Online Repository Materials

Supplementary Tables

Table E1. Subject data.

Table E2. Highly variable genes among $CD8^+$ T cells from healthy and psoriatic skin.

Table E3. Cluster markers.

Table E4. Support vector machine classification of psoriatic and healthy skin samples from two previous studies.

Table E5. Highly variable genes among melanoma-infiltrating $CD8^+$ T cells from Tirosh et al.

Table E6. Differentially expressed genes between exhausted and naïve clusters in psoriasis and melanoma $CD8^+$ T cells.

Table E7. Clonotypes identified across all input CD8+ T cells.

Table E8. TCR sequences shared between subjects.

Table E9. Genes used for functional and cell cycle scoring.

Table E10. Dyes used for RNAscope.

Supplementary Figures

Figure E1. Density-preserving UMAP plot of $CD8^+$ T cells from healthy and psoriatic skin.

Figure E2. Top 5 differentially upregulated and downregulated marker genes of each cluster.

Figure E3. Expression of IL17A and CXCL13 in a subset of Tc17 clusters.

Figure E4. Clustering and exhaustion scoring of melanoma-infiltrating $CD8^+$ T cells from Tirosh et al. 2016.

Figure E5. Comparable degrees of clonotype expansion and TCR sharing among psoriasis and healthy subjects.

Figure E6. MAIT cells in psoriatic and healthy skin.

Supplementary Methods

Patient recruitment

Study subjects were recruited between 2015 and 2017 from Bay Area patients visiting the dermatology department at the University of California, San Francisco. Patients were adults age 18 or above and provided written informed consent. Participants with psoriasis carried a diagnosis of psoriasis for at least 6 months prior to enrollment and had at least 5 percent of their body surface area covered by psoriasis, with at least one plaque on the extremities greater than 10 cm^2 that would be amenable to biopsy. We also required psoriasis subjects to be free of systemic anti-psoriatic treatments including retinoids, methotrexate, cyclosporine, corticosteroids, and cyclophosphamide for at least 4 weeks; biologic immunomodulators including infliximab, etanercept, and adalimumab for 16 weeks; and ustekinumab for 6 months. At the time of study commencement, patients could use topical treatments such as corticosteroids, vitamin D analogs, pimecrolimus, retinoids, salicylic acid, lactic acid, tacrolimus, tar, and urea, but the plaque targeted for biopsy was free of topicals for at least 2 weeks prior to enrollment. Control subjects had no previous diagnosis of psoriasis or other inflammatory skin conditions and had no known history of HIV positive status.

$CD8^+$ T cell sorting

Cells were stained with viability dye Ghost Dye™ Violet 510 (Tonbo biosciences 130870) and the following antibodies: anti-hCD3 PerCP (Biolegend; SK7), anti-hCD45 APC-eFluor 780 (eBioscience; 2D1), and anti-hCD8a eVolve 605 (eBioscience; RPA-T8). We then performed multi-parameter flow cytometry using the LSR Fortessa (BD Biosciences) flow cytometer and FlowJo software (Tree Star Inc.). To identify T cell populations of interest, a lymphocyte gate was taken and then doublets were excluded. Next, live $CD45^+$ cells were gated, then $CD45^+$ $CD3^+$ cells, and then $CD45^+$ $CD3^+$ $CD8^+$ cells. These cells were directly sorted into 10X Reaction Buffer prepared from the SMART-Seq v4 Ultra Low Input RNA Kit (Takara Clontech cat. 634894) and snap frozen in liquid nitrogen for storage at −80 °C.

Highly variable gene selection

Highly variable genes were selected according to the "mean.var.plot" method in *Seurat* 3.1.1. Briefly, the mean-variance relationship of the features in the dataset was estimated by calculating the log of the mean (log-mean) counts for each gene across all cells and the log of the variance to mean ratio (log-VMR). Genes were then grouped into 20 bins based on their log-mean values, and final estimates of mean and dispersion for each gene were calculated by normalizing the log-mean and log-VMR values of each bin by, respectively, the mean of the log-means and the standard deviation of log-VMRs within that bin. These mean and dispersion estimates were plotted and used to select genes with higher than average dispersion. We avoided selecting genes with dispersion values close to the mean of each bin as well as genes with extreme mean values under the rationales that (1) there are too few genes with highly mean values to obtain an accurate estimate of dispersion at this expression level and (2) high variability and dropout rates may affect the dispersion estimates of genes with very low mean expression.

Analysis of public microarray and RNA-seq datasets

Processed expression data from GEO deposited by the Gudjonsson et al 2010[1] (GSE13355, microarray) and Li et al. 2014[2] (GSE54456, bulk RNA-seq) studies were downloaded using the *GEOquery* R package[3]. RNA-seq expression values were converted from RPKM to TPM by dividing each expression value by the sum of RPKM values within the same sample and multiplying by 10^6. Using the *e1071* R package[4], we constructed support vector machines to classify the clinical status of a sample (psoriasis lesion or healthy skin) using the mean-centered and scaled expression of specific genes of interest (or, for microarray data, all probes associated with specific genes of interest). Accuracy of support vector machine classification was calculated as the percentage of samples correctly classified. Repeated 5-fold cross validation was performed using the 'createFolds' function in the *caret* R package[5] to obtain the average classification accuracy across 50 repetitions.

