



## Data Article

# A high-resolution mass spectrometry based proteomic dataset of human regulatory T cells



Harshi Weerakoon<sup>a,b,c</sup>, John J. Miles<sup>d,e</sup>, Ailin Lepletier<sup>d,f</sup>,  
Michelle M. Hill<sup>a,g,\*</sup>

<sup>a</sup> Precision and Systems Biomedicine Laboratory, QIMR Berghofer Medical Research Institute, Herston, QLD, Australia

<sup>b</sup> School of Biomedical Sciences, The University of Queensland, Brisbane, QLD, Australia

<sup>c</sup> Department of Biochemistry, Faculty of Medicine and Allied Sciences, Rajarata University of Sri Lanka, Saliyapura, Sri Lanka

<sup>d</sup> Human Immunity Laboratory, QIMR Berghofer Medical Research Institute, Herston, QLD, Australia

<sup>e</sup> Australian Institute of Tropical Health and Medicine, James Cook University, Cairns, QLD, Australia

<sup>f</sup> Laboratory of Vaccines for the Developing World, Institute for Glycomics, Southport, QLD, Australia

<sup>g</sup> Centre for Clinical Research, Faculty of Medicine, The University of Queensland, Brisbane, QLD, Australia

## ARTICLE INFO

## Article history:

Received 18 September 2021

Revised 21 November 2021

Accepted 3 December 2021

Available online 6 December 2021

## Keywords:

Regulatory T cell

Conventional T cell

Proteomics

Tandem mass spectrometry

LC-MS/MS

## ABSTRACT

Regulatory T cells (Tregs) play a core role in maintaining immune tolerance, homeostasis, and host health. High-resolution analysis of the Treg proteome is required to identify enriched biological processes and pathways distinct to this important immune cell lineage. We present a comprehensive proteomic dataset of Tregs paired with conventional CD4<sup>+</sup> (Conv CD4<sup>+</sup>) T cells in healthy individuals. Tregs and Conv CD4<sup>+</sup> T cells were sorted to high purity using dual magnetic bead-based and flow cytometry-based methodologies. Proteins were trypsin-digested and analysed using label-free data-dependent acquisition mass spectrometry (DDA-MS) followed by label free quantitation (LFQ) proteomics analysis using MaxQuant software. Approximately 4,000 T cell proteins were identified with a 1% false discovery rate, of which approximately 2,800 proteins were consistently identified and quantified in all the samples. Finally, flow cytometry with a monoclonal antibody was used to validate the elevated abundance of the protein phosphatase CD148 in Tregs. This proteomic dataset serves as a reference point for future mechanistic and clinical T cell immunology and

\* Corresponding author.

E-mail address: [michelle.hill@qimrberghofer.edu.au](mailto:michelle.hill@qimrberghofer.edu.au) (M.M. Hill).

identifies receptors, processes, and pathways distinct to Tregs. Collectively, these data will lead to a better understanding of Treg immunophysiology and potentially reveal novel leads for therapeutics seeking Treg regulation.

© 2021 The Author(s). Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

## Specifications Table

|                                |  |
|--------------------------------|--|
| Subject                        | Immunology   |
| Specific subject area          | Regulatory T cells (Tregs) play a key role in directing adaptive immunity, particularly enforcing immune tolerance. However, few studies have examined the <i>ex vivo</i> Treg proteome. High-resolution and fully quantitative MS was used to profile the Treg and conventional CD4 <sup>+</sup> (Conv CD4 <sup>+</sup> ) T cell proteomes from healthy human blood. Bioinformatics yielded reliable, reproducible, and quantitative data. Differentially abundant proteins were identified between subsets, and flow cytometry was performed to validate elevated cell surface levels of CD148 in Tregs.   |
| Type of data                   | Tables and Figures.  |
| How data were acquired         | Orbitrap Fusion™ Tribrid™ (Thermo Fisher Scientific, USA) inline coupled to nano ACQUITY UPLC (Waters, USA) was used to acquire label-free proteomic data using data-dependent acquisition (DDA-MS). Raw and analysed data.  |
| Data format                    | CD3 <sup>+</sup> , CD4 <sup>+</sup> , CD25 <sup>high</sup> , CD127 <sup>low</sup> , FOXP3 <sup>+</sup> and CD3 <sup>+</sup> , CD4 <sup>+</sup> , CD25 <sup>-</sup> , FOXP3 <sup>-</sup> cells were identified as Tregs and Conv CD4 <sup>+</sup> , respectively. Peptides were separated using a 160 mins chromatographic gradient at 0.3 µl/min flow rate while ionised at 1900 V and 285 °C. MS1 ranged, and resolutions were 380 – 1500 m/z and 120,000 FWHM, respectively. Injection time for MS1 and MS2 were 50 ms and 70 ms, respectively. Fifteen ions were selected to trigger MS2 at 90 s dynamic exclusion. The total cycle time was two seconds. |
| Parameters for data collection |  |
| Description of data collection | Label-free proteomics data were acquired on peptide samples prepared from Treg and Conv CD4 <sup>+</sup> from peripheral blood mononuclear cells (PBMC) from three healthy volunteers (age 30-35 years) at the QIMR Berghofer Medical Research Institute (QIMRB), Brisbane, Australia.   |
| Data source location           | Raw proteomic data are available via ProteomeXchange [1].  |
| Data accessibility             | Repository name: ProteomeXchange<br>Data identification number: PXD022095<br>Direct URL to data:<br><a href="http://www.ebi.ac.uk/pride/archive/projects/PXD022095">http://www.ebi.ac.uk/pride/archive/projects/PXD022095</a><br>FTP Download:<br><a href="ftp://ftp.pride.ebi.ac.uk/pride/data/archive/2021/08/PXD022095">ftp://ftp.pride.ebi.ac.uk/pride/data/archive/2021/08/PXD022095</a>  |
| Related research article       | H. Weerakoon, J. Straube, K. Lineburg, L. Cooper, S. Lane, C. Smith, S. Alabbas, J. Begun, J.J. Miles, M.M. Hill, A. Lepletier, Expression of CD49f defines subsets of human regulatory T cells with divergent transcriptional landscape and function that correlate with ulcerative colitis disease activity., Clin. Transl. Immunol. 10 (2021) e1334.<br><a href="https://doi.org/10.1002/cti2.1334">https://doi.org/10.1002/cti2.1334</a> .   |

