Omic data-based biomarkers for cancer immunotherapy

Wei Sun

Associate Member
Biostatistics Program, Public Health Sciences Division
Fred Hutchinson Cancer Research Center, Seattle

Symposium on Advances in Genomics, Epidemiology, and Statistics (SAGES)

Jun 1st, 2018

Outline

1. Background
   - Cancer Immunotherapy
   - Biomarkers for cancer immunotherapy

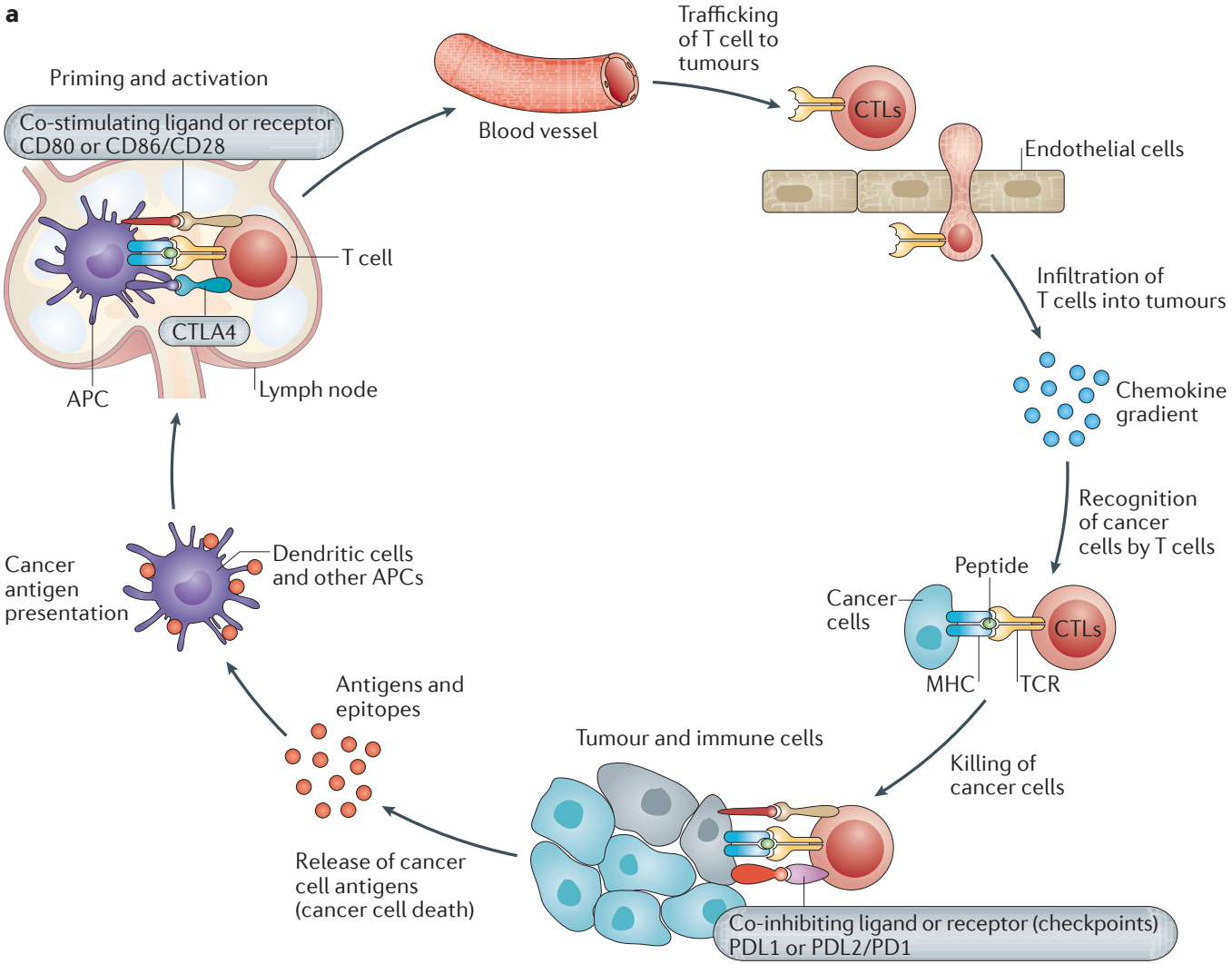
2. Methods
   - Intra-Tumor Heterogeneity and Somatic Copy Number Alteration
   - Estimation of immune cell composition
   - Association analysis for immune cell composition

Cancer-Immune Interaction

- Cancer is a genetic disease. Tumor cell growth is initiated and driven by somatic DNA mutations.
- These somatic mutations can be recognized by immune system.
  - Only a small proportion of somatic mutations may be recognized by the immune system, neoantigen.
  - This is a stochastic event. The more somatic mutations the tumor cells carry, the more likely they are recognized by the immune system.
- After initial attack on the tumor cells, the immune system get "exhausted" through negative regulatory pathways, also known as **check points**.
- Check-point inhibitor remove such check points, and thus revive immune cells' attack on tumor cells.

