A shotgun proteomic dataset of human mucosal-associated invariant T cells

Author

Weerakoon, Harshi, Miles, John J, Hill, Michelle M, Lepletier, Ailin

Published

2024

Journal Title

Data in Brief

Version

Accepted Manuscript (AM)

DOI

10.1016/j.dib.2024.110786

Data in Brief xxx (xxxx) xxx

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

A shotgun proteomic dataset of human mucosal-associated invariant T cells

Harshi Weerakoon[a,b,c], John J. Miles[a,d], Michelle M. Hill[a,e], Ailin Lepletier[a,f,*]

[a] QIMR Berghofer Medical Research Institute, Herston, QLD, Australia
[b] School of Biomedical Sciences, The University of Queensland, Brisbane, QLD, Australia
[c] Department of Biochemistry, Faculty of Medicine and Allied Sciences, Rajarata University of Sri Lanka, Saliyapura, Sri Lanka
[d] Australian Institute of Tropical Health and Medicine, James Cook University, Cairns, QLD, Australia
[e] Centre for Clinical Research, Faculty of Medicine, The University of Queensland, Brisbane, QLD, Australia
[f] Institute for Biomedicine and Glycomics, Southport, QLD, Australia

ARTICLE INFO

Article history:
Received 13 June 2024
Revised 24 July 2024
Accepted 25 July 2024
Available online xxx

Dataset link: Human mucosal-associated invariant T (MAIT) cell proteome (Original data)

Keywords:
Shotgun proteomics
Human CD8$^+$ T cells
MAIT cells
Unconventional T cells

ABSTRACT

Mucosal-associated invariant T (MAIT) cells represent a unique unconventional T cell population important in eliciting immunomodulatory responses in a range of diseases, including infectious diseases, autoimmunity and cancer. This innate-like T cell subset predominantly express CD8 in humans. Unlike conventional CD8 T cells, which recognize peptide antigen presented by polymorphic major histocompatibility complex (MHC) molecules, MAIT cells are restricted by MR1, a non-polymorphic antigen-presenting molecule widely expressed in multiple tissues. Thus, identification of proteomic signature of MAIT cells in relation to conventional T cells is pivotal in understanding it's specific functional characteristics. The high-resolution dataset presents here comprehensively describes and compare the whole cell proteomes of MAIT (TCR Vα7.2$^+$CD161$^+$) and conventional T cells (TCR Vα7.2$^-$CD161$^-$) in humans. The dataset was generated using the proteomic samples prepared from matched T cell subsets sorted from peripheral blood mononuclear cells of three healthy volunteers. Peptides obtained from trypsin-digested cell lysates were analysed using Data-Dependent

* Corresponding author.
E-mail address: a.lepletierdeoliveira@griffith.edu.au (A. Lepletier).
Social media: @AilinLepletier (A. Lepletier)

https://doi.org/10.1016/j.dib.2024.110786

Please cite this article as: H. Weerakoon, J.J. Miles and M.M. Hill et al., A shotgun proteomic dataset of human mucosal-associated invariant T cells, Data in Brief, https://doi.org/10.1016/j.dib.2024.110786

Mass Spectrometry (DDA-MS). Label-free quantitation of DDA-MS data using MaxQuant and MaxLFQ software identified 4,442 proteins at a 1 % false discovery rate. Of them, 3680 proteins that were detected with single UniProt accession and a minimum of 2 unique or razor peptides were assessed to identify differentially abundant proteins between MAIT cells and conventional T cells (total T cells and $CD8^+$ T cells). The proteomic dataset comprises high-quality label-free quantitative proteomic data that can be used to compare the expression pattern of whole cell proteomes between the above-mentioned T cell populations. Further, this can be used as a reference proteome of human MAIT cells for the in-depth understanding of the MAIT cell behaviour in health and disease and to discover potential therapeutic targets to modulate MAIT cell function.

1 Specifications Table

Subject	Immunology
Specific subject area	Mucosal-associated invariant T (MAIT) cells are unconventional T cells important in human immunity. However, few studies have examined the primary human MAIT cell proteome. As majority of MAIT cells express $CD3^+$ and $CD8^+$, this dataset compares the proteome of MAIT cells with matched conventional T cells (total $CD3^+$ and $CD8^+$ T cells) circulating in the blood of healthy volunteers to establish differentially abundant proteins.
Type of data	Tables, Figures, Raw and Processed data
Data collection	Label-free shotgun data were generated from MAIT, and conventional T cells purified from the blood of three healthy volunteers. For sorting, $CD3^+$, $CD161^{high}$, and TCR $V\alpha7.2^+$ cells were gated as MAIT cells. Peptide samples obtained from trypsin-digested cell lysates were analysed using an Orbitrap Fusion™ Tribrid™ mass spectrometer (Thermo Fisher Scientific, USA) inline coupled to nanoACQUITY ultra-performance liquid chromatography system (Waters, USA). Peptides were separated using a 160-minute chromatographic gradient at 0.3 µl/min flow rate. Raw proteomic data were analysed and normalized using MaxQuant (Release 1.6.0.16) and MaxLFQ software respectively.
Data source location	Raw proteomic data are available via ProteomeXchange [1]. Data were collected from volunteers recruited at QIMR Berghofer Medical Research Institute -Brisbane, Queensland - Australia.
Data accessibility	Repository name: ProteomeXchange via PRIDE databaseData identification number: PXD052574 *(currently deposited as a private project. Upon acceptance of the manuscript the dataset will be made publicly available)*Direct URL to data: https://www.ebi.ac.uk/pride/archive/projects/PXD052574 Instructions for accessing these data: Reviewer can access the dataset by - logging into the PRIDE website using the following account details: Username: reviewer_pxd052574@ebi.ac.uk Password: WQpDETMiTFQ8 - Unique link: https://www.ebi.ac.uk/pride/review-dataset/3d4ae97c1c1d4edd9dda94c7a9824e23 Project accession: PXD052574Token: GMEuY9oWqinI

