nature portfolio

Corresponding author(s): Ramnik Xavier

Last updated by author(s): Oct 21, 2025

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	10X Cell Ranger v3 pipeline was used to align reads and generate count matrices, with the "cellranger mkfastq" and "cellranger count" commands, respectively. EmptyDrops function from the DropletUtils v1.6.1 R package was run on the unfiltered count matrices with ignore=10, retain=800 and lower=200 parameters with an FDR 1% cutoff. Barcodes with a total number of UMIs or genes < 200 were filtered out. Percent mitochondrial UMI cutoffs of 10% and 20% were used for the samples profiled with 10X v2 and v3 chemistry, respectively. Counts were normalized with a log(TP10K+1) transformation via the sc.pp.normalize_total function of Scanpy v1.8. Preprocessing and normalization steps were performed similarly for human scleroderma and mouse (control and MC903) scRNA-seq datasets. QC steps resulted in 29,689 and 9,133 high quality cell profiles in scleroderma and mouse datasets, respectively. Cells were counted and ~8,000-12,000 cells were loaded per channel of the 10x Genomics Chromium chips using the 3'- v3 chemistry for scRNA-seq or the 5'- v1.1 chemistry for single-cell TCR- enriched V(D)J profiling. Single cells were processed through the 10x Genomics Chromium Platform according to manufacturer's instructions (10x Genomics). Fastq, h5 and JSON files representing the reads, 5' gene expression and VDJ information were obtained with Cell Ranger (v2.0.2) V(D)J pipeline commands "cellranger mkfastq" and "cellranger vdj".
Data analysis	PCA was performed with the top 2,000 highly variable genes and a k-nearest neighbor (k-NN) graph was built with k=15 on 50 PCs using sc.pp.pca and sc.pp.neighbors functions of Scanpy, respectively. The k-NN graph was clustered with the Leiden community detection method to partition cells into coarse clusters using the sc.tl.leiden in Scanpy. Marker genes were found using one-vs-rest two-tailed Welch's t-test using sc.tl.rank_genes_groups in Scanpy. The harmonize() function from the PyTorch implementation of Harmony (https://github.com/lilab-bcb/harmony-pytorch96) was used with channel IDs as a batch covariate based on the PCA representation of cells for removing unwanted inter-sample and inter-individual variation. Negative binomial regression on raw counts using the diffxpy (https://github.com/theislab/diffxpy/releases, v0.1.7) Python package. '~ 1 + sex + mt_frac + log10_n_umis + disease_status' formula was used to compare healthy vs. lesional and healthy vs. non-lesional groups within each coarse and granular subset, while correcting for the fraction of mitochondrial UMIs, total number of UMIs and sex of the donor. Wald test was performed

to estimate the effects of disease status. The dirichreg function from the DirichletReg R package (v0.7) was used for testing the significance of differences in cell subset proportions across conditions for both scRNA-seq and deconvolved bulk RNA-seq. The effect of the "disease_status" variable was tested for significance using a likelihood ratio test, and FDR values were calculated from these p-values using the Benjamini-Hochberg method. FDR values are presented in the figures. The ggplot R package with the geom_boxplot2 function from the Ipaper R package was used for Box plot visualizations. geom_boxplot2 function was run with the "width.errorbar=0" argument. Bulk RNA-seq data19 were downloaded from Gene Expression Omnibus (GEO, accession GSE121212). Psoriasis samples were discarded. The ag.optimize function from the autogenes Python package98 was run on log(TPM+1)-transformed scRNA-seq expression values with the arguments "ngen=5000, seed=0, nfeatures=400, mode='fixed', offspring_size=100". Next, the ad.deconvole function was run with the NuSVR method on log(TPM+1) transformed bulk RNA-seq expression values. A multinomial logistic ridge regression classifier was used to classify cell profiles from the published scRNA-seq skin datasets, and the 5' scRNA-seq and scleroderma scRNA-seq datasets in this study. LogisticRegressionCV class from scikit-learn Python package was instantiated with parameters "class_weight='balanced', scoring='balanced_accuracy', n_jobs=12, Cs=numpy.logspace(-6, -4, 20))" to determine the L2 regularization coefficient via stratified 5-fold cross-validation with a balanced accuracy metric.

