About

This site contains all contents, including data and code, to accompany our study titled “High-resolution and noninvasive profiling of tumor microenvironment with spatial ecotypes”.

Contents

Software

Spatial EcoTyper is implemented as an R package, with the source code hosted here and on GitHub. Comprehensive documentation, including detailed vignettes, can be accessed at https://digitalcytometry.github.io/spatialecotyper/.

Liquid EcoTyper is implemented as a python module. The source code and accompanying documentation are available for download here.

Supplementary Data

The “Supplementary Data” section hosts accompanying data too large for inclusion in the primary Supplementary Tables Excel file. In particular, we provide here files containing model feature sets for Liquid EcoTyper and their associations with each prediction class.

Files:

- **LiquidEcoTyper_feature_sets.txt**: A 38,431 × 4,000 binary matrix describing the feature set composition (1 = selected in set, 0 = not selected) of each of 4,000 feature sets. These feature sets are labeled so as to correspond with the torch .pt model files provided with the Liquid EcoTyper software. For example, feature set “M0:1_L2_S3” corresponds to layer 2 binary weight matrix column 3 of model file “0_model_for_CV_iter_1.pt”. Once the model file is loaded, the pre-binarized version of this column is accessible by *model.layers[2].weight[:,3].detach().numpy()*. Of note, feature set labels denote matching entries within the feature set associations file.
- **LiquidEcoTyper_feature_set_associations.txt**: This file contains the Pearson correlation of each feature set from above with each of the 10 prediction classes. These associations were calculated with ground truth SE levels of the simulated training cohort as described in the manuscript’s Methods section.

Datasets

The “Main Datasets” section contains newly generated data involving a total of 123 human subjects. The GEO accession number for these cohorts is pending. All human samples were collected with informed consent for research use and received approval from the Institutional

Review Boards of both Yale University School of Medicine and Washington University School of Medicine, in accordance with the principles of the Declaration of Helsinki (2013).

Cohort 1

Intact and dissociated tumor samples were collected from eight patients (four with melanoma, four with colon cancer) at the time of surgery. Each sample underwent bulk RNA sequencing, and the dissociated tumor samples additionally underwent single-cell RNA sequencing (scRNA-seq). For patient WU2109, adjacent formalin-fixed paraffin-embedded (FFPE) melanoma tissue sections were also collected to generate paired Visium and MERSCOPE datasets.

Files:

- **bulk**: bulk RNA-seq data
    - **CRC_Tumor_RNAseq_TPM.tsv**: TPM (Transcripts Per Million) matrix for eight colon cancer samples. This includes four intact tumors (sample names ending with a “T”) and four dissociated tumor samples (sample names ending with a “C”).
    - **CRC_Tumor_RNAseq_TPM.rds**: The same data as in the .tsv file, saved in RDS format for easier use in R.
    - **CRC_Tumor_RNAseq_TCGAref_scaled.rds**: The above colon cancer data integrated with TCGA data using ComBat, normalized to mean 0 and unit variance per gene across all samples. Only the eight newly generated colon cancer samples are present in this file.
    - **Melanoma_Tumor_RNAseq_TPM.tsv**: TPM matrix for eight melanoma samples. This includes four intact tumors (sample names ending with a “T”) and four dissociated tumor samples (sample names ending with a “C”).
    - **Melanoma_Tumor_RNAseq_TPM.rds**: The same data as in the .tsv file, saved in RDS format.
    - **Melanoma_Tumor_RNAseq_TCGAref_scaled.rds**: The above melanoma data integrated with TCGA data using ComBat, normalized to mean 0 and unit variance per gene across all samples. Only the eight newly generated melanoma samples are present in this file.
- **singlecell**: This folder contains paired single-cell RNA-seq data for four colon cancer samples and four melanoma samples. Each RDS file corresponds to a Seurat object for an individual sample and the Seurat object includes log-normalized expression data and cell type annotations in the metadata.
    - **WU1519_seurat_obj.rds**
    - **WU1735_seurat_obj.rds**
    - **WU2109_seurat_obj.rds**
    - **WU2130_seurat_obj.rds**
    - **WU2160_seurat_obj.rds**
    - **WU2161_seurat_obj.rds**
    - **WU2204_seurat_obj.rds**
    - **WU2321_seurat_obj.rds**

Cohort 2

Matched tumor and plasma cfDNA samples were collected from 23 metastatic melanoma patients, with matched peripheral blood mononuclear cells (PBMCs) also collected for seven. Tumor samples for each patient were profiled by spatial transcriptomics and/or whole-genome Enzymatic Methyl sequencing (EM-seq) depending on availability (Visium, Visium HD, and/or EM-seq). An additional 23 plasma cfDNA samples were collected from healthy individuals. All PBMC and plasma cfDNA samples were profiled by whole-genome EM-seq.

