Elucidating gene networks precipitating relapse in CAR T-Cell treated diffuse large B-cell lymphoma

A case for multiple CCA and heirarchical tree trimming

Will Patterson Bo Hu (Advisor)

Cleveland Clinic Lerner College of Medicine

8/28/2020

Outline

Hypothesis: Integration of multi-omics signals will elucidate gene networks involved in DLBCL relapse after CAR T-cell therapy.

1. **What**: Diffuse Large B-Cell Lymphoma
2. **Why**: Approximately half of patients with DLBCL treated with CAR T-cells will relapse
3. **How**: Multi-omics integration using sparse multiple cannonical correlation analysis

What: Diffuse Large B-Cell Lymphoma

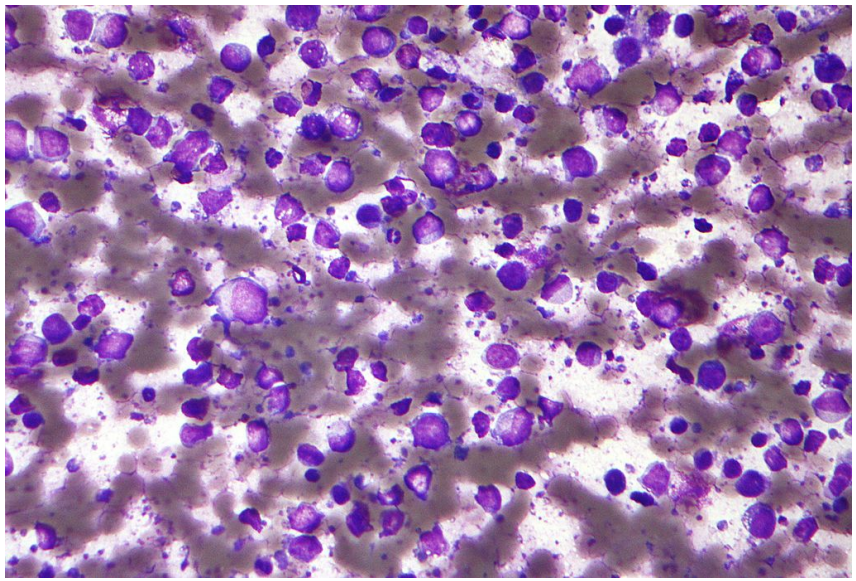

Figure 1: Micrograph of DLBCL (Field stain)

- Most common non-Hodgkin lymphoma (5.6 per 100,000 persons per year), arises from mature B lymphocytes
- Average 5-year relative survival rate of 63% (72% for all NHL)
- **Treatment:** R-CHOP/EPOCH $\rightarrow$ salvage/HCT $\rightarrow$ CAR T-cell

Multiple Canonical Correlation Analysis: Definition

Canonical Correlation Analysis (CCA) finds the relationship between sets of variables by finding their maximally correlated linear combinations.

Given: K sets of observations on same n observations, $\mathbf{X}_1, ..., \mathbf{X}_K$ of dimensions $n \times p_k$, all standardized to mean zero and SD of one

Find: Weights $\mathbf{w}_1, ..., \mathbf{w}_k$, where $\mathbf{w}_k \in \mathbb{R}^{p_k}$, such that the objective function below is maximized

Multiple CCA objective function

$$\text{maximize}_{\mathbf{w}_1,...,\mathbf{w}_K} \sum_{i<j} \mathbf{w}_i^T \mathbf{X}_i^T \mathbf{X}_j \mathbf{w}_j \text{ subject to } ||\mathbf{w}_i||^2 \leq 1, P_i(\mathbf{w}_i) \leq c_i, \forall i$$

(where P_i is the L_1 penalty for i^{th} set)

Extension of sparse mCCA to binary outcomes

Witten and Tibshirani (2009) suggest an extension of sparse mCCA that allows for the incorporation of a two-class outcome. Their method simply treats this $\mathbb{R}^{n \times 1}$ matrix as a third data set. Their objective function takes the form:

Sparse mCCA objective function with binary variables

$$\text{maximize}_{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3} \mathbf{w}_1^T \mathbf{X}_1^T \mathbf{X}_2 \mathbf{w}_2 + \mathbf{w}_1^T \mathbf{X}_1^T \mathbf{y} \mathbf{w}_3 + \mathbf{w}_2^T \mathbf{X}_2^T \mathbf{y} \mathbf{w}_3$$

$$\text{subject to } ||\mathbf{w}_i||^2 \leq 1, P_i(\mathbf{w}_i) \leq c_i, \forall i$$

Extracting gene networks from multiple CCA

Three step process for gene network extraction:

1. Compute the similarity matrix based on the outer products of absolute canonical correlation weights.
2. Apply hierarchical tree cutting to the similarity matrix and extract modules that contain all -omics data types.
3. Visualize networks.

Curse of dimensionality