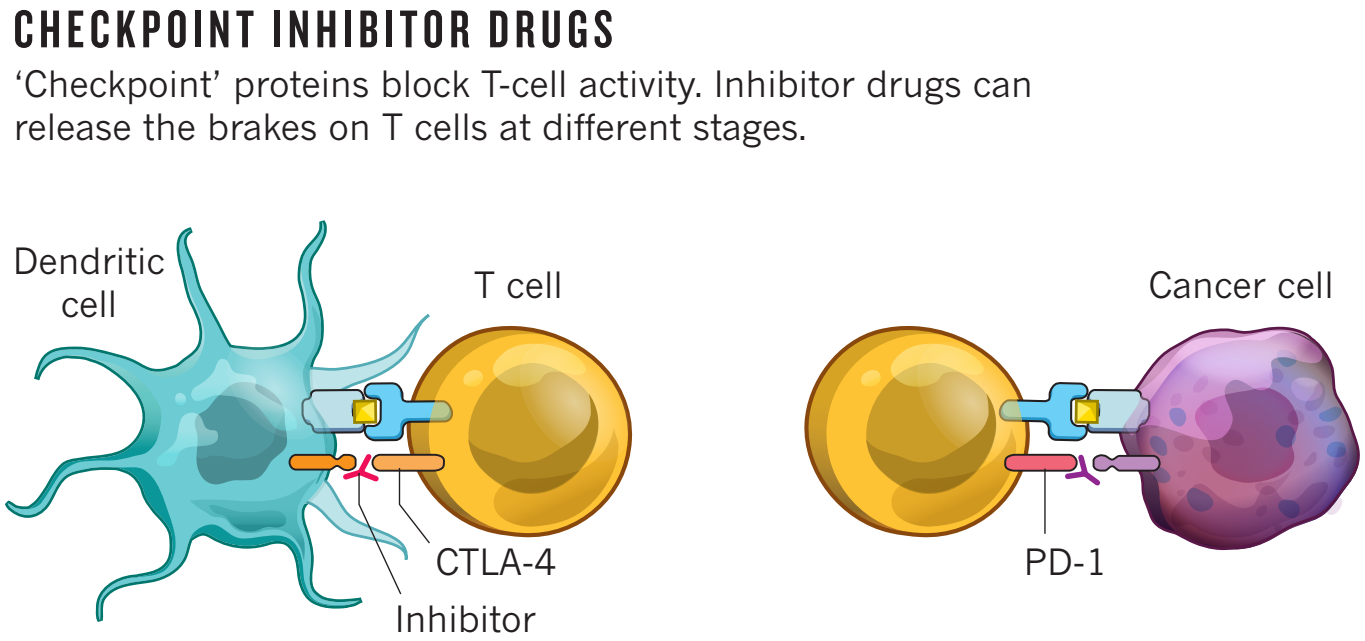
Chan et al. (2017) Nature 541: 321-330

Cancer-Immune Interaction



Hackl et al. (2016) Nature review genetics 17:441-458

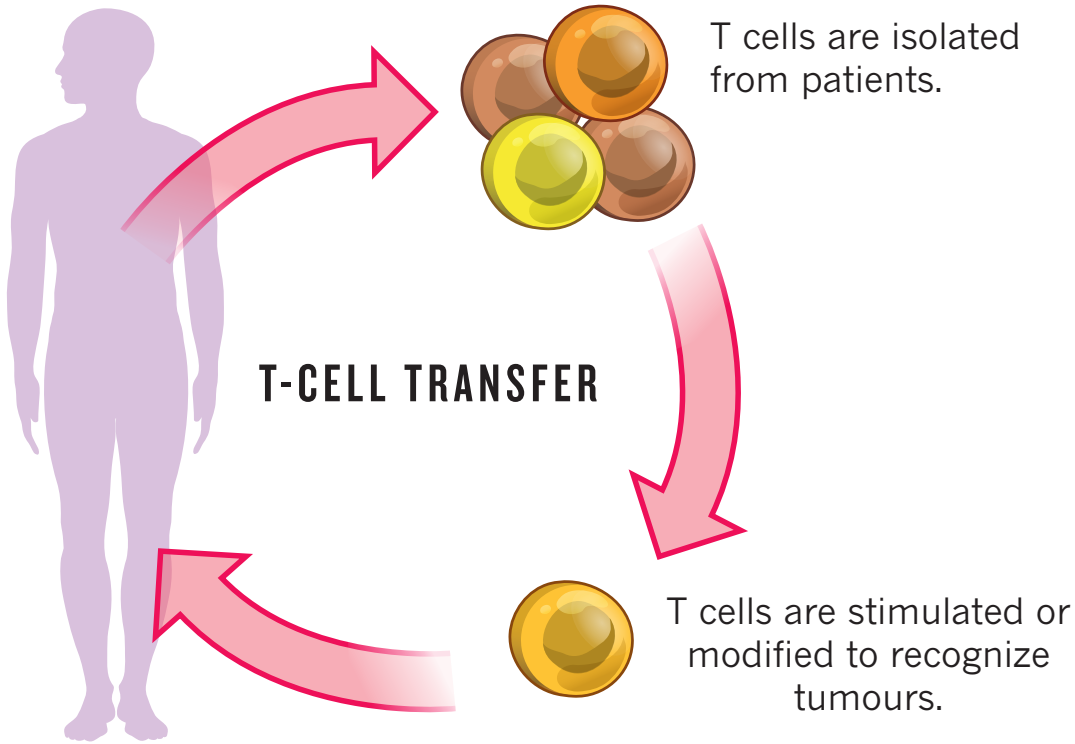
Check-point inhibitors



Ledford (2014) Nature 508:24-26

Adoptive cell transfer

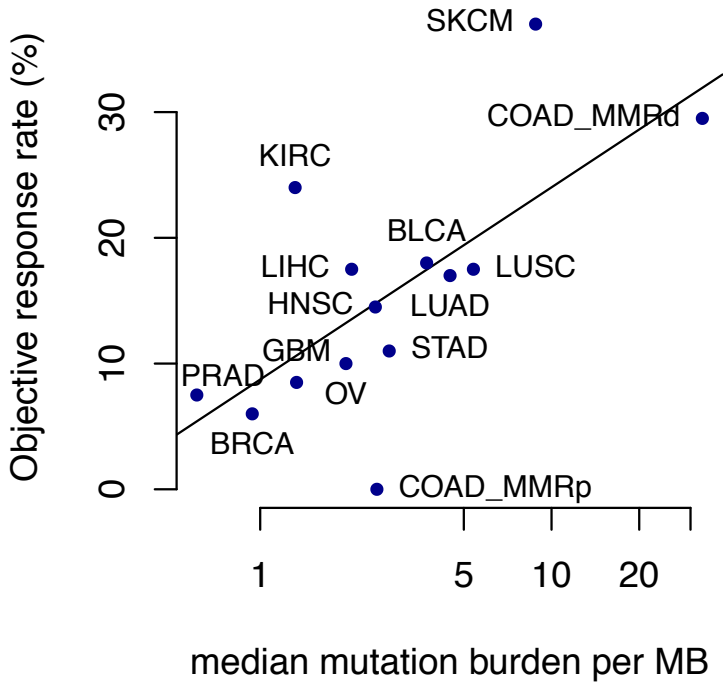
Administration of immune cells with anti-tumor, e.g., chimeric antigen receptor (CAR) T cell therapy.



Ledford (2014) Nature 508:24-26

Response rate of cancer immunotherapy

Despite the success of cancer immunotherapy, only a small proportion of patients respond to immunotherapy.



Yarchoan et al. (2017). New England Journal of Medicine, 377(25), 2500-2501.
COAD_MMRp: colon cancer mismatch repair proficient. COAD_MMRd: colon cancer mismatch repair deficient. SKCM: melanoma. KIRC: kidney cancer.

Some potential biomarkers for cancer immunotherapy

- PD1/PD-L1 expression for anti-PD1/PDL1 check point inhibitors.
