



OPEN

Differentially methylated and expressed genes in familial type 1 diabetes

Mohammed Dashti^{1,4}, Rasheeba Nizam^{1,4}, Prashantha Hebbar¹, Sindhu Jacob¹, Sumi Elsa John¹, Arshad Channanath¹, Hessa Al-Kandari^{2,3}, Thangavel Alphonse Thanaraj¹✉ & Fahd Al-Mulla¹✉

There has recently been a growing interest in examining the role of epigenetic modifications, such as DNA methylation, in the etiology of type 1 diabetes (T1D). This study aimed to delineate differences in methylation patterns between T1D-affected and healthy individuals by examining the genome-wide methylation of individuals from three Arab families from Kuwait with T1D-affected mono-/dizygotic twins and non-twin siblings. Bisulfite sequencing of DNA from the peripheral blood of the affected and healthy individuals from each of the three families was performed. Methylation profiles of the affected individuals were compared to those of the healthy individuals. Principal component analysis on the observed methylation profiling based on base-pair resolution clustered the T1D-affected twins together family-wide. The sites/regions that were differentially methylated between the T1D and healthy samples harbored 84 genes, of which 18 were known to be differentially methylated in T1D individuals compared to healthy individuals in publicly available gene expression data resources. We further validated two of the 18 genes—namely *ICA1* and *DRAM1* that were hypermethylated in T1D samples compared to healthy samples—for upregulation in T1D samples from an extended study cohort of familial T1D. The study confirmed that the *ICA1* and *DRAM1* genes are differentially expressed in T1D samples compared to healthy samples.

A dramatic increase in the incidence and prevalence of type 1 diabetes (T1D) has been witnessed in the past few decades. This poses a major challenge to the health care system worldwide. T1D accounts for 10% of the total cases of diabetes, with a global annual estimated incidence of 128,900 and prevalence of 1,110,100 in children under the age of 20 years¹. Despite this, relatively little is known regarding the etiopathogenesis of the disease. T1D is defined as a lifelong autoimmune disease characterized by an immune response against pancreatic beta cells, resulting in little or no insulin production.

Multiple lines of evidence indicate the collective role of genetic, immunological, and environmental factors in the onset and progression of T1D. The genetic risk of T1D has been evidenced by concordance and heritability rates > 50% in monozygotic twins in long-term follow-up studies². To date, more than 60 susceptibility gene loci have been implicated in the pathogenesis of T1D³. The human leukocyte antigen (HLA) region accounts for 50% of the genetic risk of T1D. Presence of specific HLA DR/DQ haplotypes tend to increase the susceptibility to T1D⁴. An autoimmune pathology of T1D is largely evidenced by a series of T and B cell-mediated events that lead to intolerance to self-antigens, resulting in chronic beta-cell inflammation. In support of this notion, the four major autoantibodies that characterize T1D include islet cell antibodies (ICA), antibodies to glutamic acid decarboxylase (GAD-65), insulin autoantibodies (IAA), and IA-2A, antibodies to protein tyrosine phosphatase⁵. Environmental factors, such as rapidly evolving lifestyle changes, viral infections, toxins, and dietary habits, are often implicated in the pathogenesis of T1D⁶; however, their specific roles have been poorly studied.

Literature reports indicate the need for deciphering the complex intertwining between the genetic and environmental architecture of T1D to get an improved knowledge of its onset and progression. The incomplete concordance of T1D incidence in monozygotic twins convincingly indicates the role of the gene-environment interactions and epigenetics in the pathogenesis of the disease⁷. Paul et al.⁸, by way of performing an epigenome-wide association study on 52 monozygotic twin pairs discordant for T1D, observed a substantial enrichment of

¹Department of Genetics and Bioinformatics, Dasman Diabetes Institute, Kuwait City, Kuwait. ²Department of Population Health, Dasman Diabetes Institute, Kuwait City, Kuwait. ³Department of Pediatrics, Farwaniya Hospital, Ministry of Health, Kuwait City, Kuwait. ⁴These authors contributed equally: Mohammed Dashti and Rasheeba Nizam. ✉email: alphonse.thangavel@dasmaninstitute.org; fahd.almulla@dasmaninstitute.org

Family	Members	Symbol	Phenotype	Age at recruitment (years)	Age at onset of T1D (years)
A	Father	I_1_A	healthy	46	–
	Mother	I_2_A	healthy	34	–
	Daughter	II_1_A	healthy	14	–
	Son	II_2_A	T1D	7	5
	Daughter *	II_3_A	T1D	12	6
	Daughter *	II_4_A	T1D	12	6
B	Mother	I_1_B	healthy	37	–
	Father	I_2_B	T1D	39	12
	Daughter	II_1_B	healthy	6	–
	Son	II_2_B	healthy	9	–
	Daughter	II_3_B	T1D	13	11
	Daughter †	II_4_B	T1D	11	10
	Daughter †	II_5_B	T1D	11	10
C	Father	I_1_C	healthy	43	–
	Mother	I_2_C	healthy	42	–
	Son	II_1_C	T1D	9	2
	Daughter	II_2_C	T1D	14	1

Table 1. Clinical characteristics of the study cohort. Twins; dizygotic *, monozygotic †.

differentially variable CpG positions in T1D twins when compared with their healthy co-twins or with unrelated healthy individuals. Stefan et al.⁹ observed significant differences in methylation patterns in known T1D-associated genes, such as HLA, INS, IL-2RB, and CD226, and further suggested that abnormalities in DNA methylation patterns in genome regions known to regulate gene transcription are possibly involved in the pathogenesis of T1D. In addition, Paul et al.⁸ found evidence even from cord blood of new-borns who progress to overt T1D to suggest that differentially variable CpG positions (DVP) among twins discordant for T1D likely emerge soon after birth. Johnson et al.¹⁰ substantiated the role of DNA methylation in T1D by establishing that methylation differences are present before both islet autoimmunity (IA) onset and diabetes diagnosis. The incidence and prevalence of T1D have been reported to be alarmingly high in the Arabian Peninsula possibly due to its unique genetic makeup, high rate of consanguinity (54.3%), extreme climatic conditions, and rapidly evolving lifestyle changes¹¹. Familial aggregation of T1D is high among Arabs. Studies from Kuwait, which has a high rate of consanguineous marriages, have reported a familial form of T1D in 33% of patients with T1D and the incidence of T1D has doubled in Kuwaiti children in the last 20 years¹². Perhaps, the gene-environment interactions and environmental conditions in the Arabian Peninsula are different from those in Europe where earlier methylation studies were mostly initiated. Hence, the present study aimed to understand the extent to which epigenetic factors contribute to T1D pathogenesis. We performed genome-wide methylation on a cohort consisting of three Arab families from Kuwait with mono-/dizygotic twins and non-twin siblings concordant for T1D phenotype. Additionally, we integrated publicly available gene expression data with the methylation profile to explore the causality between DNA methylation and transcriptional activities and thereby to prioritize the candidate methylome markers obtained in this study.