For the undirected and directed PAGA graphs, sc.tl.paga and scv.tl.paga functions were used from Scanpy and scVelo packages, respectively. Velocyto.py command line interface (v0.17.15)24 was used to count spliced and unspliced reads using the "velocyto run possorted_genome_bam.bam cellranger-GRCh38-3.0.0-genes.gtf -o sample -m hg38_rmsk.gtf -b barcodes_sample.csv -@ 30 -vv --samtools-memory 100000 -e sample" command. Repetitive elements were masked using the RepeatMasker track downloaded from UCSC Genome Browser in GTF format. Downstream analysis was performed using the scv.pp.moments, scv.tl.recover_dynamics, scv.tl.velocity, scv.tl.paga, and scv.tl.velocity_pseudotime functions of scVelo (v0.2.1)25. scv.tl.velocity was run in stochastic and dynamical modes for DCs and KCs, respectively.

The sc.queries.enrich function from Scanpy was used for gene set enrichment analysis with REACTOME gene sets and the arguments "gprofiler_kwargs={'no_evidences': False, 'sources':['REAC'], 'all_results': True}". FDR < 0.1 and log2(fold change) > 0.5 cutoffs were used for fibroblast and KC marker analyses. For differentially expressed genes between healthy vs lesional KCs, FDR < 0.1 and log2 fold change > 1 cutoffs were used. A stochastic block model-based hierarchical topic modeling approach49 (https://github.com/martingerlach/hSBM_Topicmodel) was used. The number of topics (253) was determined automatically during the SBM inference using the minimum description length-based Bayesian model selection approach. Differential expression analysis of topic expression was performed using Welch's two-sided t test (sc.tl.rank_genes_groups function in Scanpy). Benjamini-Hochberg FDR values are used in the figures. ggplot function from the plotnine plotting package was used for visualizations of the DE results.

The Python implementation of Dorothea (https://github.com/saezlab/dorothea-pycommit=5e3ee0e) was used to calculate the activity scores of 118 transcription factors, which are in A and B confidence categories of Dorothea, for each cell. dorothea.run function was used with "dorothea.load_regulons(['A', 'B']), center=True, scale=True, use_raw=False" arguments.

Scirpy102 (v0.5.0) was used for the TCR analysis. First, for each patient, TCR data were loaded in json format using the ir.io.read_10x_vdj function with "patient_tcr.json', filtered=True" arguments. Graph- based Leiden clustering was run with resolution=1.0 parameter on the k-NN graph built with k=15 on 50 PCs to annotate the T cell cluster which showed high CD3D and CD3E expression. Only cells in this cluster that also have TCR information were retained for downstream analysis, yielding 1,138 cells from two patients. Granular cell subset annotations were predicted using both the cell subset classifier trained on the 3' scRNA-seq data and a pretrained SCimilarity foundation model103 (https://zenodo.org/records/10685499). These annotations were used to subset the 7,256 total T cells to 6,047 CD4+ T cells, which were used in the subsequent topic 190 analysis.

Two-tailed Welch's t-test via sc.tl.rank_genes_groups function from the Scanpy Python package was used to identify genes differentially expressed between cells from expanded (n>=2) vs. non- expanded (n=1) clones. sc.tl.score_genes function in Scanpy was used with genes that have non- zero probabilities in Topic 190

Granular cell type proportions (including the IL13+/IL22+/IL26+ T cell subset) of each channel were 'centered log-ratio' (CLR)-transformed using the 'clr' function from the 'composition_stats' Python package.

Putative cell-cell interactions were identified between cell subsets that are significant in at least one lesional channel using the command line interface of CellPhoneDB v2.1104 ("cellphonedb method statistical_analysis meta.tsv counts.tsv --counts-data hgnc_symbol --project-name ad -- threads 90 --subsampling --subsampling-num-cells 3000 --subsampling-log false").

GWAS-nominated genes for 2,963 diseases and traits were downloaded from the Open Targets Genetics (OTG) GraphQL API (https://github.com/opentargets/genetics-api, v20.02). Additional AD GWAS-nominated genes were obtained from a genome-wide meta-analysis of one-million individuals (Table 1 and Supplementary Table S178). Fisher's exact test, implemented in Fisher python package, was used to test gene sets nominated by the Locus2Gene (L2G) framework of OTG for a particular phenotype were enriched in markers of each of the 86 cell subsets and of the IL13+/IL22+/IL26+ T cell subset, for a total of 87 cell subsets tested.

