Communications Medicine

https://doi.org/10.1038/s43856-026-01563-0

Article in Press

Single-cell transcriptomics identifies regulatory T cell heterogeneity in gestational diabetes mellitus

Received: 9 January 2024

Accepted: 10 March 2026

Cite this article as: Mensah, N.E., Efthymiou, A., Mureanu, N. *et al.* Single-cell transcriptomics identifies regulatory T cell heterogeneity in gestational diabetes mellitus. *Commun Med* (2026). https://doi.org/10.1038/s43856-026-01563-0

Nana E. Mensah, Athina Efthymiou, Nicoleta Mureanu, María Teresa Martín Monreal, Heli Vaikkinen, Shichina Kannambath, Amanda Bowman, Athul Menon, Tim Tree, Giovanna Lombardi, Pawan Dhami, Kypros H. Nicolaides, Cristiano Scottà & Panicos Shangaris

Single-cell transcriptomics identifies regulatory T cell heterogeneity in Gestational Diabetes Mellitus

Nana E. Mensah[a,b*], Athina Efthymiou[b*], Nicoleta Mureanu[b*], María Teresa Martín Monreal[c*], Heli Vaikkinen[a], Shichina Kannambath[a], Amanda Bowman[d], Athul Menon[a], Tim Tree[c], Giovanna Lombardi[c], Pawan Dhami[a], Kypros H. Nicolaides[b,d], Cristiano Scotta[c$], and Panicos Shangaris[b,c,d$].

[a] Guy's and St Thomas' NHS Foundation Trust and King's College London NIHR Biomedical Research Centre 10th Floor Tower Wing, Guy's Hospital, Great Maze Pond, London SE1 9RT, UK
[b] Harris Birthright Research Centre for Fetal Medicine, King's College London, London, United Kingdom.
[c] Peter Gorer Department of Immunobiology, School of Immunology & Microbial Sciences, Faculty of Life Sciences & Medicine, King's College London, London, UK.
[d] School of Life Course & Population Sciences, Kings College London, London, 10th Floor North Wing St Thomas' Hospital, London, UK, SE1 7EH

*Co-first authors
[$]Co-senior authors

Corresponding Author: **Panicos Shangaris, panicos.shangaris@kcl.ac.uk**

Abstract

Background

Gestational diabetes mellitus (GDM) is a common pregnancy complication associated with hyperglycaemia, chronic inflammation and adverse health outcomes. Regulatory T cells (Tregs) are thought to contribute to GDM due to their role in suppressing inflammation. However, whether specific Treg subsets are transcriptionally dysregulated in patients with GDM remains unclear.

Methods

To investigate Treg transcriptional variation in GDM, we applied single-cell RNA sequencing to Tregs and CD4+ T cells isolated from the blood of 13 healthy pregnant women and 10 female patients with GDM.

Results

We observed no significant differences in Treg cluster proportions with disease status, however, Memory CD4+ T cells were more abundant in patients diagnosed with GDM, substantiated by mass cytometry. We report Treg subsets altered in GDM, including naive Tregs with reduced expression of AP-1 transcription factor subunits and effector Tregs with increased signalling of genes associated with angiogenesis. Expression levels of genes dysregulated in GDM Tregs were informative of GDM status in pseudobulk and whole blood mRNA from independent cohorts. *TXNIP*, which regulates glucose levels, emerged as the most significant discriminator of GDM status from bulk mRNA.

Conclusions

This study uncovers transcriptional differences of Treg cell subsets from GDM patients and transcriptional markers informative of GDM status.

Plain language summary

Gestational diabetes mellitus (GDM) is a common pregnancy condition linked to high blood sugar and increased inflammation, which can affect the health of both mother and baby. Immune cells called regulatory T cells (Tregs) help control inflammation, and their activity in a mother's blood may be linked to GDM. To understand how Tregs behave in patients with GDM, we captured these cells from blood samples of pregnant women diagnosed with GDM and pregnant women without a GDM diagnosis. We profiled the expression of RNA in individual Tregs from these patients. We found that, while overall Treg numbers are similar, the activity of specific genes varies in Tregs from women with GDM. Disrupted RNA levels of one gene related to glucose control (*TXNIP*) may be an informative marker for GDM in blood. Our findings enhance the understanding of immune changes in GDM and may inform future approaches for early detection and monitoring.

Introduction

Gestational diabetes mellitus (GDM) is broadly defined as hyperglycaemia first recognised during pregnancy [1]. GDM is one of the most common complications in pregnancy, affecting 9-26% of pregnancies worldwide, with a rapidly increasing global incidence [2,3]. Diagnosis of GDM is linked to clinical risk factors such as obesity, age, ancestry, and a family history of type 2 diabetes. However, no diagnostic threshold has been adopted globally [1]. Furthermore, hyperglycaemia may remain undiagnosed in patients who do not meet specific thresholds [4]. GDM is associated with an increased risk of postpartum Type 2 Diabetes [5,6], an increased risk of cardiovascular disorders [3] and a greater risk of metabolic syndrome in offspring [7]. As such, GDM represents an ongoing global health challenge.

Chronic inflammation is a hallmark of GDM, whereby maladaptation of the maternal immune response may contribute to the disease pathophysiology [1]. Regulatory T cells (Tregs) are a subset of CD4+ T cells responsible for maintaining immune homeostasis and inhibiting unwanted immune responses [8]. Tregs are characterised by the intracellular marker *FOXP3,* which is required for lineage specification [9,10]. Although Tregs can be identified by a combination of cell surface markers (CD4+, CD25+, CD127-), they are heterogeneous and can be divided into phenotypic subsets, namely naive and effector cells, based on markers such as *CTLA-4, GITR*, *HLA-DR,* and *CCR7* [9,11,12].

In healthy pregnancies, Tregs prevent the rejection of the fetus by the maternal immune system [13]. Lower Treg percentages have been observed with specific adverse pregnancy outcomes [14], however, existing literature on the role of Tregs in GDM is conflicting [15]. Several studies have assessed Treg percentages

in GDM patients relative to healthy controls with conflicting conclusions [15–19]. A meta-analysis of seven publications concludes that Tregs are significantly lower in women with GDM however, the mechanism of this association remains unclear [20]. Functional assays suggest that Tregs in GDM patients may inefficiently regulate immune responses. For example, GDM Tregs are less effective at suppressing IFN-γ and TNF-α production in effector T cells [16] and the activity of CD4+ T cells [18]. Transcriptional networks may therefore be altered in Tregs as a result of GDM.

Single-cell RNA sequencing (scRNA-seq) can provide high-resolution maps of transcriptional states in single cells. scRNA-seq has been used to identify T cell subsets based on differential expression of molecular markers, such as *CRR7 or TCF7* in naive T cells and *CCL5* in effector T cells [21]. In comparisons between GDM and healthy patient cohorts, scRNA-seq has identified altered estrogen signalling in T and NK cells [22] and a greater abundance of lymphocytes with upregulated reactive oxygen species and oxidative phosphorylation pathways [23]. Although evidence for a pro-inflammatory T cell phenotype in GDM patients has mounted in recent years, the role and function of Tregs in GDM pathology remains ambiguous [15].

Blood glucose measurements are used to diagnose GDM, however, earlier detection of at-risk patients could inform timely management and treatment, such as the use of metformin to prevent obstetrical complications [24]. Novel molecular biomarkers may inform our understanding of GDM disease mechanisms [25]. For example, serum TNF-alpha levels are elevated in GDM patients [17], and this has been attributed to a failure of Tregs to suppress effector T cell activity [16]. Identifying GDM-driven variation in Treg expression profiles, therefore, provides a foundation from which translational research can develop.

This study aims to investigate the Treg transcriptional landscape in GDM through scRNA-seq. We identify genes differentially expressed within Treg subsets from GDM patients and detect gene sets altered in GDM Tregs. We explore the potential for these cell-type-specific signals to serve as clinical biomarkers by modelling gene expression and GDM status across cohorts. This work provides a compendium of dysregulated genes within *FOXP3*+ Tregs and expands our understanding of their contribution to the pathophysiology of GDM.

Methods

PBMC isolation and FACS

Venous blood was collected from 23 participating pregnant women at 35-36 weeks of gestation, of whom 10 were diagnosed with GDM and 13 were recruited as healthy controls. Participant characteristics can be found in **Table 1** and **Supplementary Data 1-8: Table S1**. Blood was captured in BD Vacutainer® EDTA Tubes with a volume of 10 mL, and the samples were kept at room temperature (18–22°C) while being agitated using an orbital shaker at 90-100 rpm. Peripheral blood mononuclear cells (PBMCs) were isolated as previously described [54]. Briefly, Ficoll-Paque PLUS medium was placed into a SepMate™-50 tube, and the collected blood was transferred to a separate 50 mL Falcon tube where it was diluted 1:1 with PBS-EDTA containing 2% FBS. This diluted sample was added to the SepMate™ tube and then centrifuged at 1200 x g for 20 minutes. After isolation, cells were counted and resuspended in 1 mL of freezing medium containing DMSO. Cryovials containing the cells were kept on wet ice for less than 5 minutes before commencing freezing. The cryovials were then placed in a Corning CoolCell chamber at temperatures below -70°C. After a minimum of 12 hours, these were transferred to cryo boxes and stored in a -80°C freezer for up to 24 hours before their final transfer to a liquid nitrogen tank.