Results

Clinical characteristics of the study cohort. The study cohort consisted of three families. There were at least two family members afflicted with T1D in each family (Table 1). The homozygosity between the two siblings in Family B and dizygosity between the two siblings in Family A were confirmed for genetic correlations by analysis of the whole exome sequence data for frequencies of heterozygous and homozygous exome variants (Supplementary Information). In addition, HLA typing profile of the T1D affected individuals from the three families (Supplementary Information) were positive for HLA-DRB1*03:01 in haplotypic association with DQA1*05:01-DQB1*02:01 indicating high risk of developing T1D⁴. The two children of Family C (who were aged 9 and 14 years at the time of recruitment) had their age at onset of T1D recorded as 2 and 1 years. Nevertheless, it is to be noted that they were diagnosed for T1D clinically and they tested positive for high-risk HLA DR/DQ haplotypes.

DNAm events across the three familial T1D. The targeted bisulfite sequencing facilitated deriving the DNAm profiling in peripheral blood samples of the individuals from the three families with T1D aggregation. Quality control metrics measured on the generated data (Supplementary Information) revealed (a) a consistent percentage methylation pattern which showed that bases were either methylated or unmethylated uniformly across all the samples (Supplementary Figure S1) and (b) a consistent read coverage per base of CpG sites featuring no significant difference among the samples (Supplementary Figure S2). These observations indicate that our experiments did not suffer from any bias secondary to PCR duplication. After fulfilling the criteria for calling DNAm events (Methods and Material section), the methylKit tool managed to identify 85,922 CpG single sites, 70,625 tiles, and 65,494 truseq regions across the three families.

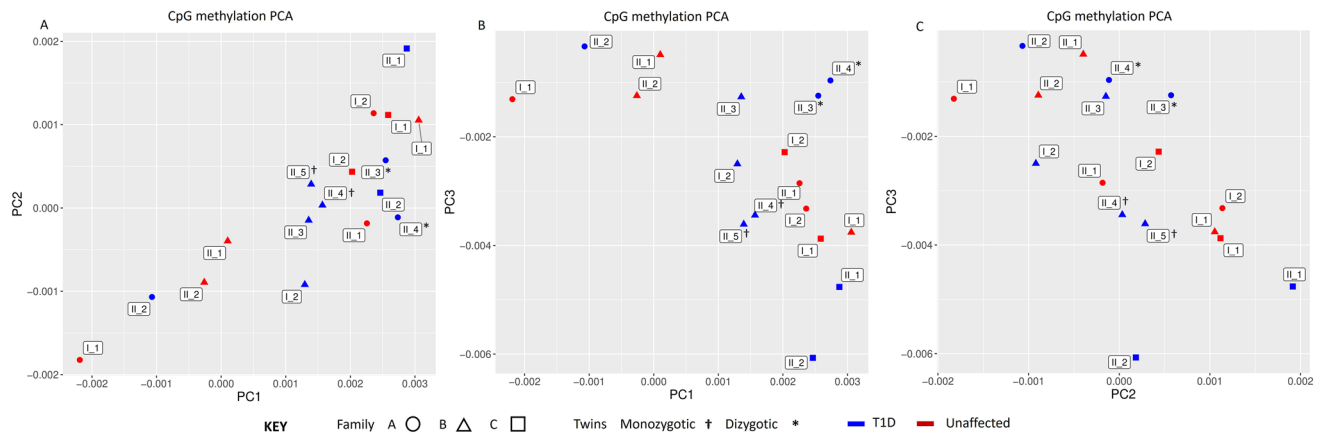


Figure 1. Illustration of inter-individual/family CpG methylome variation observed from PCA plot of the three T1D affected families. † refers to parents; II refers to children. Blue colour represents T1D individuals while red colour represents unaffected individuals. (A) depicts methylome differences observed among samples across the principal components of PC1-PC2; (B) depicts across the principal components of PC1-PC3; and (C) depicts across the principal components of PC2-PC3. The T1D individuals (in Blue) tend to cluster together, familywise. Similarly, monozygotic (†) and dizygotic (*) twins tend to cluster together suggesting a closer methylome architecture among them.

The screen plot of unsupervised PCA model of samples suggested that the principal component 1 (PC1) could explain 40% of the methylome variation observed (among samples), while the PC2 and PC3 could explain 34% and 32%, respectively (Supplementary Figure S3). Figure 1A–C presents the distribution of methylome variation observed among the individuals from the first three principal components. All the affected individuals clustered together familywise in one or the other of the PC1-PC2, PC2-PC3, and PC3-PC1 plots. Interestingly, the affected monozygotic (indicated by † in the plots) and dizygotic (indicated by * in the plots) twins clustered together in all the three PCs, suggesting a closer methylome architecture among twins. This observed close clustering of the twins is in line with our expectation, as twins share a common environmental exposure. It further rules out spurious methylation variation contributed by confounding factors.

Differential DNAm events. The methylKit tool was able to detect, localize, and annotate differential DNAm events (differentially methylated CpG sites, DMC and differentially methylated regions, DMR) along with their methylation direction (hyper- or hypo- methylation in T1D versus healthy individuals) across the three T1D families (Supplementary Table S1).

Supplementary Figure S4 presents the results of the analysis for mapping the genome locations of the DMC and DMR through *genomation* tool. The results indicate that a majority of the observed differentially methylated events occurred at intronic and intergenic regions, rather than at promoter and exon regions. Similarly, differentially methylated CpG events were more prevalent outside of CpG islands and shore regions.

The observed DMC and DMR (tiles and truseq regions) mapped to 84 unique genes (Supplementary Table S2), including coding, non-coding, and uncharacterized genes, as well as microRNAs (miRNAs). None of these 84 genes overlapped with published T1D association signals in the GWAS Catalog (www.ebi.ac.uk/gwas/).