## Value of the Data

- This label-free quantitative dataset provides a high-resolution landscape of Treg and Conv CD4<sup>+</sup> T cell proteomes from primary T cell subsets in circulating human blood.
- T cell immunologists can use this dataset to explore Treg and Conv CD4<sup>+</sup> T cell proteomic landscapes and provide novel insights into the fundamental workings of T cells from different lineages.

- Clinical researchers can use this dataset to explore Treg protein dysfunction in different clinical settings. These studies will help correct dysfunction through new interventions and result in better patient management and outcomes across Treg associated diseases.
- Researchers in T cell biology and immunology can use the described methods to achieve further knowledge gain in the important field of immunoproteomics. The methodologies will also aid researchers working with small amounts of volunteer or patient material.

## 1. Data Description

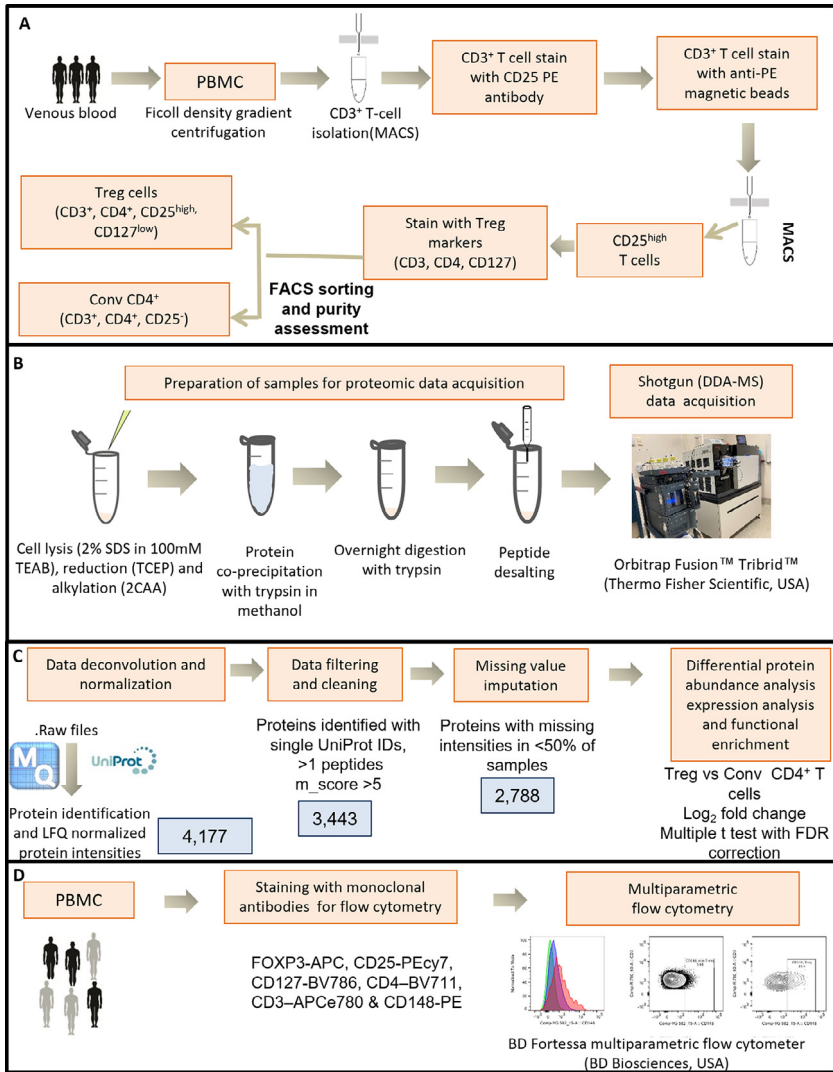
The proteomic dataset described here was generated from regulatory T cells (Tregs), and conventional CD4<sup>+</sup> T cells (Conv CD4<sup>+</sup>) purified from peripheral blood mononuclear cells (PBMC) of three healthy volunteers (age 30-35 years). The workflow across sample collection, protein processing, and data-dependent acquisition mass spectrometry (DDA-MS) analysis are depicted in Fig. 1 and detailed in the experimental design, materials, and methods section. The DDA-MS data were analysed using MaxQuant software (Release 1.6.0.16) [2] against UniProt/SwissProt human reviewed database [3] using 1% false discovery rate (FDR) cut-off to filter peptide spectral matches and the proteins. The .raw files obtained from DDA-MS analysis and the MaxQuant search results are available through the ProteomeXchange data repository (PXD022095) as detailed in Table 1. These publicly available raw data can be re-analysed using different parameters and databases to retrieve more specific information depending on the research interest. The consistency of proteomic data across donors are shown in Fig. 2, including MS/MS spectral count per sample (Fig. 2A) and per protein (Fig. 2B), the number of peptides per sample (Fig. 2C) and per protein (Fig. 2D), percentage of amino acid sequence coverage in peptides (Fig. 2E) and per sample (Fig. 2F). Of the identified total proteins (n=4,177), 92% were identified with single UniProt IDs in UniProt/SwissProt database (Fig. 2G). A similar number of proteins was identified for each donor in both the Treg and Conv CD4<sup>+</sup> T cells (Fig. 2H).

