Improving the Resolution of Single-Cell TCR-seq

Kelly Street
Assistant Professor
Division of Biostatistics
Keck School of Medicine
July 4, 2023

USC University of Southern California

The Team

David Braun

Mercedeh Movassagh

Jill Lundell

Miya Hugaboom

Mingzhi Ye

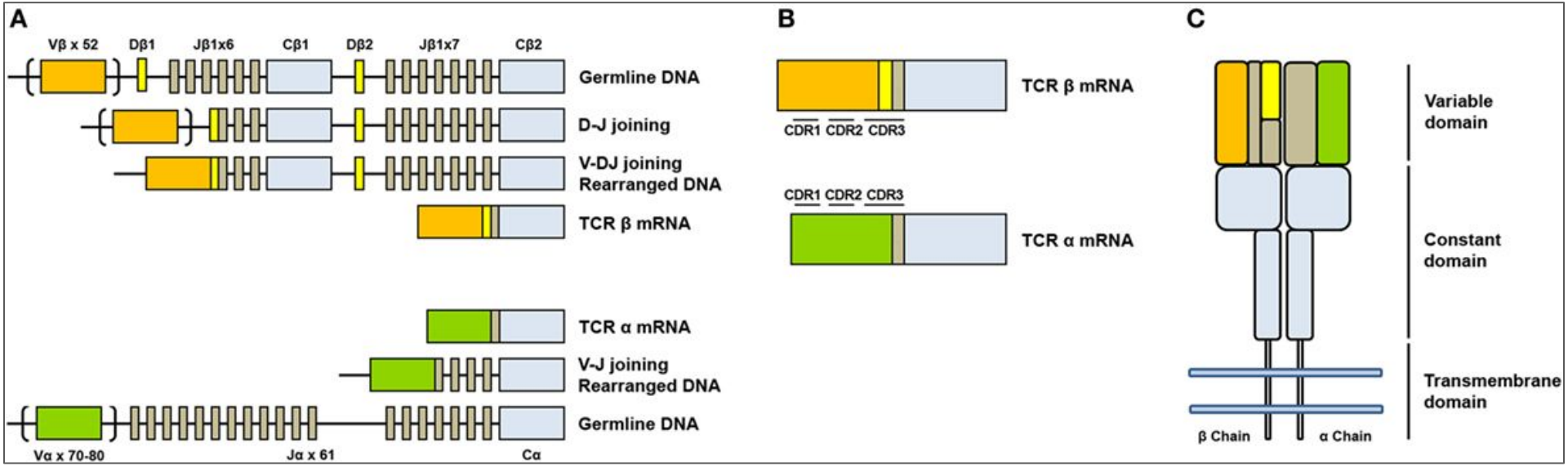
What is a T cell?

T cell

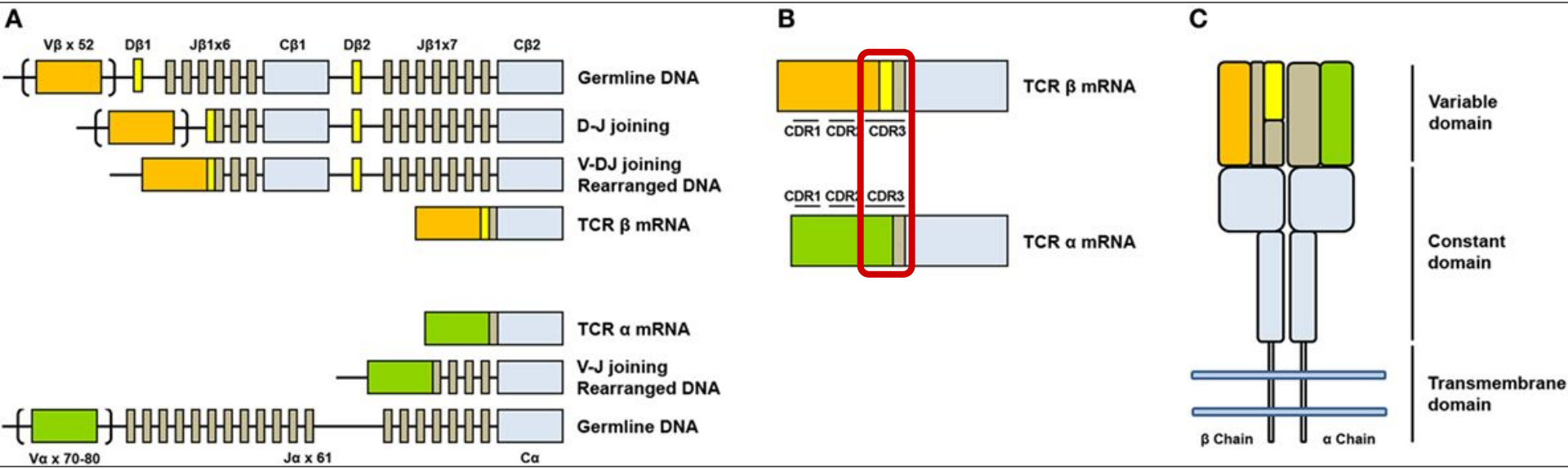
From Wikipedia, the free encyclopedia

A **T cell** is a type of lymphocyte. T cells are one of the important white blood cells of the immune system and play a central role in the adaptive immune response. T cells can be distinguished from other lymphocytes by the presence of a T-cell receptor (TCR) on their cell surface.

