INTEGRATED CLUSTERING ANALYSIS OF UPCI 08-144 CORRELATIVE STUDIES

by

Surui Hou

BS, MS, University at Buffalo, 2015, 2015

Submitted to the Graduate Faculty of

the Department of Biostatistics

the Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Surui Hou

It was defended on

June 14th, 2018

and approved by

Dianxu Ren, MD, PhD, Associate Professor, School of Nursing, University of Pittsburgh

Hong Wang, PhD, Assistant Professor, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

Ahmad Tarhini, MD, PhD, Cleveland Clinic

Thesis Director: Yan Lin, PhD, Associate Professor, Department of Biostatistics Graduate School of Public Health, University of Pittsburgh

Yan Lin, PhD

INTERGRATED CLUSTERING ANALYSIS OF UPCI 08-144 CORRELATIVE STUDIES

Surui Hou, MS

University of Pittsburgh, 2018

ABSTRACT

Melanoma, which is developed from uncontrolled growth of melanocytes, is one of the most dangerous form of skin cancer and causes thousands of deaths every year. Cure rate for early stage melanoma is fairly high (99% 5-year survival rate), while prognosis of locally and distant metastatic melanoma is very poor. Immunotherapy has shown promising efficacy in late-stage tumors. However, only a portion of the patients respond to the treatment and some experience severe immune-related side effects. Identifying patient subgroups that response differently is of great interest.

UPCI 08-144 is a single arm neoadjuvant ipilimumab clinical trial for patients with American Joint Committee on Cancer (AJCC) stage IIIB-C melanoma. Different types of biomarker data (cytokine data, flow cytometry data, mRNA data and microRNA) were obtained. We applied traditional single level analysis as well as integrative clustering analysis to identify subgroups with distinct genetic character. Although we did not identify subgroups that are significantly associated with either the prognosis or the occurrence of adverse events in patients treated with ipilimumab, partially due to small sample size, we gained valuable experiences in the application of integrated clustering analysis which will guide future study design and practice.

Public health significance: Heterogeneous response to cancer treatment has been a main obstacle in treating cancer patients. Identification of patient subgroups with regard to treatment efficacy and toxicity using various biological data has been a challenging task. Integrative clustering analysis could prove to be a promising tool in solving this piece of puzzle, which could provide a valuable tool in personalized cancer treatment, which will significantly improve the survival of cancer patients.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

My sincere thanks go to Prof. Lin for her patient guidance and support of my work. Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Wang, Prof. Ren, and Dr. Tarhini, for kindly serving as my committee.

1.0 INTRODUCTION

1.1 MELANOMA

1.1.1 Melanoma

In 2015, more than one million people were estimated living with melanoma of the skin in the United States[1]. In 2018, there would be an estimated 91,270 new cases and 9,320 deaths in the United States [1]. Melanoma is developed from uncontrolled growth of melanocytes, which is a melanin-producing cell located in bottom layer of skin and many other parts of human body such as eye, bones and heart [2]. American Joint Committee on Cancer (AJCC) divides melanoma into five stages: Stage 0 to II tumors are confined to skin and have no lymph nodes involvement. During these early stages, tumor stage is differentiated based on the size or thickness. Stage III melanoma cancer cells have spread to nearby lymph nodes, but not to distant organs. Stage IV melanoma is defined as the cancer cells have spread beyond the skin and regional lymph nodes to distant organs such as the liver, lungs or brain, or distant lymph nodes and areas of the skin [3].

Early stage melanoma patients usually have great prognosis while late stage is very deadly. Five-year survival rate for stage I patients is above 90%. However, the 5-year survival rate for stage III C patients dropped to 40% and that for stage IV patients is only about 15% to 20% [3].

Surgery, chemotherapy, radiation therapy, targeted therapy and immunotherapy are currently used treatment for melanoma. Generally, surgery is to excise the local tumor with a 3 to

5 cm circumferential margin of skin [3]. However, late stage melanoma is very unlikely to be curable by surgery. Chemotherapy drugs decrease the tumor growth by interfering with the replication process of cancer cells. Efficacy of traditional chemo therapy on late stage melanoma is limited. Side effects of chemotherapy usually involves vomiting, loss of weight, and even leukemia. Meanwhile, chemotherapeutic resistance is another big concern in melanoma treatment [4]. Target therapy drugs inhibit specific genes or enzymes that are essential to cancer growth and metastasis, and is becoming a promising venue in cancer treatment.

1.1.2 Immunotherapy in melanoma

More than a century ago, people started to explore treatment for cancer with immunostimulants. In 1976, IFNα was reported as the first observed immune component produce melanoma regression. Ten years later IFNα was approved by US FDA. It was the first immunotherapy approved by FDA. Immunotherapy stops the tumor growth by boosting a patient's own immune system. Although T-cell could be activated by some tumor associated antigen and mediate the tumor death process, some tumor cell could still "escape" from the watching of immune system. However, activated T cells could recognized tumor-specific antigens (TSA), the antigens that are present on tumor cell [4]. As a result, the capacity of blocking the immune-escape mechanism by activate T lymphocytes is important [5]. Cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) is one of the best understood inhibitory receptors found on the surface of T-cells downregulates T-cell activation by binding to corresponding molecules to maintain immune homeostasis [6]. Ipilimumab is CTLA4 inhibitor that is for melanoma treatment.

1.2 UPCI 08-144 CLINICAL TRIAL

UPCI 08-144 is a single arm neoadjuvant clinical trial for patients with AJCC stage IIIB-C melanoma. The primary objective of UPCI 08-144 trail was to study the effects of neoadjuvant ipilimumab therapy upon the host immune response in nodal metastatic melanoma and in the peripheral blood comparing baseline and 6 weeks post-treatment. Thirty-five (35) patients met the eligibility requirements were treated with ipilimumab within 1-2 weeks of baseline biopsy. Complete lymph node dissection was performed after second dose of ipilimumab no later than the 6 weeks from baseline time. Two additional doses of ipilimumab were given 3 weeks apart starting 2-4 weeks following surgery. Serum and peripheral blood mononuclear cell (PBMC) was collect at baseline, after 6 weeks of ipilimumab and 3, 6, 9 months. Evidence of immunomodulating on regulatory T cell (Treg), myeloid-derived suppressor cell (MDSC) and effector T cells in the circulation and tumor microenvironment was monitored. [7]

1.2.1 Clinical Results

In 30 months from the last treatment day, 2 complete response, 1 partial response, 8 progressed disease and 22 stable disease were observed as best clinical response. Among 33 evaluable patients, median follow up was 18 months: 16 deaths were observed. 24 progressed disease were used to estimate the relapses free survival. 25 patients experienced clinical benefit defined as complete response, partial response or stable disease.

1.2.2 Correlative Endpoints

The key endpoint of immune response was the peripheral blood intracellular Interferon gamma (IFNγ) frequency characterized by flow cytometry among CD8+, CD4+ and CD3+ cells. IFNγ expression is associated with autoinflammatory and autoimmune diseases, and one of the predominant generate position is CD4 Th1 and CD8 cytotoxic T lymphocyte effector T cells [8]. Therefore, the frequency of IFNγ could be an effective measurement of immune response. Also, MDSC level was evaluated. A significant decrease in MDSC was observed. Protein profiling data was used to evaluate the utility changes in the levels of candidate disease serum biomarker.

The association of biomarker and clinical response was also characterized. Decrease of MDSC Lin12/HLA-DR2/CD33+/CD11b+ and significant increase in Treg was observed and associated with improved relapse-free survival [7]. In addition, mRNA and microRNA microarray data are also available in a subset of the patient samples.

1.3 CHALLENGES AND MOTIVATIONS

One of the biggest concerns of immunotherapy was activation of T cells that cross-react with normal cells [9]. This can cause a severe side effect in some patients. Both risk of side effect and response to the treatment could be very heterogeneous. As a result, identification of patients who are more likely to respond the treatment or experience severe side effect is very important. It is believed that genetics plays an important role in both aspect. Clustering of genetic data has been successfully applied in identifying subgroup with distinct clinical character [10].

In this study, we applied clustering analysis on patient's protein and genetic profiling data from trial UPCI 08-144 to explore the association of patient's genetic character and their clinical performance. UPCI 08-144 trial is an early phase trail with very limit patients. Multiple levels of genetic data were collected in UPCI 08-144. The small sample size and large inter-subject variability was the greatest challenge. Therefore, in addition to the clustering analysis of each type of data, we also attempted to borrow information from different levels of genetic data through integrative clustering analysis to decipher the substructure of the patients. Three different integration strategies were used: a) simply merge all the markers in one big dataset; b) ad-hoc integration of single level clustering analysis; and c) integrative analysis using a latent model [11]. Finally, we explored clinical indication of the population structure by testing the association between the cluster membership and patient prognosis and treatment side effects.

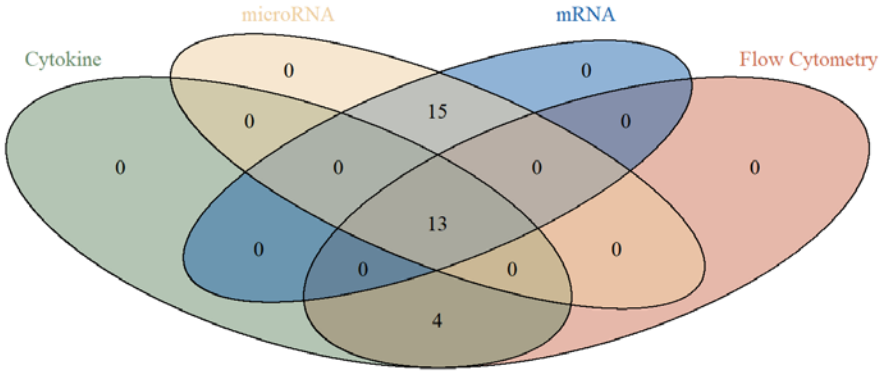
Although none of the clusters was significantly associated with the clinical endpoints, we did learn some important lessons in this practice. We observe that strategy a) and c) results are consistent to each other. This means, although, given the small sample size, the model-based method didn't gain much more power comparing to the simple analysis, it is reassuring that it did agree with simpler method, meaning that the method is applicable to small samples. With the ability of handling different types of endpoints (i.e. a mixture of categorical and continuous measurements), this method is still appealing in the integrative clustering of small data. Interestingly, although the pre- and post- treatment clustering pattern differ, the feature selected by the lasso- based method have quite some overlap. This may indicate that these are the biomarkers with some biological significance.

2.0 METHOD

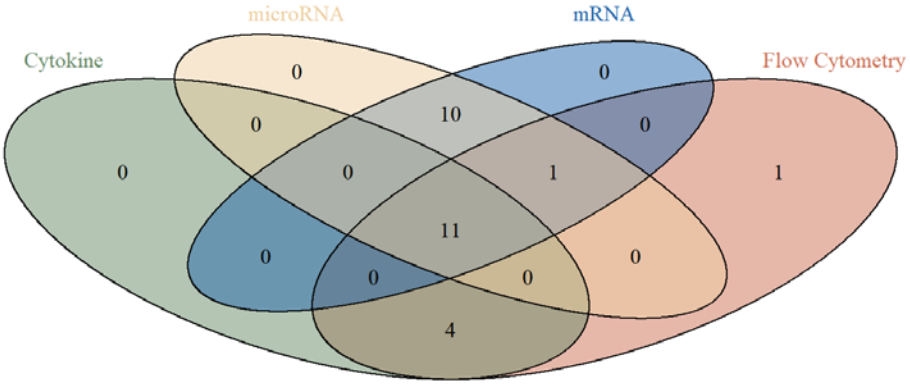
2.1 SAMPLE

Clinical trial UPCI 08-144 tumor biopsy data were collect at baseline and 6 weeks. we obtained cytokine, flow cytometry, mRNA and microRNA data sets. As shown in Figure 1, at baseline, 17 patients had records on both cytokine data and flow cytometry data; 28 patients had both mRNA and microRNA data. Only 13 patients had all four levels data sets. At week 6, 15 patients had both cytokine data and flow cytometry data; 22 patients had both mRNA and microRNA data. Only 11 patients had all four levels of data.

All data sets were normalized before analysis. Log intensity of mRNA and microRNA data were obtained from Robust Multiarray Average (RMA). We further applied coefficient of variation (CV) to exclude genes with little information. Based on the distribution of feature's CV we arbitrarily selected different cut-offs for each dataset. Approximately 5000 mRNA probesets (4992 and 5000 for pre- and post-treatment respectively) were selected for each timepoint. A total of 5433 miRs and 4992 miRs whose coefficient of variance lager then 0.4 and 0.5 respectively were included in the clustering analysis at baseline and week 6.



(A) Pre-treatment



(B)Post-treatment

Figure 1: Venn diagram of subject in each data type.

2.2 CLUSTERING METHOD

2.2.1 K-Means Clustering

k-means is one of the most commonly used unsupervised clustering methods. It clustered *M* points in *N* dimensions into *K* clusters by minimize the within cluster sum of squares (WCSS) [12]:

$$WCSS = \sum_{k=1}^{K} \frac{1}{n_k} \sum_{i,i' \in C_k} \sum_{j} \left(X_{ij} - X_{i'j}\right)^2$$

where n_k is the number of subjects in cluster k and $C_k$is the set for index of cluster k. In our work, we used the k-means function from R stats package.

2.2.2 Model Based Clustering

We applied model-based clustering based on a finite Gaussian mixture model.[13] Observations are considered coming from a mixture of more than one Gaussian distributions. Each cluster is characterized by its mean vector (μ_k), covariance matrix (Σ_k) and each observation have a probability of belonging to each cluster (π_k). For $n \times p$ dimensional observations $x_1, \dots, x_n$ the mixture model is

$$f(x|\pi, \mu, \Sigma) = \prod_{i=1}^{n} \sum_{p=1}^{G} \pi_k f_k(x_i|\mu_k, \Sigma_k)$$

Where G is the number of components, $f_k(x_i|\mu_k, \Sigma_k), k = 1, \dots G$ is the probability density function of each components. The model parameters are then estimated using the Expectation-Maximization (EM) algorithm [14]. Observations are assigned to the component has the maximum posterior probability. When assuming spherical covariance structure (i.e. independent and equal

variance between clusters), model-based clustering is equivalent to the k-means clustering. The Bayesian Information Criterion (BIC) could be used to select the best model and number of clusters.[14] In our case, however, due to the small data size, we set the number of clusters to 2 and 3 for our analysis. In our work, R package *mclust* were used for model-based clustering analysis.

2.2.3 Sparse k-means clustering

Sparse k-means clustering is a cluster method that allows simultaneous feature selection and cluster identification. To achieve this, sparse k-means optimizing a weighted WCSS subjecting to both L1 and L2 penalty terms [15]. Based on the contribution on maximization of between cluster sum of square, each feature is assigned a weight. Features given a large between cluster sum of square will be assign a large weight [15]. It is very attractive when the number of features is much more than number of observations (p>>n): firstly, it reduced the "noise" by selecting only the features that "drove" the clustering pattern; secondly, it provides a list of potential biological biomarkers. In our work, R package *sparcl* were used for sparse k-means clustering analysis.