Normalised protein intensity values obtained from MaxLFQ [4] in MaxQuant were then used in differential abundance analysis to identify proteins enriched in Tregs relative to paired Conv CD4<sup>+</sup>. First, data cleaning retrieved proteins with (i) at least two unique razor peptides, (ii) an m-score of five and (iii) less than 50% missing intensity values across samples. The missing intensity values were imputed using the maximum likelihood estimate algorithm (R package) [5]. Next, differentially abundant proteins were identified using multiple t-test with FDR correction from a published two-stage linear step-up procedure [6] (Table S1, Fig. 1C). Analysis

**Table 1**

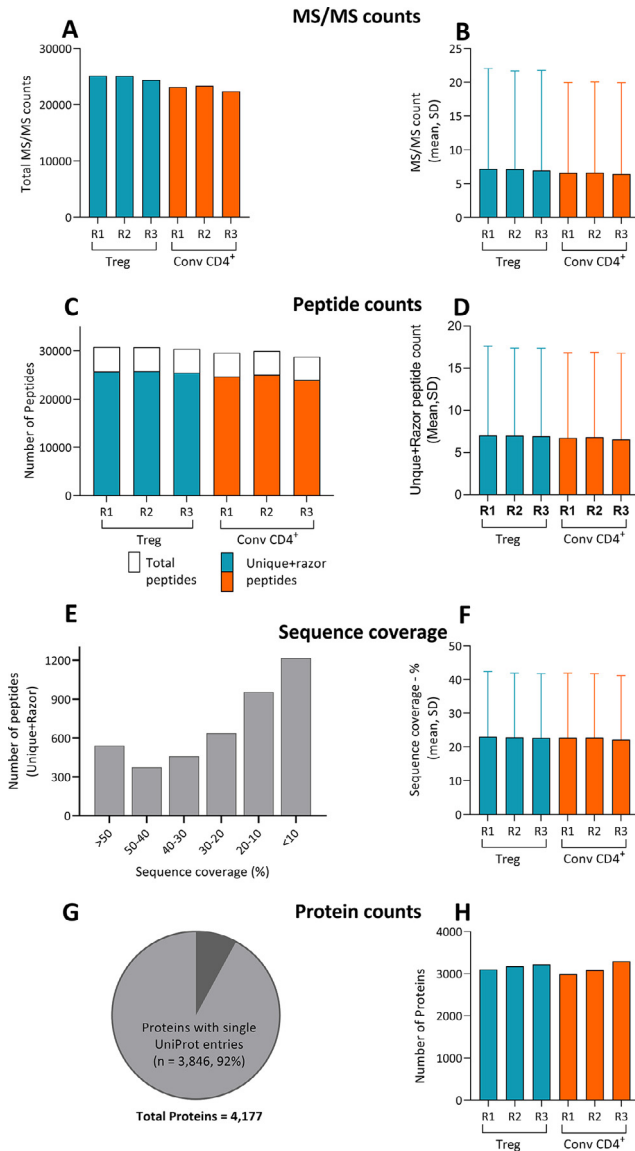
Description of the data files deposited in ProteomeXchange data repository under the data identification number of PXD022095.

|    | Title of the file/folder                | Description  |
|----|---|--|
| 1  | Rep1_Treg.raw                           | .raw file of Treg cells - Replicate 1  |
| 2  | Rep2_Treg.raw                           | .raw file of Treg cells - Replicate 2  |
| 3  | Rep3_Treg.raw                           | .raw file of Treg cells - Replicate 3  |
| 4  | Rep1_nonTreg.raw                        | .raw file of Conv CD4 <sup>+</sup> T cells - Replicate 1   |
| 5  | Rep2_nonTreg.raw                        | .raw file of Conv CD4 <sup>+</sup> T cells - Replicate 2   |
| 6  | Rep3_nonTreg.raw                        | .raw file of Conv CD4 <sup>+</sup> T cells - Replicate 3   |
| 7  | search.zip                              | MaxQuant output files resulted from the analysis of the above .raw files against UniProt/SwissProt human reviewed proteome |
| 8  | parameters.txt                          | Parameters used in the data analysis through MaxQuant search engine  |
| 9  | human_proteome_reviewed_25102017.fasta  | UniProt/SwissProt proteome database used in the analysis   |
| 10 | Treg_nonTreg_protein_quantification.txt | MaxQuant output files giving the protein quantification data and LFQ normalised protein intensities                        |