- Mutation load (the number of SNVs or indels) for check point inhibitors.
  - Neoantigen: novel immunogenic peptides due to somatic mutations. One may select neoantigen from all mutated peptides by asking whether it is bound to MHC (major histocompatibility complex).
  - Intra-tumor heterogeneity [McGranahan et al. (2016) Science, 351(6280), 1463-1469].
- Genome-wide load of somatic copy number alteration (SCNA) [Davoli et al. (2017) Science, 355(6322), eaaf8399].
- Different characteristics of tumor infiltrating immune cells, e.g., their cell type proportions and cell type-specific expression.

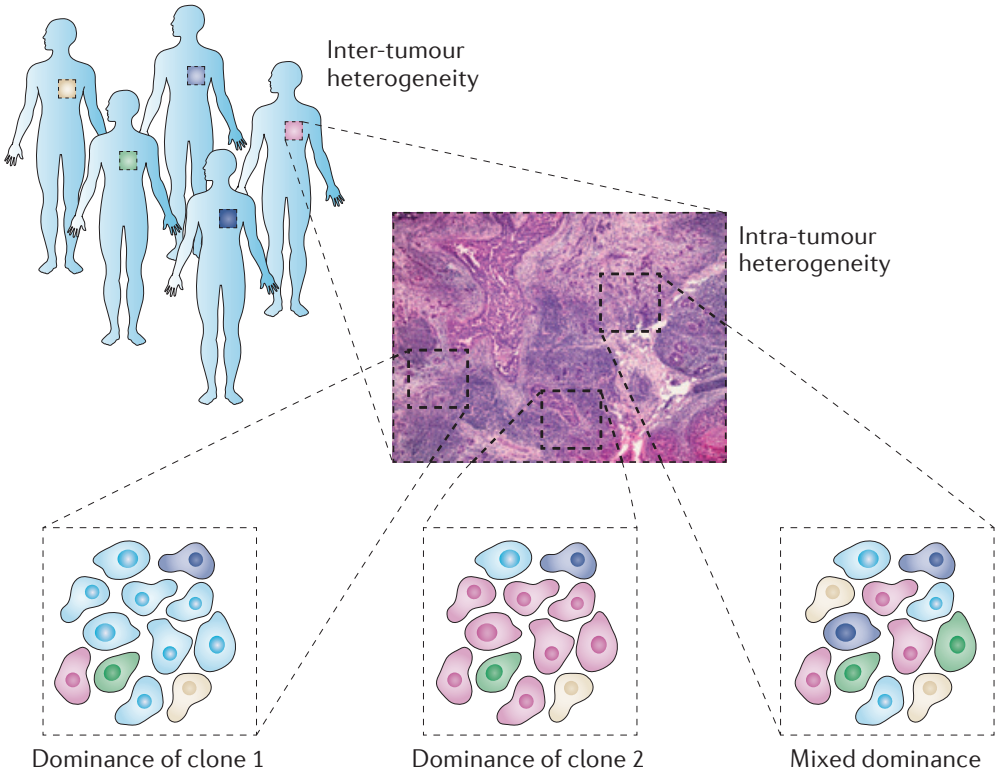
Outline

1. Background
   - Cancer Immunotherapy
   - Biomarkers for cancer immunotherapy

2. Methods
   - Intra-Tumor Heterogeneity and Somatic Copy Number Alteration
   - Estimation of immune cell composition
   - Association analysis for immune cell composition