Pathway enrichment analysis. Results of pathway enrichment analysis pointed to 10 pathways (including T1D pathways) that are enriched by the differentially methylated genes (Fig. 2). However, none of the identified pathways passed the significance level defined by an *FDR* of <0.05. The inability to identify any significantly enriched pathway is probably because of the small size of the list of examined genes.

Altered DNAm and differential expression of T1D genes—prioritizing candidate genes. Upon comparing the differentially methylated genes from our study against a list of published datasets of differentially expressed T1D genes of eight gene expression datasets (Supplementary Table S3), it was found that 18 of the 84 unique genes overlapped in both the datasets. These 18 genes (Table 2) could be considered as candidate genes that are more likely to be involved in the T1D pathogenesis. Functional regions that are differentially methylated from these genes consist of five CpG islands, 13 introns, eight shores, and six promoters, as well as exons. Of these 18 genes, Islet Cell Autoantigen 1 (*ICA1*) and Matrix Metalloproteinase 26 (*MMP26*) were significantly differentially methylated in DMC and DMR (tiles and truseq regions) analysis, while DNA Damage Regulated Autophagy Modulator 1 (*DRAM1*) and TIMP Metalloproteinase Inhibitor 3 (*TIMP3*) were significantly differentially methylated in DMR (tiles and truseq regions). Further, a subset of five differentially methylated genes, namely Ubiquitin Specific Peptidase 22 (*USP22*), Chaperonin Containing TCP1 Subunit 7 (*CCT7*), RAN Binding Protein 3 (*RANBP3*), Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Gamma (*PIK3CG*), and Protocadherin Gamma Subfamily B, 6 (*PCDHGB6*), overlapped with more than one T1D gene expression datasets that were from pancreatic and blood samples.

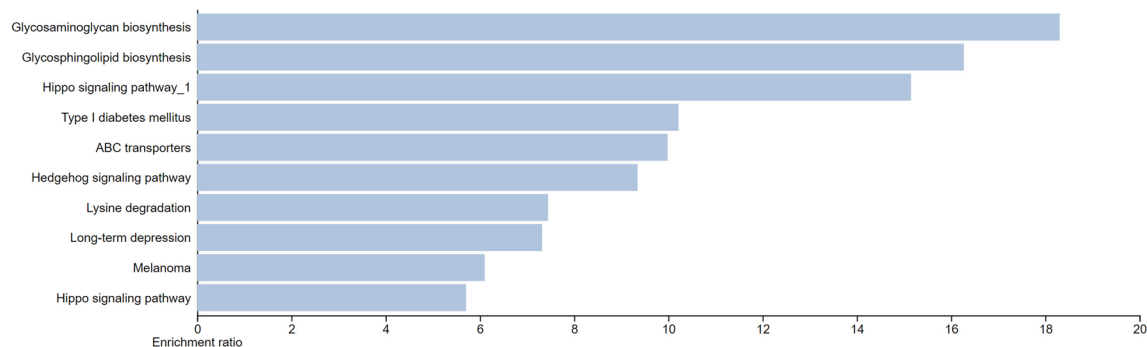


Figure 2. KEGG pathways identified as enriched by genes that are differentially methylated in T1D versus healthy individuals in our study. All the identified functional biological processes and pathways were in light blue indicating an FDR p-value of > 0.05.

Genes	Methylation direction	DMC	DMR (tiles, truseq regions)	Functional regions	Expression direction	Expression accession ID
<i>EXT1</i>	Hypo	8:119,080,372–119,080,372	-,-	intron	Control up vs T1D	GSE33440
<i>ICA1</i>	Hyper	7:8,171,773–8,171,773 7:8,171,803–8,171,803	7:8,171,001–8,172,000, 7:8,171,680–8,171,904	intron	Control down vs T1D clinical onset	E-MEXP-1140
<i>PIK3CG</i>	Hypo	7:106,513,011–106,513,011	-,-	exon	Control down vs T1D recent onset	E-MEXP-1140
					Control down vs T1D	GSE33440
<i>PCDHGB6</i>	Hypo	5:140,789,268–140,789,268	-,-	CpGi,exon,intron	Control up vs T1D clinical onset	E-MEXP-1140
					Control up vs T1D	GSE33440
<i>ST3GAL6</i>	Hyper	3:98,504,779–98,504,779	-,-	intron	Control up vs T1D	GSE33440
<i>VGLL4</i>	Hypo	3:11,610,728–11,610,728	-,-	shores,prom,intron	Control down vs T1D clinical onset	E-MEXP-1140
<i>CCT7</i>	Hypo	2:73,465,970–73,465,970	-,-	intron	Control down vs T1D clinical onset	E-MEXP-1140
					Control down vs T1D	GSE33440
<i>RANBP3</i>	Hyper	19:5,962,749–5,962,749	-,-	intron	Control down vs T1D clinical onset	E-MEXP-1140
					Control down vs T1D	GSE33440
<i>OLFM2</i>	Hypo	19:10,027,408–10,027,408	-,-	shores,intron	Control up vs T1D clinical onset	E-MEXP-1140
<i>TGIF1</i>	Hyper	18:3,412,523–3,412,523	-,-	prom,intron	Control down vs T1D	GSE33440
<i>PRPSAP1</i>	Hypo	17:74,350,980–74,350,980	-,-	shores,prom	Control up vs T1D	GSE33440
<i>ZNF778</i>	Hypo	16:89,300,215–89,300,215	-,-	shores,exon	Control up vs T1D clinical onset	E-MEXP-1140
<i>MMP26</i>	Hyper	11:5,012,592–5,012,592	11:5,012,001–5,013,000, 11:5,012,435–5,012,758	intron	Control up vs T1D	GSE43488
<i>YAP1</i>	Hypo	11:102,028,829–102,028,829	-,-	intron	Control up vs T1D clinical onset	E-MEXP-1140
<i>CNP</i>	Hyper	-	17:20,947,001–20,948,000,-	CpGi,shores,prom,exon,intron	Control down vs T1D clinical onset	E-MEXP-1140
<i>USP22</i>	Hyper	-	17:40,119,001–40,120,000,-	CpGi,shores,prom	Control up vs T1D recent onset	E-MEXP-1140
					Control up vs T1D 4 month after onset	GSE9006
<i>DRAM1</i>	Hyper	-	22:33,119,001–33,120,000, 22:33,119,131–33,119,362	CpGi,shores,prom	Control down vs T1D recent onset	E-MEXP-1140
<i>TIMP3</i>	Hypo	-	8:33,062,001–33,063,000, 8:33,062,053–33,062,289	intron	Control down vs T1D clinical onset	E-MEXP-1140

Table 2. Differentially methylated events across three familial T1D overlapped with differentially expressed genes of published T1D datasets.