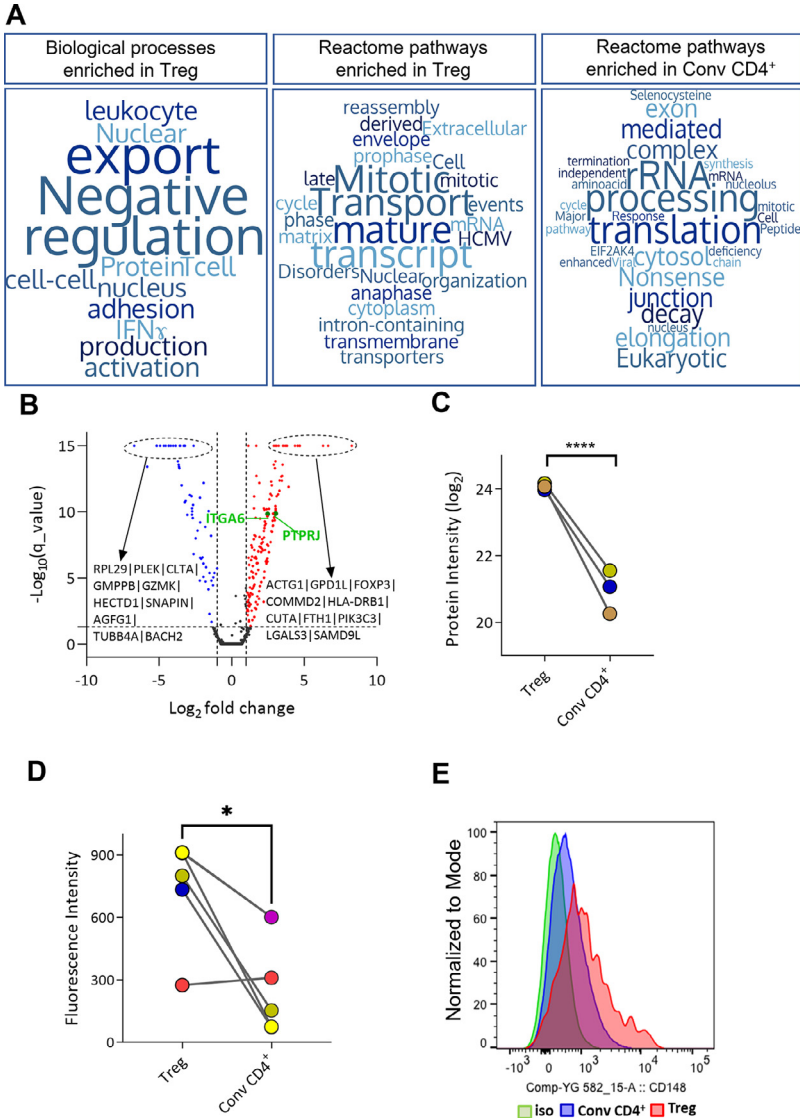
**Fig. 1.** Study workflow. **A.** Workflow for sorting Tregs and Conventional (Conv) CD4<sup>+</sup>. PBMC obtained from three volunteers underwent two rounds of magnetic-activated cell sorting (MACS) followed by flow cytometry based cell sorting (FACS), yielding Tregs and Conv CD4<sup>+</sup> at high purity. **B.** Proteomic sample preparation and mass spectrometry (LC-MS/MS) analysis. Peptide samples were prepared from whole cell lysate using protein co-precipitation with trypsin in methanol. The resulting tryptic peptides were desalted before LC-MS/MS analysis on Orbitrap Fusion Tribrid inline coupled to nano ACQUITY UPLC. **C.** DDA-MS data were deconvoluted using the MaxQuant search engine against the UniProt/SwissProt human proteome. Differential expression analysis was performed on only high-quality label-free protein intensity data. **D.** Orthogonal validation of CD148 (*PTPRJ*) enrichment in Treg cells.

returned 227 differentially abundant proteins between Treg and Conv CD4<sup>+</sup> T cells, with 157 (69% of total) proteins significantly enriched in Tregs. The enriched biological processes in each cell type were determined using gene ontology (GO) and pathway analysis comprising Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis [7], and Reactome pathways [8]. Functional enrichment analysis using the STRING network [9] revealed no enriched biological processes for Conv CD4<sup>+</sup> T cells, while Treg cells were enriched with negative regulation of interferon-gamma



**Fig. 2.** Summary of DDA-MS proteomic dataset in human Tregs. Label-free DDA-MS data were analyzed using MaxQuant search engine against UniProt/SwissProt human reviewed proteome database. MS/MS spectral data determined at 1% FDR were selected for further analysis across Tregs (blue) and Conv CD4<sup>+</sup> (orange). **A.** The number of MS/MS spectra detected in each sample. **B.** Mean MS/MS spectra per protein per sample. **C.** Total number of peptides and unique+razor peptides detected in each sample. **D.** Mean unique+razor peptides per protein per sample. **E.** Percentage of amino acid sequence coverage from the peptide dataset (unique+razor). **F.** Mean percentage of amino acid sequence coverage per peptide per sample (unique+razor). **G.** Total number of proteins quantified with single or multiple UniProt entries. Here, 92% of proteins had single entries (light grey), and 8% of proteins had multiple peptide entries (dark grey). **H.** The number of proteins quantified per sample. Error bars with standard deviation are shown.

production and T cell activation and leukocyte cell-cell adhesion and nuclear protein export (Fig. 3A). Further, extracellular matrix organisation, mitotic prophase and nuclear envelope re-assembly were the most enriched pathways in Tregs. Conv CD4<sup>+</sup> showed enrichment of viral mRNA translation, selenocysteine synthesis and peptide chain elongation pathways. Detailed