In the thawing process, the cryovial containing the frozen cells was submerged in a 37°C water bath for approximately one minute. Once thawed, 1 mL of pre-warmed medium was promptly added to the cryovial using a transfer pipette. The contents were then transferred to a 15 mL conical Falcon™ tube, which was prefilled with 5 mL of medium also warmed to 37°C. The empty cryovial was then rinsed with an additional 1 mL of medium to ensure complete transfer of cells. Following this, the Falcon™ tube was incubated in the 37°C water bath for 5 minutes. After incubation, the tube was centrifuged at 300×g at room temperature for another 5 minutes. The supernatant was discarded. PBMCs were washed twice in PBS and resuspended. Cells were counted using a hemocytometer and trypan blue to achieve a density of 700-1200 cells per ul. The cells were then stained for CD3 AlexaFLuor 700, CD4 BUV395, CD25 PE-CF594 and CD127 BV786, all used diluted 1:100. DAPI was used to exclude dead cells. The Tregs and CD4+ T cells were sorted in a BD FACS Aria III in 1.5ml Eppendorf using the gating strategy shown in **Supplementary Figure S1**.

Ethics statement

This study was approved by the King's College Hospital Research Ethics Committee, REC number 02-03-033, dated 01/04/2003. All the experiments conform to the relevant regulatory standards, including proper recruitment and obtaining informed consent.

Single-cell RNA sequencing

Single-cell mRNA, dual-indexed, sequencing libraries were generated following the Chromium Single Cell 5' v2 protocol with Feature Barcode technology for Cell Surface Protein and Immune Receptor Mapping (CG000330, 10x Genomics). For each 10x run, patient samples were hash-tagged and pooled, using one antibody-derived tag per patient following the CITE-Seq protocol [55] (Biolegend) and then flow-sorted to generate separate compartments of CD4+ T-cells and CD25+CD127lo T-cells before proceeding with the 10x assay. Gel beads-in-emulsion (GEMs) were generated on the chromium controller using reagents specific for the 5' assay. 10x barcoded, full-length cDNA from poly-adenylated mRNA and DNA from protein Feature Barcode were generated and amplified. The amplified cDNA was used to generate gene expression and V(D)J sequencing libraries. To generate the V(D)J sequencing libraries, full-length, 10x barcoded V(D)J segments were enriched from amplified cDNA via further amplification using primers specific to TCR constant regions and converted into sequencing libraries. Cell surface protein Feature Barcode-sequencing libraries were generated from amplified DNA to detect the antibody barcodes used for cell hashing. The dual indexed libraries thus generated for 5' gene expression, V(D)J and Cell Surface Protein were pooled and sequenced on Illumina NextSeq 2000 and NovaSeq 6000 platform to generate paired-end reads (28-10-10-90) following 10x recommended sequencing depth.

scRNA-seq data processing and quality control

The cellranger (v6.1.1, 10X Genomics) *mkfastq* command was used to demultiplex raw sequencing data into FASTQ files and reads were aligned to the GRCh38 and converted to raw feature counts using the cellranger *multi* command. Count tables were imported into R (v4.1.3) and analysed using Seurat (v4.1.0). For all downstream analyses, gene expression counts for isolated Treg and CD4+ T cells were processed separately.

Cells were assigned to hashtag sequences using the Seurat *HTODemux* function, and only cells identified as Singlets were retained. To detect sample mix-ups, transcript BAM files were demultiplexed with custom scripts and germline SNPs were identified per patient using cellSNP (v0.3.2). Sample mixups were labelled if pairwise SNP allele frequency Pearson correlations exceeded a threshold of 0.8. This procedure correctly identified one known duplicate sample, which was removed from subsequent analysis.

To detect outlier cells, cells with UMI counts >4x the absolute deviation were removed from downstream analysis. Furthermore, cells with transcript counts for fewer than 350 genes were excluded due to low capture. Cells were further excluded if fewer than 20% of known housekeeping genes or over 5% of

mitochondrial genes were expressed. Supervised detection of cell types was performed using SingleR (v1.8.1) and the cell dex database (v1.4.0), from which all barcodes assigned the 'T cell' label were retained for downstream analysis.

Cell cycle scores were estimated using the Seurat function *CellCycleScoring*. Transcript counts were normalised and then integrated using the Seurat sctransform workflow. First, we ran *SCTransform* with parameters: vst.flavor = "v2", verbose = FALSE, and vars.to.regress=c("S.Score", "G2M.Score"). *SelectIntegrationFeatures* was run with nfeatures = 3000*,* followed by *PrepSCTIntegration, FindIntegrationAnchors and IntegrateData* using default parameters and normalization.method = "SCT". This integration procedure is designed to align cell transcriptomes data while adjusting for batch effects.

We captured 22,218 Treg cells (mean: 966, SD:383.5) and 23,737 CD4+ T cells (mean: 1032, SD: 538.8) per patient following quality control. Across conditions, 43.57% of Tregs and 49.5% of CD4+ T cells were derived from GDM patients. Seurat objects for Treg and CD4+ T cell populations were processed separately in all downstream analyses.

T cell subpopulation detection and visualisation

Cells were clustered by running Seurat's *RunPCA* function with 50 principal components, followed by the functions *FindNeighbors* and *FindClusters* (Treg resolution=0.75, CD4+ resolution=0.9). T cell receptor (TCR) genes were filtered from variable features in the *Integrated* dataset to reduce their influence on cell clustering. For visualisation, UMAP and tSNE embeddings were derived from the Seurat *RunUMAP* (Treg components=2; Treg neighbors=50; CD4+ components=2; CD4+ neighbors=7) and *RunTSNE* functions (Treg perplexity=15; CD4+ perplexity=35).

Clusters detected in CD4+ and Treg Seurat objects were assigned broad groups of Naive, Effector or Memory by computational gating on the expression of phenotypic marker genes previously reported for conventional T cells and Tregs [11,56]. Clusters were further distinguished by the top marker genes as detected using the Seurat function *FindAllMarkers*.

Differential abundance

The propeller function from the R package speckle (v0.99.7) was used to test for differences in cell type proportions between GDM and control patients, separately for CD4+ T cells and FOXP3+ Tregs. Propeller calculates patient-level cell type proportions from cell annotations, performs a transformation on the counts (logit transform in this analysis), and applies a moderated *t*-test or ANOVA to find significant differences between cell type proportions [31].

Differential expression

Within-cluster differential expression testing was performed using the Seurat *FindMarkers* with the MAST (v1.25.2) testing framework applied to fit generalised linear models adapted for the bimodality and zero-inflated nature of single-cell gene expression data [57]. Cells from each unique cluster were filtered from Seurat SCT assay objects prior to MAST analysis. Genes with fewer than 10% of cells expressing transcripts in either GDM or control populations, Humanin+ clusters and ribosomal proteins were excluded from downstream analysis. Differentially expressed genes were defined as genes where the absolute average Log2 fold-change was more than 0.25 and the adjusted *p-value* (false discovery rate adjustment) was below 0.05. This effect size threshold on the Log2FC metric was selected to detect smaller, but significant changes in gene expression between groups of scRNA cell clusters, many of which share identical cell lineage commitments.

Gene set enrichment analysis

Gene set enrichment analysis was performed by testing all genes analysed in the differential expression analysis for their enrichment in the 50 Human MSigDB Hallmark gene sets (v7.5.1). For each within-cluster comparison, genes were ranked by the product of their average Log2Fold-change and adjusted p-value before significance testing against Hallmark gene sets using fgsea (v1.24.0). CD4+ T cells and Treg populations were treated separately. Adjusted p-values are calculated by fgsea using an adaptive multi-level split Monte-Carlo scheme.

CellPhoneDB

Seurat objects with scRNA-sequencing data from GDM-patient Tregs were exported to plain text files with R for analysis with the python CellPhoneDb package (v5.0.1), database version 5.0. Genes differentially expressed in GDM Tregs were passed to the *cpdb_degs_analysis_method* function with default parameters for ligand-receptor interaction at known differentially expressed genes.

TCR Clonotype analysis

Immune profiling data from the VDJ sequencing component of the 10X 5' scRNA-seq platform was loaded with R and processed with scRepertoire version 1.7.0 (https://github.com/ncborcherding/scRepertoire). Unique clonotypes were identified across samples using *quantContig* and contig abundance was calculated with *abundanceContig* prior to plotting with the R ggplot2 package.