2.2.4 iClusterPlus

iClusterPlus is a package offering integrative cluster analysis of multi-type genomic data. A joint latent variable model-based clustering method was applied. A latent variable $z = (z_1 \dots z_{k-1})'$ connect different set of data sets and representing the clusters. *Z* was jointly estimated using different data.

$$X_1 = W_1 Z + \varepsilon_1$$

$$X_2 = W_2 Z + \varepsilon_2$$

...

$$X_m = W_m Z + \varepsilon_m,$$

where m is the number of genomic data types available for the same set of samples. The independent error terms in which each has mean zero and diagonal covariance matrix, that is $\varepsilon \sim N(0, \Psi)$, $\Psi = (\Psi_1, \cdots \Psi_{\Sigma_i P_i})$. Ψ_i represents the remaining variances unique to each data type after accounting for the correlation across data type. $w = (w_1, \ldots, w_m)$ is the coefficient matrices. Under the assumption of $Z^* \sim N(0, I)$, $X = (X_1, \ldots, X_m)$ follows a multivariate normal distribution with mean zero and covariance matrix $\Sigma = ww' + \Psi$. The log likelihood function is[11]:

$$l(W, \sum) = -\frac{n}{2}(\sum_{i=1}^{m} p_i \ln(2\pi) + \ln \det(\sum) + tr(\Sigma^{-1} G)),$$

where G is the sample covariance matrix. EM algorithm is used to obtain the MLE estimates. In addition, a L1 penalty term is used to select features:

$$l_{c,p}(w_1, \psi) = l_C(w_1, \psi) - J_\lambda(W),$$

$$J_\lambda(W) = \lambda \sum_{i=1}^{m} \sum_{k=1}^{K-1} \sum_{j=1}^{p_i} |w_{ijk}|.$$

2.2.5 **Co-membership Matrix**

Co-membership Matrix is an intuitive way of summarized cluster membership of multiple clustering arrangement into one. For a *n* observation × *d*-dimensional data *X*, with fixed *k* > 0, let ρ (*X*, *k*) be a *k*-cluster arrangement of *X*. A $n \times n$ co-membership matrix *D [*ρ (X, *k*), X*]* corresponding to one cluster arrangement has entries

$$D_{ij}=\begin{cases}1 & if\ Xi\ and\ Xj\ were\ in\ the\ same\ cluster\\ 0 & otherwise\end{cases}$$

For more than one arrangement, D_{ij} would be the total number of Xi and Xj in the same cluster among all the cluster arrangements.

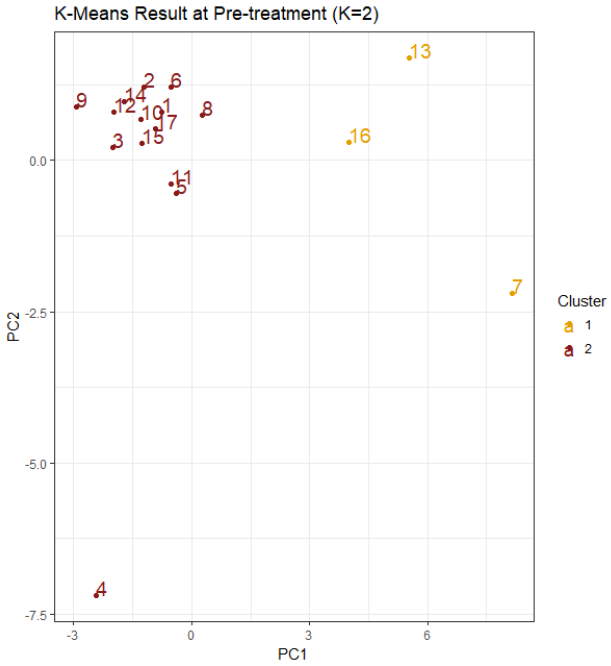
3.0 RESULTS

3.1 SINGLE LEVEL CLUSTERING

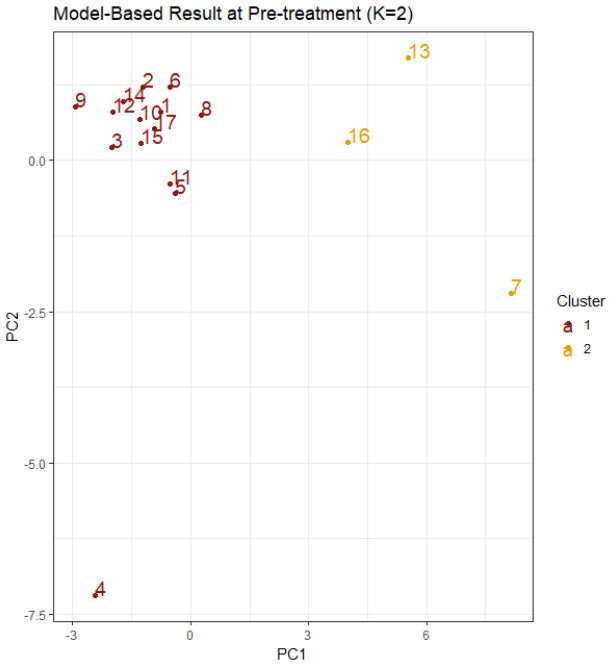
First, we cluster the subjects using each type of biomarker data separately at each time point. All available data (not only the overlapped subjects) at each level were used. For data with very high dimensions, i.e. the microarray data, markers were pre-selected by the observed information, measured by coefficient of variation, in the dataset.

3.1.1 Clustering of Cytokine Data

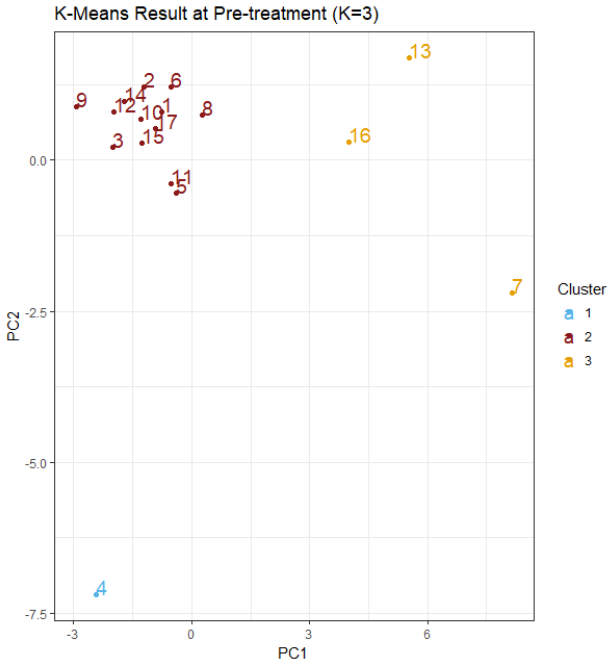
We applied k-means and model-based clustering algorithm on both pre-treatment and post-treatment cytokine data. Based on the expression of 34 markers, 17 patients (patient ID 1 to 17) were clustered into three groups pre-treatment. These clusters are shown in Figure 2. Gap statistics were initially applied for choosing the number of clusters, however, the results are not conclusive due to the extreme small number of subjects. We arbitrarily set k to 2 and 3 for the analysis. K-means and model-based methods resulted in the same clusters at both timepoint when k=2 (Figure 2A and 2B). When k is set to be 3, the results are slightly different at baseline (Figure 2C and 2D) due to the assumption of different variance structure. The clustering results of the post-treatment samples, on the other hand, are consistent between the two methods for both k=2 and k=3 (Figure 2E and 2F). For simplicity, we used k-means methods for the remaining analysis.


K-Means Result at Pre-treatment (K=2)
PC2
PC1
Cluster
1
2

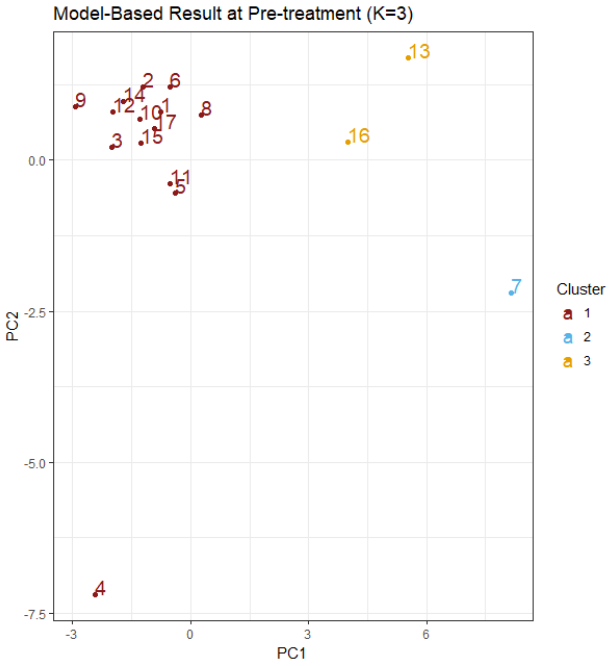
(A)


Model-Based Result at Pre-treatment (K=2)
PC2
PC1
Cluster
1
2

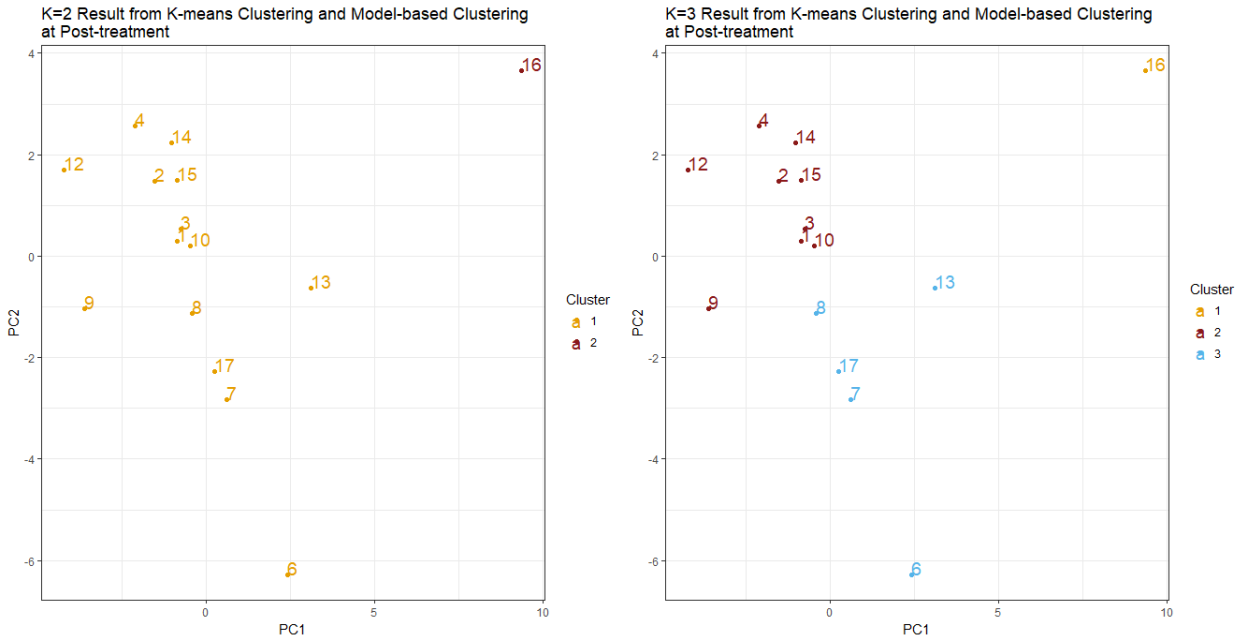
(B)


K-Means Result at Pre-treatment (K=3)
PC2
PC1
Cluster
1
2
3

(C)


Model-Based Result at Pre-treatment (K=3)
PC2
PC1
Cluster
1
2
3

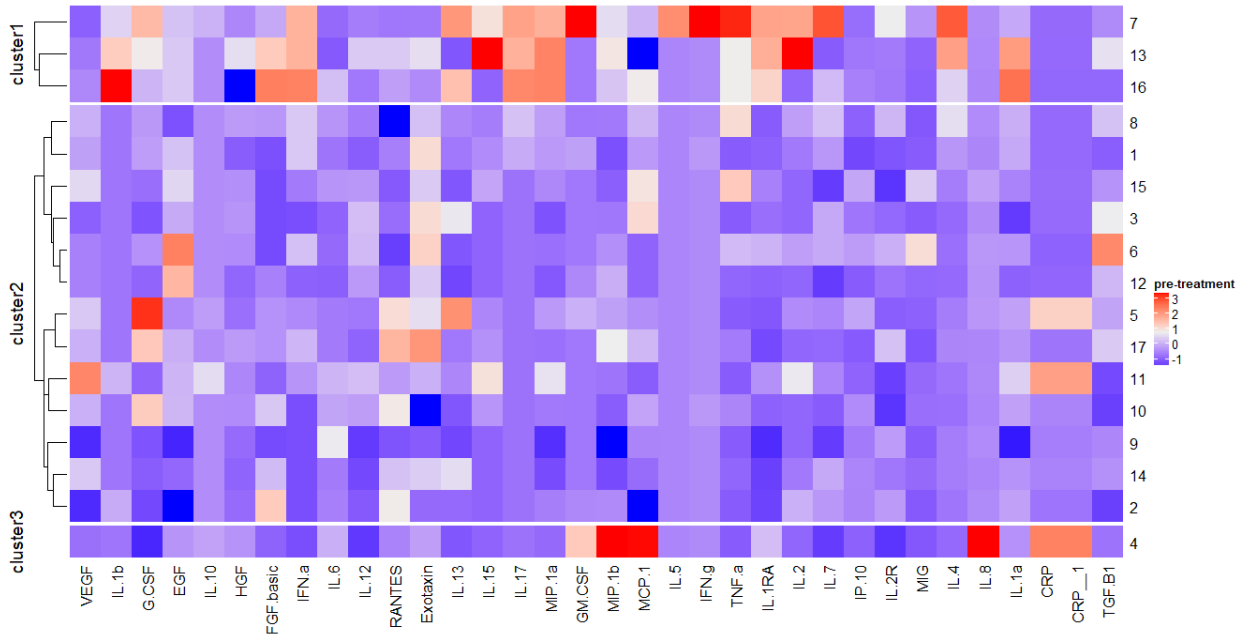
(D)



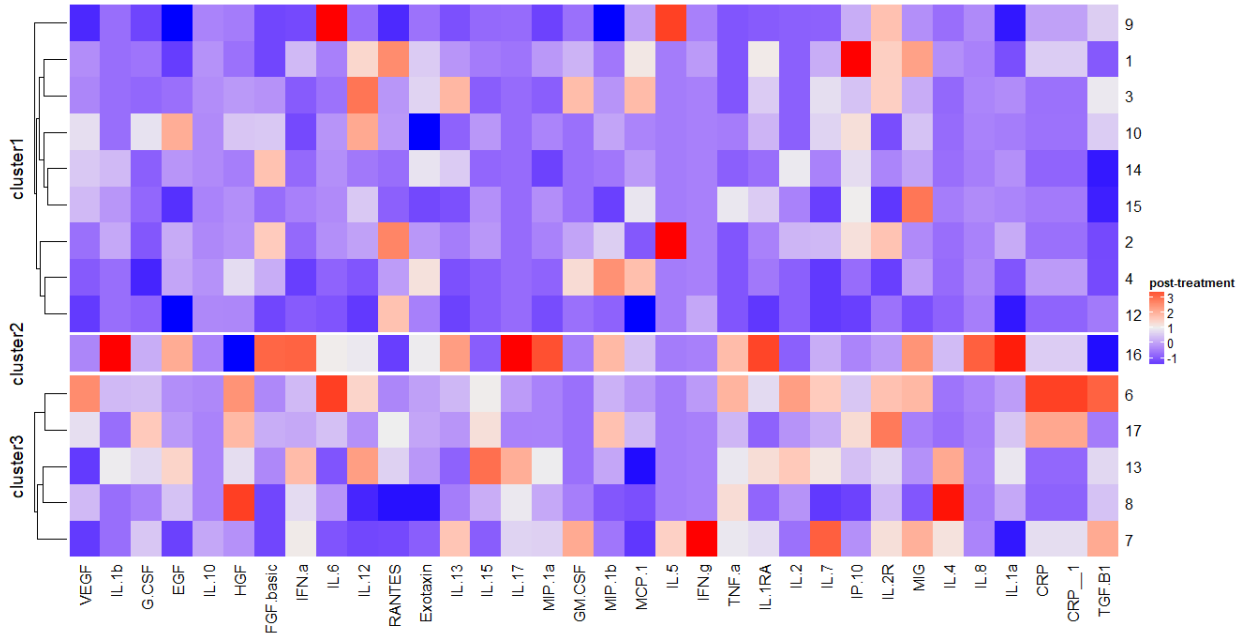
(E) (F)

Figure 2: PCA plots of the cytokine clustering results.