Intra-Tumor Heterogeneity

- Tumor initiation and progression involves gradual evolution from normal cells to cancer cells.
- A group of passenger somatic mutations may be "attached" to a few driver mutations within a group of tumor cells. Such group of tumor cells sharing similar somatic mutations forms a subclone.



Marusyk et al. (2012). Nature Reviews Cancer, 12(5), 323

Infer Intra-Tumor Heterogeneity for Association Analysis

It is important to jointly estimate the somatic point mutations and somatic copy number alterations (SCNA) of each subclone. Many methods have been developed for ITH estimation, but most of them are not ideal for (large-scale) association studies because

- Only study SCNAs
- Can only be applied to regions without SCNA events
- Assume SCNAs are known or are clonal
- Are only applicable to the settings with multiple samples per patient

One exception is Canopy (Y Jiang, ..., & N Zhang, 2016, PNAS, 113 (37) E5528-E5537). Canopy is mainly designed for multiple sample study.

We have developed a new method named SHARE (Statistical method for Heterogeneity using Allele-specific REads and somatic point mutations).

How to quantify ITH.

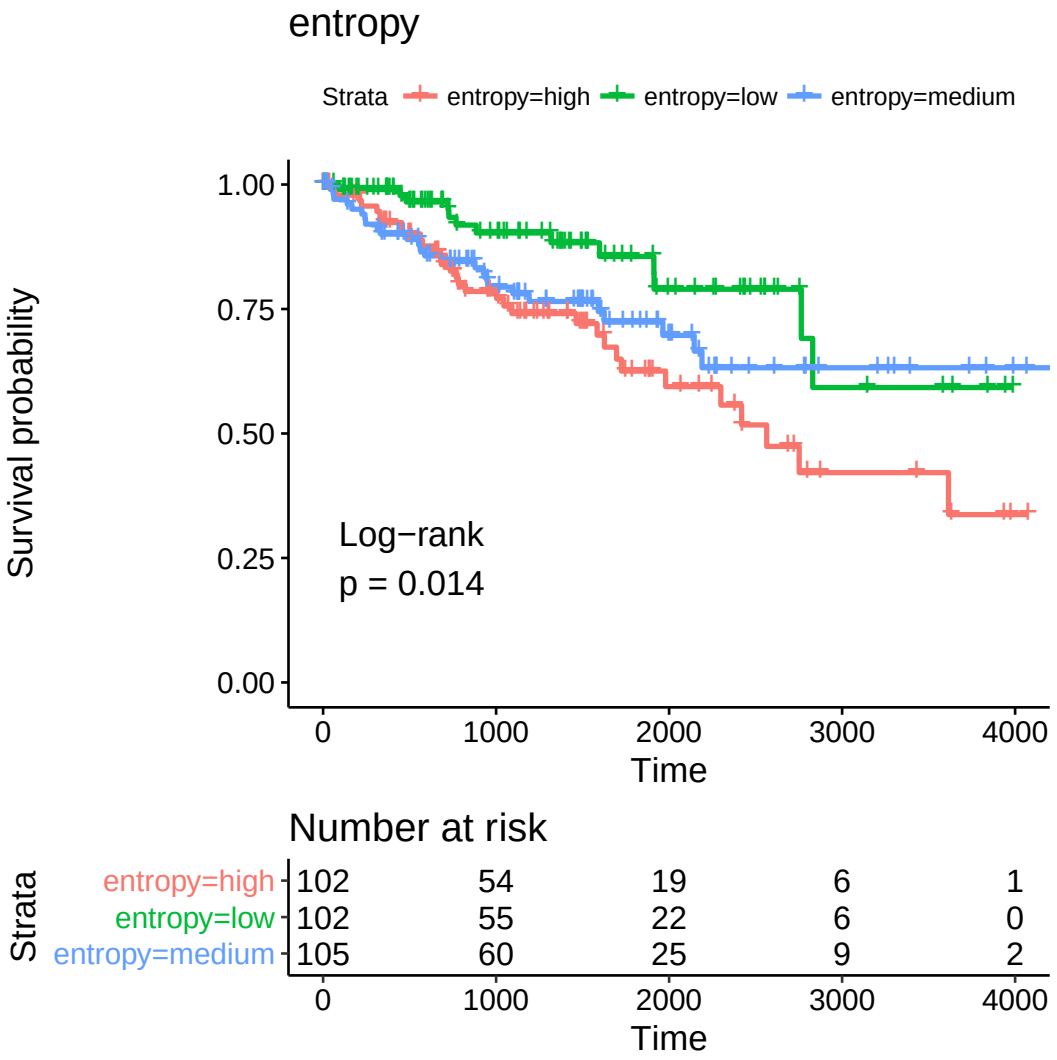
Assume there there are 3 subclones. Consider two examples of subclone proportions (the proportion of cells belonging to each subclone)