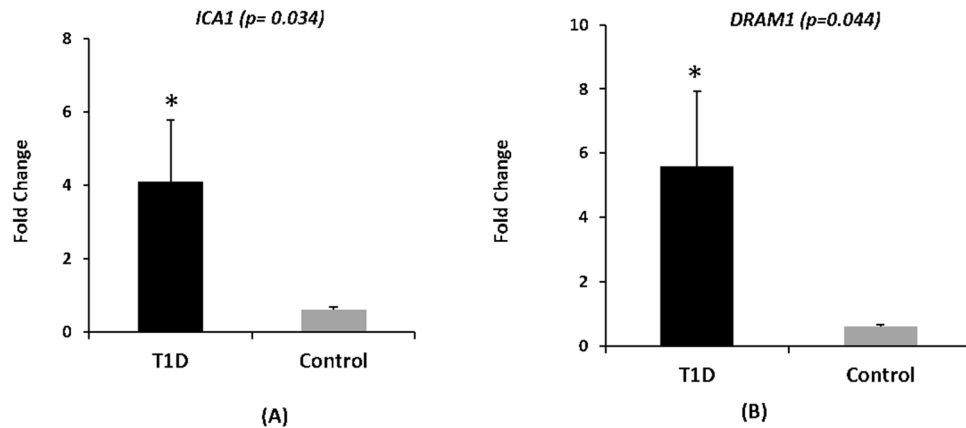


Figure 3. Bar plot indicating the expression level of (A) *ICA1* and (B) *DRAM1* in T1D versus control samples.

Gene expression validation in extended cohort of familial T1D. Quantitative real-time PCR was performed to validate two of the shortlisted candidate genes, namely *ICA1* and *DRAM1* (Fig. 3). These two genes were comparatively biologically more relevant to the pathophysiology of T1D. Analysis of individual families revealed an increase in the expression of *ICA1* ($FC \geq 2$) in 12 of the 33 tested T1D families from the extended study cohort. Similarly, an increased expression of *DRAM1* ($FC \geq 2$) was observed in 10 of the 33 tested T1D families. Compiled analysis revealed a significantly increased expression of *ICA1* ($p = 0.034$) and *DRAM1* ($p = 0.044$) in individuals afflicted with T1D compared to their unaffected family members.

Discussion

This study conducted methylation profiling of three HLA-identical T1D families in order to explore epigenetic contribution to the onset of T1D. Exploring methylation differences between the affected and unaffected individuals from twins and families minimize variation due to environment or genes and reduce residual noise variance due to factors such as population stratification. An integrative approach involving relevant human transcriptomics data that accelerated the prioritization of candidate genes for T1D was employed.

The PCA of methylation profiling of single CpG sites across the three T1D families demonstrated a familywise clustering of affected individuals, which might suggest a common, but family-specific methylation pattern in T1D. Both the monozygotic and dizygotic twins showed a tighter cluster in PCA analysis as opposed to that seen with non-twinning siblings, suggesting a similar methylation pattern in twins; thus, such a design is better suited as it helps to reduce spurious differences in methylation events because of confounders. Moreover, a higher degree of closeness could be observed between the monozygotic twins in family B than in the dizygotic twins in family A. This observation adheres to the finding of Hannon et al.¹³ who demonstrated that monozygotic twins have higher correlation of site-specific levels of DNAm compared to dizygotic twins.

Differential DNAm events were more prominent within intergenic regions and genes, compared to regulatory regions such as promoters, CpG islands, and CpG shores. Similar observations were reported by a study from the Encyclopedia of DNA Elements project that investigated the dynamics of DNAm patterns and relationships to regulatory elements¹⁴. Furthermore, the observed high percentages of differentially methylated events at promoter regions compared to exons may suggest that there is capture enrichment for promoters by the methylation kits target design¹⁵. Although the pathway enrichment analysis revealed the T1D pathway, it was not statistically significant going by FDR value. The inability to identify any significantly enriched pathway is possibly because of the small size of the inputted gene list.

The genomic location of DMC and DMR is crucial in the up-/downregulation of gene expression. Moreover, methylation over CpG functional regions (such as CpG island of around 1 Kb in length)—often located near or in the gene promoter region close to the TSS and/or CpG island shore (within 2 Kb from islands) near promoters—may reduce or inhibit gene expression¹⁶, while methylation at sites within a gene, such as the region after the first exon, may result in an increase in gene expression¹⁶.

This study prioritized 18 genes by comparing a set of 84 differentially methylated genes with a published set of differentially expressed genes in T1D. Literature reports supported the role of some of these 18 genes in T1D as listed below:

ICA1: This gene is involved in membrane trafficking at the Golgi complex, including transport of insulin in immature secretory granules^{17,18}. *ICA1* is an autoantigen in insulin-dependent diabetes mellitus^{19–21}; however, it has a low number of detectable ICA in Chinese patients compared to Caucasians²². The expression of *ICA1* is high in human pancreatic islets²³ and is upregulated in pancreatic tissue at the clinical onset of T1D (Table 2). The identified hypermethylation events of *ICA1* at the gene body correlate with its upregulation expression in our T1D patients (Fig. 3), providing evidence for DNAm mediated regulation of gene expression. However, it is difficult to discern its causality until a fully integrated functional genomics strategy is performed²⁴. Nevertheless, it is interesting that our study identifies differential methylation events in one of the four major autoantibodies

(ICA, GAD, IAA, and IA-2A) that characterize T1D. It is interesting to note that Rakyan et al.²⁵ observed T1D-associated methylation variable positions (T1D-MVPs) in one of these four autoantibodies, namely GAD2.

MMP26: This gene has been associated with higher fasting plasma glucose levels²⁶ and impaired wound healing in diabetic patients because of its role in extracellular matrix (ECM) degradation²⁷.