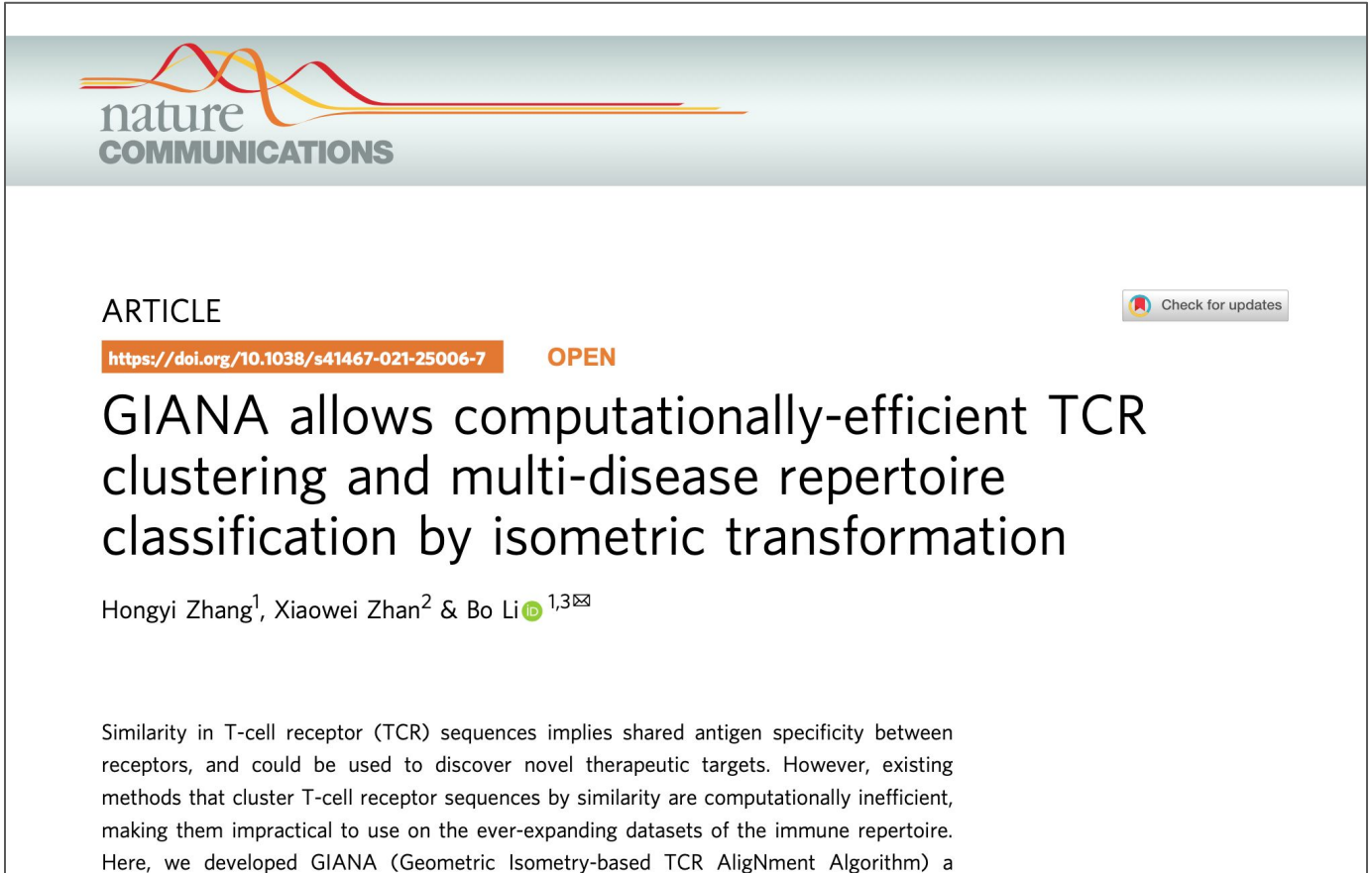
T Cell Structure



T Cell Structure



Why should we be interested?



nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-021-25006-7 OPEN

GIANA allows computationally-efficient TCR clustering and multi-disease repertoire classification by isometric transformation

Hongyi Zhang[1], Xiaowei Zhan[2] & Bo Li[1,3]✉

Similarity in T-cell receptor (TCR) sequences implies shared antigen specificity between receptors, and could be used to discover novel therapeutic targets. However, existing methods that cluster T-cell receptor sequences by similarity are computationally inefficient, making them impractical to use on the ever-expanding datasets of the immune repertoire. Here, we developed GIANA (Geometric Isometry-based TCR AlignMent Algorithm) a