GDM risk factor analysis

GDM risk factors were collected from patients at the time of recruitment, including ethnicity, body mass index (BMI), diastolic and systolic blood pressure, and interventions such as aspirin and metformin. Mean arterial pressure (MAP) was calculated using the commonly used formula summing the diastolic pressure and one-third of the pulse pressure, the difference between systolic and diastolic pressure [58]. The R *glm* function was used to fit logistic regression models to BMI, MAP and ethnicity as explanatory variables of GDM status. Leave-one-out cross-validation was implemented. The area under the receiver operating characteristic curve (AUC) was calculated from the results of each fold to evaluate the model.

GDM classification from RNA-seq

Pseudobulk expression profiles were generated from CD4+ T cell populations by averaging cell transcript counts from the Seurat RNA assay for each patient using Seurat *AverageExpression* function. Bulk mRNA profiles were acquired from previously published studies on GDM and control patients [33,34], and GDM status was converted to binary outcome in all datasets prior to downstream analysis.

Models were built using marker genes from scRNA-seq differential expression results. Specifically, GDM marker gene sets were defined by combining: differentially expressed genes in sorted Treg clusters, to capture specific signals of Treg malfunction.

Logistic regression with leave-one-out cross-validation was applied to mRNA data using the R function g*lm*. A single model was built for each gene as an explanatory variable for GDM status and the performance of the model was compared across mRNA and CD4+ T cell pseudobulk datasets.

Mass cytometry staining and data analysis

PBMCs from 2 GDM and 8 healthy pregnant women from the 23 participants recruited were stained with the MaxPar Direct Immune Profiling Panel and with the MaxPar Direct T Cell Expansion Panel 3 (both from Standard BioTools) following manufacturer's instructions. The cells were then suspended with EQ Calibration Beads (Fluidigm) and acquired in a Helios mass cytometer (Fluidigm). The data analysis was performed in RStudio (v. 4.4.1) using the packages *CATALYST*, *flowCore*, *flowWorkspace*, *ggcyto*, and *openCyto*. In brief, the .fcs data files produced were imported into R and the individual channel signals were normalised to that of beads. After that, beads were manually gated out and live intact cells were identified. CD4+ T cells were manually gated on the basis of the following markers: CD45+ CD3+ CD4+, and CD8-, and CD4+ T cell clusters were identified with unsupervised clustering using the FlowSOM workflow (clustering and metaclustering) followed by manual merging. The markers used for unsupervised clustering are detailed in **Supplementary Data 1-8: Table S7**.

Quantification of Humanin in Serum by ELISA

Serum concentrations of Humanin-like protein 8 (MTRNR2L8) were measured using the Abbexa Humanin ELISA Kit (Catalog No: abx527245) following the manufacturer's instructions. All ELISA procedures, including sample optimisation and analysis, were carried out by Boster Biotechnology (Pleasanton, CA, USA). In Phase I, five serum and five plasma samples were tested at eight serial dilutions to determine optimal loading concentrations. In Phases II and III, GDM serum samples and controls were run in triplicate using the validated protocol. For each assay, a 96-well microplate pre-coated with anti-Humanin antibody was used. 100 µL of standards, blank, and appropriately diluted serum samples were added to designated wells and incubated at 37°C for 2 hours. After removing the contents, 100 µL of biotin-conjugated Detection Reagent A was added and incubated for 1 hour at 37°C, followed by three washes with 1× Wash Buffer. Then, 100 µL of HRP-conjugated Detection Reagent B was added and incubated for an additional hour at 37°C, followed by five washing steps to remove excess conjugate. Substrate development was performed by adding 90 µL of TMB solution, incubating in the dark at 37°C for 10–20 minutes, and then stopping the reaction with 50 µL of Stop Solution, resulting in a yellow colour. Absorbance was measured at 450 nm using a microplate reader. A standard curve was generated from known concentrations, and sample concentrations were interpolated and corrected for dilution factors. All data analysis and reporting were performed by Boster using validated curve-fitting approaches.

Sensitivity analysis

Single cell RNA sequencing data from sample L215-A-AHH05 was held out from the dataset and unsupervised clustering was performed to identify T cell subpopulations in CD4+ and Treg datasets as described above. The resulting clusters in the sensitivity analysis were assigned the cluster label from the original analysis to which most cells belong. The percentage of cells retained in each cluster was calculated as the percentage of cells from the original analysis which retained the same cluster assignment in the sensitivity analysis.

Statistics and reproducibility

All analyses were performed using R version 4.2.2 (2022-10-31)**.** Differential analysis between scRNA-seq clusters was performed using MAST (v1.25.2), which fits a model and performs a two-sided likelihood ratio test with correction. Differential abundance testing was performed using the propeller function from the R speckle package, which uses a moderated *t*-test (two samples) or ANOVA (multiple samples). Continuous variables between two or more groups were analysed using the Wilcoxon rank sum test, unless otherwise specified. Mentioned sample sizes reflect the number of unique cells or patients

included in the analysis. Regression models were fit with R's glm function. Statistics derived from modelling were calculated using the R ModelMetrics package (v1.2.2.2).

Results

Single-cell RNA-seq identifies Treg subsets in GDM and healthy controls

To investigate Treg transcriptional states in GDM, we acquired peripheral blood mononuclear cells from 10 GDM patients and 13 healthy controls (**Table 1**). We performed cell sorting by FACS using cell surface markers $CD3^+$, $CD4^+$, $CD25^+$ and $CD127^-$ (**Methods, Supplementary Figure S1**), followed by single-cell RNA sequencing (10X Genomics 5' VDJ). We detected high-quality transcriptional profiles for 9,681 Tregs from GDM patients (mean: 968, SD: 468) and 12,537 cells from healthy controls (mean: 964, SD: 325) (**Methods**). Unsupervised clustering of single-cell expression profiles revealed seven Treg subsets (**Figure 1A**).

Table 1. Participant characteristics. SD, standard deviation; GDM, gestational diabetes; BMI, body mass index.

	Control (n = 13)	**GDM (n = 10)**
Age (years) (average, SD)	32.77 (2.42)	37.30 (4.03)
Ethnicity (number, %)		
White	10 (76.9%)	7 (70%)
Black	3 (23.1%)	3 (30%)
Weight (kg) (average, SD)	76.14 (9.16)	89.13 (17.11)
BMI (units) (SD)	27.88 (3.49)	32.29 (5.13)
Treatment:		
Diet		4
Metformin		3
Insulin + Metformin	1	3

