nature portfolio

Corresponding author(s): Hsiang-Ying Lee

Last updated by author(s): Oct 15, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	1.Single-cell capture was achieved by the BD Rhapsody system. Whole transcriptom libraries were prepared according to the BD Rhapsody single-cell whole-transcriptome amplification workflow and sequenced using HiSeq Xten (Illumina) on 150 bp paired-end run. Bulk-RNA Sequencing libraries were prepared by Novogene, and 150 bp paired-end sequencing of each condition with two biological replicates was obtained in a HiSeq X-Ten (PE150, Illumina) by Novogene (Beijing). The peptide sample of proteomics sequencing were prepared by Jingjie PTM Biolab. The separated peptides were analyzed in Orbitrap Astral with a nano electrospray ion source. Acquisition methods in LC-MS/MS was data-independent acquisition(DIA).
Data analysis	ScRNA-seq: Raw data were processed using fastQ to filter adaptor sequences and remove low-quality reads. Cell barcode whitelist was identified by UMI tools. The UMI-based clean data were mapped to human reference genome (Ensemble V.91) using the STAR algorithm. UMI count matrices were generated for each sample, and imported into the Seurat R toolkit (V. 4.1.3). Cluster analysis was performed by Seurat. The Seurat package (with default settings) was used for normalization and scaling of the expression matrix. Mitochondrial contamination was regressed by adjusting the "vars.to.regress" parameter. To reduce expression matrix dimensionality, principal component analysis was performed based on 2,000 highly variable genes. Unsupervised cell clusters were acquired using a graph-based clustering approach (i.e., the top 20 principal components were selected with a resolution of 0.75), then visualized by uniform manifold approximation and projection (UMAP) or t-Distributed Stochastic Neighbor Embedding (t-SNE) dimensionality reduction. Clusters were annotated to known biological cell types according to the expression patterns of canonical markers. Marker genes in each cluster were identified using the FindAllMarkers function with the following criteria: log (fold change) > 0.25, min. pct > 0.25, and adjusted P-value < 0.05. Pathway analysis was used to identify significant pathways of DEGs according to the Kyoto Encyclopedia of Genes and Genomes (KEGG) and GSEA database. Bulk-RNA seq:

FASTQ files were generated by Novogene. Hisat2 (version 2.1.0) was used to align the paired-end raw data human reference genome, HTSeq was used to calculate the reads counts, and DESeq2 was used to identify the differential expression genes.
Micro-proteomics seq:
The DIA data were processed using DIA-NN search engine(v.1.8). Tandem mass spectra were searched against Homo_sapines_9606_SP_20231220.fasta concatenated with reverse decoy database.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The data in this manuscript have been deposited in the Genome Sequence Archive of the National Genomics Data Center. The assigned accession number of the submission is HRA006868. The raw data are available under controlled access due to data privacy laws related to patient consent for data sharing. The data should be used for research purposes only. According to the guidelines of GSA-human, all non-profit researchers are allowed access to the data, and the Principal Investigator of any research group can apply for the data following the guidelines at the GSA database portal (https://ngdc.cncb.ac.cn/gsa-human/). The response time for access requests is approximately 8 working days. Once access has been granted, the data will be available for download within one month. The user can also contact the corresponding author directly for enquiries. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (https://proteomecentral.proteomexchange.org) via the iProX partner repository with the dataset identifier PXD055495. Source data are provided with this paper.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	The PRCA participant contain two males. The healthy participants contain 3 males and 1 female. The multiple myeloma patients contains 3 males and 1 female.
Reporting on race, ethnicity, or other socially relevant groupings	Healthy donors ,multipe myeloma patients, and PRCA patient participants are all Han Chinese.
Population characteristics	The age of the participants ranged from 40-60 years-old. The PRCA donor characterized by absence or near absence of erythroblasts in the bone marrow and the absolute reticulocyte count is less than 10,000/mL.
Recruitment	Participants were recruited by Peking University People's Hospital and Peking Union Medical College Hospital. Since the age of two patients ranged from 55 to 59 at the time of sampling, we recruited healthy donors with an age range of 40-60 and without any hematologic disorders.
Ethics oversight	The study design and conduct in accordance with Chinese law and the criteria set by the declaration of Helsinki. The protocol concerning the use of patient samples in this study was approved by the Biomedical Research Ethics Committee of Peking University Health Science Center and People's Hospital.(approval number:2021-479), and informed consent was obtained from all participants.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes used of each experiment has been indicated in the figures, figure legends, main text. When it is possible, minimum of three biological replicates (except for bulk-RNA and Single cell RNA sequencing) has been used, standard and significance levels have been shown, The bulk-RNA assay were performed of two biological replicates, which are determined by general variability of the data in literature and the sample sizes used are sufficient to draw conclusions. Additionally, for single-cell RNA sequencing, we collected four healthy donor samples, but due to the rarity of the patients, we were only able to collect samples from one patient at two different time points.
Data exclusions	Data were not excluded from the analysis.