TIMP3: This gene is functionally related to ECM regulation. Our study identified this gene from both the tiles and targeted DMR analysis. It had been shown that *TIMP3* expression was decreased in the kidney of STZ treated-mice, a well-known model for hyperglycemia and glucotoxicity, reproducing T1D²⁸.

DRAM1: We observed hypermethylation of this gene in DMR analysis in the CpGi and promoter regions. Hypermethylation of the CpGi and promoter regions is usually associated with gene silencing. Nevertheless, blood samples from our T1D patients revealed that it was upregulated (Fig. 3). Other research groups observed similar results after performing integrated analysis in which DNAm in functional regions was both concordantly and discordantly associated with gene expression^{29,30}. Smith et al.³¹ demonstrated that increasing levels of promoter methylation do correlate directly with increased gene expression in a selection of contexts. It is possible to present a context-dependent model whereby epigenetic contributions to transcriptional regulation occur through diverse mechanisms in a more complex and dynamic manner^{31,32}. Functionally, the loss of *DRAM1* has been demonstrated to inhibit the mechanistic target of rapamycin complex 1 (mTORC1) activation that affects insulin signaling, glycemic balance, and adipocyte differentiation³³.

USP22: This gene suppresses high glucose-induced apoptosis in podocytes when silenced³⁴.

PIK3CG: This gene is known to suppress autoimmune diabetes when suppressed³⁵.

EXT1: The Exostosin Glycosyltransferase 1 gene encodes an enzyme that is critical for Heparan Sulfate synthesis in β -cells that regulate insulin secretion and glucose homeostasis³⁶.

The identified DNA methylation events in our study have little in common with those reported in literature. Reasons for this disparity include the following: Apart from one that adapted MZ concordance for T1D⁹, reports from literature adapted twins that were T1D-discordant^{8,25,37,38}. In addition, most of these DNAm studies were array-based which can have an impact on the overall results when compared to NGS-based DNAm profiling. For instance, the methylation profiling studies of Rakyan et al.²⁵ and Stefan et al.⁹ revealed less than 133 CpG sites displaying significant methylation changes using Illumina HumanMethylation 27 K array. The methylation profiling studies on T1D of Disanto et al.³⁷ and Elboudwarej et al.³⁸ revealed more than 1,000 CpG sites using Illumina HumanMethylation 450 K array. Thus, the results can vary among the same array-based DNAm profiling studies. Moreover, we used HLA-identical (Arab) families, an approach which was used only in one of the above-listed previous studies namely that of Elboudwarej et al.³⁸ on western population.

The present study has certain limitations. Firstly, we explored specific types of families having mono- or dizygotic twins concordant for T1D phenotype in order to identify novel candidate genes for T1D using NGS-based technology from a small size population in Kuwait. This made it difficult to have a large cohort size when compared to other array-based DNA methylation studies in the larger populations^{8,10}. Secondly, gene expression profiles derived from blood samples from the three families in our study to integrate with DNA methylation events is probably not the best prioritization approach for T1D candidate genes. To compensate, we utilized (a) publicly available relevant transcriptomics data from blood and pancreatic tissue, and (b) larger cohort to validate the prioritized T1D candidate genes. Thirdly, peripheral blood sample is not the ideal biological sample as (a) it contains a diverse mixture of cells and DNAm is cell-specific³⁹ (b) it may harbour DNA methylation events due to pathological differences relating to autoimmune processes that we could not statistically adjust for¹⁰. Additionally, Paul et al.⁸ demonstrated an increased DNAm variability in T1D across the three immune cells. Despite this, Disanto et al.³⁷ observed a high correlation of methylation events in immune cell types between controls and T1D quadruplets, suggesting a common driving mechanism in more than one immune cell type. Furthermore, blood as biological sample has been used in other DNA methylation studies on T1D^{10,37,40} as well as in publicly available gene expression data used to prioritize T1D candidate genes in our study. Nevertheless, our future plans include use of enriched cell types from the peripheral blood to focus on identifying cell type-specific gene regulatory circuits involved in immune cell metabolism and the cell cycle.

Conclusion

DNAm profiling of Arab familial T1D in Kuwait revealed that several genes are related to T1D. The integrative approach of using differential methylation events with numerous publicly available transcriptomics data allowed us to shortlist functionally relevant genes that can be differentially methylated in T1D pathogenesis. Validation of the candidate genes in an independent cohort of T1D families from the same ethnicity helped to identify *ICA1* and *DRAM1* as the most plausible candidates. Monozygotic twins are tightly clustered in the PCA scatter plot suggesting a similar methylation pattern. Finally, in order to discern the methylation events that cause or respond to the disease, a large cohort (of prediabetes or T1D versus healthy individuals) with gene expression profiling is needed to replicate the findings of our study.

Materials and methods

Ethics statement and study cohort. The Ethical Review Committee of Dasman Diabetes Institute reviewed and approved the study protocol as per the guidelines of the Declaration of Helsinki, and of the United States Federal Policy for the Protection of Human Subjects.

T1D families were recruited from the outpatient pediatric diabetes clinic at Dasman Diabetes Institute based on physician referrals from Ministry of Health, Kuwait. Three families in each of which at least two children (twinned or non-twinned) were diagnosed with T1D were considered in this study. The T1D diagnosis criteria used are detailed in¹². In brief: Diagnosis was established according to WHO criteria⁴¹, which include fasting

hyperglycemia and absolute insulin deficiency as measured by C-Peptide level. The date of first insulin injection was taken as the date of onset.

Prior to participating in the study, every adult participant signed the informed consent form. As regards the pediatric subjects, informed consent was obtained from the parents/legal guardians and assent was obtained from children aged 7 years or more.

Blood sample collection and processing. Blood samples were collected in ethylenediaminetetraacetic acid (EDTA)-treated tubes and the genomic DNA was extracted using a QIAamp Blood DNA kit (Qiagen, Germany). A Qubit Fluorometer (ThermoFisher, USA) was used to quantify spectrophotometry according to the manufacturer's protocol. Absorbance values at 260–280 nm was checked for adherence to an optical density range of 1.8–2.1.