Motivation

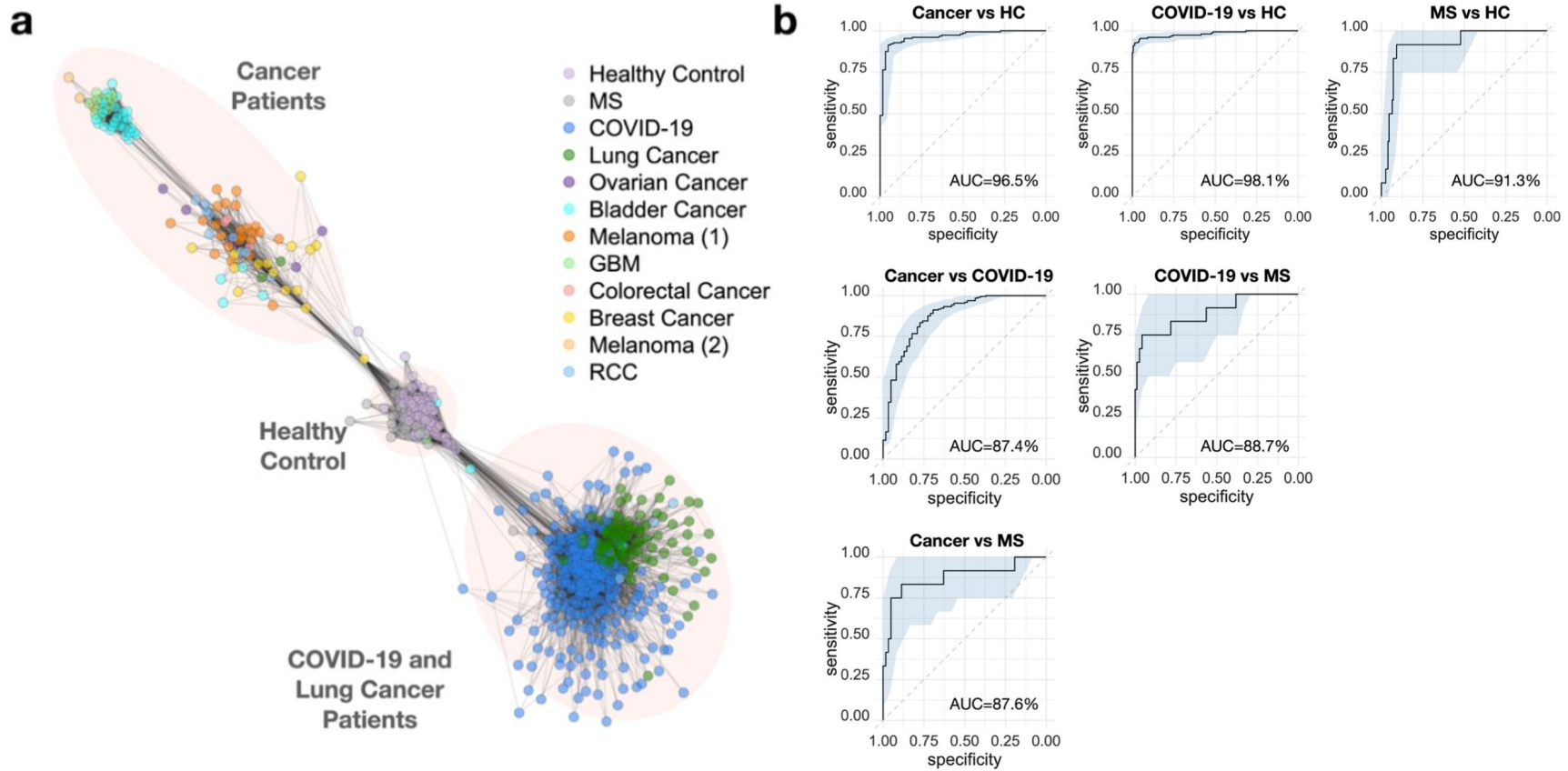


Fig. 4 Disease-specific grouping of TCR repertoire samples via ultra-large-scale clustering.

Motivation

As demonstrated in autoimmune and infectious diseases, antigen-specific public TCRs shared at low frequencies are potentially important biomarkers[20,41,42], which can be detected by comparing large amount of TCRs from thousands of individuals. Methods have been developed to individually detect cancer[17,18], COVID-19[20], or multiple sclerosis[43] using immune repertoire, but none has been able to simultaneously diagnose and separate different diseases. In contrast, our effort could be developed into a unified platform to diagnose infectious disease, autoimmune disorders and cancer. Such a platform has been

Motivation

As demonstrated in autoimmune and infectious diseases, antigen-specific public TCRs shared at low frequencies are potentially important biomarkers[20,41,42], which can be detected by comparing large amount of TCRs from thousands of individuals. Methods have been developed to individually detect cancer[17,18], COVID
repertoire, but none
separate different
developed into a un
autoimmune disord

We believe this is potentially a significant advance because: first, disease diagnosis is mainly symptom-driven for decades, with each disease requiring a distinct set of signatures obtained from diverse clinical assays, such as radioactive imaging, liquid biopsy, invasive endoscopy, surgery, etc. The feasibility of using the immune system as a single biomarker to indicate multiple diseases could shift the paradigm from symptom-driven to immune-response-driven, which provides a universal solution to many immune-related disorders. Additionally, differential diag-