Pre- and Post-treatment clustering patterns are quite different (Figure 3). When cluster number is set to 3, at baseline (n=17), subject 4 turned out to be an outlier and formed a cluster by himself, subject 7, 13, 16 was grouped together forming a cluster, the rests 13 subjects were tightly clustered together and formed the biggest cluster. Subjects 5 and is 11 do not have post-treatment cytokine data. Therefore n=15 for post-treatment clustering: subject 16 formed one cluster, subjects 6, 7, 8, 13, 17 are clustered together as a second cluster, while the rest 9 subjects formed the more homogenous cluster. When k is set to 2, at baseline, subject 7, 13, 16 again formed a cluster, the rests 14 subjects were tightly clustered together. At post-treatment, subject 16 formed one cluster by itself, and the rests 14 subjects were clustered together (Figure 4). When k is set to 3, cluster patterns were not striking at both time point. Therefore, k=2 would be a more reasonable setting.

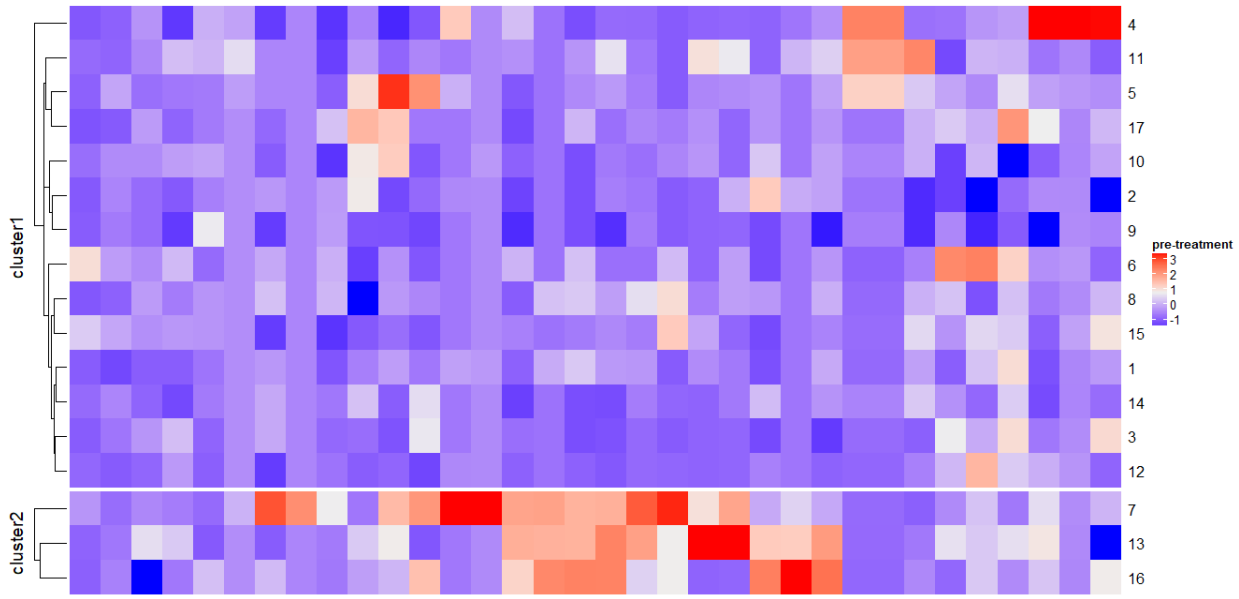


(A) Pre-treatment clustering result

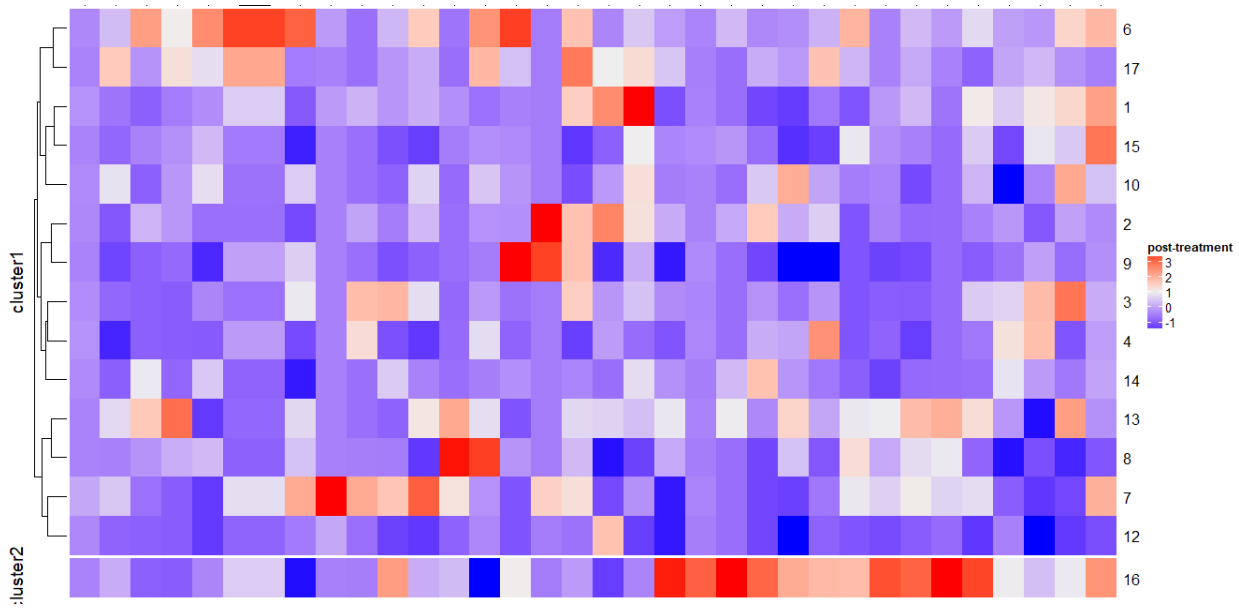


(B) Post-treatment clustering result

Figure 3: Heatmaps of k-means clustering results (k=3) using all cytokine features.



(A) Pre-treatment clustering result



(B) Post-treatment clustering result

Figure 4: Heatmaps of k-means clustering results (k=2) using all cytokine features.

We next used sparse k-means approach which allows us to simultaneously select features while clustering the samples, to cluster the cytokine data at each time point. Many features returned a non-zero weight, however, quite a lot of them are fairly close to 0 (Figure 5). Therefore, we

further used 0.18 as a cutoff point to screen out further noises (Figure 5). The clustering results of the sparse k-means and k-means are consistent to each other, indicating that the clusters are driven by a small number of cytokines. Interestingly, although the cluster patterns of the two time points are quite different, the selected features largely overlapped. Among the 10 and 8 selected features at pre- and post-treatment, 6 features are selected in both time points: IL.1b, IFN.a, IL.17, MIP.1a, IL.1RA, IL.1a. Using these selected cytokines, we re-clustered the samples at baseline (Figure 6A) and post-treatment (Figure 6B), the results are the same as using the whole panel of 34 cytokines.

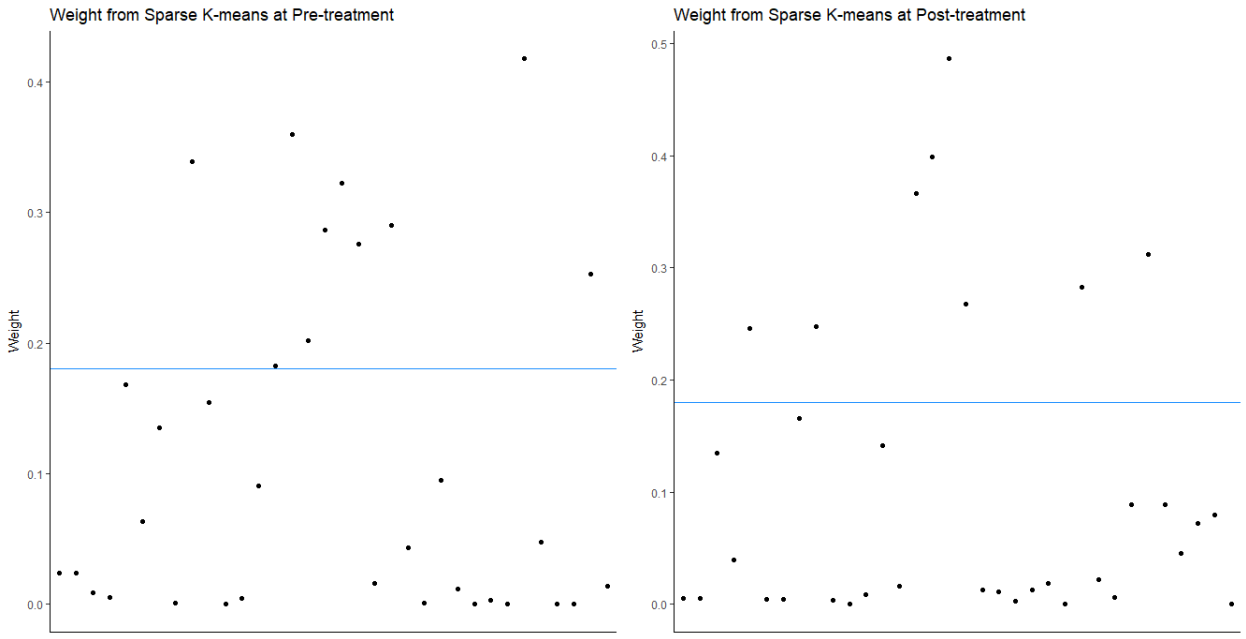
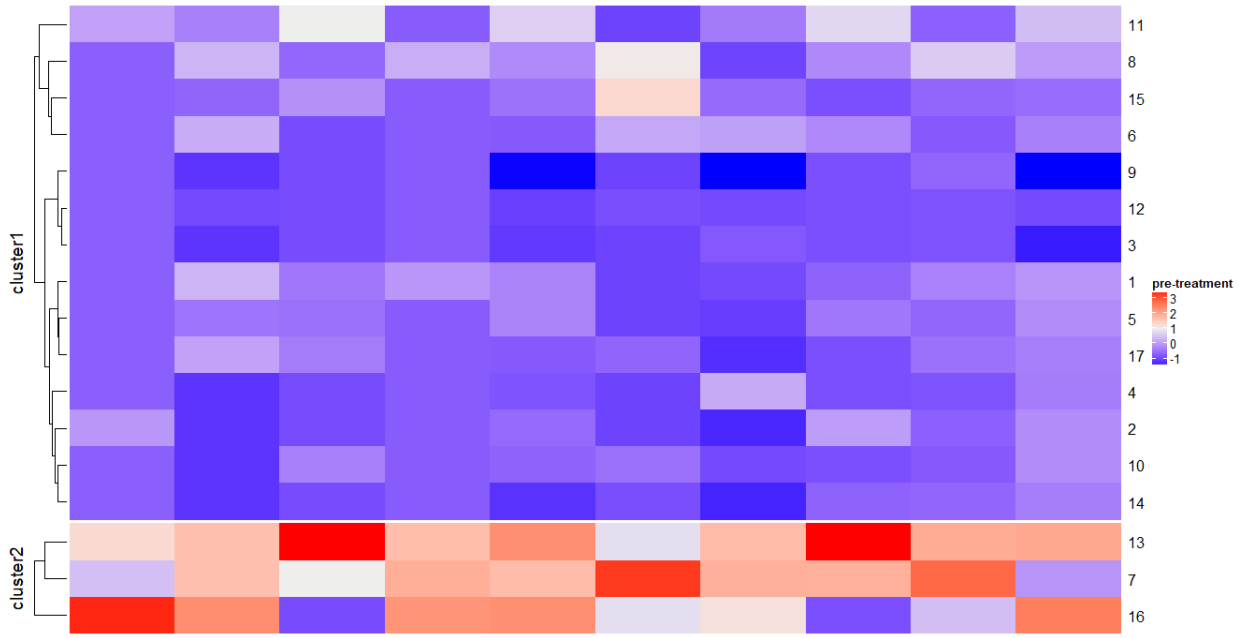
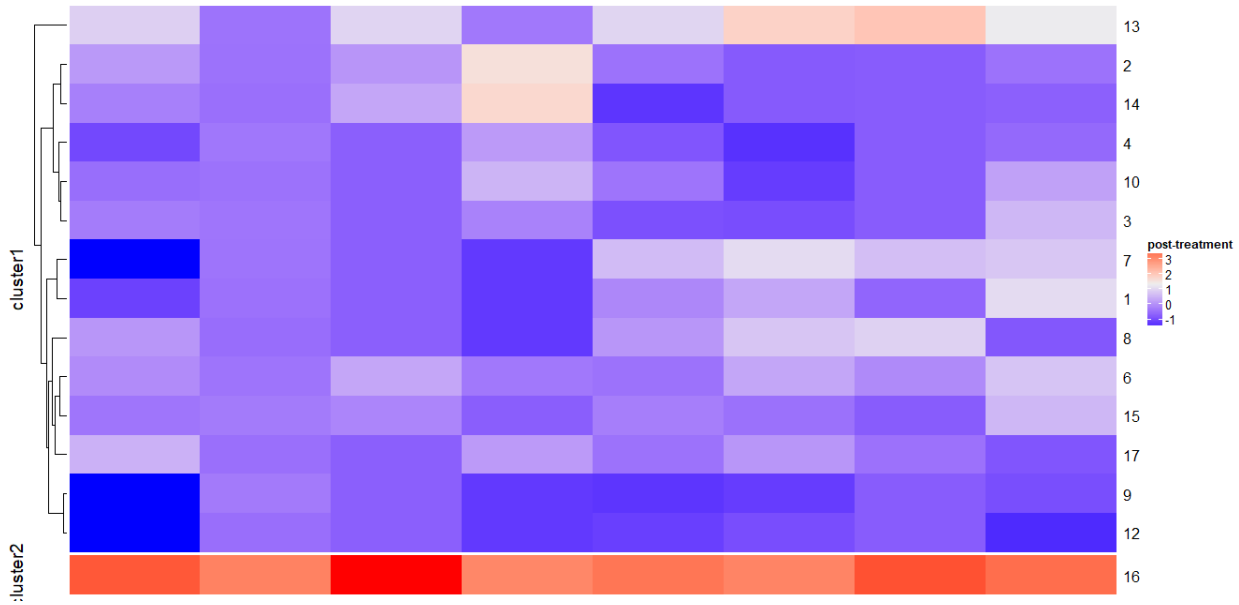


Figure 5: Cytokine feature weight given by sparse k-means clustering.