Bisulfite sequencing using Illumina Truseq Methyl Captures EPIC kit. Genomic DNA (500 ng) was subjected to fragmentation using Covaris E220 (Covaris Inc, USA). Purified DNA fragments were then used with Truseq Methyl Capture EPIC Library Prep Kit reagents (Illumina Inc, USA) for library preparation, according to the Illumina protocol. The protocol involved adapters ligation, hybridization of the libraries with Illumina adapters, followed by bisulfite conversion and amplification. Additionally, KAPA HiFi HotStart Uracil + ReadyMix Kits (Roche, Switzerland) were used to enrich the resultant library. The enriched libraries were later pooled and multiplexed for up to six samples and clustered by Truseq Paired Cluster Kit V3 (Illumina Inc, USA). Paired-end sequencing was performed using the Illumina HiSeq 2500 sequencing platform.

DNA methylation (DNAm) bioinformatics analysis. Raw Illumina paired-end bisulfite treated reads were converted into FastQ format by executing bcl2fastq2 Conversion Software version 2.20 (Illumina, USA) with the option of enabling sequence adaptors trimming. The quality of the sequenced data from FastQ files was assessed using FastQC version 0.11.9 (Babraham Bioinformatics, UK). Low quality reads and the first six base pairs of both the forward and reverse reads, were trimmed using *Trim Galore!* version 0.3.3 (Babraham Bioinformatics, UK), which utilizes Cutadapt version 1.1⁴². The trimmed reads were then aligned to in silico bisulfite-converted human genome hg19 using Bismark version 0.22.3⁴³ and Bowtie2 version 2.3.5.1⁴⁴. The resultant outputs of sequence alignment/map (SAM) files were subsequently sorted and were not subjected to a deduplication step as recommended for targeted capture sequencing.

SAM files were imported to the R console version 3.6.2 (<https://www.R-project.org/>) and were processed using methylKit package version 1.2.4⁴⁵ to make methylated cytosine calling and to derive statistics on coverage of CpG bases, on per base methylation statistics, as well as to perform clustering and differential methylation calculations. CpG sites with a minimum read quality of at least 20 and a minimum read coverage of at least 10, as well as a maximum coverage of 99.9%, were considered for differential methylation analysis. In the single base CpG resolution analysis, methylation calls were optimized by combining methylated sites from both the forward and reverse reads using the `destrand = TRUE` option. For target region CpG resolution analysis (called as truseq analysis), we used the target regions of epigenetic interests as downloaded from the manifest file of the Illumina Truseq Methyl Capture EPIC kit. Tiles analysis was performed with a tiling window of 1000 bp. Clustering analysis based on levels of CpG methylation at base-pair resolution was performed on all the samples using Ward's minimum variance method. Differentially methylated single CpG site (DMC) and regions (DMR) were detected if the methylation difference level between the compared groups was at least 25%, and methylKit q-value (equivalent to false discovery rate adjusted p-value) was ≤ 0.01 . Genome annotations of the DMC and DMR regions (from both the truseq and tiles analyses) were performed using the methylKit package and the hg19 refseq files on genes and CpG islands as downloaded from the UCSC genome browser (<http://genome.ucsc.edu/UCSC>). Additionally, annotation of CpG sites—performed using genomation⁴⁵ tool from R package—defined promoters as 1 kilo base (kb) upstream of transcription start sites (TSS), and CpG shores as 2 kb flanking CpG islands.

Environmental factors such as lifestyle, diet, and use of medication often correlate with disease status and influence epigenetic mechanisms⁴⁶. Hence, exploring methylation differences between affected and unaffected individuals from twins and families can provide an estimation of contribution from such plausible factors to methylome variation and reduce residual noise variance. To understand this, unsupervised principal component analysis (PCA) and Hierarchical clustering analysis (HCA) with Ward's hierarchical clustering approach were performed using functions available from the methylKit package.

Pathway enrichment analyses. WEB-based GENE SeT AnaLysis Toolkit (WebGestalt;⁴⁷) was used to perform pathway analysis through the method of over-representation analysis based on Kyoto Encyclopedia of Genes and Genomes (KEGG) that integrates genomic, chemical, and systemic functional information⁴⁸. A p -value < 0.05 was used as a cut-off for false discovery rate (FDR) to detect significantly enriched pathways.

Publicly available gene expression datasets for T1D. Gene Expression Omnibus and ArrayExpress databases were searched for published gene expression datasets on T1D by using “type 1 diabetes” as keyword. Genes that are differentially expressed with statistical significance were analyzed using the LIMMA package⁴⁹ with a value of < 0.05 as threshold for FDR p -value (while adjusting for covariates when available). Significant differentially expressed genes were integrated with DNAm results to prioritize candidate genes for T1D.

Validation of differentially expressed genes in an extended cohort of T1D families. Gene expression analysis was performed in an extended cohort of 33 T1D families including families A and B consisting of 63 affected and 65 unaffected family members. This cohort was collected as part of another study using similar ethical guidelines and recruitment protocols. RNA was extracted from peripheral blood using Qiagen RNA blood mini kit (Qiagen, Hilden, Germany) and then reverse transcribed using ABI reverse transcriptase kit (Applied Biosystem, USA) according to the manufacturer's protocol. Quantitative real-time PCR was performed using pre-designed ready to use Taqman gene expression assays obtained from Applied Biosystem (Life technologies, NY, USA). Assays were selected based on maximum transcript variant detection: *ICA1* (#Hs00245256: NM_001136020.2, NM_001276478.1, NM_004968.3, NM_022307.2) and *DRAM1* (#HS00218048: NM_018370.2, XM_005269004.2, XM_017019578.1). Gene expression analysis of each target was individually carried out relative to *GAPDH* as endogenous control on ABI 7500 real time PCR system (Life technologies, NY, USA) following manufacturer's protocol. The fold change (FC) for each sample relative to the control sample was calculated using the $2^{-\Delta\Delta CT}$ method, with a cut-off value of ≥ 2 . Differences in the expression level of candidate markers between the two tested groups were characterized using unpaired t-test.

Data availability

The DNA methylation NGS data generated is deposited at the NCBI Sequence Read Archive (BioProject ID PRJNA783490) and will be released when the manuscript is accepted for publication.