- 1%, 49%, 50%
- 30%, 32%, 38%

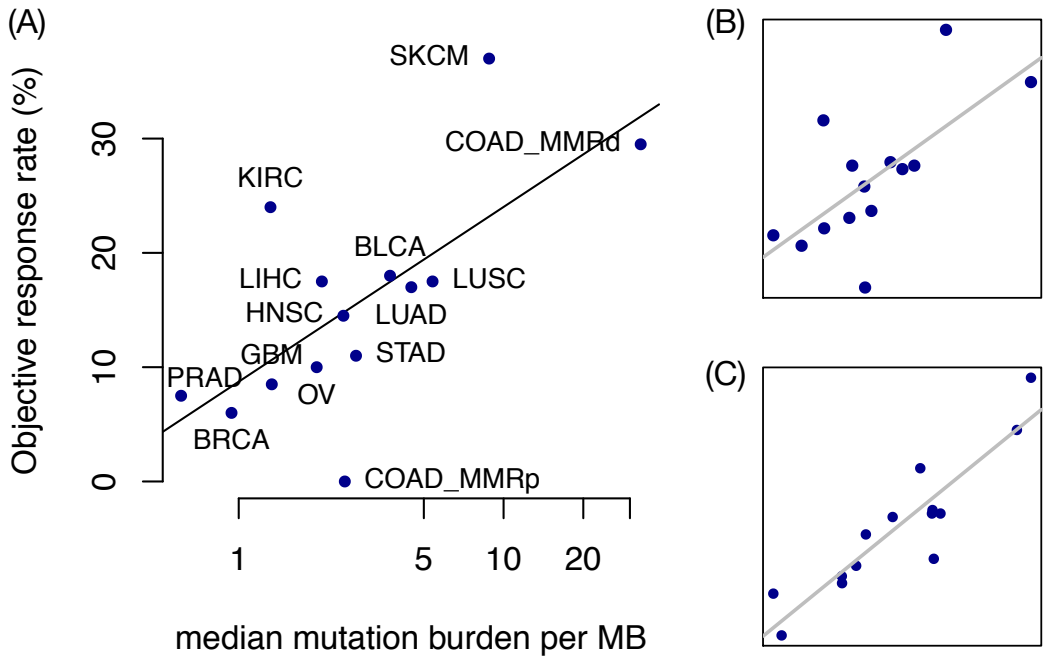
We may quantify ITH using the number of subclones, but then the information of subclone proportions is lost. If we have to use the number of subclones, the first example is better to be considered as two subclones instead of three.

We propose to use an entropy measurement: $-\sum_{s=1}^{S} \vartheta_s \log \vartheta_s$, where S is the number of subclones, and ϑ_s is the proportion of tumor cells in the s-th subclone.