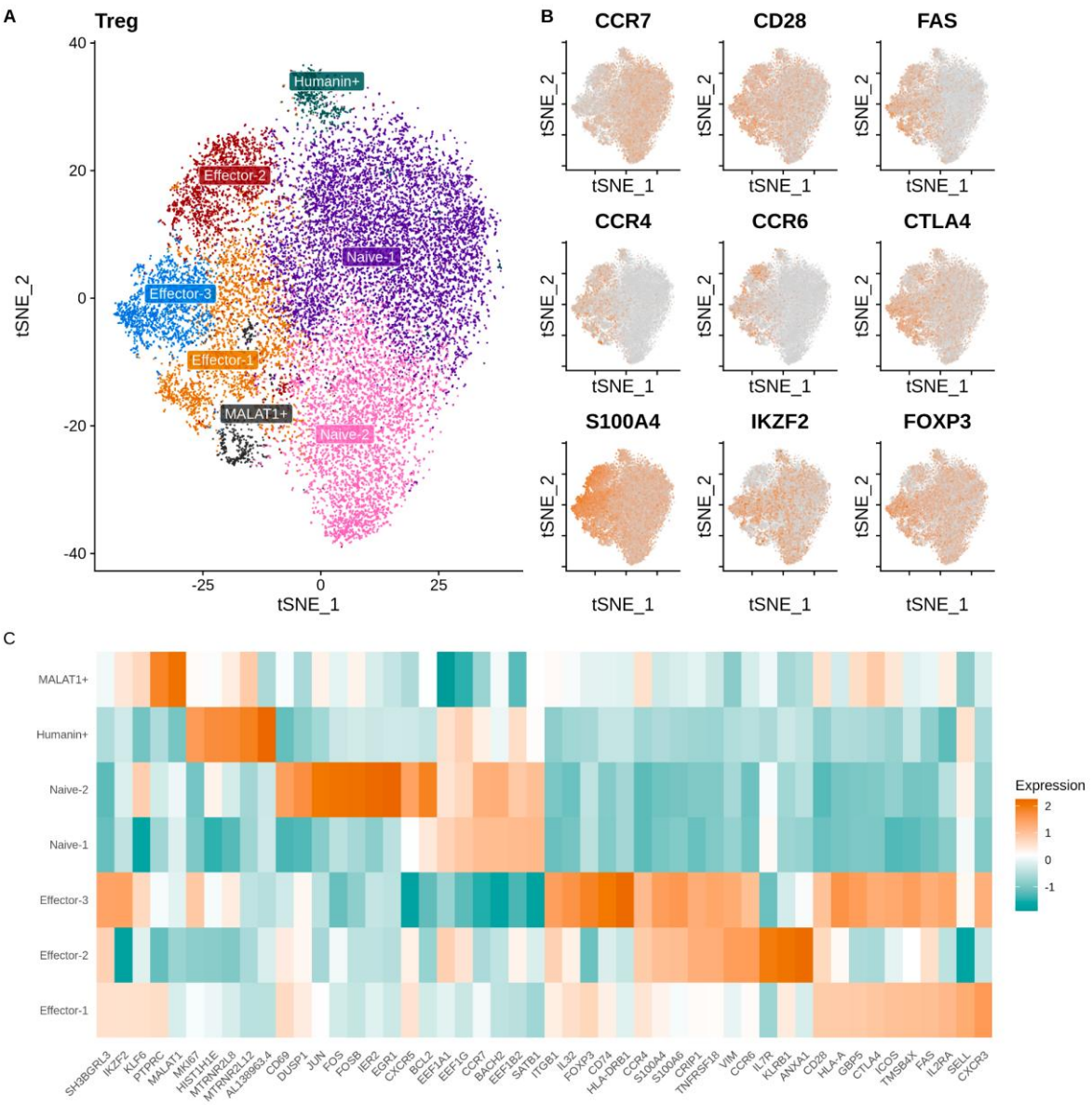

Figure 1. Treg subsets identified in GDM patients and healthy controls (N=22,218 cells). A) Treg cell clusters visualised using a t-stochastic neighbour embedding (t-SNE). B) Log-transformed normalized expression values of marker genes informative of Treg subset labels. C) Log-transformed normalized expression values of marker genes distinguishing Treg clusters.

To identify and label each Treg subset, we performed differential expression analysis comparing cells within each cluster to cells from all other clusters (**Methods, Figure 1C**). We observed high CCR7 expression in two clusters, which indicated the presence of naive Tregs. The Naive-1 cluster showed no unique markers, whereas the Naive-2 cluster uniquely overexpressed *EGR1, DUSP1, FOS, JUN,* and *IER2*. *JUN* and *FOS* are subunits of the AP-1 transcription factor involved in regulating Treg differentiation and Treg proliferation [26–28], which suggests that the two Naive subsets may present distinct cellular phenotypes. Next, we identified three effector Treg clusters among *CCR7* low cells expressing *GBP5, KLRDB1* and *S100A4,* respectively. An additional Treg cluster was defined by high *MALAT1*

expression, which is an indicator of low-quality cells in single-cell RNA sequencing experiments [29,30]. Lastly, we identified a subset of Tregs expressing Humanin nuclear isoform genes *MT-RNR2L8* and *MT-RNRL12*. Humanin+ cells were marked by the proliferation marker *MKi67* even after adjusting for the expression of cell cycle markers; however, no significant increase was observed for cell cycle scores from this cluster (**Supplementary Figure S2**). A complete list of marker genes distinguishing these Treg subsets is summarised in **Table 2**.

Table 2. Treg cell subtype assignment. Tregs from all patients were assigned to clusters by gene expression profile similarity. Clusters were assigned subtype names based on previous literature and marker genes identified by between-cluster differential expression analysis. N/A: Not applicable.

Cluster	Cells	Canonical Markers	Cluster-specific genes (Top 3)
Naive-1	9,206	*CCR7+, CD28-, CD95(FAS)-*	*JUN-, DUSP1-, KLF6-*
Naive-2	5,173	*CCR7+, CD28-, CD95(FAS)-*	*JUN+, FOS+, IER2+, EGR1+*
Effector-1	2,890	*CCR7-, CD28+, CD95(FAS)+*	*HLA-A+, IL32+, GBP5+*
Effector-2	1,956	*CCR7-, CD28+, CD95(FAS)+*	*ANXA1+, KLRB1+, CRIP1+*
Effector-3	1,765	*CCR7-, CD28+, CD95(FAS)+*	*CD74+, IL32+, S100A4+*
Humanin+	623	*N/A*	*MT-RNR2L12+, MT-RNR2L8+, AL138963.4+*
MALAT1+	605	*N/A*	*MALAT1+, MT-RNR2L12+, EEF1B2-*

To determine whether these marker genes are unique to Treg subsets, we performed scRNA-seq on sorted CD4+ T cells from the same patients, followed by unsupervised clustering (**Methods**). We identified 10 CD4+ T cell subsets including four naive clusters (*CCR7*+), two memory cell clusters (*CCR6*+), an effector cell cluster (*CXCR3*+), a Humanin+ cluster (*MT-RNR2L12*+), a *MALAT1*+ cluster and one *FOXP3*+ cluster expected to represent Tregs amongst the sorted CD4+ T cells (**Supplementary Figure S3A-C, Table S2**).

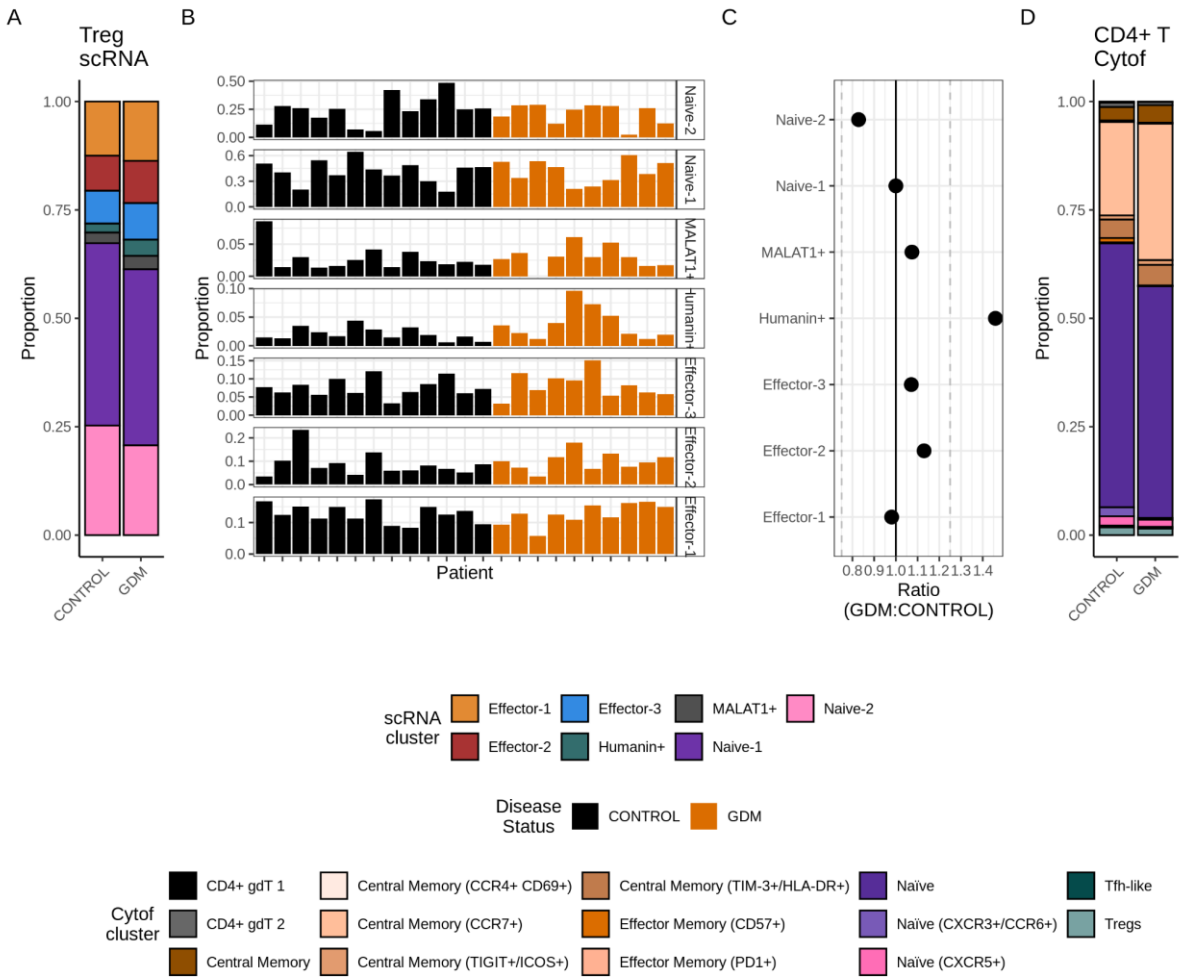

Figure 2. Treg and T cell abundance in GDM and Control patients. A) Proportion of cells assigned to each Treg cluster from scRNA-seq analysis. Control N = 12,537, GDM N = 9,681. B) Patient-level proportion of cells assigned to each Treg cell cluster from scRNA-seq analysis. C) Propeller analysis of the ratio of GDM:Control cells in Treg scRNA-seq clusters. Vertical dashed lines show effect size thresholds for ratios (<0.75 or >1.25). The total number of cells per cluster is reported in Table 1. D) Proportion of CD4+ T cells assigned to clusters from Cytof analysis. Control N = 1,534,988, GDM N = 461,131.