Received: 15 March 2022; Accepted: 22 June 2022

Published online: 30 June 2022

References

- Patterson, C. C. *et al.* Worldwide estimates of incidence, prevalence and mortality of type 1 diabetes in children and adolescents: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. *Diabet. Res. Clin. Practice*, <https://doi.org/10.1016/j.diabres.2019.107842> (2019).
- Redondo, M. J., Jeffrey, J., Fain, P. R., Eisenbarth, G. S. & Orban, T. Concordance for islet autoimmunity among monozygotic twins. *N. Engl. J. Med.* **359**, 2849–2850. <https://doi.org/10.1056/NEJMc0805398> (2008).
- Bakay, M., Pandey, R., Grant, S. F. A. & Hakonarson, H. The genetic contribution to type 1 diabetes. *Curr. Diabet. Rep.* <https://doi.org/10.1007/s11892-019-1235-1> (2019).
- Cao, N., Varney, M. D., Harrison, L. C. & Morahan, G. Definition of high-risk type 1 diabetes HLA-DR and HLA-DQ types using only three single nucleotide polymorphisms. *Diabetes* **62**, 2135–2140. <https://doi.org/10.2337/db12-1398> (2013).
- Ziegler, A. G. *et al.* Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children. *J. Am. Med. Assoc.* **309**, 2473–2479. <https://doi.org/10.1001/jama.2013.6285> (2013).
- Quinn, L. M., Wong, F. S. & Narendran, P. Environmental determinants of type 1 diabetes: From association to proving causality. *Front. Immunol.* <https://doi.org/10.3389/fimmu.2021.737964> (2021).
- Zhang, J. *et al.* Implication of epigenetic factors in the pathogenesis of type 1 diabetes. *Chin. Med. J.* **134**, 1031–1042. <https://doi.org/10.1097/cm9.0000000000001450> (2021).
- Paul, D. S. *et al.* Increased DNA methylation variability in type 1 diabetes across three immune effector cell types. *Nat. Commun.* <https://doi.org/10.1038/ncomms13555> (2016).
- Stefan, M., Zhang, W., Concepcion, E., Yi, Z. & Tomer, Y. DNA methylation profiles in type 1 diabetes twins point to strong epigenetic effects on etiology. *J. Autoimmun.* **50**, 33–37. <https://doi.org/10.1016/j.jaut.2013.10.001> (2014).
- Johnson, R. K. *et al.* Longitudinal DNA methylation differences precede type 1 diabetes. *Sci. Rep.* **10**, 3721. <https://doi.org/10.1038/s41598-020-60758-0> (2020).
- Zayed, H. Genetic epidemiology of type 1 diabetes in the 22 Arab Countries. *Curr. Diabet. Rep.* <https://doi.org/10.1007/s11892-016-0736-4> (2016).
- Shaltout, A. A. *et al.* Incidence of type 1 diabetes has doubled in Kuwaiti children 0–14 years over the last 20 years. *Pediatr. Diabetes* **18**, 761–766. <https://doi.org/10.1111/peidi.12480> (2017).
- Hannon, E. *et al.* Characterizing genetic and environmental influences on variable DNA methylation using monozygotic and dizygotic twins. *PLoS Genet.* <https://doi.org/10.1371/journal.pgen.1007544> (2018).
- Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74. <https://doi.org/10.1038/nature11247> (2012).
- Lin, N. *et al.* Genome-wide DNA methylation profiling in human breast tissue by illumina TruSeq methyl capture EPIC sequencing and infinium methylationEPIC beadchip microarray. *Epigenetics* **16**, 754–769. <https://doi.org/10.1080/15592294.2020.1827703> (2021).
- Jjingo, D., Conley, A. B., Yi, S. V., Lunyak, V. V. & Jordan, I. K. On the presence and role of human gene-body DNA methylation. *Oncotarget* **3**, 462–474 (2012).
- Spitzenberger, F. *et al.* Islet cell autoantigen of 69 kDa is an arfaptin-related protein associated with the Golgi complex of insulinoma INS-1 cells. *J. Biol. Chem.* **278**, 26166–26173. <https://doi.org/10.1074/jbc.M213222200> (2003).
- Buffà, L., Fuchs, E., Pietropaolo, M., Barr, F. & Solimena, M. ICA69 is a novel Rab2 effector regulating ER-Golgi trafficking in insulinoma cells. *Eur. J. Cell Biol.* **87**, 197–209. <https://doi.org/10.1016/j.ejcb.2007.11.003> (2008).
- Pietropaolo, M., Towns, R. & Eisenbarth, G. S. Humoral autoimmunity in type 1 diabetes: prediction, significance, and detection of distinct disease subtypes. *Cold Spring Harbor Perspect. Med.* <https://doi.org/10.1101/cshperspect.a012831> (2012).
- Lampasona, V. *et al.* Sera from patients with IDDM and healthy individuals have antibodies to ICA69 on western blots but do not immunoprecipitate liquid-phase antigen. *J. Autoimmun.* **7**, 665–674. <https://doi.org/10.1006/jaut.1994.1051> (1994).
- Martin, S. *et al.* Autoantibodies to the islet antigen ICA-69 occur in IDDM and in rheumatoid-arthritis. *Diabetologia* **38**, 351–355. <https://doi.org/10.1007/bf00400641> (1995).
- Thai, A. C., Ng, W. Y., Lui, K. F. & Cheah, J. S. Islet-cell and thyroid autoimmunity in Chinese patients with IDDM. *Diabetes Care* **18**, 586–587. <https://doi.org/10.2337/diacare.18.4.586> (1995).
- Karges, W., Pietropaolo, M., Ackerley, C. A. & Dosch, H. M. Gene expression of islet cell antigen p69 in human, mouse, and rat. *Diabetes* **45**, 513–521. <https://doi.org/10.2337/diabetes.45.4.513> (1996).
- Mill, J. & Heijmans, B. T. From promises to practical strategies in epigenetic epidemiology. *Nat. Rev. Genet.* **14**, 585–594. <https://doi.org/10.1038/nrg3405> (2013).
- Rakyan, V. K. *et al.* Identification of type 1 diabetes-associated DNA methylation variable positions that precede disease diagnosis. *PLoS Genet.* <https://doi.org/10.1371/journal.pgen.1002300> (2011).