(A) Pre-treatment clustering result

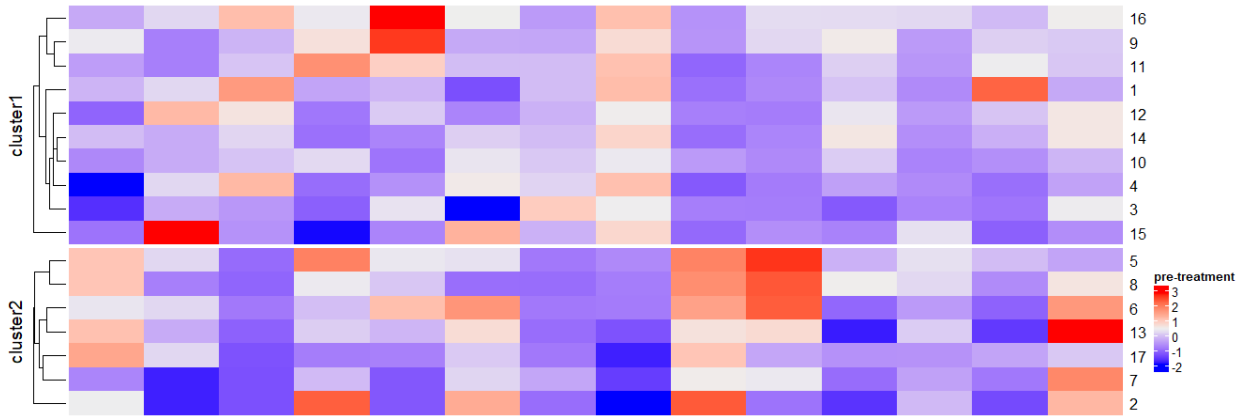


(B) Post-treatment clustering result

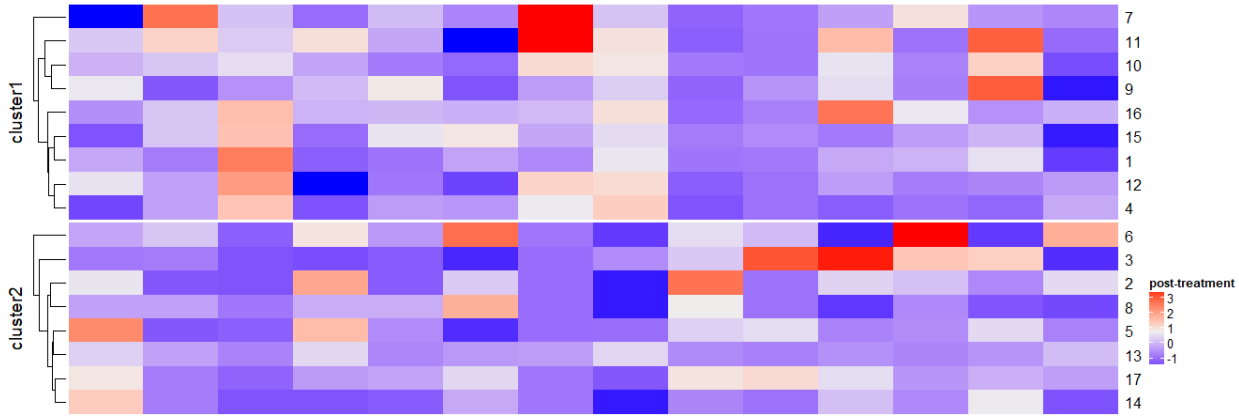
Figure 6: Heatmap of k-means clustering using high weight cytokines.

3.1.2 Clustering of Flow Cytometry Data

Using 16 markers, 17 subjects were clustered in two groups at both timepoints by k-means clustering. Similar to the cytokine clustering results, the pre-treatment and post-treatment cluster pattern are very different (Figure 7). Subject 1, 3, 4, 9, 10, 11, 12, 14, 15, 16 were clustered together, and subject 2, 5, 6, 7, 8, 13, 17 were clustered together at pre-treatment, while subject 2, 3, 5, 6, 8, 13, 14, 17 were clustered together and subject 1, 4, 7, 9, 10, 11, 12, 15, 16 clustered together post-treatment.



(A) Pre-treatment clustering result



(B) Post-treatment clustering result

Figure 7: Heatmaps of k-Means clustering results using all flow cytometry markers.

We also applied sparse k-means and used the arbitrary cut off point of 0.18 to select the most informative markers (Figure 8). Again, we observe that k-means and sparse k-means gave almost identical results with one subject (id=9) moved from one cluster to the other at post-treatment (Figure 9). Similar to the case of the cytokine data, most of the informative features at the two time points overlapped. Out of 5 selected features at both each time point, the following 4 features were overlapped: Lymphocytes, lym_Lin1/HLA/DR/CD33, mono_Lin1/HLA/DR/CD33, and mono_Lin1/HLA/DR/CD11b.

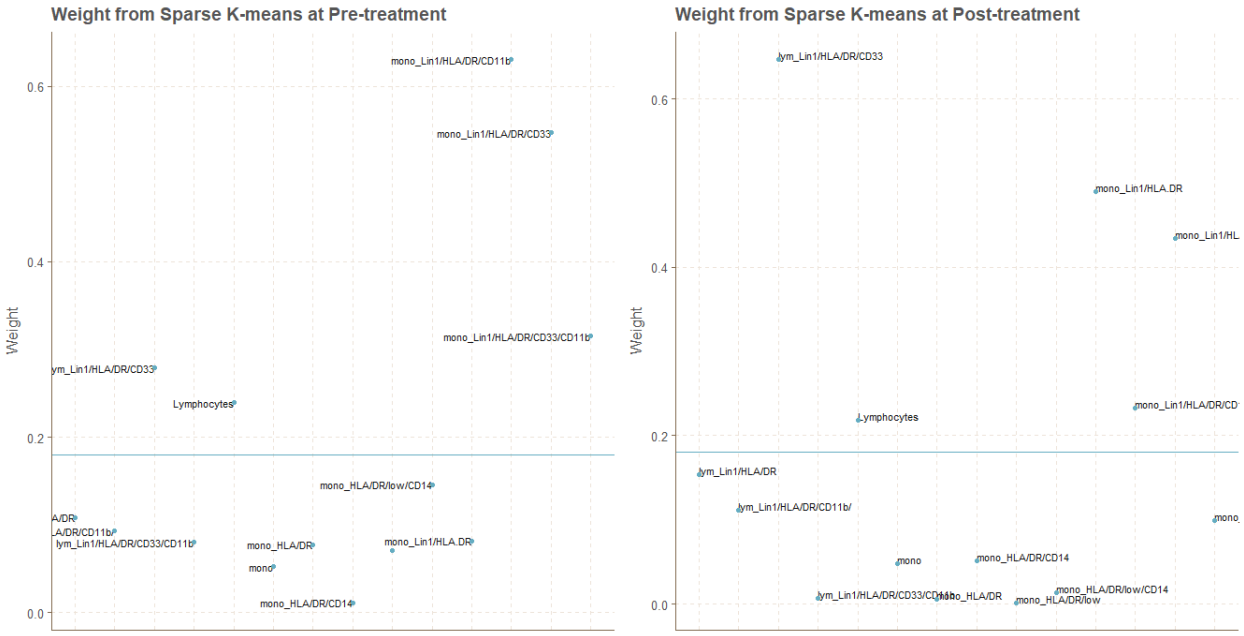
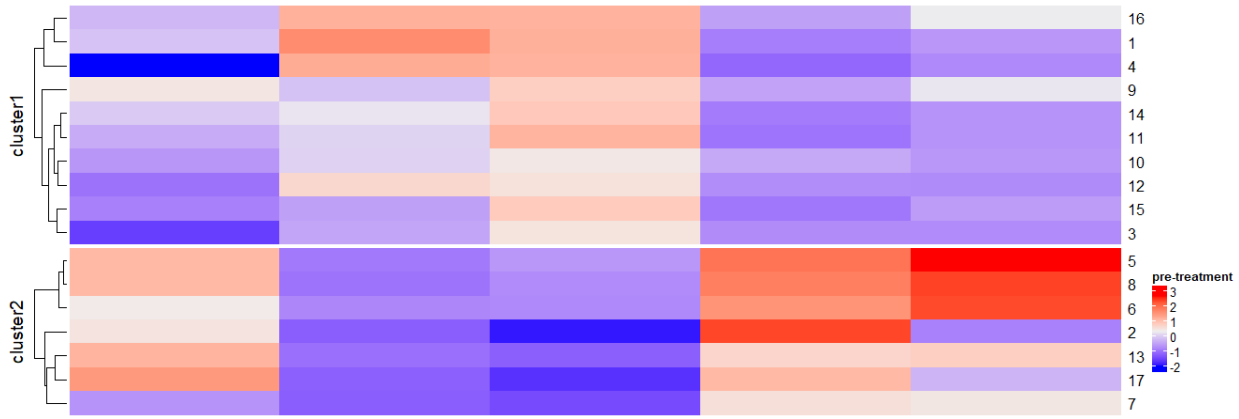
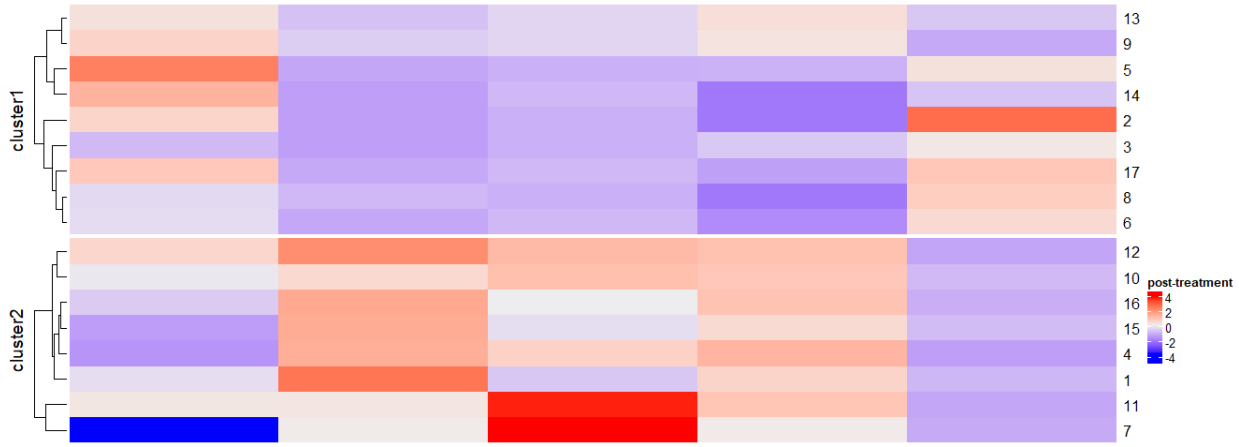


Figure 8: Flow cytometry feature weight given by sparse k-means clustering.



(A) Pre-treatment clustering result.



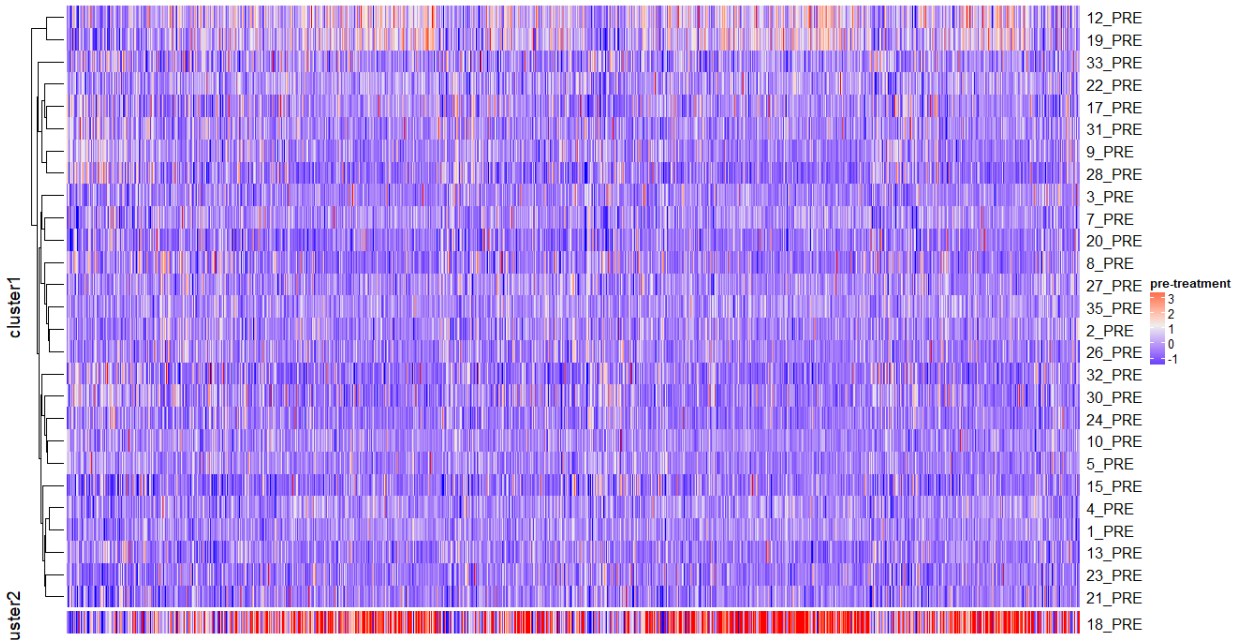
(B) Post-treatment clustering result

Figure 9: Heatmap of k-means clustering using high weight flow cytometry features.

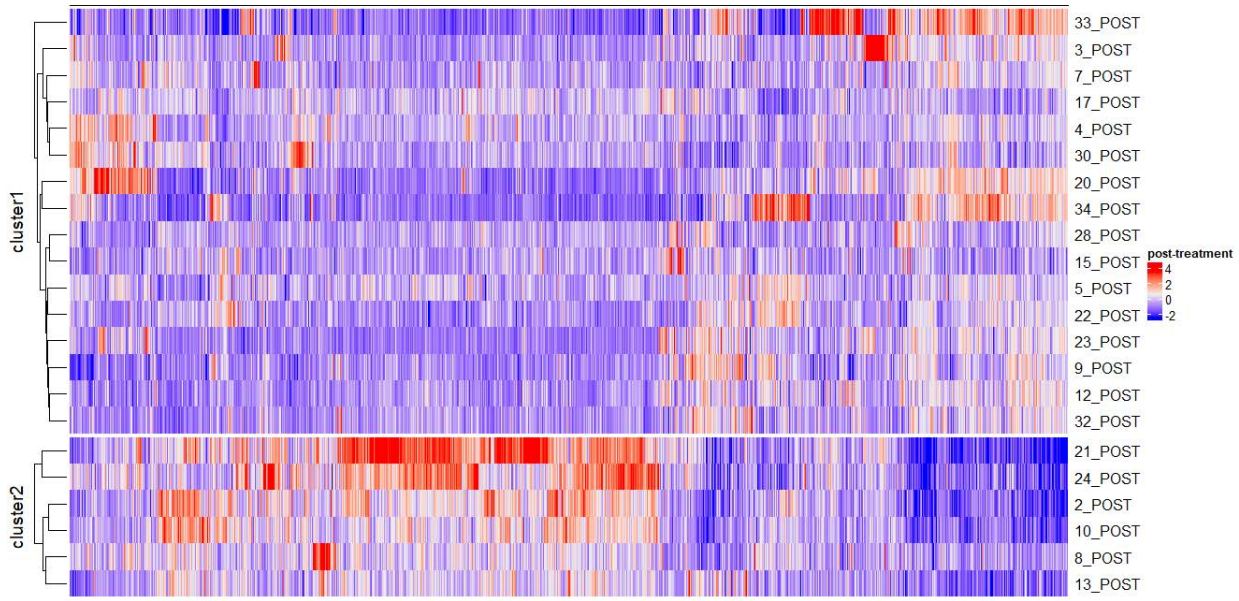
3.1.3 **Clustering of mRNA Data**

U133A arrays contain over 22000 probesets. To reduce background noise, we first removed all probesets with coefficient of variation (CV) under 0.2 in our data. Approximately 5000 probesets (4992 and 5000 for pre- and post- treatment respectively) were left for each timepoint.