**Fig. 3.** Tregs and Conv CD4<sup>+</sup> T cell exhibit divergent proteomes. **A.** Word clouds of biological processes and reactome pathways enriched in Treg versus Conv CD4<sup>+</sup>. **B.** Volcano plot showing differential protein abundance of proteins between Treg and Conv CD4<sup>+</sup>, using  $q < 0.05$  and  $\log_2FC > 1$  or  $-1$  as cut-off (dotted lines). When comparing Conv CD4<sup>+</sup>, proteins significantly upregulated in Treg (red) and downregulated (blue) are shown, along with CD49f (*ITGA6*) and CD148 (*PTPRJ*) which are shown in green. Gene lists of the top 10 most upregulated and top 10 most downregulated proteins are shown. **C.** Proteomic data quantifying CD148 levels in Treg and Conv CD4<sup>+</sup>. Pairwise comparison showed higher abundance of CD148 in Treg cells in all donors ( $q < 0.0001$ , multiple t-test with false discovery rate correction). **D.** Flow cytometry data for cell surface CD148 in Treg and Conv CD4<sup>+</sup> T cells in six healthy donors denoted by different colors. Note some of the data points overlap (\* $p < 0.05$ , Mann Whitney U test). **E.** Representative histogram from one donor displaying the fluorescence intensity difference of CD148 between two T cell populations.

results, including enrichment score, the number of proteins detected, and gene nomenclature, are provided in Table S2.

To validate the proteomics dataset by orthogonal methods, we selected two Treg-enriched cell surface proteins with available monoclonal antibodies for flow cytometry, specifically CD49f (*ITGA6*) and CD148 (*PTPRJ*). The relative abundance of these and other proteins are shown in volcano plot (Fig. 3B). We have recently reported the flow cytometry and functional validation of CD49f in Tregs [10]. CD148 is a protein tyrosine phosphatase that regulates T cell receptor (TCR) signalling through Src family kinases (SFK), and has dual inhibitory/activatory functions and immune cell-specific patterning [11]. In agreement with the markedly enriched by DDA-MS ( $\text{Log}_2\text{FC}=3.02$ ,  $p < 0.001$ , Fig. 3C), flow cytometry validation confirmed CD148 enrichment in Tregs (Fig. 3D and 3E).

## 2. Experimental Design, Materials, and Methods

### 2.1. Experimental design

The experimental phases are depicted in Fig. 1, including T cell isolation (Fig. 1A), proteomic sample preparation and data acquisition (Fig. 1B), proteomics data deconvolution and analysis (Fig. 1C), and the orthogonal validation (Fig. 1D).

### 2.2. Human T cell isolation

Sequential magnetic-activated cell isolation (MACS) and flow cytometry-based cell sorting (FACS) were used to isolate Tregs and Conv CD4<sup>+</sup> T cells from PBMC to high purity (Fig. 1A).

#### 2.2.1. Isolation of human PBMC from venous blood

PBMC were isolated from fresh venous blood from volunteers at QIMRB, Brisbane, Australia. Ficoll-Paque™ density gradient medium (GE-Healthcare, USA) was used to separate PBMC from peripheral blood. Samples were centrifuged at 434xg for 20 mins with a no-brake deceleration. PBMC were then washed three times with Roswell Park Memorial Institute Medium (RPMI-1640) medium.

#### 2.2.2. MACS purification of T cell subsets

CD3<sup>+</sup> T cells next purified from PBMC ( $n = 3$ ) using MACS. Here, a human pan T cell isolation kit (Miltenyi Biotec, Germany) was used for negative cell selection as per the manufacturer's instructions. The unlabeled CD3<sup>+</sup> T cell flow-through was collected and then moved to another round of MACS cell isolation. Here, purified CD3<sup>+</sup> T cells were labelled with CD25-PE (BioLegend, USA) for 20 mins at 4°C, washed with cold MACS buffer (Miltenyi Biotec, Germany) and then labelled with anti-PE magnetic beads (Miltenyi Biotec, Germany) and purified per the manufacturer's instructions using LS columns (Miltenyi Biotec, Germany). CD25<sup>high</sup> (column bound) and CD25<sup>low</sup> (flow-through) T cells were collected, washed in R10 medium (RPMI-1640 containing 10% FCS) and transferred into 5 ml tubes for FACS-based sorting. This dual MACS method reduces FACS time and increases sort yields.

#### 2.2.3. FACS purification of Tregs and Conv CD4<sup>+</sup> T cells

MACS sorted CD25<sup>high</sup>, and CD25<sup>low</sup> T cells were next stained with LIVE/DEAD® Fixable Aqua (Life Technologies, USA), CD3-APCe780 (eBioscience, Thermo Fisher, Scientific, USA), CD4-VB711 and CD127-BV786 (BD Biosciences, USA) for 20 mins at 4°C, washed with cold FACS buffer three times for sorting. Single cells were sorted on a BD FACSAria III (BD Biosciences, USA) gated using FSC-W and FSC-H, and non-viable cells were dumped. In sorting, first the viable single cells were gated to select CD3<sup>+</sup> and CD4<sup>+</sup> T cell population. Of them Treg CD25<sup>high</sup>, CD127<sup>low</sup> cells were



sorted as Treg cells while CD25<sup>-</sup> cells were sorted as Conv CD4<sup>+</sup> T cells [10]. Low flow rates were used to increase yield through a reduction in stream abort events. Approximately 10<sup>6</sup> T cells were sorted for Treg and Conv CD4<sup>+</sup> in tandem for each sample. Each sorted T cell sample was washed three times with phosphate buffered saline (pH = 7.2) and then lysed in 100  $\mu$ l lysis buffer containing 1% sodium dodecyl sulphate (Biorad, USA) in 100 mM Triethylammonium bicarbonate (Sigma-Aldrich, USA) and 1 x Roche complete protease inhibitor cocktail (Sigma-Aldrich, USA) and stored at -80°C for proteomic sample preparation.