Please cite this article as: H. Weerakoon, J.J. Miles and M.M. Hill et al., A shotgun proteomic dataset of human mucosal-associated invariant T cells, Data in Brief, https://doi.org/10.1016/j.dib.2024.110786

1. Value of the Data

- The dataset generated by label-free shotgun proteomic approach allows comparing the expression of ~3600 proteins between human MAIT and conventional T cells (including total $CD3^+$ T cells, and $CD8^+$ T cells).
- Researchers can use this dataset to explore the phenotypic and functional characteristics of human MAIT cells and differentiate them from conventional T cells.
- As the dataset was generated from peripheral blood mononuclear cells (PBMC) collected from normal healthy adults, it can be used as an exploratory proteome when characterizing changes in MAIT cell proteome associated with multiple physiological and disease conditions.

2. Background

Mucosal-associated invariant T (MAIT) cells are evolutionary conserved, unconventional T cells characterized by the expression of semi-invariant T cell receptor (TCR) with a canonical TRAV1-2/TRAJ33 (Vα7.2/Jα33) that can recognize vitamin B metabolites, derived from some bacteria and fungi [2]. Their immunomodulatory functions are mainly associated with secretion of cytotoxic molecules [3] and cytokines [4–6]. In humans, MAIT cell are found in mucosal tissues [7,8], peripheral blood and liver [9,10]. MAIT cells represent ~ 10 % of circulating T cells and present a memory phenotype that allow them to rapidly respond to stimulus in a range of pathological conditions [9,11]. Since they were first described 15 years ago, omics analysis of human MAIT cells have evidenced their phenotypic and functional characteristics [12–14]. As MAIT cells are classified under the common T cell antigen, CD3, and primarily express CD8 in humans, which is a canonical marker for conventional cytotoxic T cells [7], describing the proteomic demarcation of MAIT cells in relation to conventional T cell populations is pivotal in identifying their unique functional and phenotypic properties.

3. Data description

The dataset presented in this article contains label-free quantitative proteomic data of human MAIT ($CD3^+$TCR Vα7.2$^+$CD161$^+$) and conventional T cells (including total $CD3^+$ and $CD8^+$ T cells bearing a TCR Vα 7.2$^-$CD161$^-$ phenotype) generated from Data-Dependent Acquisition approach (DDA-MS) using an Orbitrap Fusion™ Tribrid™ mass spectrometer (Thermo Fisher Scientific, USA) inline coupled to nanoACQUITY ultra performance liquid chromatographic (Waters, USA) system. The method used to isolate the cell populations, and the key steps for proteomic sample preparation and data acquisition are summarized in Fig. 1. Raw data were analyzed using MaxQuant software (Release 1.6.0.16) [15] against UniProt human-reviewed proteome while MaxLFQ was used to normalize the protein expression data for label-free quantification [16]. All raw and processed data as summarized in Table 1 are deposited and publicly available through the ProteomeXchange data repository (PXD052574).

The parameter file deposited with the dataset guides the researchers on the criteria used in the identification and quantification of peptides and proteins. The current analysis has led to the detection and quantification of 4440 protein groups at a peptide and protein false discovery rate (FDR) of 1 %. In future applications, the raw data can be reanalyzed with different parameters depending on the study objectives. Of the identified protein groups in the current analysis, 4110 (93 % of the total identified proteins) had a single UniProt accession name and 3680 (83 % of the total identified proteins) were detected with a minimum of 2 unique or razor peptides (Fig. 2A). To assess the quality of these selected proteins, the data were further analysed to identify the normalized protein intensity distribution (Fig. 2B), the number of peptide ions detected per protein (Fig. 2C), and the percentage of missing protein intensity values in each sample (Fig. 2D). Results of the principal component analysis and unsupervised hierarchical clustering of proteomic data generated from the different T cell subsets are shown in Fig. 2E