The clustering pattern were very different between pre- and post-treatment (Figure 10). At baseline, mRNA data were available for 28 subjects. Using all 4992 probesets, subject 18 turned out to be an outlier and formed a cluster by himself, the rest 27 subjects formed a homogenous cluster (Figure 10A). However, subject 12, and 19 have slightly different pattern from the rest subjects in the large cluster. Twenty-two (22) post-treatment samples were available for mRNA profiling, subject 2, 8, 10, 13, 21, 24 were clustered together and; subject 3, 4, 5, 7, 9, 12, 15, 17, 20, 22, 23, 28, 30, 32, 33, 34 were cluster together (Figure 10B).



(A) Pre-treatment clustering result



(B) Post-treatment clustering result

Figure 10: Heatmaps of K-Means Clustering Results Using all mRNA probesets.

Sparse k-means gave the same clustering result as k-means at post-treatment, but moved subject 12, and 19 from one cluster to the other at baseline, which reconfirmed our observation in k-means result. We did not observe any clear gap for feature selection at either at pre- or post-treatment (Figure 11). Therefore, we use the third quantile of feature weight (0.019 and 0.018) as the cut-off (Figure 11). K-means clustering results using the selected features at each time points are presented in Figure 12.

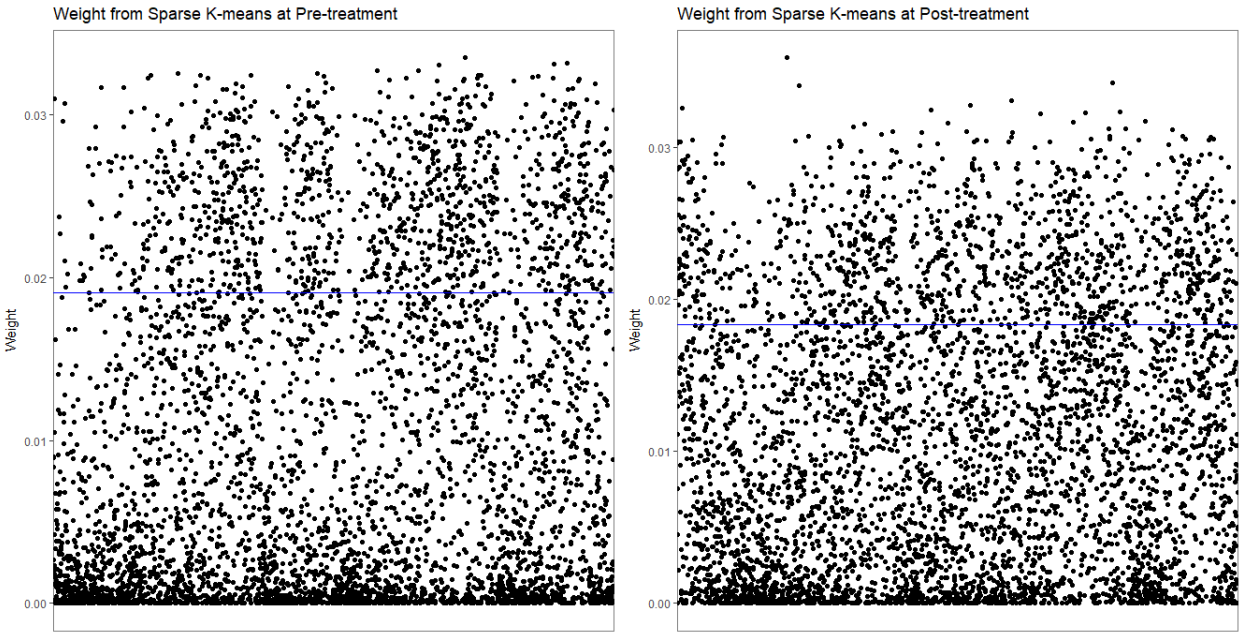
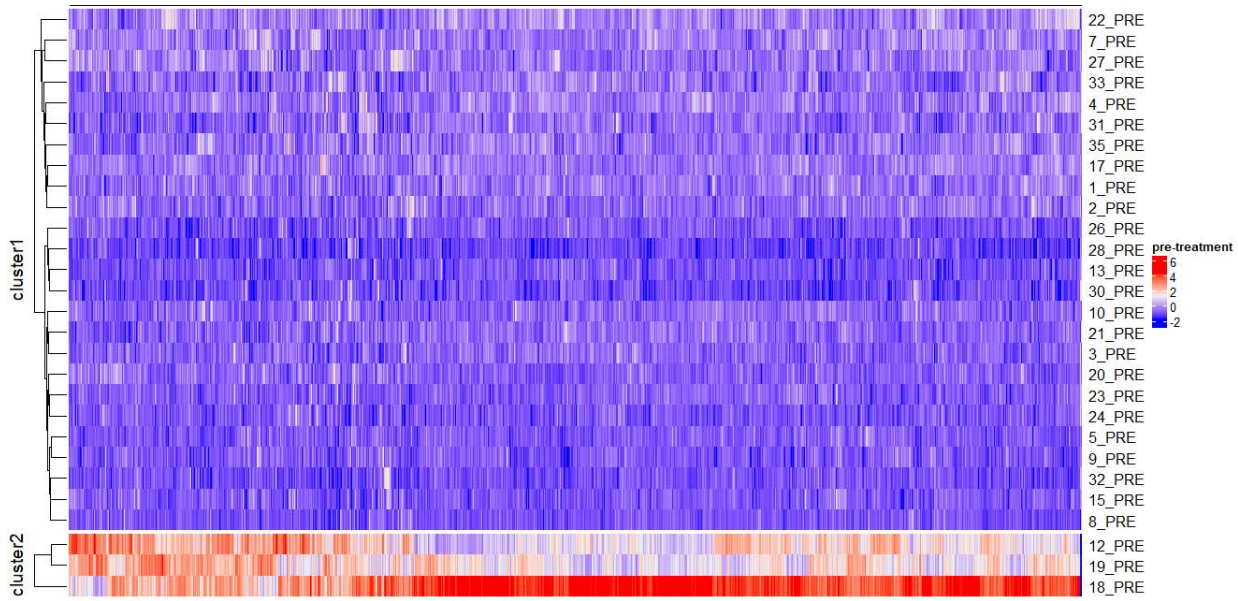
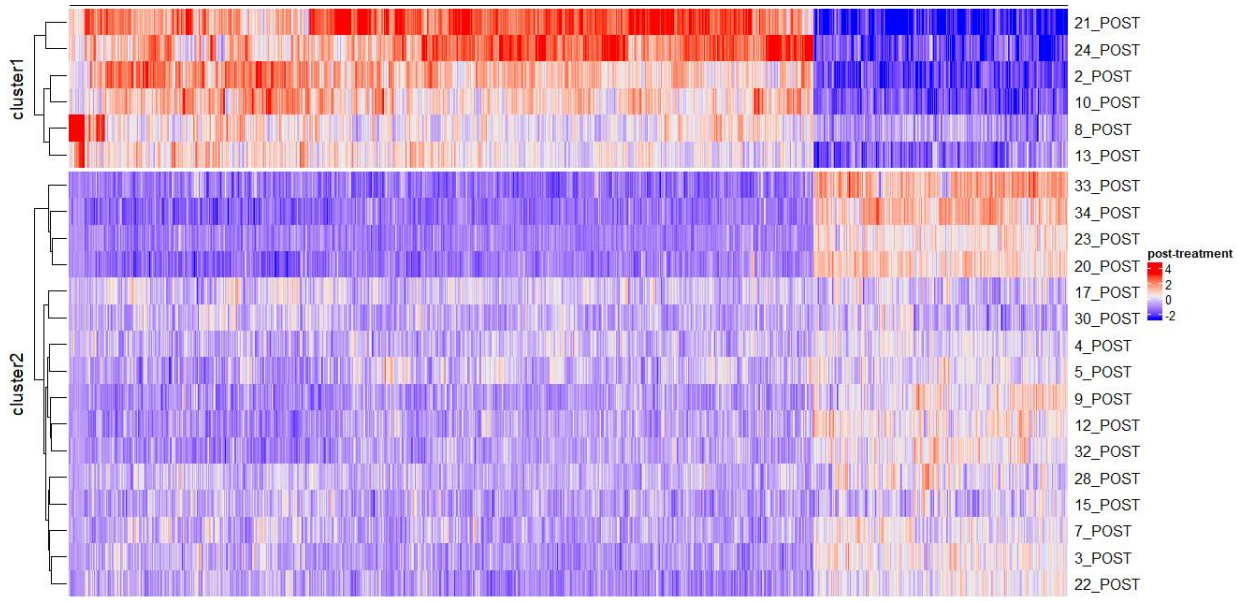


Figure 11: mRNA feature weight given by sparse k-means clustering.



(A) Pre-treatment clustering result

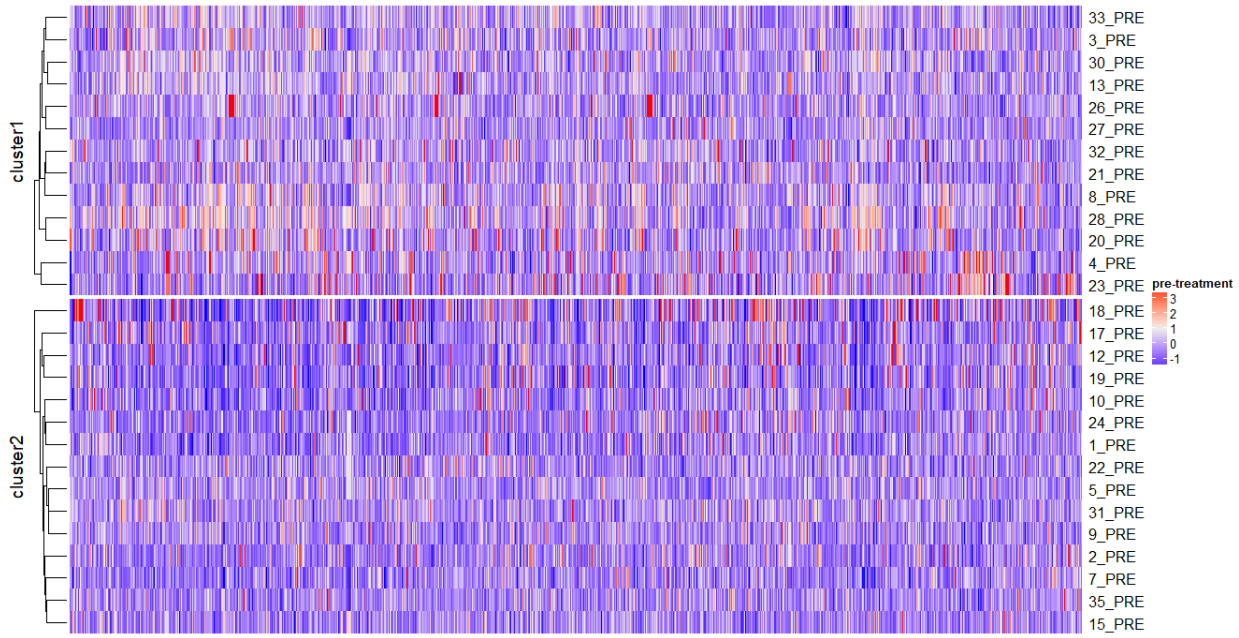


(B) Post-treatment clustering result

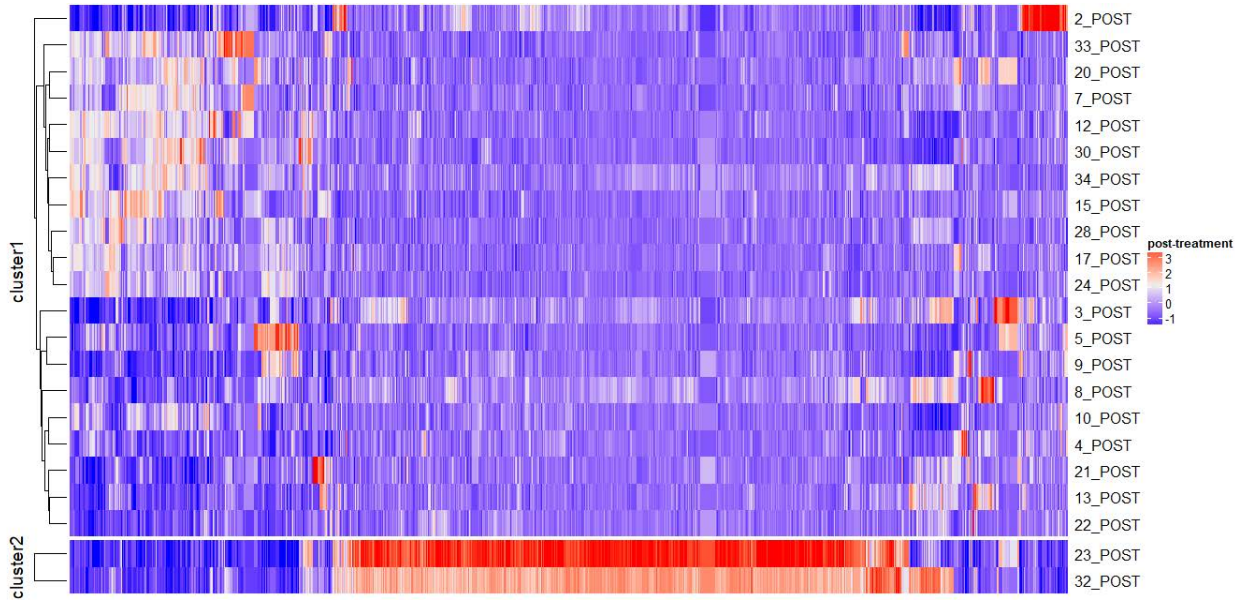
Figure 12: Heatmap of k-means clustering using high weight mRNA features.

3.1.4 **microRNA (miR) Clustering Analysis**

A total of 5433 miRs and 4992 miRs whose coefficient of variance lager then 0.4 and 0.5 respectively were included in the clustering analysis at baseline and post-treatment. At baseline, Out of 28 subjects, k-means and sparse k means both clustered subject 3, 4, 8, 13, 20, 21, 23, 26, 27, 28, 30, 32, 33, together and the rest 15 subjects formed the second cluster (Figure 13A). Among the 22 subjects with post-treatment microRNA data. Subject 23 and 32 were clustered together and subject 2, 3, 4, 5, 7, 8, 9, 10, 12, 13, 15, 17, 20, 21, 22, 24, 28, 30, 33, 34 were cluster together (Figure 13B).



