or post-PCR. *Nucleic Acids Res*. 1996;24(13):2623–5.
120. Boomsma DI, Wijmenga C, Slagboom EP, Swertz MA, Karssen LC, Abdellaoui A, et al. The Genome of the Netherlands: design, and project goals. *Eur J Hum Genet*. 2014;22(2):221–7. <https://doi.org/10.1038/ejhg.2013.118>.
121. Cañadas-Garre M, Smyth LJ, Neville C, Woodside JV, Kee F, McKnight AJ. Chapter 7, biomarkers. In: NICOLA Health Assessment Report. 2020.
122. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *J Open Source Softw*. 2018;3(25):731.
123. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, et al. The reactome pathway knowledgebase. *Nucleic Acids Res*. 2020;48(D1):D498–503.
124. Breeze CE, Reynolds AP, Van Dongen J, Dunham I, Lazar J, Neph S, et al. EFORGE v2.0: updated analysis of cell type-specific signal in epigenomic data. *Bioinformatics*. 2019;35(22):4767–9.
125. Weil EJ, Fufaa G, Jones LI, Lovato T, Lemley KV, Hanson RL, et al. Effect of losartan on prevention and progression of early diabetic nephropathy in American Indians with type 2 diabetes. *Diabetes*. 2013;62(9):3224–31.
126. Berthier CC, Zhang H, Schin M, Henger A, Nelson RG, Yee B, et al. Enhanced expression of janus kinase-signal transducer and activator of transcription pathway members in human diabetic nephropathy. *Diabetes*. 2009;58(2):469–77.
127. Schmid H, Boucherot A, Yasuda Y, Henger A, Brunner B, Eichinger F, et al. Modular activation of nuclear factor-κB transcriptional programs in human diabetic nephropathy. *Diabetes*. 2006;55(11):2993–3003.
128. Fioretto P, Kim Y, Mauer M. Diabetic nephropathy as a model of reversibility of established renal lesions. *Curr Opin Nephrol Hypertens*. 1998;7(5):489–94.
129. Mauer M, Zinman B, Gardiner R, Suissa S, Sinaiko A, Strand T, et al. Renal and retinal effects of enalapril and losartan in type 1 diabetes. *N Engl J Med*. 2009;2(361):40–51.
130. Ibrahim HN, Jackson S, Connaire J, Matas A, Ney A, Najafian B, et al. Angiotensin II blockade in kidney transplant recipients. *J Am Soc Nephrol*. 2013;24(2):320–7.
131. Mauer M, Caramori ML, Fioretto P, Najafian B. Glomerular structural-functional relationship models of diabetic nephropathy are robust in type 1 diabetic patients. *Nephrol Dial Transplant*. 2015;30(6):918–23.
132. Luiza Caramori M, Kim Y, Huang C, Fish AJ, Rich SS, Miller ME, et al. Cellular basis of diabetic nephropathy: 1. Study design and renal structural-functional relationships in patients with long-standing type 1 diabetes. *Diabetes*. 2002;51(2):506–13.
133. Klein R, Zinman B, Gardiner R, Suissa S, Donnelly SM, Sinaiko AR, et al. The relationship of diabetic retinopathy to preclinical diabetic glomerulopathy lesions in type 1 diabetic patients: the renin-angiotensin system study. *Diabetes*. 2005;54(2):527–33.
134. Najafian B, Mauer M. Quantitating glomerular endothelial fenestration: an unbiased stereological approach. *Am J Nephrol*. 2011;33(Suppl 1):34–9.
135. Najafian B, Tøndel C, Svarstad E, Sokolovkiy A, Smith K, Mauer M. One year of enzyme replacement therapy reduces globotriaosylceramide inclusions in podocytes in Male adult patients with Fabry disease. *PLoS ONE*. 2016;11(4):e0152812.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