#### 2.2.4. Protein extraction and trypsin digestion

Each cell lysate was next thawed on ice, and 200 ng of ovalbumin (Sigma-Aldrich, USA) was added as an internal standard. The amount of protein in each cell lysate was quantified at a wavelength of 562 nm using Pierce bicinchoninic acid (BCA) protein assay (Thermo Fisher Scientific, USA), following the manufacturer's instructions. For trypsin digestion, 20  $\mu$ g protein from each sample was reduced with 10 mM of Tris (2-carboxyethyl) phosphine hydrochloride (Thermo Fisher Scientific, USA) at 60°C for 30 mins followed by alkylation with freshly prepared 40 mM chloroacetamide (Sigma, USA) at 37°C in the dark for 45 mins in a volume of 100  $\mu$ l. Detergents in the samples were then removed using protein co-precipitation with trypsin in methanol [12,13] using 1  $\mu$ l of 1  $\mu$ g/ $\mu$ l sequencing grade modified porcine trypsin (Promega, USA) and 1 ml of 100% cold (-20°C) chromAR grade methanol (Honeywell Research Chemicals, USA) to obtain a final lysate: methanol ratio of 1:10. Samples were mixed thoroughly by vortexing and kept overnight at -20°C for protein precipitation. The next day, samples were centrifuged for 15 mins at 16,100xg at 4°C, and the supernatant was aspirated carefully without disturbing the protein pellet. The protein pellets were washed (15 mins at 16,100xg at 4°C) two times with 1 ml of 90% and 100% cold methanol, respectively. Methanol was removed after the second centrifugation through careful pipetting and air drying for 2-3 mins. Protein pellets were then resuspended in 50 mM TEAB containing 5% acetonitrile (ACN, Honeywell research chemicals, USA) and mixed thoroughly by repeated pipetting and vortexing. Samples were next incubated at 37°C for 2 hrs with shaking. One  $\mu$ l of 1  $\mu$ g/ $\mu$ l (1:100) trypsin was then added, vortexed and incubated for 12 hrs at 37°C. The enzymatic reaction was next inhibited by adding 25  $\mu$ l of 5% formic acid (FA, Sigma, USA) and desalted using strata-x polymeric reversed-phase 10 mg/ml C18 cartridges (Phenomenex, USA). Conditioning, equilibrating, and washing steps in the desalting procedure were performed as follows. First, the C18 cartridges were conditioned through 2 washes of 100% ACN and equilibrated via 2 washes with 0.1% FA. Samples were then incubated in conditioned cartridges for 1 min to allow adsorption of ionised peptides. Cartridges were washed twice with 0.1% FA, and adsorbed peptides were eluted into new 1.5 ml microcentrifuge tubes using 500  $\mu$ l 80% ACN in 0.1% FA. In each of these steps 1 ml of the relevant solution was used. Peptide samples were finally dried in a speed vacuum at 35°C and stored at -80°C until mass spectrometry (LC-MS/MS) analysis.

#### 2.2.5. Peptide quantification and preparation for LC-MS/MS analysis

For LC-MS/MS analysis, each sample was resuspended in 20  $\mu$ l of 0.1% FA, and peptide quantification was performed using Pierce microBCA protein estimation kit (Thermo Fisher Scientific, USA) as per the manufacturer's instructions. After adjusting the peptide concentration 0.2  $\mu$ g/ $\mu$ l, 5  $\mu$ l of solution (1  $\mu$ g peptide) from each sample was injected into the LC-MS/MS system.

#### 2.2.6. LC-MS/MS proteomic sample analysis

Hybrid LC-MS/MS system; Orbitrap Fusion<sup>TM</sup> Tribrid<sup>TM</sup> (Thermo Fisher Scientific, USA) was used to analyse the peptide samples using DDA-MS. The LC-MS/MS parameters used in sample acquisition are detailed in Table 2.

#### 2.2.7. DDA-MS data deconvolution, protein identification and intensity normalisation

MaxQuant (Release 1.6.0.16) software was used in the analysis of .raw files from DDA-MS, by searching against the UniProt/SwissProt human reviewed proteome database containing 20,242 entries (October, 2017). Parameters used in peptide and protein identification in this study are



**Table 2**

Chromatographic and mass spectrometry parameters used in sample analysis.

| Parameter                               | Description/Settings  |
|---|---|
| Mass Spectrometer                       | Orbitrap Fusion™ Tribrid™ (Thermo Fisher Scientific, USA)   |
| LC system                               | nanoACQUITY UPLC (Waters, USA)  |
| Total LC gradient                       | 175 minutes   |
| Buffers                                 | A<br>0.1% FA<br>B<br>100% ACN + 0.1% FA   |
| LC gradient<br>(buffer B concentration) | 5% at 3 minutes, 9% at 10 minutes, 26% at 120 minutes, 40% at 145 minutes, 80% at 152 minutes, 80% at 157 minutes and 1% at 160 minutes |
| Trap column                             | Symmetry C18 trap, 2G VM trap (Waters, USA), 100Å, 5 µm particle size, 180 µm × 20 mm   |
| Column                                  | BEH C18 (Waters, USA), 130Å, 1.7 µm particle size, 75 µm × 200 mm   |
| Flow rate                               | 0.3 µl/ minutes   |
| Ion source                              | EASY-Max NG™ ion source (Thermo Fisher Scientific, USA)   |
| Ion spray voltage                       | 1900 V  |
| Heating temperature                     | 285°C   |
| Data acquisition method                 | DDA-MS  |
| MS1 mass range                          | 380 – 1500 m/z  |
| Injection time                          | 50 milliseconds   |
| Resolution                              | 120,000 FWHM  |
| Charge state                            | Rejecting +1, Selecting +2 to +7  |
| No. of ions selected to trigger MS2     | 15  |
| Fragmentation                           | Higher Energy C-trap Dissociation (HCD)   |
| Injection time                          | 70 milliseconds   |
| Resolution                              | 30,000 FWHM   |
| Dynamic exclusion                       | 90 seconds  |
| Cycle time                              | 2 seconds   |