Survival Analysis in TCGA Kidney Cancer

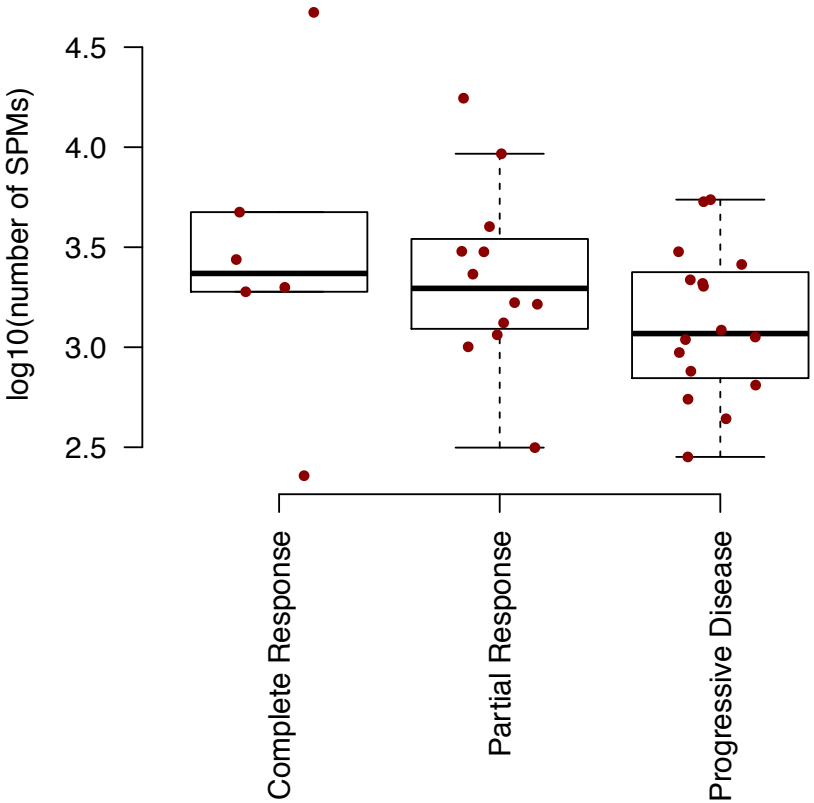


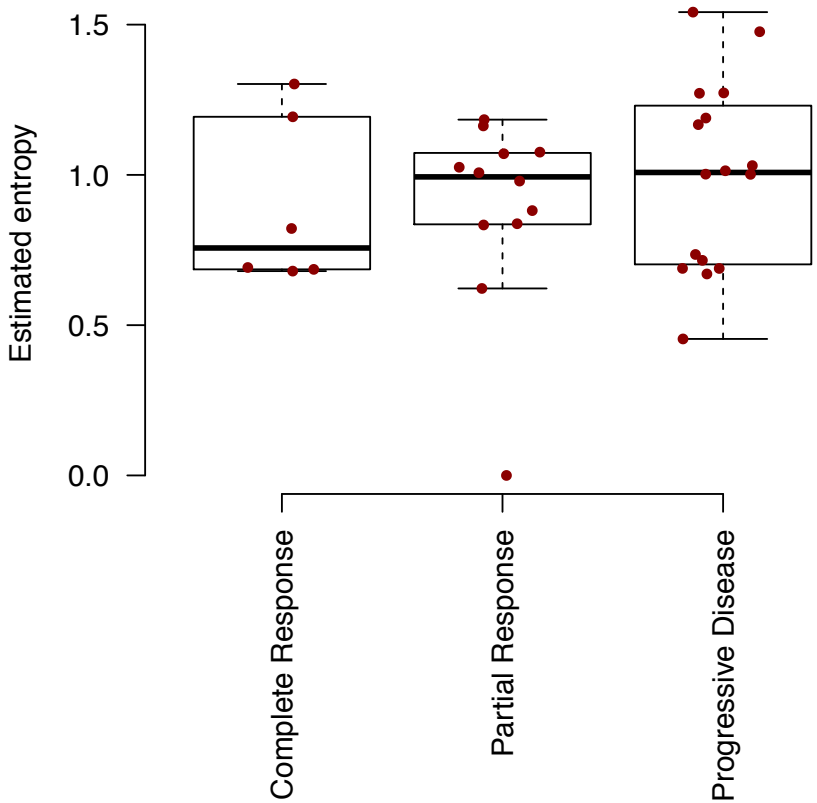
Immunotherapy



- (A-B) Association between median mutation burden and objective response rate of anti-PD1 immunotherapy.
- (C) Observed response rate (Y-axis) versus the fitted values (X-axis) from a linear regression of objective response rate vs. log(mutation burden), tumor purity, ploidy, proportion of clonal mutations, and entropy.

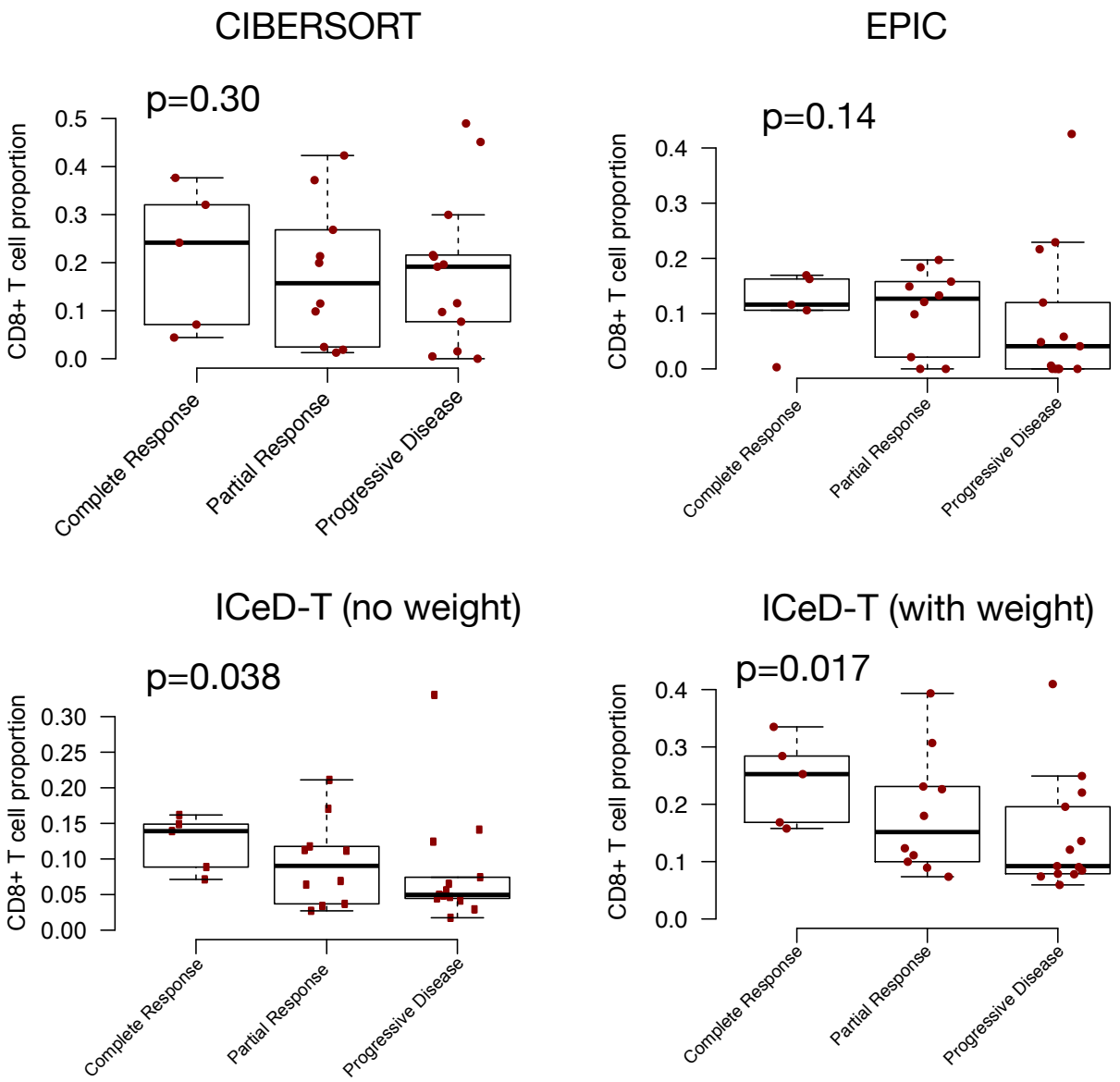
Immunotherapy: Anti-PD1 in melanoma patients