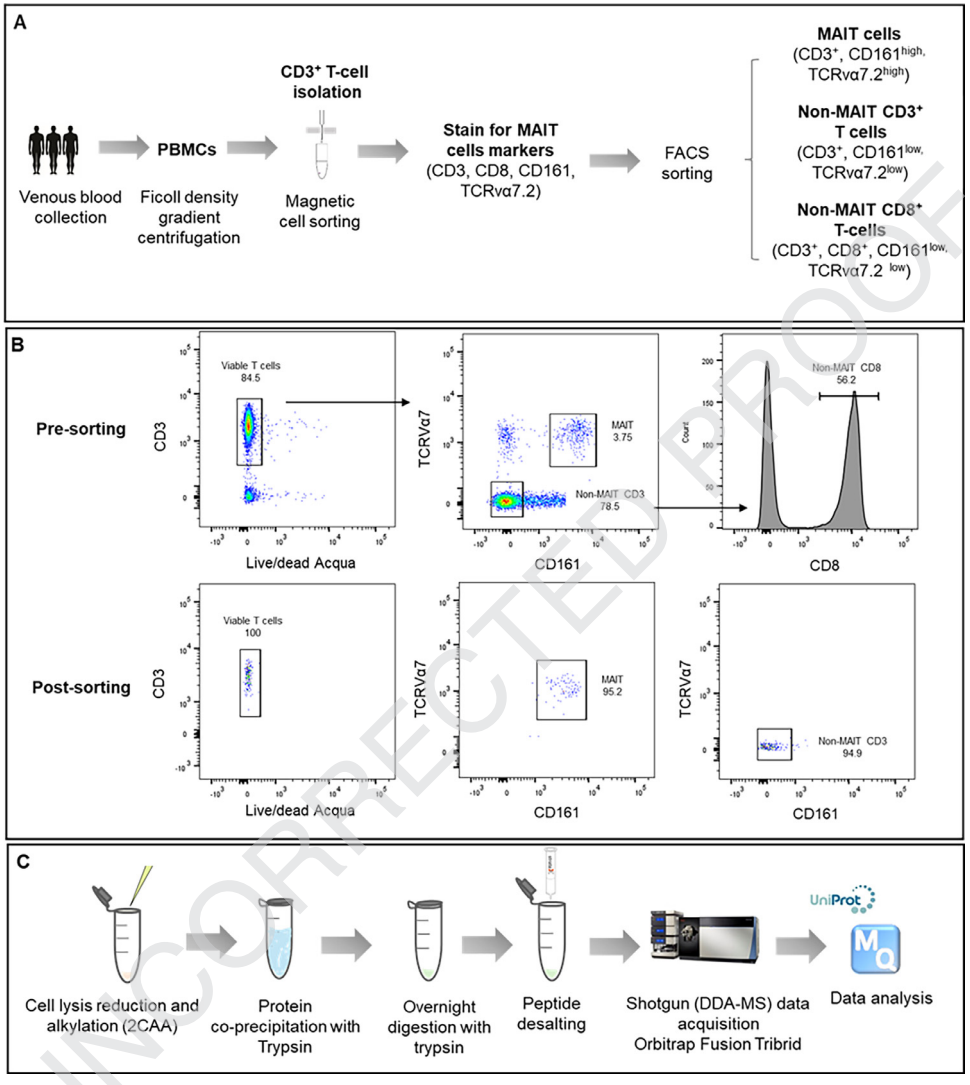

Fig. 1. The experimental design used to generate the proteomic data A. Key steps followed for isolating three T cell populations **B.** Gating strategy used in flow cytometry-based cell sorting to obtain MAIT cells and non-MAIT T cell populations at a high purity**.** Gating strategy is shown for cells pre and post-flow cytometry cell sorting. **C.** Key steps followed for the obtention of trypsin digested peptide samples and proteomic data acquisition.

and 2F, while the protein expression variation (fold change) across different donors is showed in Fig. 2G.

The current analysis removed the proteins with missing expression data in > 50 % of samples and quantified at m-score of < 5 when calculating the differential expression across the three T cell populations. As per the subcellular (Fig. 3A) and functional group (Fig. 3B) analysis, about half (~54 %) of these selected proteins ($n = 1566$) were mainly present in the cytoplasm and 28 % were classified as enzymes ($n = 798$). As expected, differential expression analysis revealed significant overexpression of CD161, one of the main signature surface proteins of MAIT cells (Fig. 3C). In total 243 (~8 %) and 285 (~10 %) proteins were differentially expressed (DE) in MAIT

Please cite this article as: H. Weerakoon, J.J. Miles and M.M. Hill et al., A shotgun proteomic dataset of human mucosal-associated invariant T cells, Data in Brief, https://doi.org/10.1016/j.dib.2024.110786

Table 1
Data files available through the ProteomeXchange data repository (PXD052574).

	File/folder	Description
1	Rep1_MAIT.raw	.raw file of MAIT cells – Biological Replicate 1
2	Rep2_MAIT.raw	.raw file of MAIT cells – Biological Replicate 2
3	Rep3_MAIT.raw	.raw file of MAIT cells – Biological Replicate 3
4	Rep1_nonMAIT_CD3.raw	.raw file of nonMAIT $CD3^+$ T cells – Biological Replicate 1
5	Rep2_nonMAIT_CD3.raw	.raw file of nonMAIT $CD3^+$ T cells - Biological Replicate 2
6	Rep3_nonMAIT_CD3.raw	.raw file of nonMAIT $CD3^+$ T cells – Biological Replicate 3
7	Rep1_nonMAIT_CD8.raw	.raw file of nonMAIT $CD8^+$ T cells – Biological Replicate 1
8	Rep2_nonMAIT_CD8.raw	.raw file of nonMAIT $CD8^+$ T cells - Biological Replicate 2
9	Rep3_nonMAIT_CD8.raw	.raw file of nonMAIT $CD8^+$ T cells – Biological Replicate 3
10	search.zip	MaxQuant ouput files resulted from the analysis of the above .raw files against UniProt/SwissProt human reviewed proteome
11	parameters.txt	Parameters used in the data analysis through MaxQuant, MaxLFQ search engine
12	human_proteome_reviewed_25102017.fasta	UniProt/SwissProt proteome database used in the analysis
13	MaxQuant_MaxLFQ_Output_protein group file.txt	MaxQuant ouput files giving the protein quantification data and LFQ normalised protein intensities

cells compared to conventional $CD3^+$ and $CD8^+$ T cells, respectively. The top 20 DE proteins in MAIT cells compared to $CD3^+$ and $CD8^+$ are shown in Fig. 3D and 3E, which evidence that the majority of DE proteins were over-expressed in MAIT cells (Fig. 3D and 3E). Further, the DE proteins and canonical pathways are summarized in Fig. 4 and Table 2 respectively.

4. Experimental design, materials and methods

4.1. Purification of primary human MAIT, $CD3^+$, and $CD8^+$ T cell populations

Human circulating MAIT cells were isolated from peripheral blood mononuclear cells (PBMCs) obtained from three healthy young volunteers aged between 30-35 years (2 males and 1 female). To isolate MAIT cells, first $CD3^+$ T cells were negatively enriched from PBMCs using a pan-human T cell isolation kit (Miltenyi Biotec, USA) and magnetic activated cell sorting. $CD3^+$ T cells were surface stained with live/dead Fixable Aqua (Life Technologies, USA), CD3-APCe780 (clone SK7; eBioscience, Thermo Fisher Scientific, USA), CD161-APC (clone HP-3G10; eBioscience, Thermo Fisher Scientific, USA), TCR Vα7.2-FITC (clone 3C10; Biolegend, USA) and CD8-Percp/cy5.5 (clone SK1, Biolegend, USA) by incubating the cells for 20 minutes at 4°C in dark. After washing three times with cold FACS buffer the stained cells were sorted using a FACS Aria III flow cytometer (BD bioscience, USA) to obtain $\sim 1\times10^6$ $CD3^+$, $CD8^+$, and MAIT cells from each donor. In the FACS sorting, $CD3^+$, $CD161^+$, and TCR Vα7.2$^+$ cells were sorted as MAIT cells while $CD3^+$, $CD161^-$ and TCR Vα7.2$^-$ and $CD3^+$, $CD8^+$, $CD161^-$ and TCR Vα7.2$^-$ were collected as $CD3^+$ and $CD8^+$ conventional T cells, respectively (Fig. 1A and B). Collected cells were washed three times with cold PBS, pelleted, and stored at – 80°C for proteomic sample preparation.