26. Cauchi, S. *et al.* Analysis of novel risk loci for type 2 diabetes in a general French population: the DESIR study. *J. Mol. Med.-Jmm* **86**, 341–348. <https://doi.org/10.1007/s00109-007-0295-x> (2008).
27. Ayuk, S. M., Abrahamse, H. & Houreld, N. N. The role of matrix metalloproteinases in diabetic wound healing in relation to photobiomodulation. *J. Diabet. Res.* <https://doi.org/10.1155/2016/2897656> (2016).
28. Florentino, I. *et al.* Loss of TIMP3 underlies diabetic nephropathy via FoxO1/STAT1 interplay. *EMBO Mol. Med.* **52**, 441–455. <https://doi.org/10.1002/emmm.201201475> (2013).
29. Chen, Z. *et al.* Epigenomic profiling reveals an association between persistence of DNA methylation and metabolic memory in the DCCT/EDIC type 1 diabetes cohort. *Proc. Natl. Acad. Sci. USA* **113**, E3002–E3011. <https://doi.org/10.1073/pnas.1603712113> (2016).
30. Guo, K. *et al.* Genome-wide profiling of DNA methylation and gene expression identifies candidate genes for human diabetic neuropathy. *Clin. Epigenet.* <https://doi.org/10.1186/s13148-020-00913-6> (2020).
31. Smith, J., Sen, S., Weeks, R. J., Eccles, M. R. & Chatterjee, A. Promoter DNA hypermethylation and paradoxical gene activation. *Trends Cancer* **6**, 392–406. <https://doi.org/10.1016/j.trecan.2020.02.007> (2020).
32. Rauluseviute, I., Drablos, F. & Rye, M. B. DNA hypermethylation associated with upregulated gene expression in prostate cancer demonstrates the diversity of epigenetic regulation. *BMC Med. Genomics* <https://doi.org/10.1186/s12920-020-0657-6> (2020).
33. Beaumatin, F. *et al.* mTORC1 activation requires DRAM-1 by facilitating lysosomal amino acid efflux. *Mol. Cell* <https://doi.org/10.1016/j.molcel.2019.07.021> (2019).
34. Shi, J.-X., Wang, Q.-J., Li, H. & Huang, Q. Silencing of USP22 suppresses high glucose-induced apoptosis, ROS production and inflammation in podocytes. *Mol. Biosyst.* **12**, 1445–1456. <https://doi.org/10.1039/c5mb00722d> (2016).
35. Azzi, J. *et al.* PI3K gamma Deficient NOD-Mice Are Protected from Diabetes by Restoring the Balance of Regulatory to Effector-T-Cells. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0169695> (2017).
36. Matsuzawa, T. *et al.* Heparan sulfate in pancreatic beta-cells contributes to normal glucose homeostasis by regulating insulin secretion. *Biochem. Biophys. Res. Commun.* **499**, 688–695. <https://doi.org/10.1016/j.bbrc.2018.03.213> (2018).
37. Disanto, G. *et al.* DNA methylation in monozygotic quadruplets affected by type 1 diabetes. *Diabetologia* **56**, 2093–2095. <https://doi.org/10.1007/s00125-013-2972-3> (2013).
38. Elboudwarej, E. *et al.* Hypomethylation within gene promoter regions and type 1 diabetes in discordant monozygotic twins. *J. Autoimmun.* **68**, 23–29. <https://doi.org/10.1016/j.jaut.2015.12.003> (2016).
39. Varley, K. E. *et al.* Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res.* **23**, 555–567. <https://doi.org/10.1101/gr.147942.112> (2013).
40. Agardh, E. *et al.* Genome-wide analysis of DNA methylation in subjects with type 1 diabetes identifies epigenetic modifications associated with proliferative diabetic retinopathy. *BMC Med.* <https://doi.org/10.1186/s12916-015-0421-5> (2015).
41. Gabir, M. M. *et al.* The 1997 American Diabetes Association and 1999 World Health Organization criteria for hyperglycemia in the diagnosis and prediction of diabetes. *Diabetes Care* **23**, 1108–1112. <https://doi.org/10.2337/diacare.23.8.1108> (2000).
42. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12. <https://doi.org/10.14806/ej.17.1.200> (2011).
43. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572. <https://doi.org/10.1093/bioinformatics/btr167> (2011).
44. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–U354. <https://doi.org/10.1038/nmeth.1923> (2012).
45. Akalin, A. *et al.* methylKit: A comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* <https://doi.org/10.1186/gb-2012-13-10-R87> (2012).
46. Alegria-Torres, J. A., Baccarelli, A. & Bollati, V. Epigenetics and lifestyle. *Epigenomics* **3**, 267–277. <https://doi.org/10.2217/epi.11.22> (2011).
47. Wang, J., Duncan, D., Shi, Z. & Zhang, B. WEB-based Gene Set Analysis Toolkit (WebGestalt): Update 2013. *Nucleic Acids Res.* **41**, W77–W83. <https://doi.org/10.1093/nar/gkt439> (2013).
48. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30. <https://doi.org/10.1093/nar/28.1.27> (2000).
49. Smyth, G. K. Limma: Linear models for microarray data. *Bioinf. Comput. Biol. Sol. Using R Bioconductor* https://doi.org/10.1007/0-387-29362-0_23 (2005).

Acknowledgements

The authors acknowledge the Kuwait Foundation for Advancement in Sciences for the institutional funding. We also thank all clinicians, patients and their family members for their interest and willingness to contribute to the study. We extend our sincere gratitude to National Dasman Biobank staff Ms. Daisy Thomas, Ms. Aditi Mathur and Mr. Sriraman Devarajan for their extended support throughout the study. We also thank Dr. Azza Shaltout for her support to the initial stages of the project.

Author contributions

F.A.-M., M.D., R.N., & T.A.T. designed the study. H.A.-K. carried out family recruitment and collected data on clinical characteristics. R.N. carried out genome wide methylation experiments & M.D. carried out bioinformatic analysis and interpretation. M.D. & F.A.-M. shortlisted the target genes. R.N., A.M. & S.J. performed the target validation. M.D., F.A.-M., T.A.T., P.H. & R.N. carried out analysis and interpretation of the resulting data. S.E.J and A.C. contributed to the data analysis. M.D., T.A.T., R.N. and F.A.-M. wrote the manuscript. F.A.-M. & T.A.T. directed the work and critically reviewed/edited the manuscript. All authors read and approved the final version of the manuscript.

Funding

Kuwait Foundation for Advancement in Sciences funding RA-2014–024.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-15304-5>.

Correspondence and requests for materials should be addressed to T.A.T. or F.A.-M.

Reprints and permissions information is available at www.nature.com/reprints.

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



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 Author(s) 2022