(A) Pre-treatment clustering result



(B) Post-treatment clustering result

Figure 13: Heatmaps of k-means clustering results using all microRNA probesets.

A gap among feature's weight around 0.01 was observed at post-treatment (Figure 14B). However, no such pattern was observed pre-treatment (Figure 14A). Therefore, at post-treatment,

miRs with >0.01 weight were selected and baseline miRs with weight > the 3rd quantile (0.013) were selected. Heatmaps of the clustering results using selected miRs at each time point is displayed in Figure 15. At baseline, we observed a striking difference between the two clusters, one with much higher expression of the selected miRs than the other (Figure 15A). Post-treatment expression levels of the selected miRs are more homogenous except for two outliers (Figure 15B).

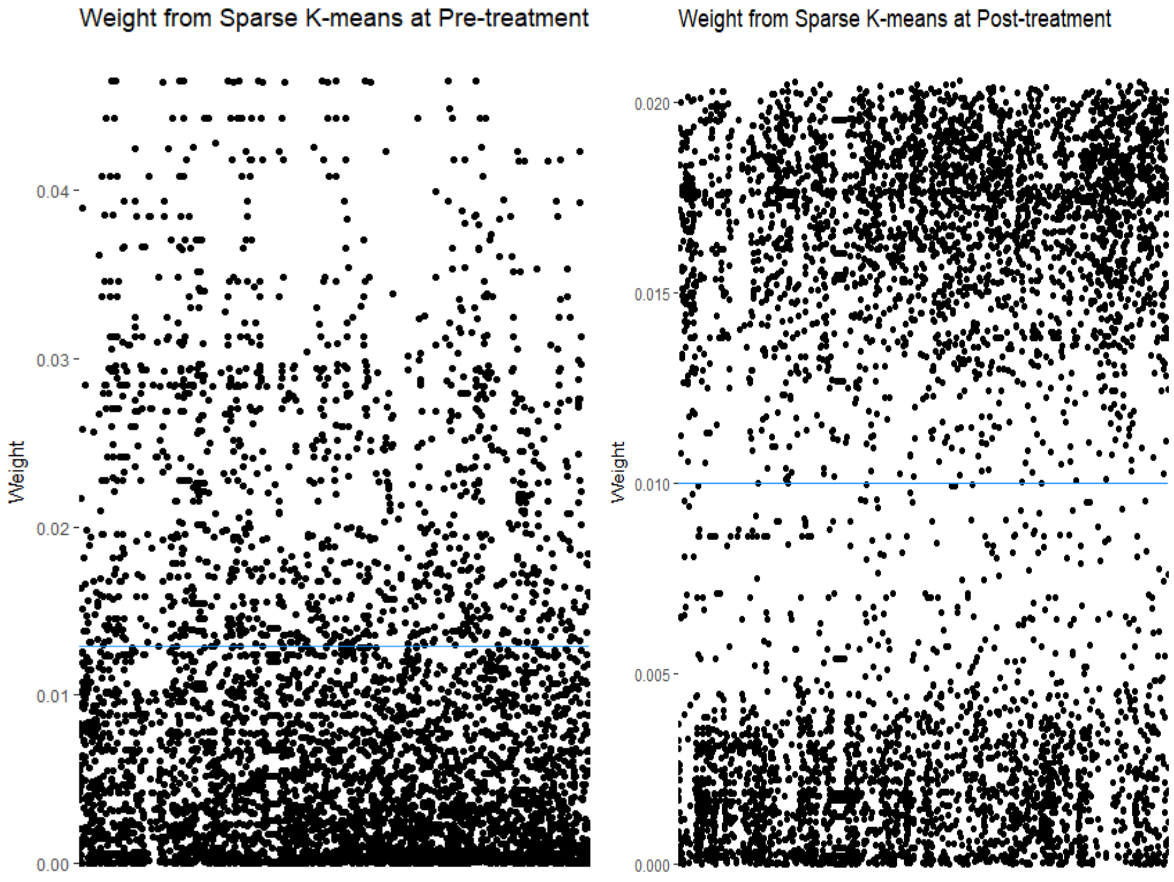
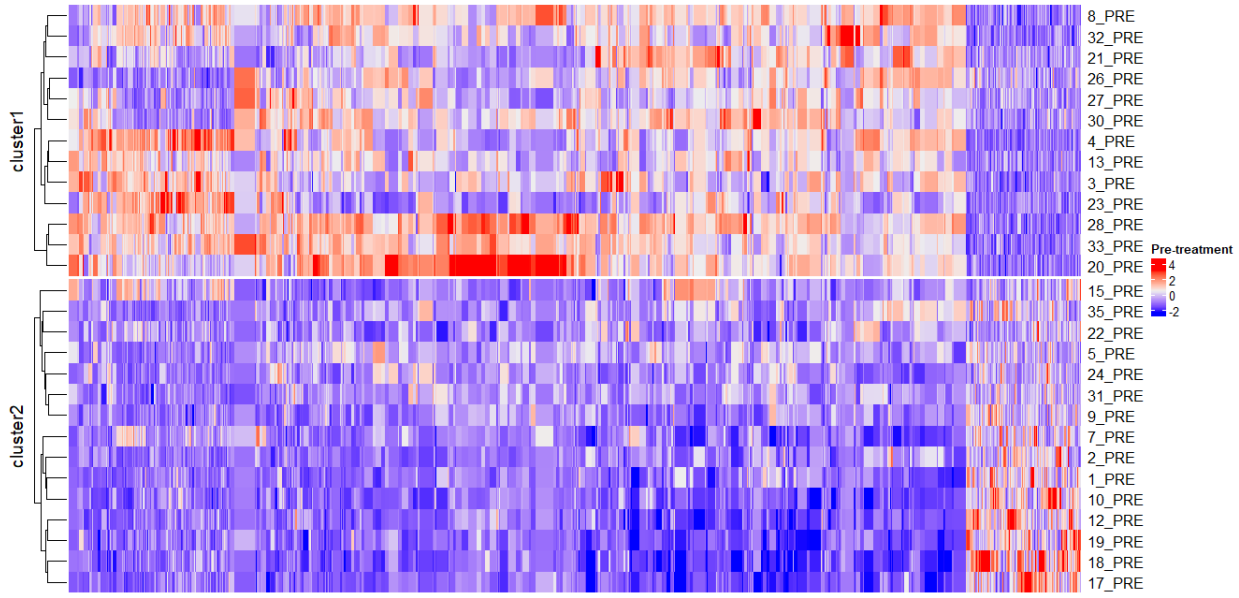
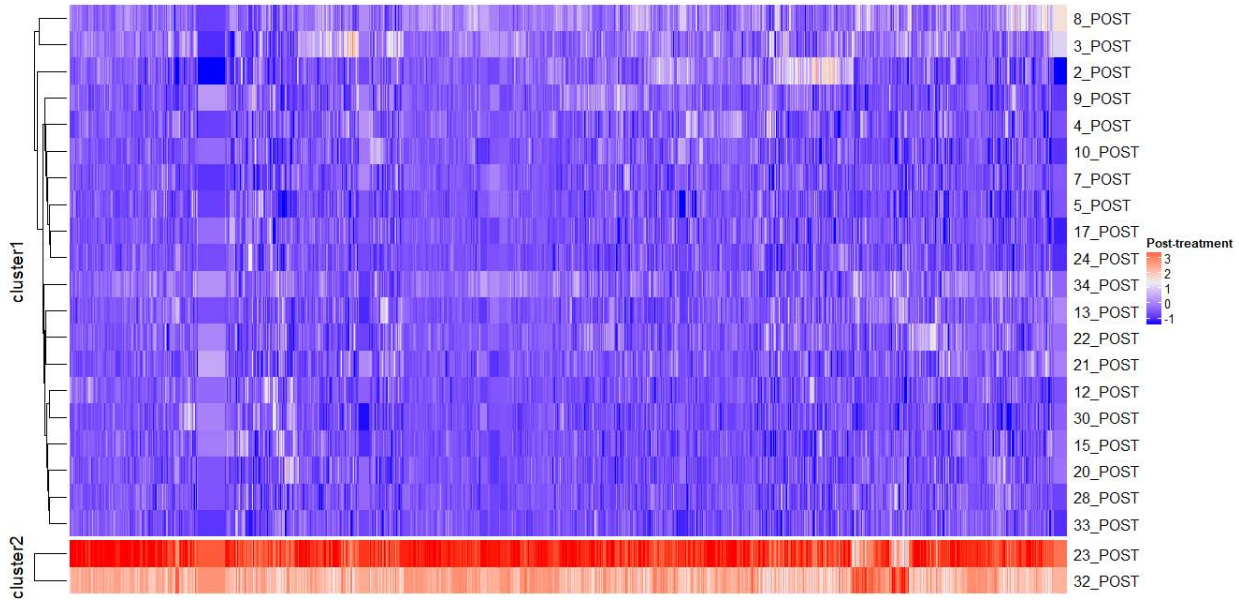


Figure 14: microRNA feature weight given by sparse k-means clustering.



(A) Pre-treatment clustering result

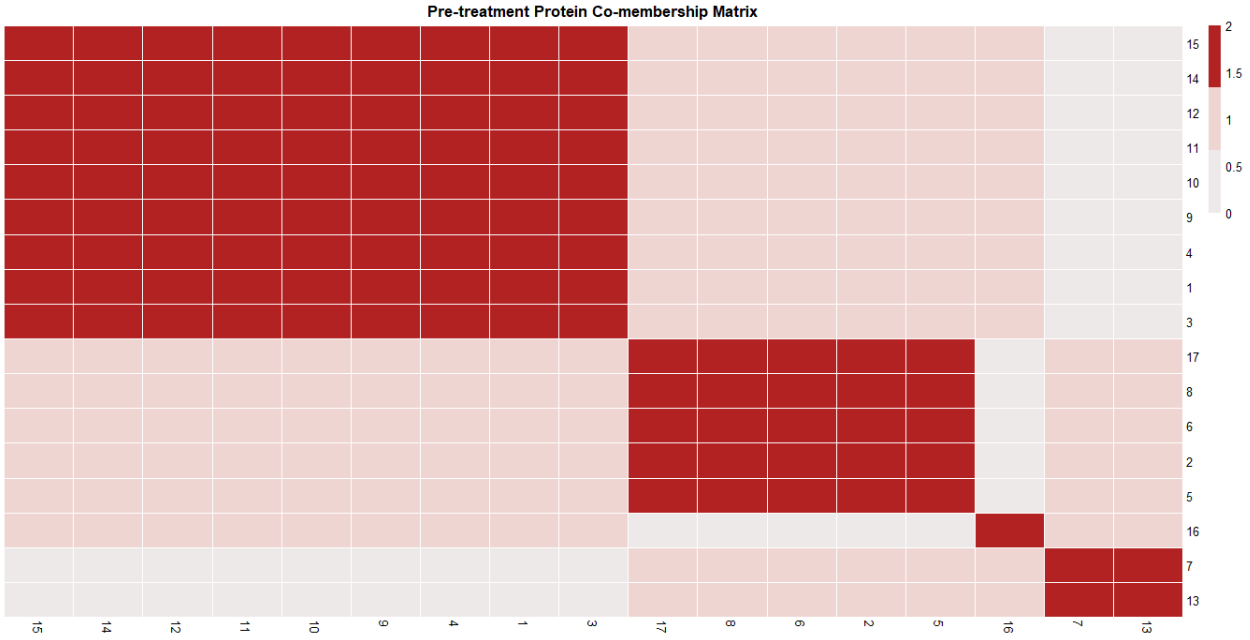


(B) Post-treatment clustering result

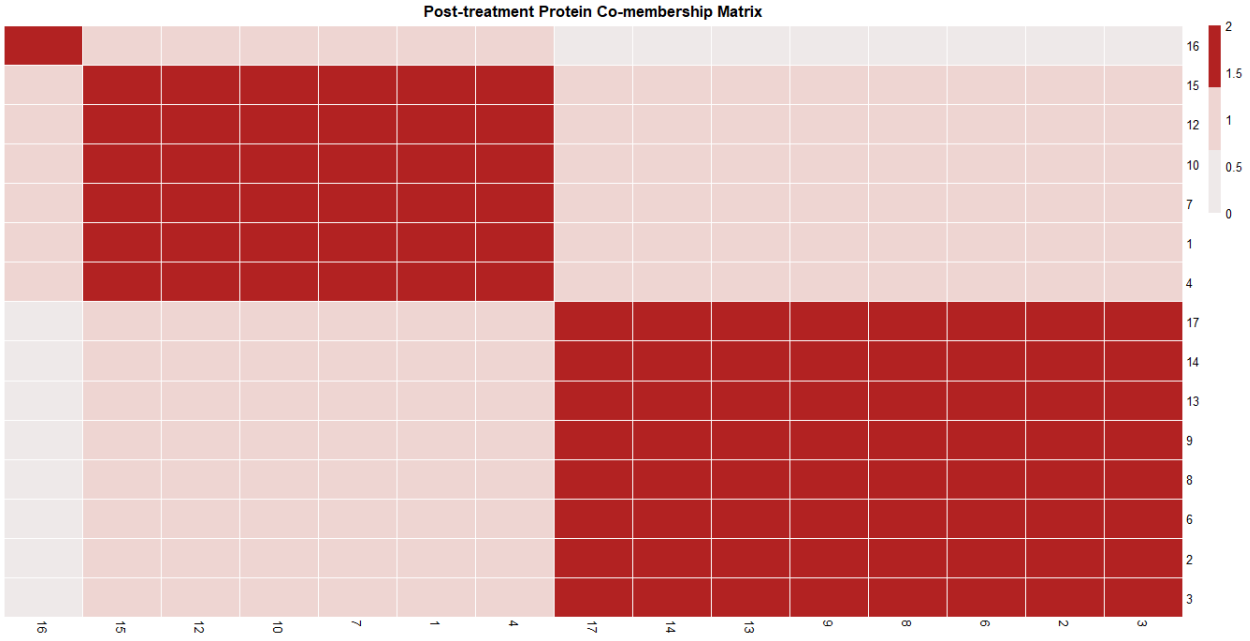
Figure 15: Heatmap of k-means clustering using high weight microRNA features.