Hugo et al. 2016, *Cell*, **1**65(1), 35–44

Estimation of immune cell composition



Hugo et al. 2016, *Cell*, **165**(1), 35–44

ICeD-T

ICeD-T (Immune Cell Deconvolution in Tumor tissues)

- Assume a regression based framework with gene expression from a tumor sample as response (sample size is the number of genes) and gene expression of each cell type as covariates. The regression coefficients are proportional to cell type proportion.
- Gene expression needs to be log-transformed to stabilize variance, but regression should be conducted using un-transformed expression.
  - Model gene expression using log-normal distribution.
- ICeD-T automatically identifies aberrant genes whose expression are inconsistent with the deconvolution model and down-weights their contributions to cell type abundance estimates.
  - A mixture model to separate the genes into two groups: aberrant genes and consistent genes.

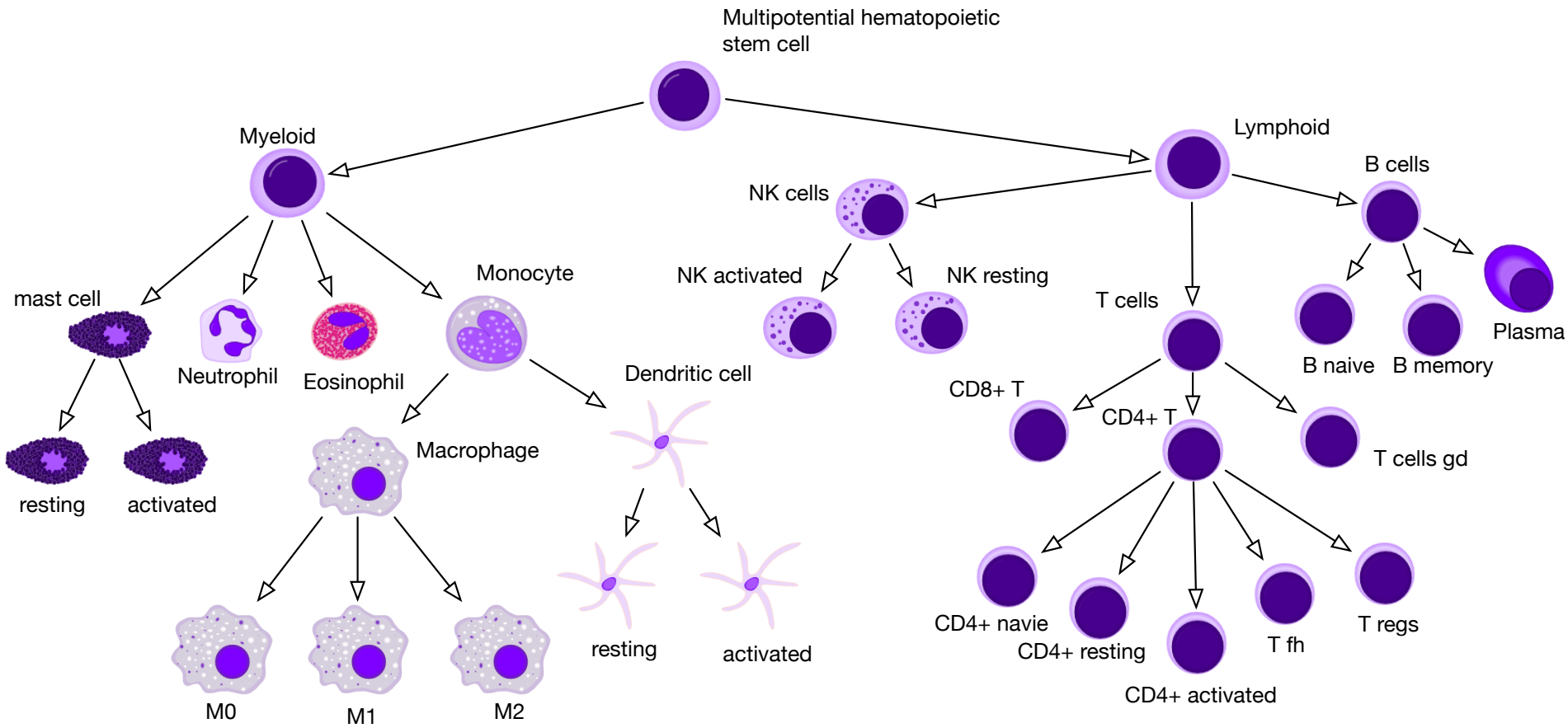
Association analysis for immune cell composition

- We may associate the proportion of each cell type with a response variable, but it is also desirable to consider all cell types together.
  - Account for similarities across cell types.
  - Borrow information of weak associations across cell types.

