nature research

Corresponding author(s): J. Christopher Love and Alex K. Shalek

Last updated by author(s): Oct 17, 2019

Reporting Summary

Nature Research 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 Research policies, see Authors & Referees 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	Illumina bcl2fastq2.17.1.14 was used to generate fastqs from bcl files. Gene expression data was processed using Drop-seq tools 1.12 (http://mccarrolllab.com/dropseq), and mapped against hg19 or mm9 genome using STAR 2.5.3a then processed using samtools 0.1.19, picard 2.9.0, bwa 0.7.12, Matlab 2016b, and the Matlab Bioinformatics Toolbox 4.10 were used to process TCR sequencing data.
Data analysis	R version > 3.0, Seurat 2.3.4 (and all associated functions), FlowJo v8, Tidyverse, Monocle 3.0 (and all associated functions), MSigDB (http://software.broadinstitute.org/gsea/), viridis, circlize, ggseqlogo, RColorBrewer, and pheatmap were used to analyzed processed TCR and gene expression data. R scripts for generating all analysis, Matlab scripts for processing TCR sequencing data, as well as all updates will be made available on http://shaleklab.com/resources/ or upon request.

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/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

FASTQ file format data related to murine samples will be available through GEO and BioProject under accession numbers GSE136028 and PRJNA560970. FASTQ file format data related to human samples will be available through dbGaP under accession number phs001897.v1.p1. Source data files and associated metadata tables for Figures 2-4 will be made available on http://shaleklab.com/resources/, https://github.com/mitlovelab/, or upon request. MsigDB results for Extended Data 1c, Extended Data 5a, and Supplementary Fig. 3h are available as Supplementary Table 12. Full results from differential gene expression comparison shown in Extended Data 2c are available as Supplementary Table 13. All recovered CDR3 sequences, and their frequencies, of TCR alpha and beta chains from E7-immunized mice and

peanut allergy patients are available as Supplementary Tables 6, 10. Gene expression matrices for E7-immunized mice and peanut allergy patients are available as Supplementary Data 1, 2.

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	As our study was exploratory and method-centric, we did not predetermine sample size. Given the number of cells we recovered from each Seq-Well chip, we determined that one chip was sufficient for each OT-I spiked-in sample to successfully detect spiked-in cells. Similarly, given the clonal selection of MHC-tetramer reagent and the common genetic background of the mice, we determined that four animals (processed on one Seq-Well chip each) was sufficient to identify common T cell clones across multiple animals. Due to the scarcity of CD154+ T cells in peanut-allergy patients, we included four patients to demonstrate sufficient throughput and recovery of TCR sequences from human samples.
Data exclusions	Described in detail in Methods. Cells were pre-filtered during alignment, and further filtered with no cell accepted with less than 500 genes and 1000 UMIs. TCR data was filtered as described in Methods. UMIs in TCR data with less than 10 reads were filtered. UMIs with no TCR consensus, defined as plurality agreement of sequencing reads attributed to the UMI, were filtered. Filtering criteria were pre-determined before collection of data.
Replication	Reproducibility of the presented method was confirmed using murine samples spiked-in with OT-1 T cells. We detected previously published matching of TCR alpha and beta chains. Furthermore, we performed two technical replicates of the TCR recovery method using the same starting amplified transcriptome libraries. As shown in Supplementary Table 2, the replicates resulted in highly similar data.
Randomization	Randomization was not applicable to the study, as no correlation to outcome was studied.
Blinding	Blinding was not necessary, as no correlation to outcome was studied.

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
☐	☒ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

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

Antibodies

Antibodies used	The following antibodies were used to identify CD154+ cells from peanut-allergy samples via flow cytometry: Anti-CD154-PE (clone TRAP1; BD Bioscience; dilution 1:5), anti-CD3-AF700 (UCHT1; BD Bioscience; dilution 1:20), anti-CD4-APC-Cy7 (RPA-T4: BD Bioscience; dilution 1:20), anti-CD45RA-FITC (HI100; BD Biosciences; dilution 1:5), anti-CD69-AF647 (FN50; BioLegend; dilution 1:20). The following antibodies were used to identify tetramer positive cells from E7-immunized mice samples via flow cytometry: anti-CD8-APC (clone 53-6.7; BioLegend; 1:100 dilution) and E7-MHC-tetramer-PE (Cat.No.TB-5008-2; MBL; 1:50 dilution).
Validation	All antibodies from BioLegend and BD Bioscience have been individually validated by the manufacturer (data available on respective websites). We also validated the human staining panel by comparing staining of primary human cells after stimulation at different time points to isotype controls and fluorescence minus one controls. CD154 antibody has been described and validated previously in publications such as Chattopadhyay, et al., 2006 Nat Protoc, to identify CD154+ T cells. In the study, the CD154 antibody was compared to isotype controls and used in both antigen-stimulated and ex vivo samples to assess specificity.

E7-MHC-tetramer was validated in Wilmschen, et al., 2018 J. Vis. Exp., to identify antigen-specific T cells. In the study, E7-MHC-tetramer was compared to naive and immunized mice to show specificity. We also validated the mouse CD8 antibody by comparison to isotype control.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	B6 mice (C57BL/6NTac) were purchased from Taconic. B6 OT-I mice (C57BL/6-Tg) were purchased from Jackson Laboratories. For experiments represented in Figure 2 (and related supplements), all mice were male at age of 8-12 weeks. For experiments represented in Figure 3 (and related supplements), all mice were female at age of 8 weeks at time of immunization.
Wild animals	Study did not involve wild animals.
Field-collected samples	Study did not involve samples collected from the field.
Ethics oversight	All animal work was conducted under the approval of the Massachusetts Institute of Technology (MIT) Division of Comparative Medicine in accordance with federal, state, and local guidelines (CAC protocol #01717-076-20, #0917-092-20).

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

Human research participants

Policy information about studies involving human research participants

Population characteristics	All subjects had a previous diagnosis of peanut allergy, a history of peanut-induced reactions consistent with immediate hypersensitivity and confirmatory peanut- and Ara h 2-specific serum IgE concentrations (> 0.35 kU/l; ImmunoCAP; Thermo Fisher).
Recruitment	The human subjects in this study were all screened for participation in a peanut oral immunotherapy trial (NCT01750879) at the Food Allergy Center at Massachusetts General Hospital. The participants all had a previous diagnosis of peanut allergy, a history of peanut-induced reactions consistent with immediate hypersensitivity, and confirmatory peanut- and Ara h 2 (a dominant peanut allergen)-specific serum IgE concentrations (> 0.35 kU/l; ImmunoCAP; Thermo Fisher).
Ethics oversight	All subjects were recruited with informed consent, and the study was approved by the Institutional Review Board of Partners Healthcare (protocol no. 2012P002153) and MIT (protocol no. 1312006071).

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

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	Single-cell suspensions were prepared in FACS Buffer (PBS +1 mM EDTA +25 mM HEPES +1% FBS pH 7.5) before staining for surface antigens. Cells were stained for 30 minutes on ice in FACS buffer and then washed for immediate sorting.
Instrument	FACSAria II instrument (BD Biosciences)
Software	BD FACSDiva Software for collection and FlowJo v8 by TreeStar for analysis
Cell population abundance	Purity of the samples was determined by correlating gene expression results in RNA-Seq to antibody targets in flow cytometry. Minimal number of non-T cells (e.g. monocytes, B cells) were detected (<1%).
Gating strategy	Supplementary Figure 3a and Supplementary Figure 4a for full gating strategy

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