Application: Renal Cell Carcinoma

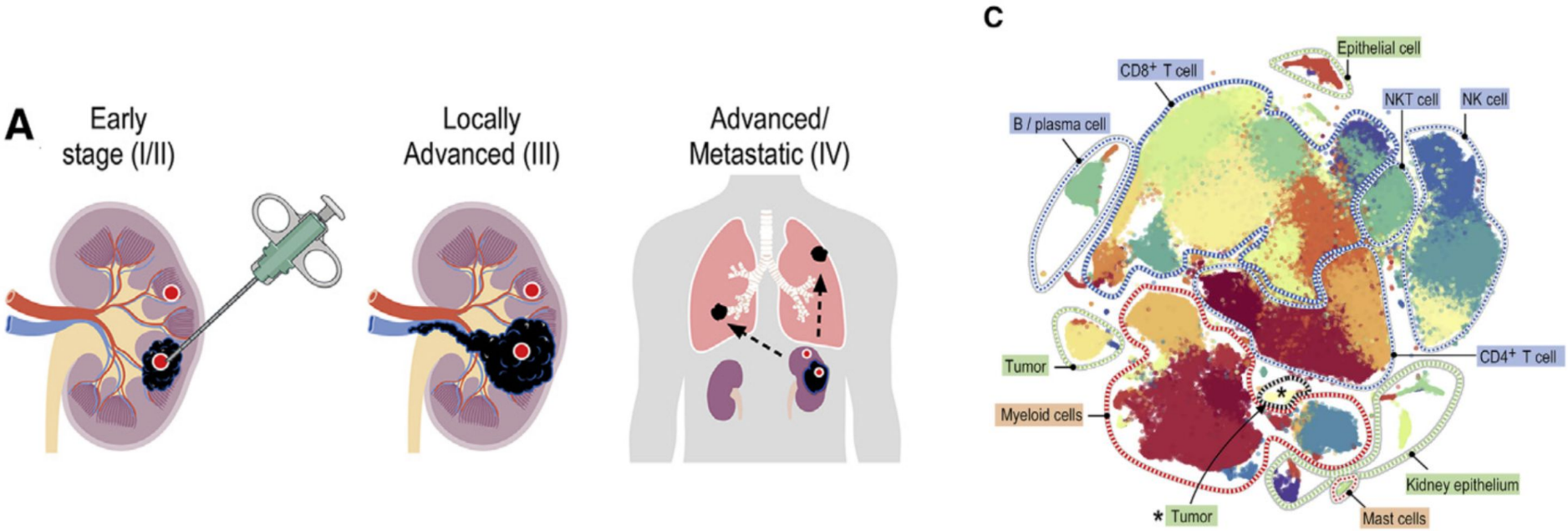


Figure 1. Single-cell profiling of clear cell renal cell carcinoma

TCR Sequencing

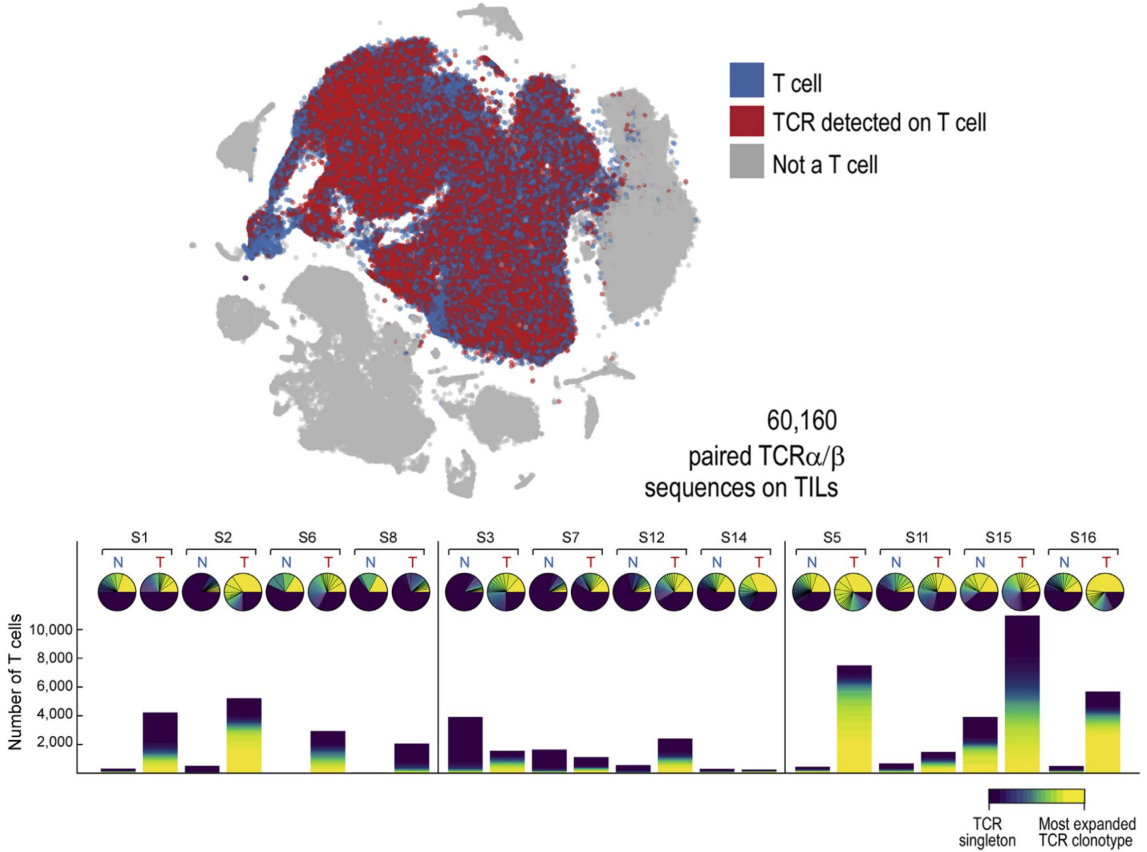


Figure 4. TCR analysis reveals lower diversity in terminally exhausted T cells

Cells can be ambiguous

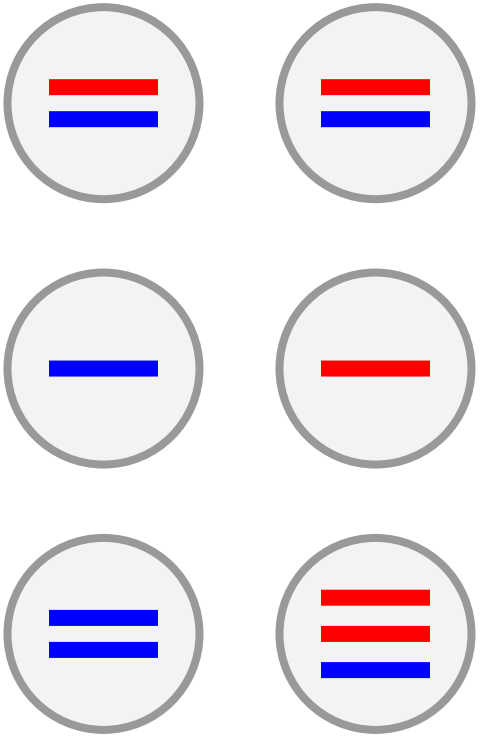
Motivation

As demonstrated in autoimmune and infectious diseases, antigen-specific public TCRs shared at low frequencies are potentially important biomarkers[20,41,42], which can be detected by comparing large amount of TCRs from thousands of individuals. Methods have been developed to individually detect cancer[17,18], COVID-19[20], or multiple sclerosis[43] using immune repertoire, but none has been able to simultaneously diagnose and separate different diseases. In contrast, our effort could be developed into a unified platform to diagnose infectious disease, autoimmune disorders and cancer. Such a platform has been

ant advance because:
-driven for decades,
f signatures obtained
ctive imaging, liquid
he feasibility of using
the immune system as a single biomarker to indicate multiple diseases could shift the paradigm from symptom-driven to immune-response-driven, which provides a universal solution to many immune-related disorders. Additionally, differential diag-