Please cite this article as: H. Weerakoon, J.J. Miles and M.M. Hill et al., A shotgun proteomic dataset of human mucosal-associated invariant T cells, Data in Brief, https://doi.org/10.1016/j.dib.2024.110786

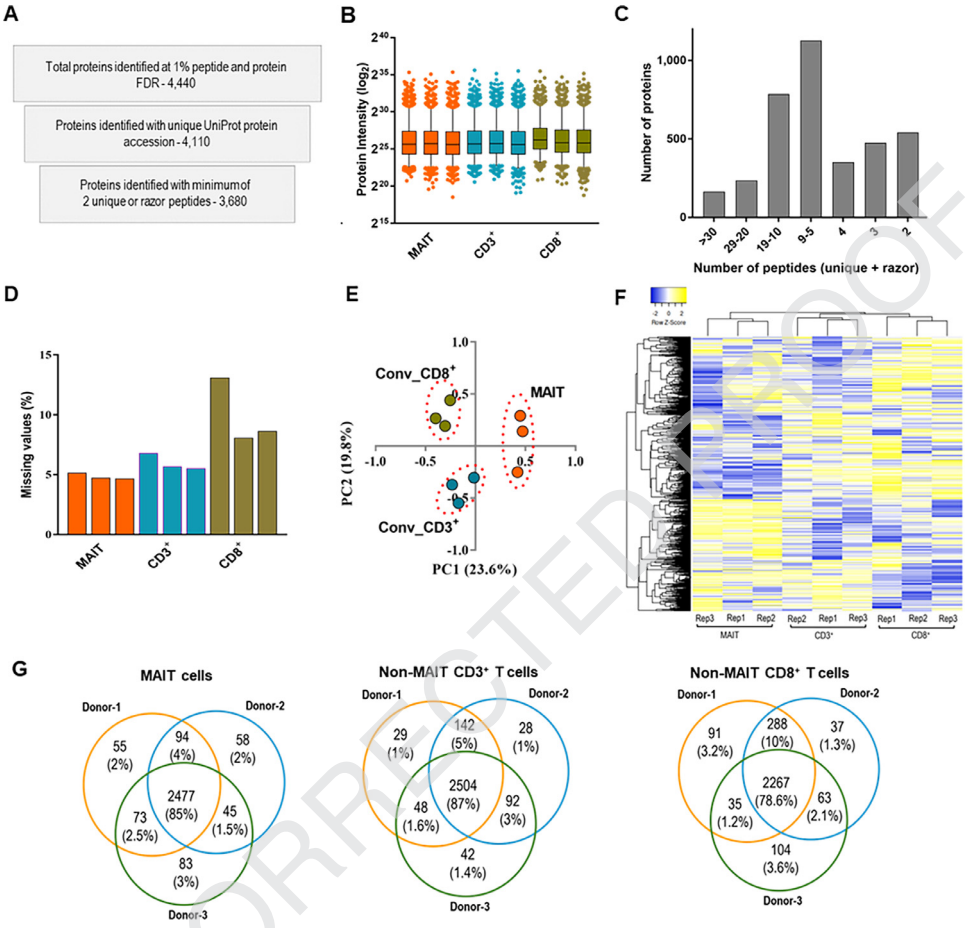

Fig. 2. An overview of the proteomic dataset obtained from MAIT and conventional T cell. A. Number of proteins obtained at different levels of data curation **B.** Distribution of normalized protein intensities across different samples (central lines and boxes represent means and 95 % confidence intervals respectively while whiskers are 2.5 to 97.5 percentiles) **C.** Number of proteins identified and quantified with different number of peptides **D.** Percentage of proteins with missing values in each (percentage from the total number of identified proteins) **E.** Principal component analysis of protein intensity data **F.** Heat map displaying the hierarchical clustering of protein intensity data of quantified proteins. **G.** Venn diagram show the variation of protein expression across three donors in three cell populations. Within each T cell group, common and differential proteins within each donor can be visualised.

4.2. Proteomic sample preparation

The steps used in the proteomic sample preparation and data acquisition are shown in Fig. 1C. The cell pellets were thawed and lysed in a lysis buffer composed of 2 % sodium dodecyl sulphate (Biorad, USA) in 100 mM triethylammonium bicarbonate (TEAB, Sigma-Aldrich, USA) and 1 x Roche complete protease inhibitor cocktail (Sigma-Aldrich, USA). Then 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. About 20 µg of protein from each cell lysate was reduced in 10 mM of tris (2-carboxyethyl) phosphine hydrochloride (Thermo Fisher Scientific, USA) at 60°C for 30 minutes and alkylated in

Please cite this article as: H. Weerakoon, J.J. Miles and M.M. Hill et al., A shotgun proteomic dataset of human mucosal-associated invariant T cells, Data in Brief, https://doi.org/10.1016/j.dib.2024.110786