**Table 3**

Parameters used in DDA-MS data analysis and label-free quantification using MaxQuant software and MaxLFQ.

| Parameter                            | Settings   |
|--------------------------------------|--|
| Digestive enzyme                     | Trypsin  |
| Maximum number of miscleavages       | 2  |
| Fixed modification                   | Carbamidomethylation (Cystiene)  |
| Variable modifications               | Oxidation (Methionine), Acetylation (Protein N-terminal)                                     |
| Precursor mass tolerance             | ± 20 ppm   |
| Product mass tolerance               | ± 40 ppm   |
| Maximum peptide charge               | +7   |
| Match between runs                   | Yes (0.7 minutes match time window and 20 minutes alignment time window)                     |
| Protein FDR                          | 0.01   |
| Peptide spectral match (PSM) FDR     | 0.01   |
| Minimum peptides                     | 1  |
| Peptide for protein quantification   | Unique + Razor   |
| Peptide selection for quantification | Unmodified peptides and only the peptides modified with oxidation and N terminal acetylation |
| No. of peptides for quantification   | ≥ 2  |

summarised in Table 3. MaxLFQ tool in MaxQuant software was used to normalise peptide and protein intensities for label-free quantification. For robustness, the internal control (chicken ovalbumin, UniProt ID P01012) and the common MS contaminant database inbuilt in MaxQuant software were used for benchmarking.

### 2.2.8. Data filtering and missing value imputation

Common contaminants and false-positive proteins detected during MaxQuant search were removed manually from further analysis. Proteins with more than one UniProt/SwissProt

accessions identified with  $\leq 1$  peptide and/or a  $m\_score$  of  $\leq 5$  were removed. The resulting protein lists were further examined for the percentage of proteins with unrecorded intensity values, and only the proteins with  $< 50\%$  missing protein intensity values were selected and imputed using maximum likelihood estimation (R package) and used for downstream analysis.

### 2.2.9. Differentially abundant protein identification and pathways analyses

LFQ normalised expression data of the selected proteins were used to identify the differentially abundant proteins. Following imputing the missing values in these selected proteins, differentially abundant proteins between Treg and Conv CD4<sup>+</sup> were analysed (Treg vs. Conv CD4<sup>+</sup>). The results were obtained as  $\text{Log}_2\text{FC}$  values, where the positive values represent highly abundant proteins in Treg cells. Multiple t-test with false discovery determination by a two-stage linear step-up procedure of Benjamini, Krieger and Yekutieli was used to calculate q values of differentially abundant proteins between Tregs and Conv CD4<sup>+</sup> T cells. Functional enrichment of the proteomic dataset was analysed using  $\text{Log}_2\text{FC}$  values using the bioinformatics software, STRING: Functional protein association network, version 11.5 to determine GO functional enrichment, KEGG and Reactome pathways with a cut-off FDR of 0.05%. The R package and GraphPad Prism (version 9.2.0 for Windows, GraphPad Software, San Diego, California USA) were used in bioinformatics analysis and graph generation. Word clouds were generated using the online tool (<https://www.wordclouds.com>, September 8th, 2021).

### 2.2.10. Validation of MS results by flow cytometry

DDA-MS data was validated using FACS. Here,  $10^6$  PBMC from six healthy individuals, including donors used in the proteomics study, were labelled with the forkhead box P3 (FOXP3) staining kit as per the manufacturer's instructions (Biolegend, USA), followed by surface staining with LIVE/DEAD® Fixable Aqua (Life Technologies, USA), CD3-APCe780 (eBioscience, Thermo Fisher, Scientific, USA), CD4-BV711 (BD Biosciences, USA) CD25-PEcy7 (BD Biosciences, USA), CD127-BV786 (BD Biosciences, USA) and CD148-PE (BioLegend, USA) for 20 mins at 4°C, and washed with cold FACS buffer three times. Single cells were analysed on a BD LSRFortessa (BD Biosciences, USA) using BD FACSDiva 8.0 software (BD Biosciences, USA). Gating comprised FSC-W and FSC-H with non-viable cells dumped. Flow data was analysed using FlowJo v10 (TreeStar, USA).

## Ethics Statement

Ethical clearance for this study was obtained from the QIMRB human research ethics committee (HREC, #P2058). Informed consent was obtained from all volunteers and the study adhered to the Declaration of Helsinki of 1975.