3.2 AD-HOC INTEGRATION OF SINGLE LEVEL CLUSTERING

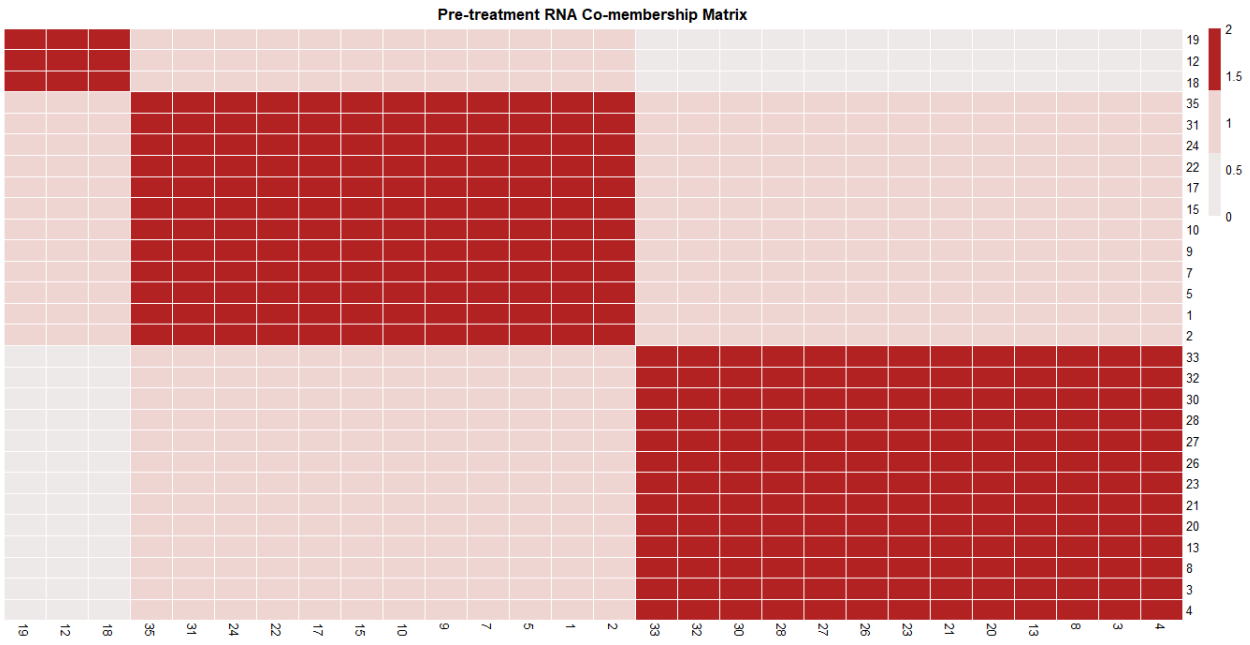
Traditionally, single level clustering analysis results were integrated in an ad-hoc fashion. Co-membership matrix was used to visualize the relationship between different subjects. In a com-membership matrix, each cell records the number of times two subjects were clustered together by different methods. Results from single level analsis were combined (Figure 16). We combined cytokine with flow cytometry data as protein data and mRNA with miR as RNA data. Subgroups of subjects were formed based on the frequency of paires of subjects being clustered together. Three subgroups could be observed according to the analysis of the protein data at baseline. Subjects 7, 13, and 16 were always grouped together. Subjects 2, 5, 6, 8, 17 were always grouped together and subject 1, 3, 4, 9, 10, 11, 12, 14, 15 were always grouped together (Figure 16A). At post-treatment, subject 1, 4, 7, 10, 12, 15, 16 were repetitively being in the same cluster, and the rest subjects remained in the other cluster (Figure 16B). In baseline RNA data integrated result, 1, 2, 5, 7, 9, 10, 15, 17, 22, 24, 31, 35were likely to be clustered together, subjects 12, 18 and 19 were very likely to be clustered together, and the rest subject were always clustered together (Figure 16C). Among post-treatment results, subject 23, 32 seemed to be outliers (mostly due to microRNA results). The frequency of the rest subjects being together depends on mRNA results: subject 2, 8, 10,13, 21, 24 were always clustered together, while subject 3, 4, 5, 7, 9, 12, 15, 17, 20, 22, 28, 30, 33, 34 were cluster together (Figure 16D).


Pre-treatment Protein Co-membership Matrix
2
1.5
1
0.5
0
15
14
12
11
10
9
4
1
3
17
8
6
2
5
16
7
13

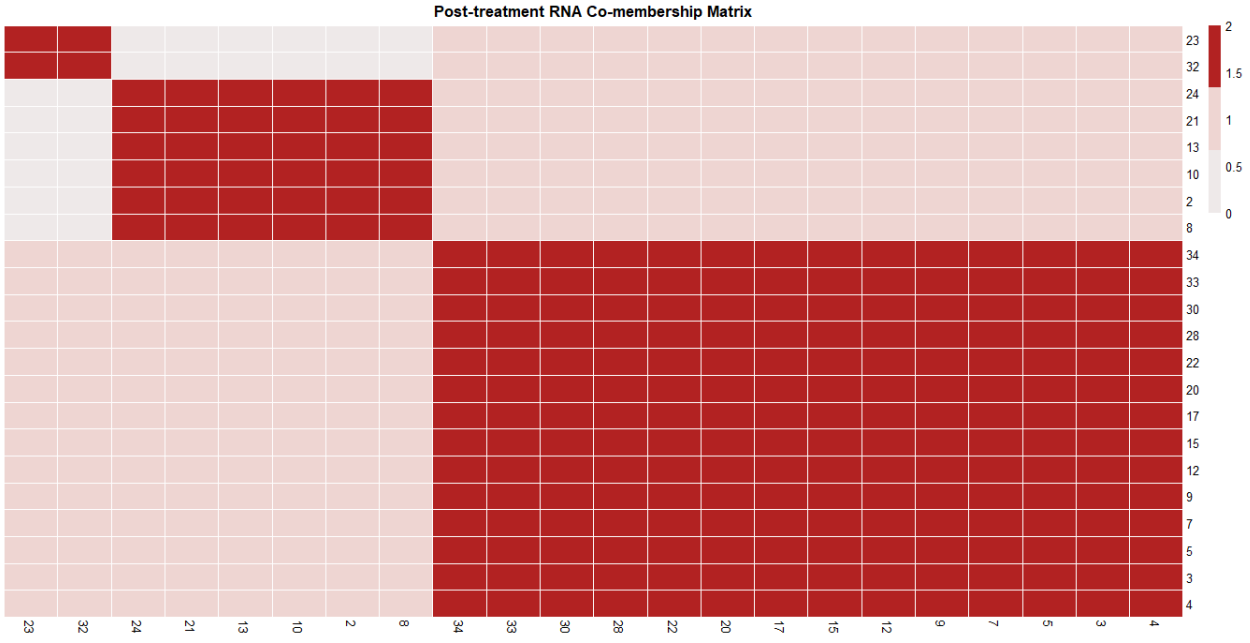
(A)


Post-treatment Protein Co-membership Matrix
2
1.5
1
0.5
0
16
15
12
10
7
1
4
17
14
13
9
8
6
2
3

(B)



(C)



(D)

Figure 16: Consensus clustering.

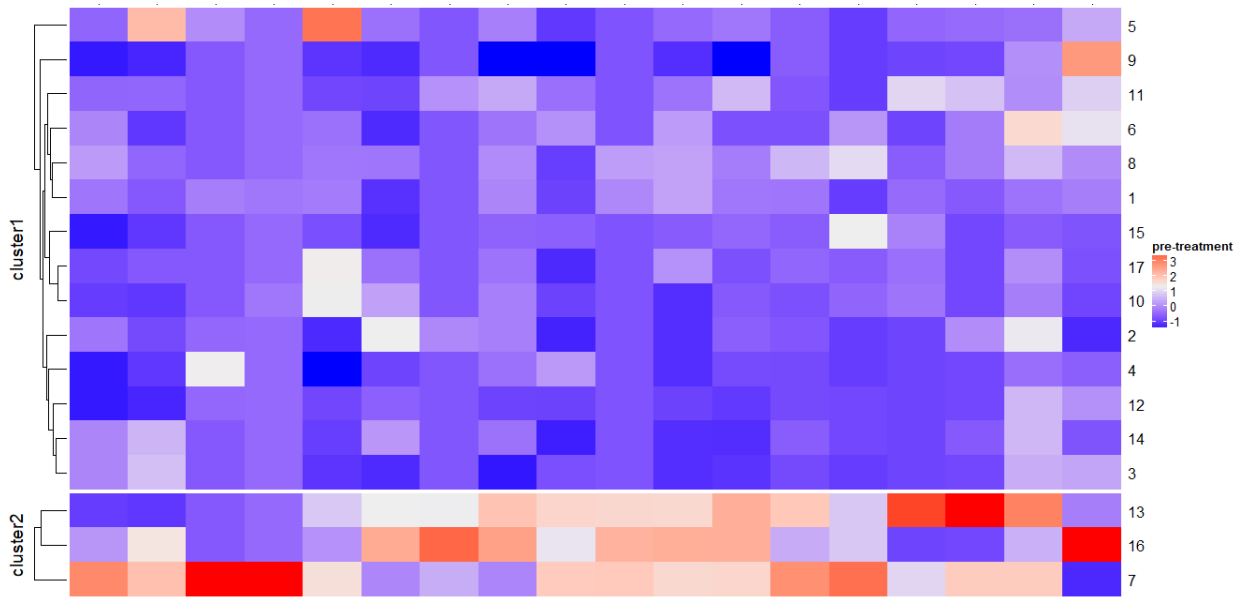
3.3 INTERGRATED CLUSTERNG ANALYSIS

In this section we attempted systematic vertical integrated clustering analysis to combine multiple layers of biomarker data to define the substructure in the study cohort. As demonstrated in Figure 1, overlap between protein and cell surface marker level data (Cytokine and flow cytometry data) is very limited. Thus, we only attempted integrating two levels of data. Cytokine and flow cytometry data were integrated together for proteomic feature analysis; mRNA and microRNA expression data were integrated together for genomic feature analysis. Two approaches were applied in both pre- and post-treatment. Firstly, we simply merged two data sets together and treat them as one data set without specifying the type. The second approach is to use iClusterPlus, which is a more sophisticated method uses a joint latent variable model to integrate different data types.

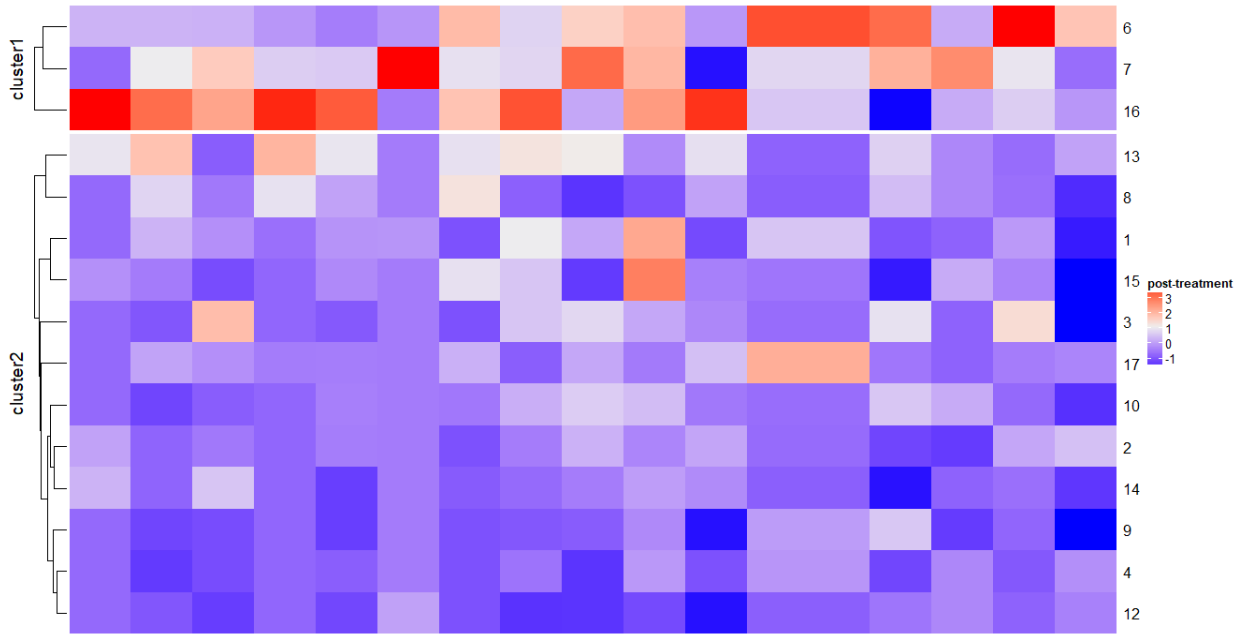
3.3.1 Integration of Cytokine and Flowcytometry Data

First, we merged cytokine and flow cytometry data together and applied sparse k-means clustering to the data at baseline and post-treatment. We obtained the results that is almost in complete agreement with the Cytokine clustering analysis at both timepoints (Figure 17). A study of the selected features explains why this is the case. As shown in Figure 18, most of the selected biomarkers are Cytokines in both time points (12/18 at baseline and 14/17 post-treatment). Features selected were listed in Table 1. Among the 18 and 17 selected features at pre- and post-treatment, 10 features are overlapped. Out of the 9 overlapped cytokine features, 5 are also selected in both time point in single level cytokine analysis.

iClusterPlus reached the exact same result as single level analysis of cytokine at both baseline and post-treatment. Interestingly, features selected by iClusterPlus, based on lasso coefficient estimates are cytokine (Table 1).



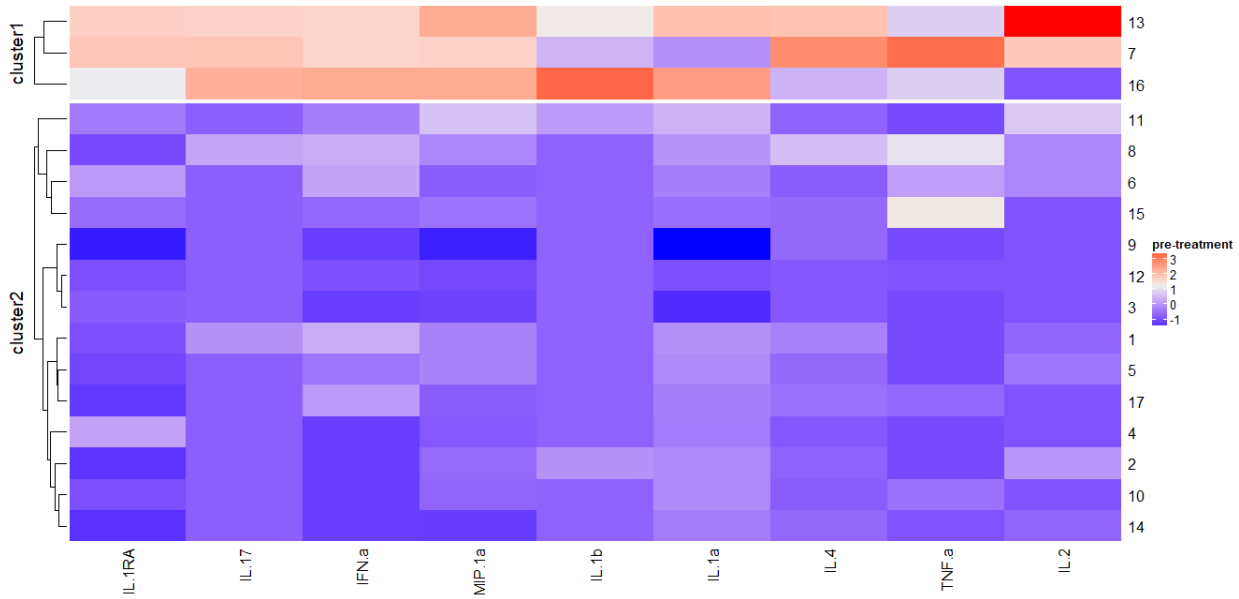
(A) Pre-treatment clustering result



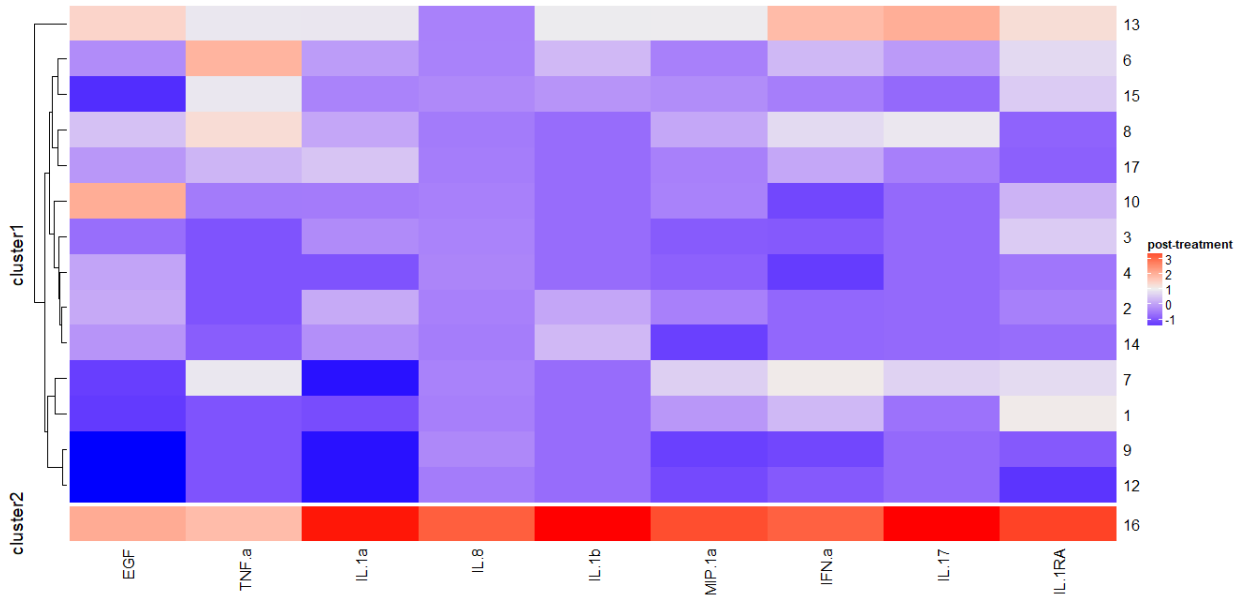
(B) Post-treatment clustering result

Figure 17: Heatmap of sparse k-means clustering using cytokine and flow cytometry features.