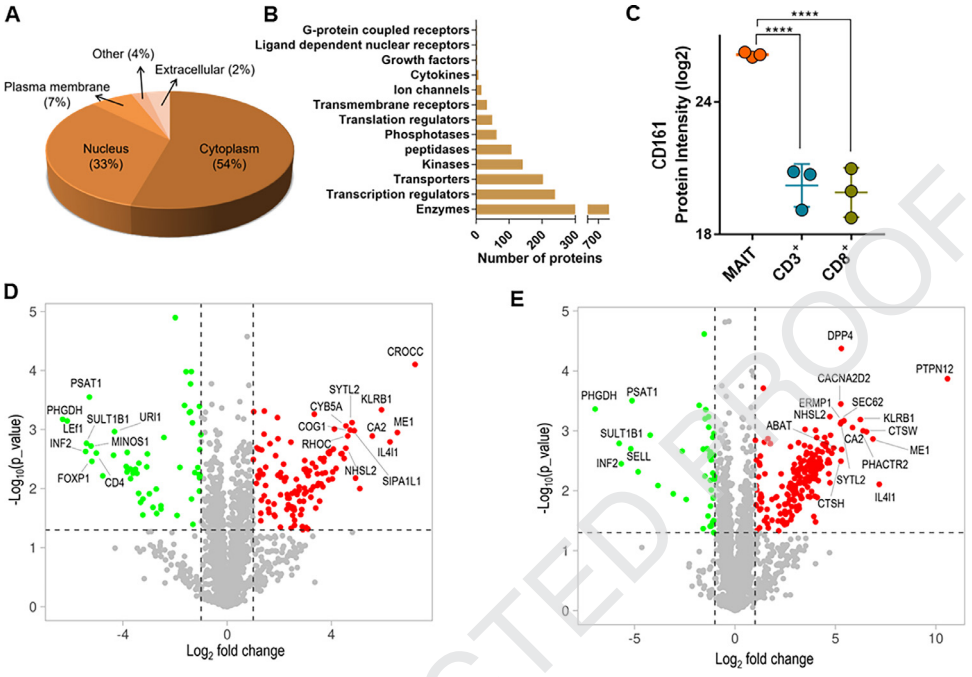

Fig. 3. Gene ontology and differential expression analysis of normalized protein intensity data. A. Distribution of proteins in different subcellular compartments as a percentage of all the selected proteins for differential expression analysis. B. Distribution of all the selected proteins for differential expression analysis in different functional groups (Qiagen, IPA). C. Expression of CD161 in three cell compartments as revealed by DDA-MS data (****q < 0.0001, multiple t-test with false discovery determination by two-stage linear step-up procedure of Benjamini, Krieger and Yekutieli) D. Volcano plots labelling names of the top 20 differentially expressed proteins in MAIT cells compared to non-MAIT $CD3^+$ T cells E. Volcano plots labelling names of the top 20 differentially expressed proteins in MAIT cells compared to non-MAIT $CD8^+$ T cells.

40 mM chloroacetamide (Sigma, USA) at 37°C in dark for 45 minutes. Cell lysates were then co-precipitated overnight with sequencing grade modified porcine trypsin (Promega, USA) at a trypsin: protein ratio of 1:100 in cold (-20°C) cold chromAR grade methanol (Honeywell Research Chemicals, USA) as described previously [17]. On the next day, samples with precipitated proteins were further cleaned by washing the pellet three times consecutively with 100 %, 90 %, and 100 %, cold chromAR grade methanol (Honeywell Research Chemicals, USA). Each centrifugation was performed for 15 min at 16,100xg at 4°C and the supernatants were aspirated carefully without disturbing the protein pellets. Resulted protein pellets were then resolubilized in 50 mM TEAB containing 5 % acetonitrile (ACN, Honeywell research chemicals, USA) and were incubated in a thermo-mixture at 37°C for 2 h at 600 rpm after adding 1 µl of 1 µg/µl sequencing grade trypsin. At the end of the incubation period, another 1 µl of 1 µg/µl (1:50) trypsin was added, vortexed to mix, and incubated overnight to obtain complete protein digestion. After 12 hours of digestion, the enzymatic reaction was inhibited by adding 25 µl of 5 % formic acid (Sigma-Aldrich, USA), and the resulting acidified tryptic digested peptides were desalted using strata-x polymeric reversed phase 10 mg/ml C18 cartridges (Phenomenex, USA). Desalted peptides were dried using a speedVac vacuum concentrator (Thermo Fisher Scientific, USA) at 35°C and stored at -80°C until tandem mass spectrometry (LC-MS/MS) based proteomic analysis.