Files:

- **EMseq**
    - **Healthy_EMseq_cfDNA.tsv**: Methylation levels (beta scores) of 38,431 CpGs across 23 plasma cfDNA samples from healthy donors.
    - **Melanoma_EMseq_PBMC_Bulk.tsv**: Methylation levels (beta scores) of 38,431 CpGs across seven PBMC samples collected from melanoma patients.
    - **Melanoma_EMseq_Plasma_cfDNA.tsv**: Methylation levels (beta scores) of 38,431 CpGs across 23 plasma cfDNA samples from the same melanoma patients.
    - **Melanoma_EMseq_Tumor_Bulk.tsv**: Methylation levels (beta scores) of 38,431 CpGs across tumor samples from 20 metastatic melanoma patients. For several patients, biological replicates sectioned from the same tumor are available. Samples are named according to patient, tissue/DNA preservation method (cryopreserved or FFPE), and replicate number.
- **Visium:** This folder contains Space Ranger outputs for 15 Visium samples (13 patients with melanoma) and 2 Visium HD samples (2 melanoma patients with melanoma).

Cohort 3

Plasma cfDNA samples were collected from 78 melanoma patients who received immune checkpoint inhibitor (ICI) monotherapy (30 receiving anti-PD1 and five receiving anti-CTLA4) or combination therapy (43 receiving anti-PD1/anti-CTLA4 combination). Samples were collected before treatment initiation (before or on the first day of ICI cycle 1) and underwent whole-genome EM-seq.

File:

- **Cohort3_EMseq_cfDNA.tsv**: Methylation levels (beta scores) of 38,431 CpGs across 78 plasma cfDNA samples from metastatic melanoma patients.

Cohort 4

Plasma samples were collected from 10 melanoma patients, including eight from Cohort 2, who received immune checkpoint inhibitor (ICI) monotherapy (7 receiving anti-PD-1) or combination therapy (3 receiving anti-PD-1/anti-CTLA-4 combination). Samples were collected before or during treatment and underwent whole-genome EM-seq.

File:

- **Cohort4_EMseq_cfDNA.tsv**: Methylation levels (beta scores) of 38,431 CpGs across 10 plasma cfDNA samples from metastatic melanoma patients.

Paired MERSCOPE and Visium data

- **paired_MERSCOPE_Visium:** This folder contains paired Visium and MERSCOPE datasets derived from adjacent melanoma tumor sections from patient WU2109.
    - **MERSCOPE_seuratobj.rds**: Seurat object for the MERSCOPE dataset, including raw counts, log-normalized expression data, spatial coordinates of individual cells, and cell type annotations stored in the metadata.
    - **Visium_seuratobj.rds**: Seurat object for the Visium dataset, including a Spatial Assay with raw counts, log-normalized expression data, spatial coordinates of spots, and the corresponding tissue image.

Vignette data for Spatial EcoTyper

The “Vignette Data” section contains datasets essential for running the SpatialEcoTyper vignettes. These datasets are organized into subdirectories based on the specific vignette they support. Detailed information about these datasets are available in the SpatialEcoTyper Documentation on GitHub.

Files:

- **SingleSample:** This directory contains files required or generated by the vignette titled “Discovery of Spatial Ecotypes from A Single Sample”.
- **MultiSampleIntegration:** This directory contains files required or generated by the vignette titled “Discovery of Spatial Ecotypes from Multiple Samples”.
- **Recovery:** This directory contains files required or generated by the vignette titled “Development of NMF Models for Spatial Ecotype Recovery” and “Recovery of Spatial Ecotypes from Single-Cell Gene Expression Data”.
- **BulkRNA:** This directory contains files required for the vignette of “Recovery of Spatial Ecotypes from Bulk Gene Expression Data”.