- Analogous to microbiome studies that use phylogenetic tree of bacteria, we model the dependence of immune cell types using a cell lineage tree, and calculate distance of two samples using the tree.

- Association testing
  - If cell type composition are covariates, we may fit a mixed effect model (kernel based model), using similarity defined by cell type composition to specify covariance structure.
  - If cell type composition are responses, we use a distance-based regression, i.e, to assess whether subjects with more similar cell type composition also have similar covariate values.

Cell lineage tree



Cell lineage tree

Given a immune cell lineage tree, generalized UniFrac distance between two samples, indexed by i and j is defined as:

$$\frac{\sum_{l=1}^{n} b_l (p_{il} + p_{jl})^{\alpha} \left| \frac{(p_{il} - p_{jl})}{(p_{il} + p_{jl})} \right|}{\sum_{l=1}^{n} b_l (p_{il} + p_{jl})^{\alpha}},$$

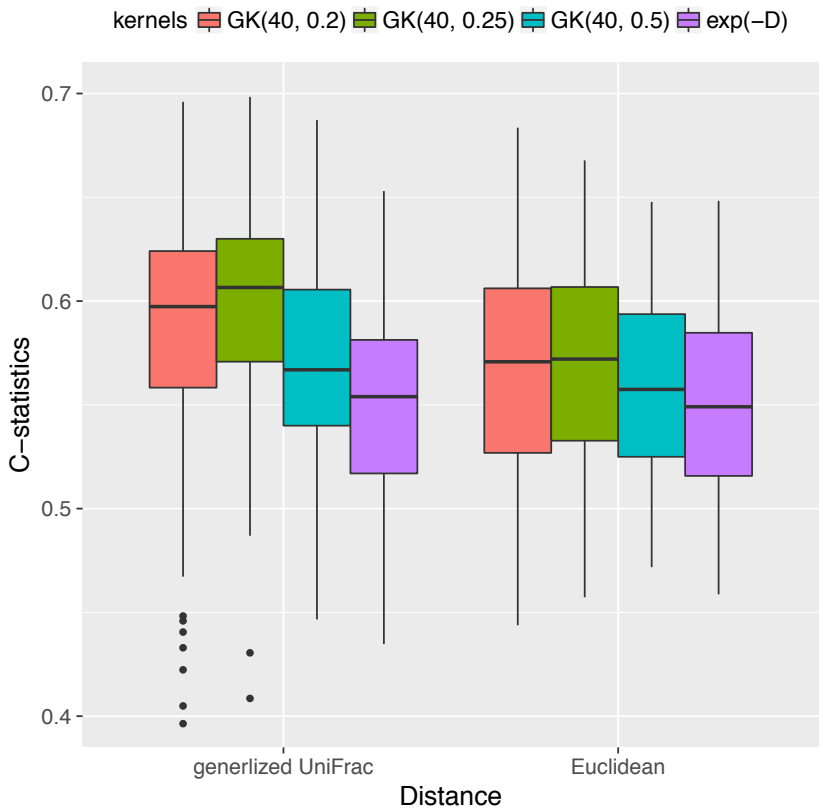
where

- l indexes branches
- p_{il} and p_{jl} denote the proportion of immune cells descending from branch l for samples i and j
- The branch length b_l is calculated based on gene expression of purified immune cells.
- α varies from 0 to 1, to adjust for the relative contribution from the immune cells with larger proportions.

Chen, J., ... & Li, H. (2012). Bioinformatics, 28(16), 2106-2113.

Use immune cell composition to predict survival time in colon cancer patients

We use kernels defined by immune cell composition to fit mixed effect model with survival outcome. Randomly split the data with 2/3 of the samples as training data and 1/3 as testing data.



Summary

Omic data-based biomarkers for cancer immunotherapy.

- DNA level data
  - Intra-tumor heterogeneity (ITH) and SCNA: SHARE, Canopy etc.
  - Neoantigen prediction: deep-learning, neural network, Python...
  - ITH of Neoantigen
  - To combine multiple features into one model
- Gene expression data
  - Immune cell composition estimation: ICeD-T, CyberSORT, TIMER etc.
  - Association with immune cell composition
  - Cell type-specific differential expression using bulk or scRNA-seq data from tumor samples. For example, to whether C8+ T cells in non-responding patients are more exhausted.

Ackowlegement

- Chong Jin, Paul Little, Doug Wilson, UNC Chapel Hill.



- Licai Huang, Fred Hutch.
- Danyu Lin, UNC Chapel Hill.
- Mengjie Chen, The University of Chicago.

Funding: NCI, NIGMS

Ackowlegement

- Chong Jin, Paul Little, Doug Wilson, UNC Chapel Hill.



- Licai Huang, Fred Hutch.
- Danyu Lin, UNC Chapel Hill.
- Mengjie Chen, The University of Chicago.

Funding: NCI, NIGMS

Thank You!