The proportion of cells in each cluster may vary between patients as a consequence of disease. Indeed, the proportion of Tregs assigned to each cluster varied between GDM and controls (**Figure 2A**) as well as within individual patients (**Figure 2B**). For example, patient L215-A-AHH03 (Control) carried the lowest proportion of Effector-2 cells (0.035), meanwhile, patient L215-A-AHH05 (Control) showed the highest proportion of Effector-2 cells (0.234) and a low relative proportion of Naive-1 cells (0.204). Therefore, to quantify the degree to which Treg proportions are altered by GDM, we performed significance testing using propeller[31] (**Methods**).

Although differences in Treg subset proportions were not statistically significant, the proportion of Humanin+ Tregs was notably higher in GDM patients (ratio=1.45, *p*=0.052) (**Figure 2B, Figure 2C**). We observed the Humanin+ cluster in Treg and CD4+ T cell populations, and we confirmed the presence of

reads mapped to the coding sequences of cluster markers *MT-RNR2L8* and *MT-RNR2L12* (**Supplementary Figure S4**). Therefore, we aimed to determine whether the Humanin protein product was enriched in GDM patients. We conducted ELISA experiments to measure serum levels of the Humanin protein in a broader cohort of GDM patients and controls (**Methods, Supplementary Figure S5**), however, we observed no significant increase in Humanin levels.

Next, we investigated the abundance of cell clusters in CD4+ T cell populations profiled by single-cell RNA-seq (**Supplementary Figure S6A**). Interestingly, Memory-1 CD4+ T cell proportions from the sorted CD4+ population were relatively higher in GDM (ratio=1.39, p=0.008), suggestive of a reduced immunosuppressive capacity of GDM Tregs (**Supplementary Figure S6B**). Mass cytometry staining confirmed a higher proportion of memory CD4+ T cells in GDM patients compared to controls (**Figure 2D, Supplementary Figure S7, Methods**). Sensitivity analysis showed that excluding patient L215-A-AHH05 reduced our ability to identify the low-quality *MALAT1*+ cells in the CD4 dataset, but not in the Treg datset **(Supplementary Figure S8)**. Furthermore, no significant difference in proportions was observed in the CD4+ T cell *FOXP3*+ subset (ratio=0.987, p=0.776) (**Supplementary Figure S6B**), suggesting that the overall proportion of Tregs in GDM and control patients may remain unchanged, as observed with Treg subtypes.

Altogether, these results indicate that Treg abundance in PBMCs may be consistent between GDM patients and healthy individuals, whereas Memory-1 CD4+ T cell abundance may be increased in GDM patients.

GDM-associated differential expression in Treg subsets

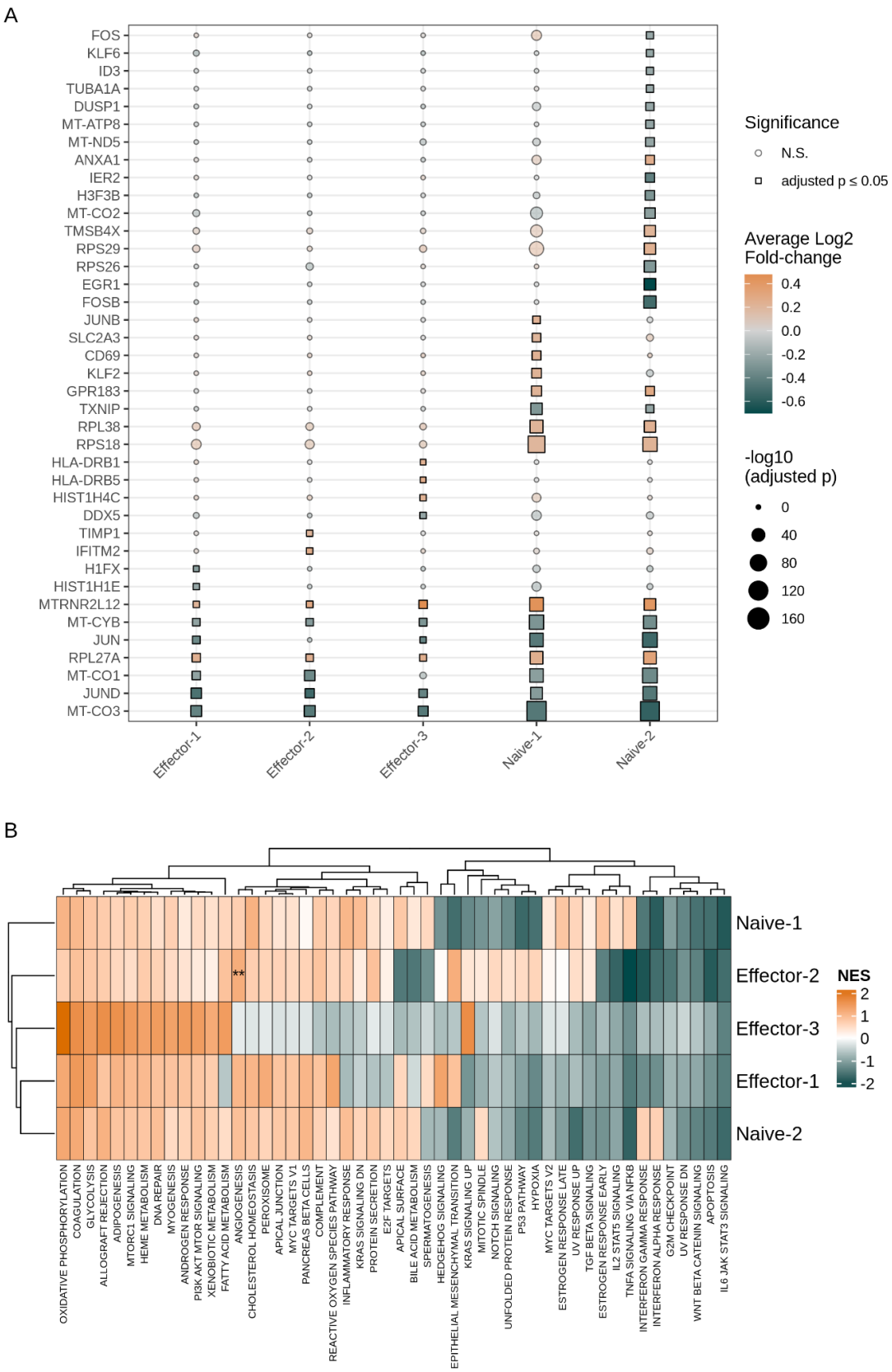

Figure 3. Differential expression between GDM and healthy controls within Treg subsets. A) Genes with significant differential expression results from MAST. The average log2 Fold-change is calculated from the expression cells from GDM patients relative to healthy controls in the same cluster. Adjusted p-values are derived from the MAST hurdle model with a two-sided likelihood ratio test with Bonferroni correction. N.S. = Not significant. N=20,990 cells. The number of cells per cluster is represented in Table 1. B) Gene set enrichment scores per Treg cluster. NES: Normalised enrichment score. '**' : fgsea *p*-value<=0.05.

To determine whether Treg subsets have altered transcriptomes in GDM, we performed differential expression analysis using Seurat and MAST (**Methods**). Here, we compared GDM and healthy controls within each Treg subset, excluding the Humanin+ and *MALAT1+* populations as they were not exclusive to the Treg cell population and their cell assignments were reliably stable in this cohort (**Supplementary Figure S8**). We detected an average of 9 and 21 differentially expressed genes in effector and naive Treg subsets, respectively (**Figure 3A**). *JUN, JUND, MT-CO1 and MT-CYB* were widely downregulated across GDM patient Tregs. The Naive-2 subset showed reduced expression of all four marker genes *JUN, FOS, EGR1* and *IER2*, which are components of the AP-1 transcription factor associated with Treg development [28]. In addition, we hypothesised that genes differentially expressed in GDM Effector Tregs might inform altered ligand-receptor interactions; however, no significant interactions were observed using CellPhoneDb (**Methods, Supplementary Data 1-8: Table S3**), nor did we observe differences in the expansion of specific TCR clonotypes from VDJ sequencing data (**Supplementary Figure S9**).

To investigate the transcriptional pathways altered in each Treg subset, we performed gene set enrichment analysis using the Hallmark Gene Sets from MSigDb[32]. We observed that Effector-2 Tregs were enriched for the Angiogenesis hallmark gene set, driven by *TIMP1*, *S100A4*, and *APP* expression (Normalised enrichment score=1.75, *p*-value=0.00474) (**Figure 3B**). Genes involved in TNF-α (Tumour Necrosis Factor Alpha) signalling via NF-kB (Nuclear Factor Kappa B) were downregulated in the Naive-2 cluster (Normalised enrichment score=-1.93, *p*-value=0.0553). This pathway includes genes *FOSB*, *EGR1*, *KLF2*, *NFKBIA*, DUSP2, and *JUNB,* several of which are uniquely expressed in the Naive-2 cluster (**Table 2**). This suggests that the transcriptional programme underlying Naive-2 Treg subsets may be dysregulated in GDM.