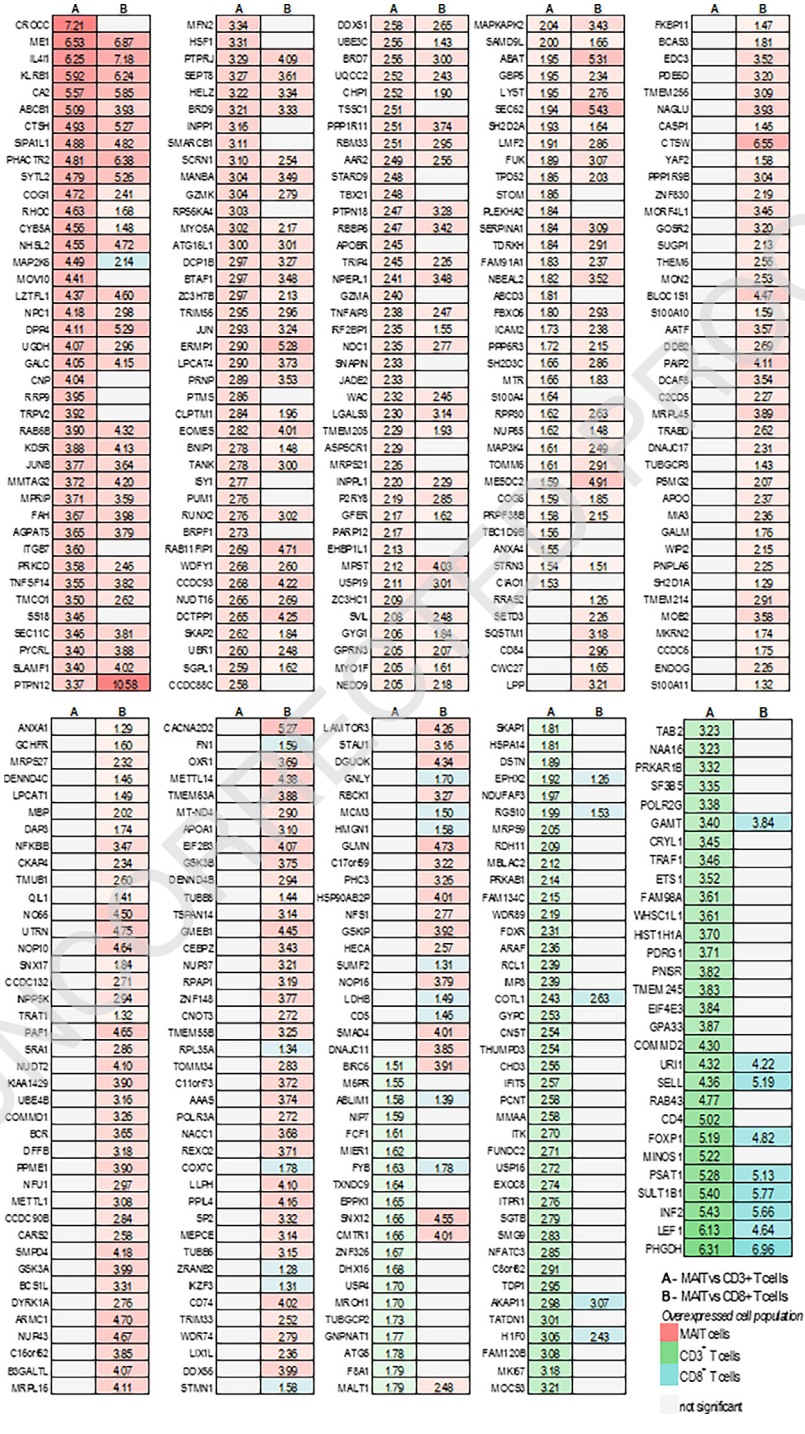

Fig. 4. Differentially expressed proteins in MAIT cells compared to $CD3^+$ and $CD8^+$ T cell populations. Proteins which showed $\pm\log_2 fc \geq 1$ at q value of 0.05 were considered as differentially expressed.

Table 2
Canonical pathways enriched by differentially expressed proteins in MAIT cells compared to Non-MAIT T cells.

	Ingenuity Canonical Pathways	Log_2 fold-change	Over-expressed	Under-expressed	Quantified proteins
MAIT cells vs non-MAIT $CD3^+$ T cells	B cell receptor signalling	1.4	9/192 (5 %)	1/192 (1 %)	MAP2K6, RRAS2, JUN, INPPL1, GSK3A, MAP3K4, INPP5K, GSK3B, MALT1, NFKBIB
	14-3-3-mediated signalling	1.4	8/131 (6 %)	0/131 (0 %)	RRAS2, JUN, TUBB6, TUBB8, PRKCD, EDC3, GSK3A, GSK3B
	CD40 signalling	1.4	5/79 (6 %)	1/79 (1 %)	MAP2K6, TANK, JUN, TNFAIP3, MAPKAPK2, NFKBIB
	TNFR2 signalling	1.4	4/30 (13 %)	0/30 (0 %)	TANK, JUN, TNFAIP3, NFKBIB
	Regulation of IL-2 in activated and anergic T-cells	1.4	6/80 (8 %)	0/80 (0 %)	RRAS2, JUN, CHP1, SMAD4, MALT1, NFKBIB
	Protein kinase A signalling	1.39	11/401 (3 %)	3/401 (1 %)	CHP1, PTPN18, PPP1R11, GSK3A, PTPN12, AKAP11, PTPRJ, PRKCD, SMAD4, LEF1, H1F0, GSK3B, NFKBIB, PDE6D
	TNRF2-mediated oxidative stress response	1.38	8/193 (4 %)	1/193 (1 %)	MAP2K6, DNAJC17, RRAS2, JUN, PRKCD, JUNB, GSK3B, SQSTM1, DNAJC11
	3-phosphoinositide degradation	1.38	7/157 (4 %)	1/157 (1 %)	PTPRJ, PIP4P1, NUDT16, EPHX2, INPPL1, INPP5K, NUDT2, PTPN12
MAIT cells vs non-MAIT $CD8^+$ T cells	CD40 signalling	1.44	5/79 (6 %)	1/79 (1 %)	MAP2K6, TANK, JUN, TNFAIP3, MAPKAPK2, TRAF1
	TNFR2 signalling	1.44	3/30 (10 %)	1/30 (3 %)	TANK, JUN, TNFAIP3, TRAF1
	Thiosulfate disproportionation III (Rhodanese)	1.44	2/3 (66 %)	0/3 (0 %)	MPST, MOCS3
	Protein kinase A signalling	1.44	6/401 (1 %)	7/401 (2 %)	NFATC3, CHP1, PTPN18, PPP1R11, ITPR1, PTPN12, AKAP11, HIST1H1A, PTPRJ, PRKCD, PRKAR1B, LEF1, H1F0
	CD28 signalling in Th cells	1.42	2/132 (2 %)	5/132 (4 %)	JUN, NFATC3, CD4, CHP1, ITPR1, MALT1, ITK