## CRediT Author Statement

**Harshi Weerakoon:** Data curation; Formal analysis; Validation; Investigation; Methodology; Writing-original draft; **John J. Miles:** Conceptualization; Methodology; Resources; Supervision; Writing-Review & Editing; Project administration; Funding acquisition; **Ailin Lepletier:** Conceptualization; Methodology; Supervision; Writing-Review & Editing; Project administration; **Michelle M. Hill:** Conceptualization; Methodology; Resources; Supervision; Writing-original draft; Project administration.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

## Acknowledgments

We thank the volunteers for blood samples for this study. We sincerely thank Dr Renee Richards and Dr Jason Mulvenna (QIMRB) for administrative and project support. HW was supported by an International Postgraduate Research Scholarship, The University of Queensland, Brisbane, Australia, and PhD Top-up Scholarship (QIMRB). JJM was supported by a National Health and Medical Research Council Career Development Level 2 Fellowship (1131732). The authors thank the staff from the flow cytometry facility and the analytical facility at QIMRB.

## Supplementary Materials

Supplementary material associated with this article can be found in the online version at doi:[10.1016/j.dib.2021.107687](https://doi.org/10.1016/j.dib.2021.107687).

## References

- [1] Y. Perez-Riverol, A. Csordas, J. Bai, M. Bernal-Llinares, S. Hewapathirana, D.J. Kundu, A. Inuganti, J. Griss, G. Mayer, M. Eisenacher, E. Pérez, J. Uszkoreit, J. Pfeuffer, T. Sachsenberg, S. Yilmaz, S. Tiwary, J. Cox, E. Audain, M. Walzer, A.F. Jarnuczak, T. Ternent, A. Brazma, J.A. Vizcaíno, The PRIDE database and related tools and resources in 2019: improving support for quantification data, *Nucleic Acids Res.* 47 (2019) D442–D450, doi:[10.1093/nar/gky1106](https://doi.org/10.1093/nar/gky1106).
- [2] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, *Nat. Biotechnol.* 26 (2008) 1367–1372, doi:[10.1038/nbt.1511](https://doi.org/10.1038/nbt.1511).
- [3] T.U. Consortium, UniProt: a worldwide hub of protein knowledge, *Nucleic Acids Res.* 47 (2019) D506–D515, doi:[10.1093/nar/gky1049](https://doi.org/10.1093/nar/gky1049).
- [4] J. Cox, M.Y. Hein, C.A. Lubner, I. Paron, N. Nagaraj, M. Mann, Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ, *Mol. Cell. Proteom.* 13 (2014) 2513–2526, doi:[10.1074/mcp.M113.031591](https://doi.org/10.1074/mcp.M113.031591).
- [5] RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>.
- [6] Y. Benjamini, A.M. Krieger, D. Yekutieli, Adaptive linear step-up procedures that control the false discovery rate, *Biometrika* 93 (2006) 491–507, doi:[10.1093/biomet/93.3.491](https://doi.org/10.1093/biomet/93.3.491).
- [7] M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes, *Nucleic Acids Res.* 28 (2000) 27–30, doi:[10.1093/nar/28.1.27](https://doi.org/10.1093/nar/28.1.27).
- [8] B. Jassal, L. Matthews, G. Viteri, C. Gong, P. Lorente, A. Fabregat, K. Sidiropoulos, J. Cook, M. Gillespie, R. Haw, F. Loney, B. May, M. Milacic, K. Rothfels, C. Sevilla, V. Shamovsky, S. Shorser, T. Varusai, J. Weiser, G. Wu, L. Stein, H. Hermjakob, P. D'Eustachio, The reactome pathway knowledgebase, *Nucleic Acids Res.* 48 (2020) D498–D503, doi:[10.1093/nar/gkz1031](https://doi.org/10.1093/nar/gkz1031).
- [9] D. Szklarczyk, A.L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N.T. Doncheva, J.H. Morris, P. Bork, L.J. Jensen, C. von Mering, STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets, *Nucleic Acids Res.* 47 (2019) D607–D613, doi:[10.1093/nar/gky1131](https://doi.org/10.1093/nar/gky1131).
- [10] H. Weerakoon, J. Straube, K. Lineburg, L. Cooper, S. Lane, C. Smith, S. Alabbas, J. Begun, J.J. Miles, M.M. Hill, A. Lepletier, Expression of CD49f defines subsets of human regulatory T cells with divergent transcriptional landscape and function that correlate with ulcerative colitis disease activity, *Clin. Transl. Immunol.* 10 (2021) e1334, doi:[10.1002/cti2.1334](https://doi.org/10.1002/cti2.1334).
- [11] O. Stepanek, T. Kalina, P. Draber, T. Skopцова, K. Svojr, P. Angelisova, V. Horejsi, A. Weiss, T. Brdicka, Regulation of Src family kinases involved in T cell receptor signaling by protein-tyrosine phosphatase CD148, *J. Biol. Chem.* 286 (2011) 22101–22112, doi:[10.1074/jbc.M110.196733](https://doi.org/10.1074/jbc.M110.196733).
- [12] K.A. Dave, M.J. Headlam, T.P. Wallis, J.J. Gorman, Preparation and analysis of proteins and peptides using MALDI TOF/TOF mass spectrometry, *Curr. Protoc. Protein Sci.* Chapter 16 (2011) Unit 16.13, doi:[10.1002/0471140864.ps1613s63](https://doi.org/10.1002/0471140864.ps1613s63).
- [13] H. Weerakoon, J. Potriquet, A.K. Shah, S. Reed, B. Jayakody, C. Kapil, M.K. Midha, R.L. Moritz, A. Lepletier, J. Mulvenna, J.J. Miles, M.M. Hill, A primary human T-cell spectral library to facilitate large scale quantitative T-cell proteomics, *Sci. Data* 7 (2020) 412, doi:[10.1038/s41597-020-00744-3](https://doi.org/10.1038/s41597-020-00744-3).