Clonotype Identification

- Many cells have a unique, identifiable clonotype
- Rare clonotypes are clinically relevant



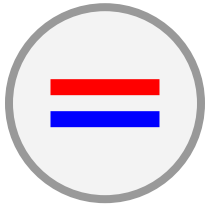
Unique clonotype

Ambiguous clonotype

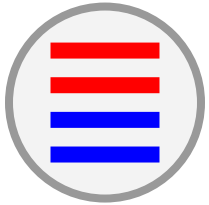
alpha

beta

Consistent Clonotypes



1 possible clonotype



4 possible clonotypes

`nBeta` possible clonotypes

alpha

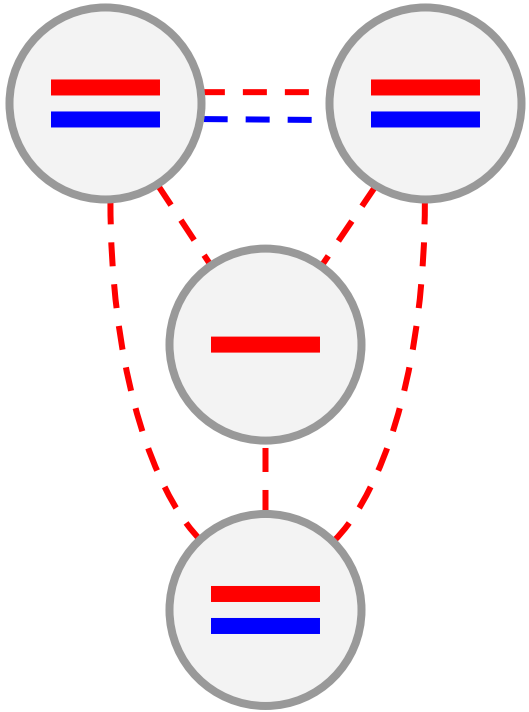
beta

Clonotype Identification

Clonotype 1:

Ambiguous:

Clonotype 2:



alpha

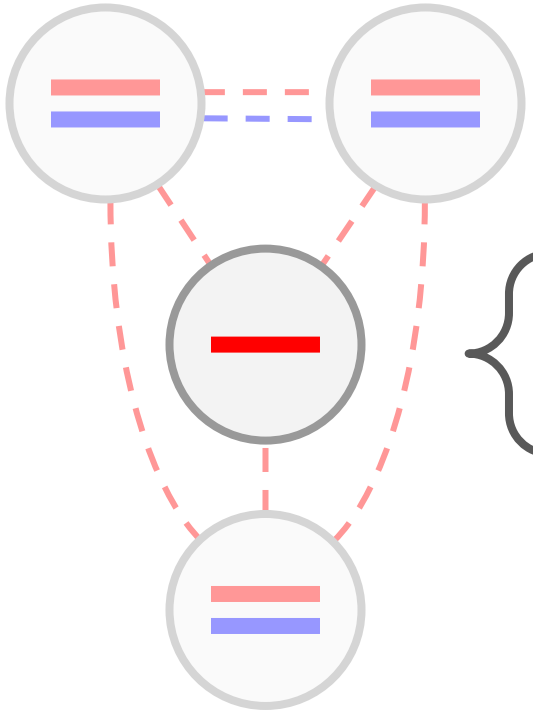
beta

common alpha

common beta

Proportional Assignment (EM)

Clonotype 1:



66.6% Clonotype 1

33.3% Clonotype 2

Clonotype 2:

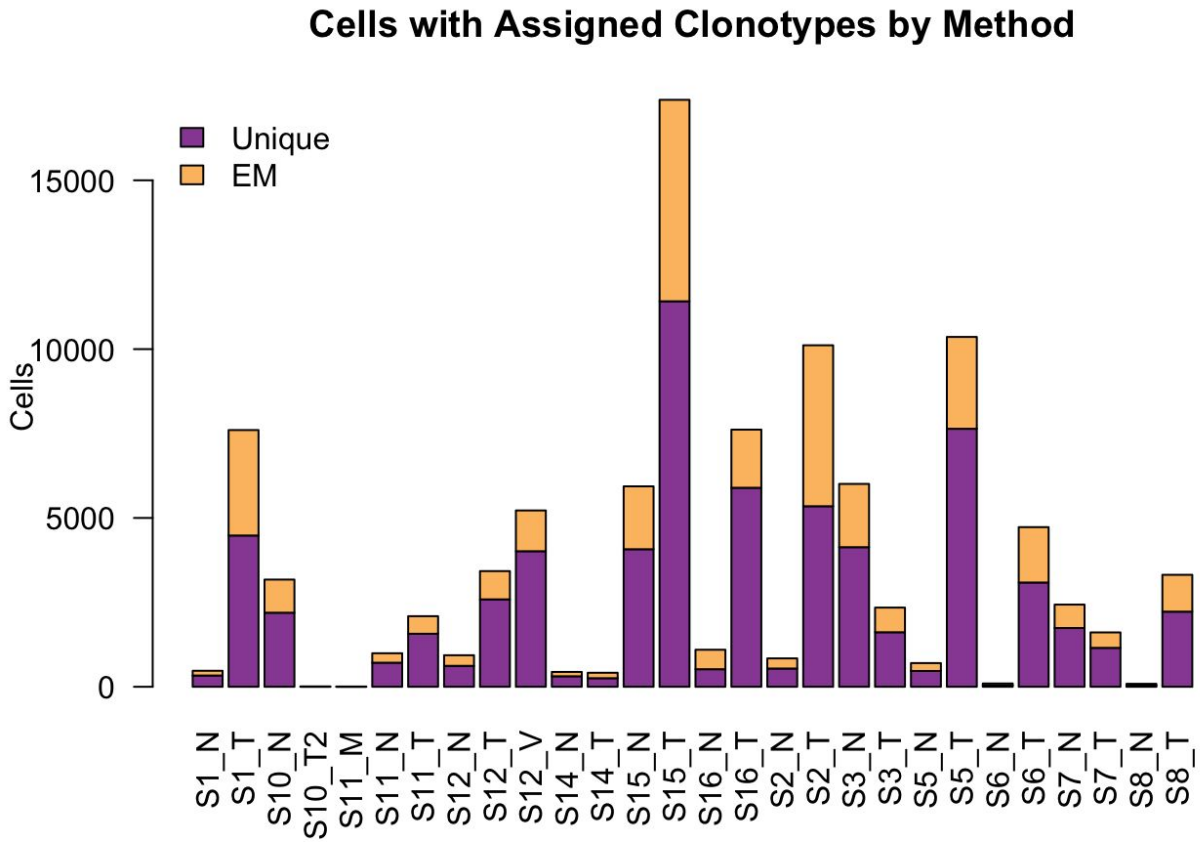
alpha

beta

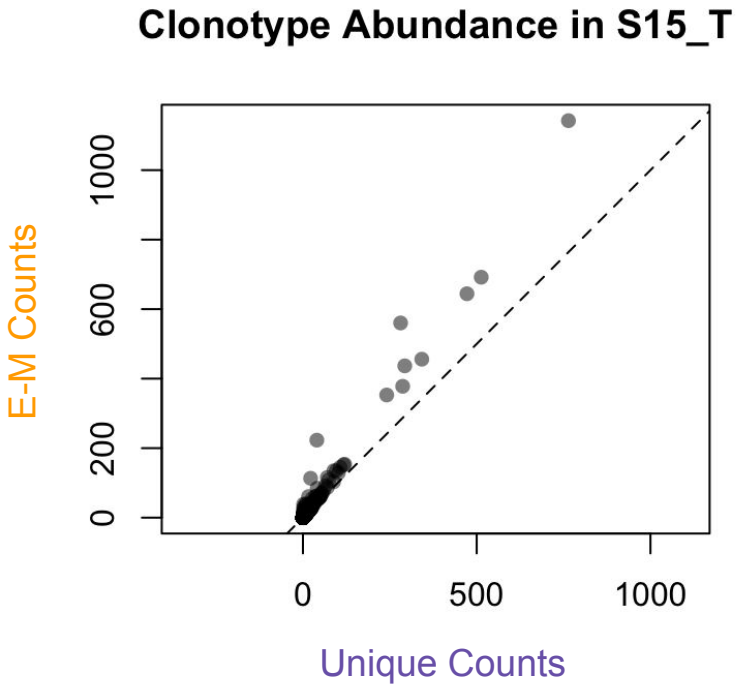
common alpha

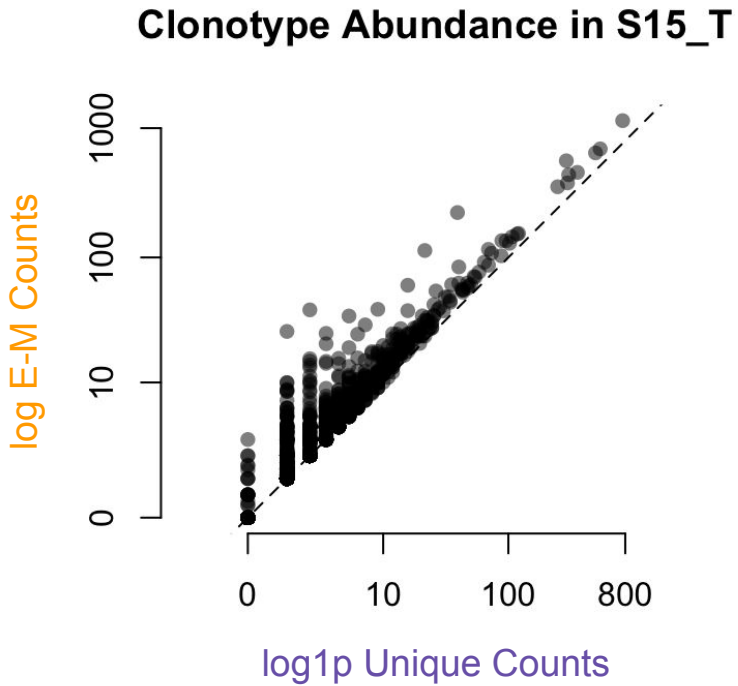
common beta

Same Data, More Cells



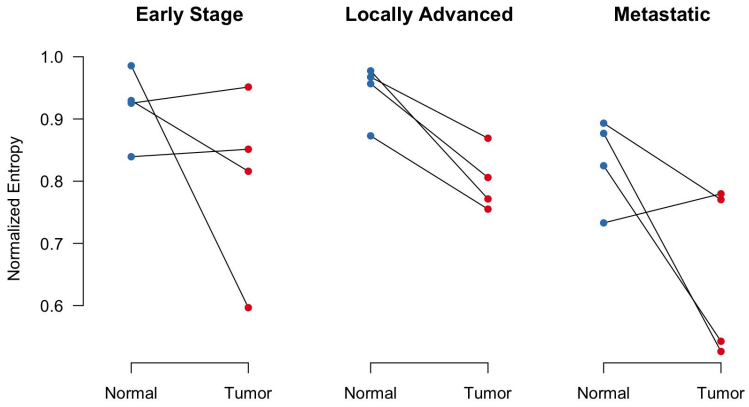
Same Data, More Cells



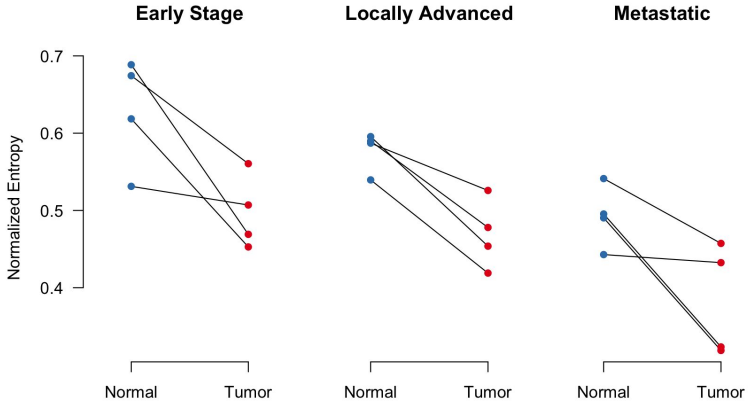


Improved Signal

Unique



E-M



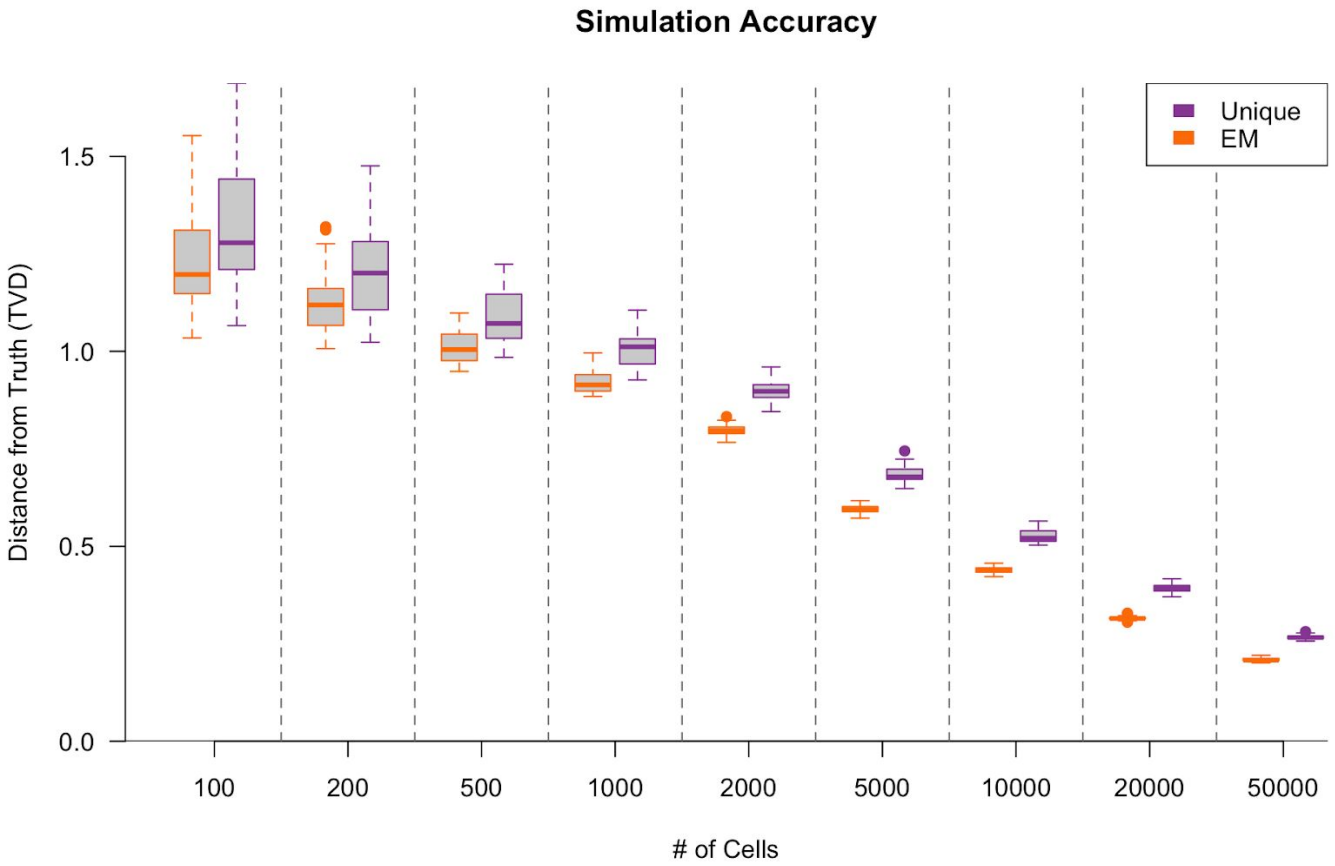
Quantifying (alpha) Diversity

- Total clonotypes
- (normalized) Shannon entropy
- (inverse) Simpson Index

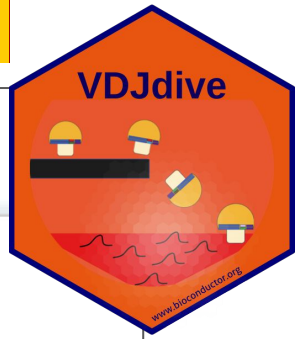
(currently) limited:

- Chao1
- Chao-Bunge
- breakaway
- breakaway_nof1

Simulation


Simulation Accuracy
Unique
EM
Distance from Truth (TVD)
1.5
1.0
0.5
0.0
100
200
500
1000
2000
5000
10000
20000
50000
# of Cells

Bioconductor Package



VDJdive

platforms some | rank 2122 / 2229 | support 0 / 0 | in Bioc 0.5 years

build ok | updated before release | dependencies 79

DOI: 10.18129/B9.bioc.VDJdive

Analysis Tools for 10X V(D)J Data

Bioconductor version: Release (3.17)

This package provides functions for handling and analyzing immune receptor repertoire data, such as produced by the CellRanger V(D)J pipeline. This includes reading the data into R, merging it with paired single-cell data, quantifying clonotype abundances, calculating diversity metrics, and producing common plots. It implements the E-M Algorithm for clonotype assignment, along with other methods, which makes use of ambiguous cells for improved quantification.