4.3. DDA-MS data acquisition

LC-MS/MS analysis of desalted peptide samples was performed on an Orbitrap Fusion™ Tribrid™ mass spectrometer (Thermo Fisher Scientific, USA) inline coupled to nanoACQUITY ultra performance liquid chromatographic (Waters, USA) system. From each sample, ~ 1 µg of peptides as quantified by micro-BCA (Thermo Fisher Scientific, USA) was loaded onto a Sym-

metry C18, 2G, VM (100Å, 5 µm particle size, 180 µm x 20 mm) trap column (Waters, USA) at a flow rate of 0.3 µL/min to separate the peptides on a BEH C18 (130Å, 1.7 µm particle size, 75 µm x 200 mm) column (Waters, USA). The mobile phase consisted of buffer A (0.1 % formic acid), and buffer B (100 % acetonitrile and 0.1 % formic acid) was used to create three consecutive linear gradients (buffer B, 5 %- 9 % between 3 and 10 min, 9 %-26 % between 10 and 120 min and 26 %-40 % between 120 and 145 min) to elute the peptides. After elution, the column was washed with buffer B at a concentration of 40 %- 80 % between 145 and 152 min, then holding it at 80 % until 157 min and at 1 % until 160 min. The eluted peptides were ionized using Nanospray Flex ion source (Thermo Fisher Scientific, USA) in which the ion spray voltage and heating temperature were held at 1.9 kV V and 285°C respectively. In DDA-MS acquisition, Chromeleon software (version 6.8, Dionex) included in Xcalibur software (version 3.0.63, Thermo Fisher Scientific, USA) was used to control the liquid chromatographic system. Peptide ions in the mass range of 380 – 1500 m/z were selected at 120,000 FWHM resolution to generate MS1 spectra. The mass spectrometer was controlled by Xcalibur software to operate "top speed" mode allowing automatic selection of positively charged (+2 to +7) top 15 peptides to trigger MS2. Higher Energy C-trap Dissociation (HCD) was used to fragment the selected peptide ions. In the acquisition of MS2 spectra, the resolution and dynamic exclusion time were set as 30,000 FWHM and 90 seconds respectively. The cycle time was 2 s.

4.3.1. Data processing and statistical analysis

MaxQuant (Release 1.6.0.16) software [15] was used to process the .raw files in which spectral data were searched against UniProt human-reviewed proteome database containing 20,242 entries (downloaded on 25th October 2017). MaxLFQ included in MaxQuant software was used to obtain the normalized label-free peptide and protein intensity data [16]. Trypsin-digested peptides with a maximum of 2 miscleavages were included in the analysis. Only carbamidomethylation of cystine (fixed modification), and oxidation of methionine and N terminal acetylation (variable modifications) peptide modifications were allowed. Precursor and product mass tolerance were set as ± 20 ppm and ± 40 ppm respectively to identify the peptides up to the maximum charge of +7. Only the peptide spectral matches and proteins detected at a 1 % of FDR were selected in which the proteins that were detected with at least one unique or razor peptide were quantified between runs.

In the downstream analysis, less reproducible proteins (expression data is missing for > 50 % of samples) and that were quantified at m-score of < 5 were removed from the final quantification and the missing values of the remaining proteins were imputed using maximum likelihood estimate (R package) [18]. In the statistical analysis, mean intensity values of each cell population were compared using multiple t-test with FDR determination by two-stage linear step-up procedure of Benjamini, Krieger and Yekutieli [19] to identify the differentially abundant proteins in MAIT cells compared to $CD3^+$ and $CD8^+$ T-cells. Expression fold change was obtained in the $\log_2$ fold change ($\log_2$fc) scale to depict proteins expressed at $\pm\log_2 fc \geq 1$ at q value < 0.05 as differentially abundant proteins. Qiagen ingenuity pathway analysis (QIAGEN Inc., https://digitalinsights.qiagen.com/IPA) was used to profile the subcellular localization, biological functions of the selected proteins, and the canonical pathways enriched by the differentially abundant proteins [20]. In IPA analysis, the *p*-value corrected for false discoveries in multiple comparisons using Benjamini-Hochberg (B-H p value) equation was used to set the cut-off, in which canonical pathways identified at or above 1.30 $-\log_{10}$ B-H p value (B-H p value = 0.05) were considered significant.

Limitations

As the present analysis aimed to characterize the proteome of human circulating MAIT cells, data were generated using the cell populations purified from only three young healthy volunteers. Proteome of MAIT cells can be different especially in elderly people and children as well as with different disease conditions However, that diversity is not included in the current dataset.

Please cite this article as: H. Weerakoon, J.J. Miles and M.M. Hill et al., A shotgun proteomic dataset of human mucosal-associated invariant T cells, Data in Brief, https://doi.org/10.1016/j.dib.2024.110786

As non-MAIT T cells were sorted from $CD3^+$ T cells, a contamination with non-MAIT unconventional T cell populations (e.g. Vδ2 $\gamma\delta$ T cells), which have some similarities to MAIT cells can be expected. Further, $CD3^+$ and $CD8^+$ conventional T cell populations contain T cells of different phenotypes (eg; naïve and memory) while MAIT cells predominantly have memory phenotype. Thus, some differentially expressed proteins will not relate to differences between MAIT and conventional T cells, but rather to differences in the ratio of naïve vs. memory cells. As indexed retention time (iRT) peptides (Biognosys AG, Switzerland) were not added to the samples during DDA-MS data acquisition, the use of this data to develop spectral libraries for data independent analysis (DIA) will be limited.

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; Project administration; Funding acquisition, **Michelle M Hill:** Conceptualization; Methodology; Resources; Supervision; Writing-original draft; Project administration. **Ailin Lepletier:** Conceptualization; Methodology; Supervision; Writing-Review & Editing; Project administration.

Data Availability

Human mucosal-associated invariant T (MAIT) cell proteome (Original data) (ProteomeXchange via the PRIDE database)

Acknowledgements

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 at QIMRB.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

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.