Treg marker gene expression in bulk mRNA samples correlates with GDM

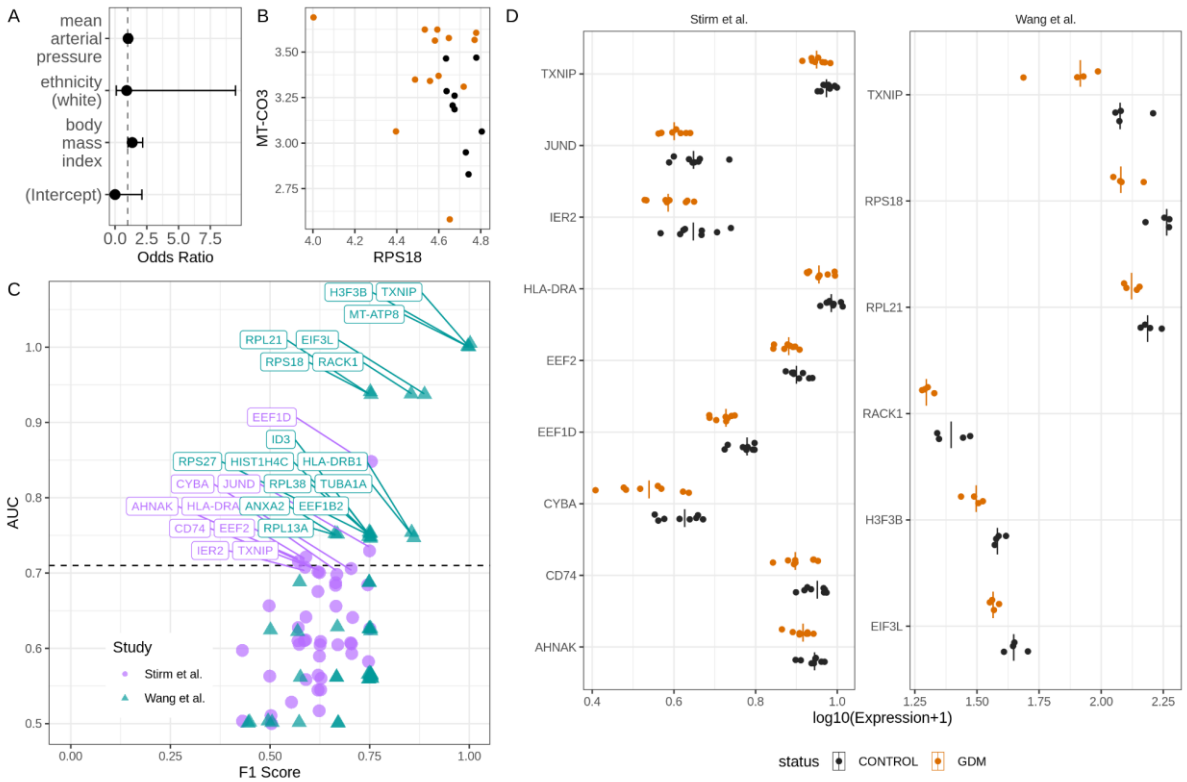

Figure 4. Evaluating the clinical utility of GDM Treg marker gene expression. A) Odds ratios from a multiple regression model for GDM status fit to mean arterial pressure, ethnicity (white or other) and body mass index. Error bars indicate 95% confidence intervals from the model. N = 23 patients. B) MT-CO3 and RPS18 gene expression in CD4+ T cell pseudobulk samples from GDM and control patients. N = 23 patients. C) Model evaluation scores for regression models fit with a single gene and evaluated with leave-one-out cross-validation. Area under the receiver operating characteristic curve (AUC) and F1 scores are displayed for Wang et al. (2021) (N = 8 patients) and Stirm et al. (2018) (N = 16 patients). Genes tested in each model were selected due to differential expression in Treg subsets. Gene names are displayed for genes with an AUC above 0.7. D) mRNA expression data from Wang et al. (2021) (N = 8 patients) and Stirm et al. (2018) (N = 16 patients) for the top associated marker genes highlighted in C). Vertical bars crossing data points represent the median of the distribution.

We hypothesised that genes differentially expressed in Tregs from patients with GDM may discern GDM status across patient cohorts. As a baseline for comparison, we determined the association between GDM and clinical features associated with elevated GDM risk. A logistic regression model using body mass index (BMI) and mean arterial pressure (MAP)[1] (**Figure 4A**) showed that BMI had a stronger independent association with GDM status (OR: 1.35, 95% CI: 1.01-2.17). Modelling BMI as an explanatory variable for GDM state by leave-one-out cross-validation (LOO-CV) returned a pooled area under the ROC curve (AUC) of 0.718, providing a baseline AUC score for a clinical measurement with a known disease association.

We selected 108 candidate marker genes differentially expressed within and between Treg subsets (**Table S4**) and performed logistic regression using gene expression as features, GDM status as a response variable and LOO-CV for evaluation (**Methods**). To approximate CD4+ T cell isolation from

blood, we included averaged scRNA-seq profiles of our CD4+ T cell populations (pseudobulk), which contains a subset of Tregs that were not used for discovery. Here, the highest AUC was achieved with *RPS18* (Treg=0.79, CD4=0.75) and *MT-CO3* (Treg=0.78, CD4=0.70) (**Figure 4B; Table S5**). These genes were broadly downregulated across multiple Treg clusters (**Figure 3A**) and may therefore reflect global metabolic alterations in GDM patients. *ANXA2* was among the markers of the Memory-1 CD4+ T cell population, and this was reflected in its AUC score for CD4+ T cells (Treg=0.52, CD4=0.72). These data suggest that genes differentially expressed between Treg subsets may discriminate GDM status from bulk RNA-sequencing of CD4+ T cells.

To investigate whether GDM Treg marker gene expression discriminates GDM status from whole blood, we analysed blood-derived bulk RNA sequencing data from two independent studies comparing patients with GDM to controls (N=8 and N=16 with 50/50 disease cases to controls in each study, respectively) [33,34]. Logistic regression models fitted to 21/108 candidate marker genes returned an AUC above 0.5 in blood (**Figure 4D**), including *TXNIP, MT-ATP8, H3F3B, RPS18, ANXA2, HLA-DRA* and *HLA-DRB1* (**Figure 4E**). In particular, *TXNIP* returned a highly discriminant model in both bulk mRNA datasets (Wang et al.=1; Stirm et al.=0.70), consistent with its observed expression across scRNA-seq clusters (**Supplementary Figure S10**). As orthogonal support for the results of the logistic regression models, we also performed recursive feature elimination using a random forest model (**Methods**). The most discriminative genes for each bulk RNA-seq dataset were also retained by the best models using this recursive approach (**Supplementary Data 1-8: Table S6**). Collectively, these results suggest that a subset of genes with dysregulated expression in GDM-derived Tregs may discriminate between patients with and without GDM from whole-blood mRNA.

Discussion

Regulatory T cell dysfunction is proposed to underlie the pathology of gestational diabetes mellitus (GDM). We present single-cell transcriptomes of Tregs and CD4+ T cells isolated from PBMCs of women with GDM and without GDM diagnosed during pregnancy. We identify naive and effector molecular Treg subsets and detect an increase in memory CD4+ T cells from GDM patients, suggesting impaired immune suppression. Differential expression analysis revealed a decrease in AP-1 transcription factor subunits in the Naive-2 cluster, and pathway analysis indicated downregulation of NF-kB signalling in Naive-2 Tregs. We evaluated the translational utility of scRNA-seq-derived expression markers from CD4+ T cell populations and their potential to discriminate GDM status in independent cohorts.

The percentage of Tregs in GDM patients is reported to be reduced in women with GDM [20]. We identified FOXP3+ cells within sorted CD4+ T cells, which we expect to carry a Treg cell population, however, their proportion did not significantly differ between conditions. FOXP3+ cells represent a low proportion of the CD4+ T cell pool. Therefore, this study may be underpowered to observe the effect at a single cell level. Of note, we observed Humanin+ and *MALAT1*+ cell clusters in both Treg and CD4+ T cell populations. *MALAT1*+ cells have been previously reported in sorted Tregs from ankylosing spondylitis patients profiled with the 10X scRNA-seq platform[35], however, these clusters likely represent poor-quality cells[29,30].

To the best of our knowledge, Humanin-expressing CD4+ T cells have not been described in the context of GDM, however, Humanin expression has been linked to hyperglycaemia in other disease contexts [36], and Humanin+ CD4+ T cells have been reported in the context of Rheumatoid Arthritis [37]. The alignment of reads to markers *MT-RNR2L12* and *-L8* may be driven by alignment of reads from the Humanin protein owing to high homology [38]. Altered Humanin concentrations have been proposed as a biomarker for diabetic conditions, as Humanin is protective under oxidative stress, which can arise from excess ROS generation from diabetic hyperglycaemia [36,39]. Although we detected a greater abundance of Humanin+ Tregs in our study, we did not observe a significant increase in the serum Humanin concentration in GDM patients. This indicates no systematic changes in Humanin protein levels in GDM, however, distinguishing whether the observed cluster results from biological or technical sources requires additional research beyond the scope of this study.