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

Data associated with this work is available at the Gene Expression Omnibusat (GEO) under accession GSE204765 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE204765]. Raw fastq files for adult samples are deposited on dbGAP as controlled access data under accession phs004337. Raw fastq files for pediatric samples cannot be deposited in a public repository because of restrictions in the informed consent, and may be requested from L.S. (Lynda.Schneider@childrens.harvard.edu). Dr. Schneider will respond to requests within two weeks; access to data will be granted after individuals are added to the

IRB, appropriate institutional approvals will typically be completed within two months. Processed scRNA-seq data of the current study and the curated Atopic Dermatitis Atlas are available at https://cellxgene.cziscience.com/collections/5e143645-177c-45b6-952d-48770e29a54b, https://singlecell.broadinstitute.org/single_cell/study/SCP2613, and https://singlecell.broadinstitute.org/single_cell/study/SCP2738. The scleroderma data and the mmDCs from our AD study and public studies are available at https://singlecell.broadinstitute.org/single_cell/study/SCP3122 and https://singlecell.broadinstitute.org/single_cell/study/SCP3120 respectively. The 5'-scRNA-Seq + scTCR-Seq dataset is available at https://singlecell.broadinstitute.org/single_cell/study/SCP3036. Raw numbers for plots displayed in figures are provided in the Source Data file.

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	Participant sex is provided in Supplementary Data Figure 1. Samples were obtained from gender-matched atopic dermatitis patients patients, scleroderma patients, and healthy individuals. When performing differential expression analysis and comparing healthy vs. lesional and healthy vs. non-lesional groups within each coarse and granular cell subset, donor sex was used as a covariate, in addition to the fraction of mitochondrial UMIs and total number of UMIs.
Reporting on race, ethnicity, or other socially relevant groupings	Race and ethnicity were not considered in our analysis.
Population characteristics	17 adult individuals, including 11 Atopic Dermatitis patients, 6 gender-matched healthy controls and 2 scleroderma patients, to a total of 39 specimens, spanning Caucasian, Black, Asian and Hispanic descent. Subjects were recruited through RALLY at Massachusetts General Hospital (MGH), an online platform that connects the public with research opportunities at MGH. Sample and clinical data collection was conducted in accordance with protocols approved by Institutional Review Boards at Massachusetts General Hospital (protocol number 2018P002325) and Boston Children's Hospital (protocol number P00001109). Inclusion Criteria 1. Subjects at least 18 years of age or older (for the adult recruitment at MGH). 2. A confirmed diagnosis of an inflammatory skin disorder such as atopic dermatitis, morphea, neutrophilic dermatosis. 3. Presence of an active lesion of the above diagnosis. 4. For the healthy subjects, those with an inflammatory skin disorder were excluded. Exclusion Criteria 1. Have any condition that, in the opinion of the investigator, would compromise the well-being of the subject or the study. 2. History of poor wound healing or blood-clotting abnormality. 3. History of keloid formation or hypertrophic scarring; 4. Regular intake of high doses of aspirin or anti-coagulant medications. 5. Hypersensitivity to local anesthetics. 6. History of poorly controlled diabetes mellitus. Pregnancy (for the adult recruitment at MGH).
Recruitment	Patients were recruited through RALLY at Massachusetts General Hospital.
Ethics oversight	Sample collection was conducted in accordance with protocols approved by Institutional Review Boards at Massachusetts General Hospital (protocol number 2018P002325) and Boston Children's Hospital (protocol number P00001109).

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	No sample size calculations were performed a priori. Comparative analysis between conditions involved hundreds or thousands of cells providing robust sample sizes in line with similar scRNA-seq comparisons and technologies.
Data exclusions	Common scRNA-seq quality metric cutoffs were applied to identify true cell barcodes as implemented in the CellRanger workflow. While metrics for exclusion were determined ahead of time, the exact thresholds were determined empirically based on the density of single cells to

	determine appropriate, data-specific thresholds.
Replication	Technical replicates of 10x scRNA-seq channels were performed. Other replicates are not relevant to this study.
Randomization	Not relevant to this study.
Blinding	Blinding is not relevant to our study, as there was no intervention, and analyses were performed in an exploratory manner where blinding is not possible.

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

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	9 week old female C57BL/6J mice were obtained from Jackson Laboratories.
Wild animals	Not relevant to this study.
Reporting on sex	Female mice were used for all experiments in this study.
Field-collected samples	Not relevant to this study.
Ethics oversight	All experiments were approved by the Massachusetts General Hospital Subcommittee on Research Animal Care (Protocol number 2016N000203).

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

Plants

Seed stocks	Not relevant to this study.
Novel plant genotypes	Not relevant to this study.
Authentication	Not relevant to this study.