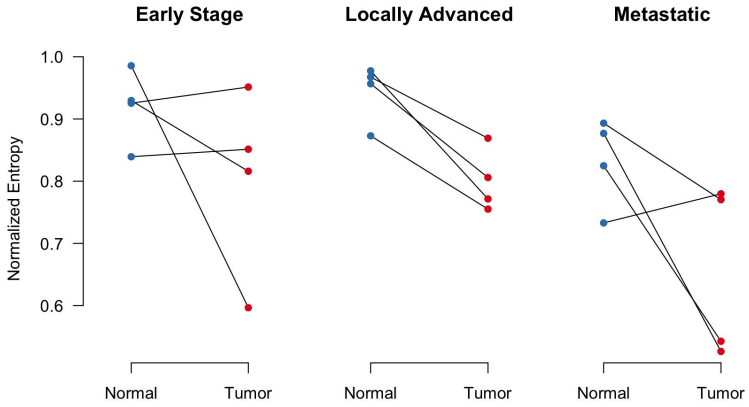
Author: Kelly Street [aut, cre], Mercedeh Movassagh [aut], Jill Lundell [aut], Jared Brown [ctb], Linglin Huang [ctb]

Cells can be ambiguous

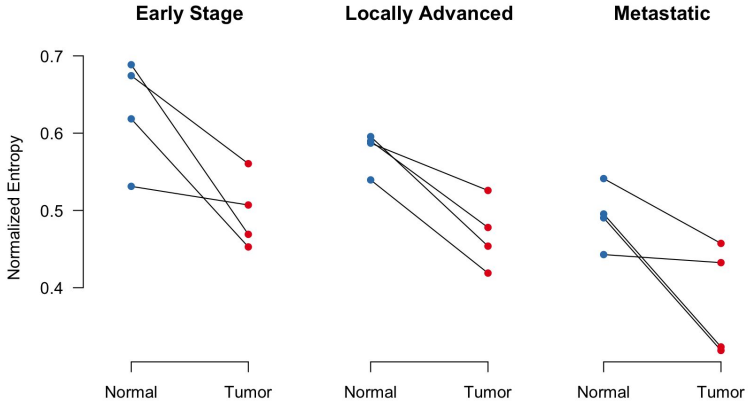
With enclone (Cell Ranger >=3.1)

Before - Improved Signal

Unique

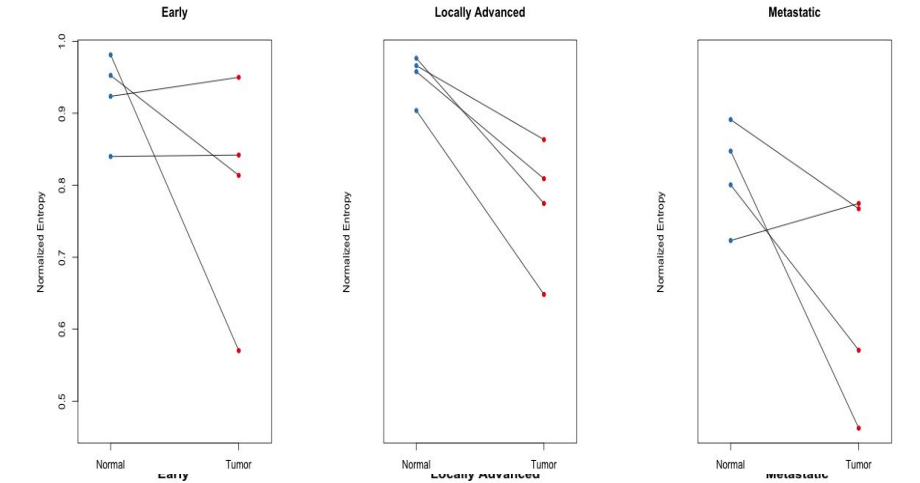


E-M

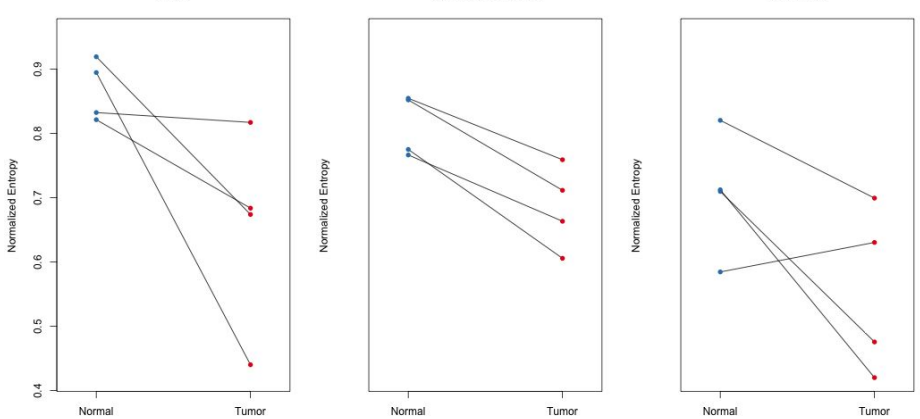


After - Same Signal

Unique



E-M



The Team

David Braun

Mercedeh Movassagh

Jill Lundell

Miya Hugaboom

Mingzhi Ye