NF-kB signalling is associated with apoptosis, insulin resistance, systemic inflammation and all major diabetic complications [40]. *NFKBIA* is upregulated in whole blood from GDM patients [41]. Similarly, the TNF-alpha signalling via NF-kb hallmark pathway was disrupted through RNA-seq of a murine GDM

model [42]. We observed a decrease in *NFKBIA* signalling specific to the Naive-2 GDM Treg subset. Dimerisation between NF-kB and FOXP3 is essential for Treg activation and disruption to this pathway leads to immune dysregulation and altered Treg frequencies [43]. Therefore, this suggests a potential mechanism by which GDM Tregs would be less effective at managing inflammatory responses. Additionally, the Effector-2 population was enriched for angiogenesis pathway genes. Chronic inflammation and angiogenesis are associated with GDM and promoted by Tregs [44,45]. Future work will require isolation and phenotypic characterisation of the Effector-2 and Naive-2 cell populations in order to discern their roles in GDM.

Body mass index (BMI) is the most significant risk factor for GDM and maintaining a BMI below 25kg/m^2 is advised in order to lower GDM risk [1]. In addition, the lack of globally accepted diagnostic criteria for GDM[4] motivates the discovery of molecular disease indicators. We evaluated whether genes differentially expressed in Tregs can discriminate GDM status in whole blood assays. Rapid protocols are readily available for CD4+ T cell isolation from whole blood [46], therefore, we emulated this profiling using pseudobulk scRNA-seq expression. We observed that genes broadly altered across Treg subsets were discriminative of GDM, such as *RPS18*, *MT-CO1* and *MT-CO3*. *MT-CO1* and *MT-CO3* are subunits of Cytochrome c oxidase II involved in ATP synthesis, and mutations in *MT-CO3* are associated with maternally inherited mitochondrial diabetes[47], possibly contributing to a hyperglycaemic environment in GDM that affects Treg signalling. Expanding our analysis to bulk mRNA from independent cohorts, the Treg marker *TXNIP* was the best independent discriminator of GDM status. *TXNIP* has a role in diabetes and inflammation and has been proposed as a candidate biomarker for GDM [48], however, care must be taken to distinguish its utility from treatment effects, as it is a target for anti-diabetic drugs due to its association with insulin uptake and Metformin activity[48].

Our study offers insights into the Treg transcriptional landscape in GDM, though not without several limitations. Our focus on CD4+ cells misses several immune subtypes that are associated with the pathophysiology of GDM, such as Neutrophils, Macrophages and Monocytes [49]. Functional studies in models of GDM may further uncover the role of immune cells in attenuating Treg activity. For example, we observed the under-expression of *JUNB*, an AP-1 transcription factor which regulates intestinal Treg development and whose ablation is associated with increased T helper cell accumulation and inflammation [50]. Chronic inflammation leading to metabolic dysregulation is a significant coupling factor driving the association between obesity, cardiovascular disease, and diabetes[51]. Although single-cell RNA-seq empowers us to discover cell type-specific expression, this analysis remains limited to protein-coding RNAs. In contrast, several non-coding RNAs such as microRNAs[52] and circular RNAs[53] have been proposed as candidate circulating biomarkers for GDM. Studying the effect of non-coding RNAs on

T regulatory cell subtypes may uncover insights into GDM pathology. Importantly, this study is unable to discern the predictive potential of GDM Treg gene expression markers in early pregnancy, as patients were recruited close to delivery and after disease onset. Future longitudinal study designs may track GDM-associated gene expression trends between sample collection and time of onset from the whole blood transcriptome. In summary, we provide a map of transcriptional dysregulation of Treg subsets in GDM, highlighting programmes that may underlie GDM physiology. We uncover candidate marker genes with Treg-associated differential expression, laying the foundation for follow-up studies on their translational potential and a refined understanding of Treg dysregulation in GDM.

Data availability

Single-cell RNA sequencing data has been submitted to Gene Expression Omnibus (Accession: GSE280975). Bulk RNA sequencing data was downloaded from Gene Expression Omnibus (Accession: GSE154414; GSE92772 - RNA sequencing data of whole blood cells of normal glucose tolerant (NGT) and gestational diabetes (GDM) pregnant women). Data underlying the figures are available on Zenodo[59].

Code availability

R scripts used to perform the analyses are available at GitHub[60].

Author contributions

N.E.M.* , N.M.*, A.E.* and M.T.M.M* contributed equally to this work. N.E.M.*: Data analysis, manuscript writing, manuscript editing. A.E.*: Consent, sample collection, and data analysis. N.M.: Consent, sample collection, and data analysis. M.T.M.M.*: Data analysis, manuscript editing. S.K.: Data analysis. H.V.: Conducted experiments. A.B.: Data analysis and manuscript editing. A.M.: Data analysis. T.T.: Provided expert opinion. G.L.: Provided expert opinion and edited the manuscript. P.D.: Data analysis and provided expert opinion. K.H.N.: Provided expert opinion and edited the manuscript. C.S.: Provided expert opinion and edited the manuscript. P.S.: Conceptualized the study, conducted experiments, collected samples, wrote and edited the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgements

This research is supported by the Fetal Medicine Foundation (registered charity 1037116 Project Number 909375), and the Carlsberg Foundation (CF23-0418). PS is supported by an NIHR Clinical Lectureship (CL-2018-17-002) an Academy of Medical Sciences Starter Grant for Clinical Lecturers (SGL023\1023). The project was also supported by the National Institute for Health Research (NIHR) Biomedical Research Center at Guy's and St Thomas' National Health Service Foundation Trust and King's College London (IS-BRC-1215–20006). We would like to thank the team and patients at King's College Hospital who donated their blood for this project. We sincerely thank the team at the Advanced Cytometry

Platform, R&D Department and the Genomics facility at Guy's and St Thomas' NHS Foundation Trust, Guy's Hospital, London.

References

1. McIntyre, H. D. *et al.* Gestational diabetes mellitus. *Nat Rev Dis Primers* **5**, 47 (2019).
2. Jiang, L. *et al.* A global view of hypertensive disorders and diabetes mellitus during pregnancy. *Nat. Rev. Endocrinol.* **18**, 760–775 (2022).
3. Sweeting, A., Wong, J., Murphy, H. R. & Ross, G. P. A Clinical Update on Gestational Diabetes Mellitus. *Endocr. Rev.* **43**, 763–793 (2022).
4. American Diabetes Association. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2020. *Diabetes Care* **43**, S14–S31 (2020).
5. Bellamy, L., Casas, J.-P., Hingorani, A. D. & Williams, D. Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis. *Lancet* **373**, 1773–1779 (2009).
6. Noctor, E. & Dunne, F. P. Type 2 diabetes after gestational diabetes: The influence of changing diagnostic criteria. *World J. Diabetes* **6**, 234–244 (2015).
7. Ornoy, A., Becker, M., Weinstein-Fudim, L. & Ergaz, Z. Diabetes during Pregnancy: A Maternal Disease Complicating the Course of Pregnancy with Long-Term Deleterious Effects on the Offspring. A Clinical Review. *Int. J. Mol. Sci.* **22**, (2021).
8. Sakaguchi, S., Yamaguchi, T., Nomura, T. & Ono, M. Regulatory T cells and immune tolerance. *Cell* **133**, 775–787 (2008).
9. Churov, A. V., Mamashov, K. Y. & Novitskaia, A. V. Homeostasis and the functional roles of CD4+ Treg cells in aging. *Immunol. Lett.* **226**, 83–89 (2020).
10. Hori, S., Nomura, T. & Sakaguchi, S. Control of regulatory T cell development by the transcription factor Foxp3. *Science* **299**, 1057–1061 (2003).
11. Rosenblum, M. D., Way, S. S. & Abbas, A. K. Regulatory T cell memory. *Nat. Rev. Immunol.* **16**, 90–101 (2015).
12. Zemmour, D. *et al.* Single-cell gene expression reveals a landscape of regulatory T cell phenotypes shaped by the TCR. *Nat. Immunol.* **19**, 291–301 (2018).
13. Krop, J., Heidt, S., Claas, F. H. J. & Eikmans, M. Regulatory T Cells in Pregnancy: It Is Not All

About FoxP3. *Front. Immunol.* **11**, 1182 (2020).