Functional and cell cycle scoring

Phenotype scoring of $CD8^+$ T cells was performed using the 'AddModuleScore' function in *Seurat* 3.1.1 with default settings. Briefly, the score for a given cell was calculated from a panel of genes by taking the average of the normalized expression of those genes in the cell and subtracting from this the average normalized expression of a set of control genes in the same cell. Control genes were selected by first grouping all detected genes in the dataset into 24 bins based on their average normalized expression across all cells. For each gene in the panel, 100 genes from the same bin were randomly selected (seed = 1), and the cumulative nonredundant set of genes were used as the final controls. The input panel for exhaustion consisted of 28 genes from the core melanoma exhaustion signature reported by Tirosh et al.[6], and the panels of metabolic and cell cycle genes were obtained from mSigDB via the *msigdbr* R package[7]. Before scoring, each panel was filtered to exclude undetected genes for each dataset (Table E9).

The cell cycle phase of each $CD8^+$ T cell was annotated using the 'CellCycleScoring' function in *Seurat* 3.1.1, which determines the phase of a cell based on the higher of two scores calculated as above using panels of S phase and G2M phase genes (Table E9), assigning a 'G1' phase to cells in which both scores are negative.

RNAscope

Hs-CD3E (ACD Biosystems, Cat. 553971-C4) was used as a lymphocyte marker, which in combination with Hs-CD8A (ACD Biosystems, Cat. 560391-C3) marks $CD8^+$ T cells. Expression of *CXCL13* and *IL17A* was measured by Hs-CXCL13 (ACD Biosystems, Cat. 311321) and Hs-IL17A (ACD Biosystems, Cat. 310931-C2). The probe signals were amplified according to the manufacturer's protocol and labeled with Opal dyes as listed in Table E10. In addition, the nucleus was stained with DAPI. The specificity and sensitivity of the staining protocol was confirmed by RNAscope®4-plex Negative Control Probe (ACD Biosystems, Cat. 321831) and RNAscope®4-plex Positive Control Probe (ACD Biosystems, Cat. 321801). Images were acquired using Perkin Elmer Vectra Whole Slide Analytical System at both 10X and 20X objective lens magnifications. Acquired images were then processed and quantified using Perkin Elmer inForm® Cell Analysis™ software. Quantification was done with 5-8 images taken at 20X objective lens magnification per subject depending on tissue section size. The punctate nature of RNAscope signals poorly resolves cell boundaries, presenting difficulties when quantifying probe signals for each cell. For this reason, we

only scored pixels that are co-localized between DAPI and marker(s) of interest. For each image, total DAPI pixels and pixels colocalized between DAPI and marker(s) of interest (Marker-DAPI pixels) were determined by the colocalization algorithm in inForm® Cell Analysis™, which binarizes fluorescence intensity signals into positive detection pixel data based on per-slide calculated thresholds of background intensity. To assess the average level of co-expressed markers for each subject, the sum of Marker-DAPI pixels from all images associated with the same subject was divided by the sum of total DAPI pixels in the same set of images to normalize for cell numbers. The average level of co-expressed markers from each subject was comparison using the Wilcoxon rank sum test.

Sampling of VOYAGE 1 patients for CXCL13 expression analysis

Lesional skin biopsies were isolated from a representative psoriatic target lesion (≥3 cm) located on the trunk or extremities with approximately 0.5 mm or more of induration and approximately 5 mm from the edge of the plaque. Four 4-mm skin biopsies were obtained from the identified target lesion at Week 0 prior to treatment and at Week 4, 24, and 48 post-treatment (one biopsy per visit). One additional punch biopsy was obtained at Week 0 from uninvolved, i.e. macroscopically normal or non-lesional skin, on a similar body site as the biopsied lesion.

References

1. Gudjonsson JE, Ding J, Johnston A, Tejasvi T, Guzman AM, Nair RP, et al. Assessment of the psoriatic transcriptome in a large sample: additional regulated genes and comparisons with in vitro models. J Invest Dermatol [Internet]. 2010;130:1829–40. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022202X15348983

2. Li B, Tsoi LC, Swindell WR, Gudjonsson JE, Tejasvi T, Johnston A, et al. Transcriptome analysis of psoriasis in a large case-control sample: RNA-seq provides insights into disease mechanisms. J Invest Dermatol [Internet]. 2014;134:1828–38. Available from: http://dx.doi.org/10.1038/jid.2014.28

3. Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics [Internet]. 2007;23:1846–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17496320

4. Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071) [Internet]. TU Wien; 2019. Available from: https://cran.r-project.org/package=e1071

5. Kuhn M. Building Predictive Models in R Using the caret Package. J Stat Softw [Internet]. 2008;28. Available from: http://www.jstatsoft.org/v28/i05/

6. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science [Internet]. 2016;352:189–96. Available from: https://science.sciencemag.org/content/352/6282/189.long

7. Dolgalev I. msigdbr: MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format [Internet].

2019. Available from: https://cran.r-project.org/package=msigdbr