Replication	The number replicates per experiment has been described in the figures, figure legends, main text. The results have been reproduced as three biological replicates (except for bulk-RNA and Single cell RNA sequencing).
Randomization	Randomization is no applicable limited by rarity of patients. Any specimen preparation or selection from a large pool of cells/DNA/RNA has always been randomized.
Blinding	The healthy donor sample and PRCA sample were labeled as sample A, sample B before the experiment. The in vitro studies of drug function, erythroid differentiation, and bulk-RNA sequencing experiments were blinded. Blinding was not performed for single cell RNA sequencing due to availability of researchers.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	The antibody,clone, manufacture, and catalog number are mentioned below: Mouse anti-human CD3-BV650 (clone SK7,BD Horizon,Cat. #563999)(dilution 1:200) Mouse anti-human CD8a-FITC (clone HIT8a, BD Pharmingen, Cat. #555634)(dilution 1:200) Purified Mouse anti-CD28 (CD28.2, BD Biosciences, Cat. #555725)(dilution 1:1000) Mouse anti-human HLA-A, B, C antibody (W6/32, BioLegend, Cat. #311402)(dilution 1:200) Mouse anti-CD71 (FITC, OKT9, eBioscience, Cat.#11-0719-42)(dilution 1:200) Mouse anti-CD235a (APC, HIR2 (GA-R2), eBioscience, Cat.# 17-9987-42)(dilution 1:200)
Validation	Validation of antibodies was performed by the manufacturer, and is provided on their websites: https://www.biolegend.com/, https://www.bdbiosciences.com/, https://www.thermofisher.cn/cn/zh/antibody/product/CD235a-Glycophorin-A-Antibody-clone-HIR2-GA-R2-Monoclonal/17-9987-42, https://www.thermofisher.cn/cn/zh/antibody/product/CD71-Transferrin-Receptor-Antibody-clone-OKT9-OKT-9-Monoclonal/11-0719-42

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJEguidelines for publication of clinical research and a completedCONSORT checklist must be included with all submissions.

Clinical trial registration	*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*
Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*
Outcomes	*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Isolation of erythroid cells were sorted by fluorescence-activated cell sorting (FACS). Briefly, at the day 6 of erythroid differentiation of BMMCs, cell were spin -down and resuspend in staining buffer (phosphate buffered saline with 2% fetal bovine serum) and added following antibodies as per 1,000,000 cells: 5ul anti-CD71 (FITC, eBioscience, Cat.#11-0819-42) and 5ul anti-CD235a (APC, eBioscience, Cat.#17-9987-42). Cell were subsequently stained with cell surface antibodies for 20 minutes at room temperature protected from light. Cells were then washed with FACS buffer. Sorting were perform on BD FACSAria III. For the depletion of T cells, BMMCs were stained as per 1,000,000 cells: 5ul anti-CD3 (BV650, BD Horizon, Cat.#563999) and 5ul anti-CD8a (FITC, BD Pharmingen, Cat.#555634). The analytical flow cytometry was performed on BD Fortessa by the same staining methods.

Instrument

Flow cytometry data was collected on a BD LSRFortessa (BD Biosciences). Cell sorting was performed on a BD FACSAria (BD Biosciences).

Software

Flow cytometry data was collected using BDFACSDiva software (version 8.0) and was analyzed using FlowJo Software (version v10.8.1).

Cell population abundance

Post-sort purity was routinely >95% as determined by FACS.

Gating strategy

A broad morphology gate was first set using FSC-A/SSC-A. Single cells were gated using FSC-H/FSC-A and SSC-H/SSC-A. DAPI chancel was used to remove dead cells. Anti-human CD3 antibody was used to remove T cells. Erythroid cells for bulk-RNA sequencing were sorted using anti-human CD71 antibody.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.