14. Green, S. *et al.* Regulatory T Cells in Pregnancy Adverse Outcomes: A Systematic Review and Meta-Analysis. *Front. Immunol.* **12**, 737862 (2021).
15. McElwain, C. J., McCarthy, F. P. & McCarthy, C. M. Gestational Diabetes Mellitus and Maternal Immune Dysregulation: What We Know So Far. *Int. J. Mol. Sci.* **22**, (2021).
16. Yang, Y. *et al.* Functional Defects of Regulatory T Cell Through Interleukin 10 Mediated Mechanism in the Induction of Gestational Diabetes Mellitus. *DNA Cell Biol.* **37**, 278–285 (2018).
17. Lobo, T. F. *et al.* Impaired Treg and NK cells profile in overweight women with gestational diabetes mellitus. *Am. J. Reprod. Immunol.* **79**, (2018).
18. Schober, L. *et al.* The role of regulatory T cell (Treg) subsets in gestational diabetes mellitus. *Clin. Exp. Immunol.* **177**, 76–85 (2014).
19. Sifnaios, E. *et al.* Gestational Diabetes and T-cell (Th1/Th2/Th17/Treg) Immune Profile. *In Vivo* **33**, 31–40 (2019).
20. Arain, H. *et al.* Regulatory T cells in the peripheral blood of women with gestational diabetes: a systematic review and meta-analysis. *Front. Immunol.* **14**, (2023).
21. Szabo, P. A. *et al.* Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nat. Commun.* **10**, 4706 (2019).
22. Yang, Y. *et al.* Transcriptomic Profiling of Human Placenta in Gestational Diabetes Mellitus at the Single-Cell Level. *Front. Endocrinol.* **12**, 679582 (2021).
23. Jiao, B. *et al.* Dissecting human placental cells heterogeneity in preeclampsia and gestational diabetes using single-cell sequencing. *Mol. Immunol.* **161**, 104–118 (2023).
24. Farahvar, S., Walfisch, A. & Sheiner, E. Gestational diabetes risk factors and long-term consequences for both mother and offspring: a literature review. *Expert Rev. Endocrinol. Metab.* **14**, 63–74 (2019).
25. Rodrigo, N. & Glastras, S. J. The Emerging Role of Biomarkers in the Diagnosis of Gestational Diabetes Mellitus. *J. Clin. Med. Res.* **7**, (2018).
26. Lee, S.-M., Gao, B. & Fang, D. FoxP3 maintains Treg unresponsiveness by selectively inhibiting

the promoter DNA-binding activity of AP-1. *Blood* **111**, 3599–3606 (2008).

27. Bahrami, S. & Drabløs, F. Gene regulation in the immediate-early response process. *Adv. Biol. Regul.* **62**, 37–49 (2016).
28. Katagiri, T., Kameda, H., Nakano, H. & Yamazaki, S. Regulation of T cell differentiation by the AP-1 transcription factor JunB. *Immunol Med* **44**, 197–203 (2021).
29. Clarke, Z. A. & Bader, G. D. MALAT1 expression indicates cell quality in single-cell RNA sequencing data. *bioRxiv* 2024.07.14.603469 (2024) doi:10.1101/2024.07.14.603469.
30. Montserrat-Ayuso, T. & Esteve-Codina, A. High content of nuclei-free low-quality cells in reference single-cell atlases: a call for more stringent quality control using nuclear fraction. *BMC Genomics* **25**, 1124 (2024).
31. Phipson, B. *et al.* propeller: testing for differences in cell type proportions in single cell data. *Bioinformatics* **38**, 4720–4726 (2022).
32. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417–425 (2015).
33. Stirm, L. *et al.* Maternal whole blood cell miRNA-340 is elevated in gestational diabetes and inversely regulated by glucose and insulin. *Sci. Rep.* **8**, 1366 (2018).
34. Wang, J., Wang, K., Liu, W., Cai, Y. & Jin, H. m6A mRNA methylation regulates the development of gestational diabetes mellitus in Han Chinese women. *Genomics* **113**, 1048–1056 (2021).
35. Simone, D. *et al.* Single cell analysis of spondyloarthritis regulatory T cells identifies distinct synovial gene expression patterns and clonal fates. *Commun Biol* **4**, 1395 (2021).
36. Boutari, C., Pappas, P. D., Theodoridis, T. D. & Vavilis, D. Humanin and diabetes mellitus: A review of in vitro and in vivo studies. *World J. Diabetes* **13**, 213–223 (2022).
37. Argyriou, A. *et al.* Single cell sequencing reveals expanded cytotoxic CD4+ T cells and two states of peripheral helper T cells in synovial fluid of ACPA+ RA patients. *bioRxiv* (2021) doi:10.1101/2021.05.28.21255902.
38. Huth, A. *et al.* Single cell transcriptomics of cerebrospinal fluid cells from patients with recent-onset narcolepsy. *J. Autoimmun.* **146**, 103234 (2024).

39. Rochette, L., Meloux, A., Zeller, M., Cottin, Y. & Vergely, C. Role of humanin, a mitochondrial-derived peptide, in cardiovascular disorders. *Arch. Cardiovasc. Dis.* **113**, 564–571 (2020).
40. Indira, M. & Abhilash, P. A. Role of NF-Kappa B (NF-κB) in Diabetes. *OT* **4**, (2013).
41. Chen, Y.-M. *et al.* Upregulation of T Cell Receptor Signaling Pathway Components in Gestational Diabetes Mellitus Patients: Joint Analysis of mRNA and circRNA Expression Profiles. *Front. Endocrinol.* **12**, 774608 (2021).
42. Paolino, M. *et al.* RANK links thymic regulatory T cells to fetal loss and gestational diabetes in pregnancy. *Nature* **589**, 442–447 (2021).
43. Cui, Y. *et al.* A Stk4-Foxp3-NF-κB p65 transcriptional complex promotes Treg cell activation and homeostasis. *Sci Immunol* **7**, eabl8357 (2022).
44. Joshi, N. P., Madiwale, S. D., Sundrani, D. P. & Joshi, S. R. Fatty acids, inflammation and angiogenesis in women with gestational diabetes mellitus. *Biochimie* **212**, 31–40 (2023).
45. Lužnik, Z., Anchouche, S., Dana, R. & Yin, J. Regulatory T Cells in Angiogenesis. *J. Immunol.* **205**, 2557–2565 (2020).
46. Hsu, C.-H., Chen, C., Irimia, D. & Toner, M. Fast sorting of CD4+ T cells from whole blood using glass microbubbles. *Technology* **3**, 38–44 (2015).
47. Tabebi, M. *et al.* A novel mutation MT-COIII m.9267G>C and MT-COI m.5913G>A mutation in mitochondrial genes in a Tunisian family with maternally inherited diabetes and deafness (MIDD) associated with severe nephropathy. *Biochem. Biophys. Res. Commun.* **459**, 353–360 (2015).
48. Masutani, H. Thioredoxin-Interacting Protein in Cancer and Diabetes. *Antioxid. Redox Signal.* **36**, 1001–1022 (2022).
49. De Luccia, T. P. B. *et al.* Unveiling the pathophysiology of gestational diabetes: Studies on local and peripheral immune cells. *Scand. J. Immunol.* **91**, e12860 (2020).
50. Wheaton, J. D. & Ciofani, M. JunB Controls Intestinal Effector Programs in Regulatory T Cells. *Front. Immunol.* **11**, 444 (2020).
51. Hotamisligil, G. S. Inflammation, metaflammation and immunometabolic disorders. *Nature* **542**, 177–185 (2017).

52. Guarino, E. *et al.* Circulating MicroRNAs as Biomarkers of Gestational Diabetes Mellitus: Updates and Perspectives. *Int. J. Endocrinol.* **2018**, 6380463 (2018).

53. Fan, W., Pang, H., Xie, Z., Huang, G. & Zhou, Z. Circular RNAs in diabetes mellitus and its complications. *Front. Endocrinol.* **13**, 885650 (2022).

54. Efthymiou, A. *et al.* Isolation and freezing of human peripheral blood mononuclear cells from pregnant patients. *STAR Protoc* **3**, 101204 (2022).

55. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).

56. Stroukov, W. *et al.* OMIP-090: A 20-parameter flow cytometry panel for rapid analysis of cell diversity and homing capacity in human conventional and regulatory T cells. *Cytometry A* **103**, 362–367 (2023).

57. Finak, G. *et al.* MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278 (2015).

58. Papaioannou, T. G. *et al.* Mean arterial pressure values calculated using seven different methods and their associations with target organ deterioration in a single-center study of 1878 individuals. *Hypertens. Res.* **39**, 640–647 (2016).

59. Mensah, N. (2026). Single-cell transcriptomics identifies regulatory T cell heterogeneity in Gestational Diabetes Mellitus [Data set]. Zenodo. https://doi.org/10.5281/zenodo.18032075

60. Mensah, N. E. rutepo_gdm_treg. GitHub repository (2025). https://github.com/NMNS93/rutepo_gdm_treg

ED SUMMARY

Mensah, Efthymiou, Mureanu, and Monreal et al. examine RNA transcripts from over 40,000 CD4+ T cells to investigate regulatory T cell states in patients with gestational diabetes mellitus (GDM). GDM was associated with the abundance of memory CD4+ T cells and altered expression within naive and effector Treg subsets with aberrant genes serving to identify GDM from whole blood mRNA datasets.

Peer review information: *Communications Medicine* thanks Jacob Friedman, Peijie Zhou and Joan Camuñas-Soler for their contribution to the peer review of this work. A peer review file is available.