Please cite this article as: H. Weerakoon, J.J. Miles and M.M. Hill et al., A shotgun proteomic dataset of human mucosal-associated invariant T cells, Data in Brief, https://doi.org/10.1016/j.dib.2024.110786

[2] D.I. Godfrey, H.F. Koay, J. McCluskey, N.A. Gherardin, The biology and functional importance of MAIT cells, Nat. Immunol. 20 (2019) 1110–1128, doi:10.1038/s41590-019-0444-8.
[3] A. Kurioka, J.E. Ussher, C. Cosgrove, C. Clough, J.R. Fergusson, K. Smith, Y.-H. Kang, L.J. Walker, T.H. Hansen, C.B. Willberg, P. Klenerman, MAIT cells are licensed through granzyme exchange to kill bacterially sensitized targets, Mucosal Immunol 8 (2015) 429–440, doi:10.1038/mi.2014.81.
[4] M. Tomura, S. Maruo, J. Mu, X.Y. Zhou, H.J. Ahn, T. Hamaoka, H. Okamura, K. Nakanishi, S. Clark, M. Kurimoto, H. Fujiwara, Differential capacities of CD4+, CD8+, and CD4-CD8- T cell subsets to express IL-18 receptor and produce IFN-gamma in response to IL-18, J. Immunol. 160 (1998) 3759–3765, doi:10.4049/jimmunol.166.2.1097.
[5] M. Dusseaux, E. Martin, N. Serriari, I. Péguillet, V. Premel, D. Louis, M. Milder, L.Le Bourhis, C. Soudais, E. Treiner, O. Lantz, Human MAIT cells are xenobiotic-resistant, tissue-targeted, CD161hi IL-17-secreting T cells, Blood 117 (2011) 1250–1259, doi:10.1182/blood-2010-08-303339.
[6] S.-J. Kang, H.-M. Jin, E.J. Won, Y.-N. Cho, H.-J. Jung, Y.-S. Kwon, H.J. Kee, J.K. Ju, J.-C. Kim, U.J. Kim, H.-C. Jang, S.-I. Jung, S.-J. Kee, Y.-W. Park, Activation, impaired tumor necrosis factor-α production, and deficiency of circulating mucosal-associated invariant T cells in patients with scrub typhus, PLoS Negl. Trop. Dis. 10 (2016) e0004832, doi:10.1371/journal.pntd.0004832.
[7] E. Martin, E. Treiner, L. Duban, L. Guerri, H. Laude, C. Toly, V. Premel, A. Devys, I.C. Moura, F. Tilloy, S. Cherif, G. Vera, S. Latour, C. Soudais, O. Lantz, Stepwise development of MAIT cells in mouse and human, PLoS Biol. 7 (2009) e54, doi:10.1371/journal.pbio.1000054.
[8] M.C. Gold, R.J. Napier, D.M. Lewinsohn, MR1-restricted mucosal associated invariant T (MAIT) cells in the immune response to Mycobacterium tuberculosis, Immunol. Rev. 264 (2015) 154–166, doi:10.1111/imr.12271.
[9] O.-J. Lee, Y.-N. Cho, S.-J. Kee, M.-J. Kim, H.-M. Jin, S.-J. Lee, K.-J. Park, T.-J. Kim, S.-S. Lee, Y.-S. Kwon, N. Kim, M.-G. Shin, J.-H. Shin, S.-P. Suh, D.-W. Ryang, Y.-W. Park, Circulating mucosal-associated invariant T cell levels and their cytokine levels in healthy adults, Exp. Gerontol. 49 (2014) 47–54, doi:10.1016/j.exger.2013.11.003.
[10] A. Kurioka, L.J. Walker, P. Klenerman, C.B. Willberg, MAIT cells: new guardians of the liver, Clin. Transl. Immunol. 5 (2016) e98, doi:10.1038/cti.2016.51.
[11] I. Magalhaes, M. Solders, H. Kaipe, MAIT cells in health and disease, in: H. Kaipe, I. Magalhaes (Eds.), MAIT Cells. Methods Mol. Biol. (Clifton, N.J.), Humana, New YorkNY, 2020 p., doi:10.1007/978-1-0716-0207-2_1.
[12] J.R. Fergusson, K.E. Smith, V.M. Fleming, N. Rajoriya, E.W. Newell, R. Simmons, E. Marchi, S. Björkander, Y.H. Kang, L. Swadling, A. Kurioka, N. Sahgal, H. Lockstone, D. Baban, G.J. Freeman, E. Sverremark-Ekström, M.M. Davis, M.P. Davenport, V. Venturi, J.E. Ussher, C.B. Willberg, P. Klenerman, CD161 defines a transcriptional and functional phenotype across distinct human T cell lineages, Cell Rep. 9 (2014) 1075–1088, doi:10.1016/j.celrep.2014.09.045.
[13] J. Dias, E. Leeansyah, J.K. Sandberg, Multiple layers of heterogeneity and subset diversity in human MAIT cell responses to distinct microorganisms and to innate cytokines, Proc. Natl. Acad. Sci. U. S. A. 114 (2017) E5434–E5443, doi:10.1073/pnas.1705759114.
[14] B. Bulitta, W. Zuschratter, I. Bernal, D. Bruder, F. Klawonn, M. von Bergen, H.S.P. Garritsen, L. Jänsch, Proteomic definition of human mucosal-associated invariant T cells determines their unique molecular effector phenotype, Eur. J. Immunol. 48 (2018) 1336–1349, doi:10.1002/eji.201747398.
[15] 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.
[16] J. Cox, M.Y. Hein, C.A. Luber, I. Paron, N. Nagaraj, M. Mann, Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ, Mol. Cell. Proteomics. 13 (2014) 2513–2526, doi:10.1074/mcp.M113.031591.
[17] 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.
[18] RStudio Team, RStudio: Integrated Development for R, R Studio, PBC, Boston, MA, 2020 URL http://www.rstudio.com/ .
[19] 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.
[20] A. Krämer, J. Green, J. Pollard, S. Tugendreich, Causal analysis approaches in ingenuity pathway analysis, Bioinformatics 30 (2014) 523–530, doi:10.1093/bioinformatics/btt703.