Figure 18: Protein feature weight given by sparse k-means clustering.



(A) Pre-treatment clustering result



(B) Post-treatment clustering result

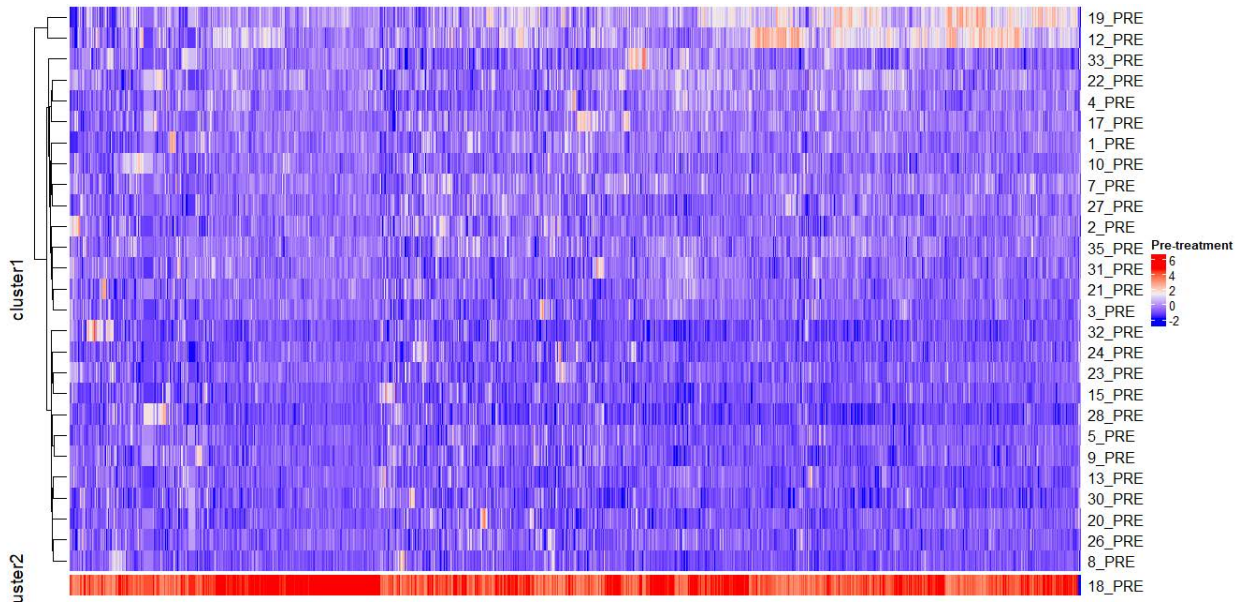
Figure 19 iClustePlus results with selected features of cytokine and flow cytometry data.

Table 1: Protein feature selected by sparse k-means and iCLusterPlus.

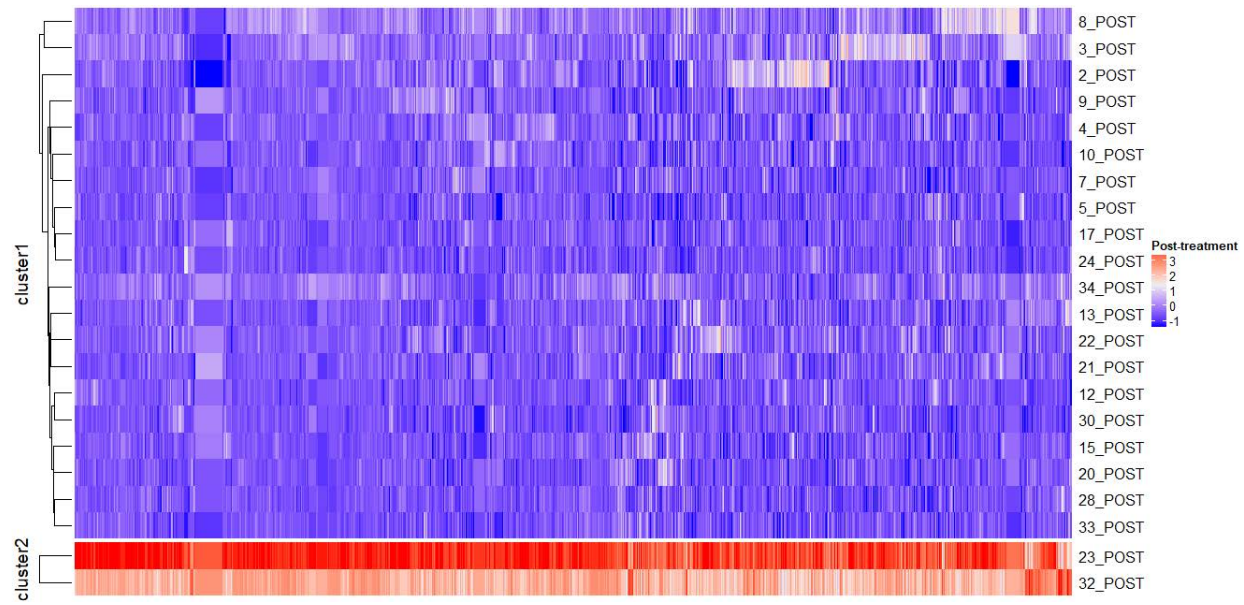
	Pre-treatment	Post-treatment
Sparse k-means	IL.1b, G.CSF, FGF.basic, IFN.a, IL.13, IL.15, IL.17, MIP.1a, IL.2, GM.CSF, IFN.g, TNF.a, IL.1RA, IL.7, IL.4, IL.1a, lymphocytes_Lin1/HLA/DR/CD33/CD11b, monocytes_HLA.DR.low.CD14.	IL.1b, IFN.a, IL.13, IL.17, IP.1a, IF.g, TNF.a, IL.1RA, IL.7, MIG, IL.1a, CRP, CRP__1, TGF.B1, lymphocytes_Lin1.HLA.DR, monocytes_HLA.DR..CD14, monocytes_HLA.DR.low.CD14.
iClusterPlus	IL.1b, IFN.a, IL.17, MIP.1a, TNF.a, IL.1RA, IL.2, IL.4, IL.1a	IL.1b, EGF, IFN.a, IL.17, MIP.1a, TNF.a, IL.1RA, IL.8, IL.1a

3.3.2 **RNA Study**

Firstly, we merged mRNA and microRNA data and applied sparse k-means. At baseline, subject 18 stood out and formed a cluster by itself, and the other 27 subjects formed a homogeneous cluster (Figure 20A). Baseline result is exactly same with mRNA single level analysis using k-means clustering (Figure 10A). Post-treatment result, on the other hand, completely agreed with miR single level analysis (Figure 20B, Figure 13B): subject 23 and 32 were clustered together and the rest subjects were clustered together. The clustering result could be explained by the feature selection (Figure 21): features were mostly coming from one data type. Therefore, the clustering pattern is driven by one single data type at each time point.



(A) Pre-treatment clustering result



(B) Post-treatment clustering result

Figure 20: Heatmap of sparse k-means clustering using miRNA and microRNA features.

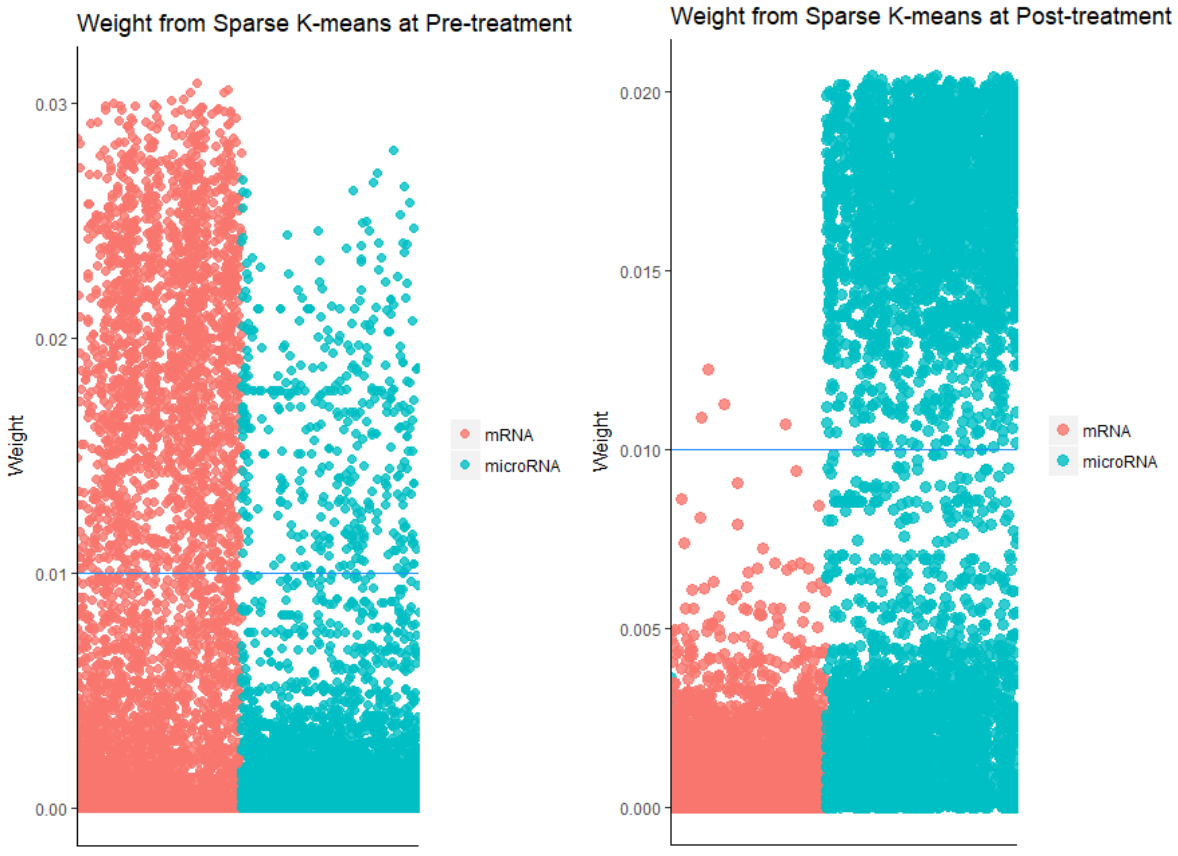
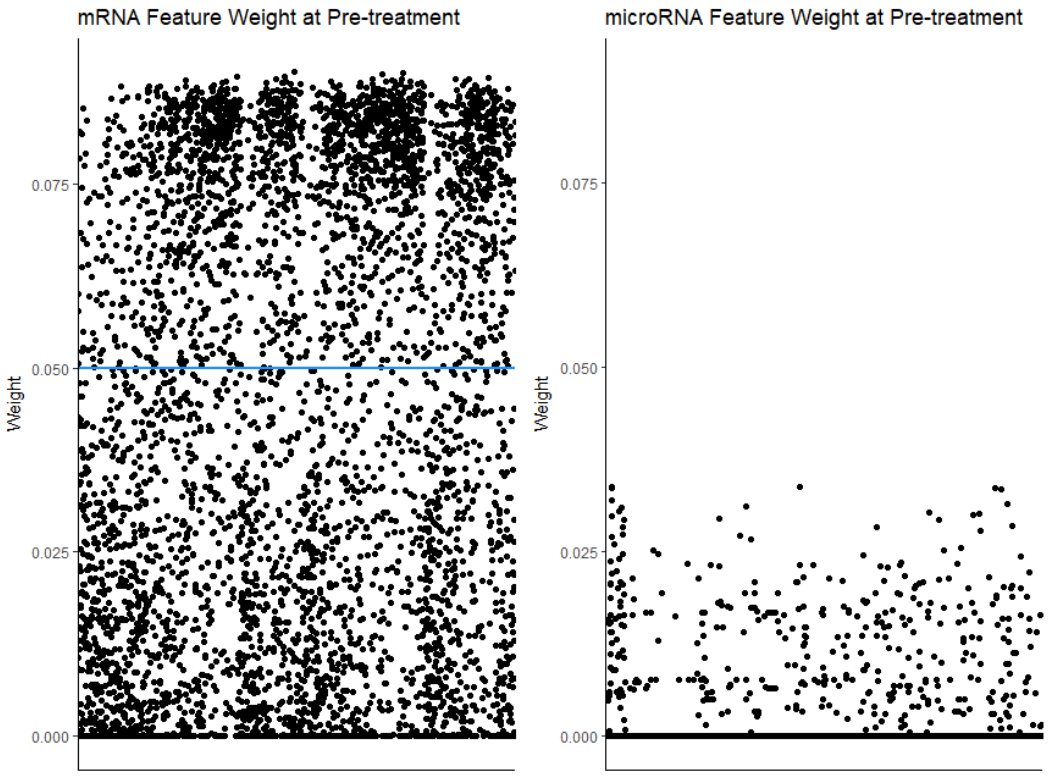


Figure 21: RNA feature weight given by sparse k-means clustering.

iClusterPlus reached the exact same results as sparse k-means at both baseline and post-treatment. Interestingly, most of the features from microRNA at baseline shrank to zero, while most features from mRNA at post-treatment were shrank to zero (Figure 22). As a result, the result was driven by mRNA at baseline, and by microRNA post-treatment. Features with top 1% lasso coefficient estimates in both data type were merged together and demonstrated a striking cluster pattern (Figure 23).



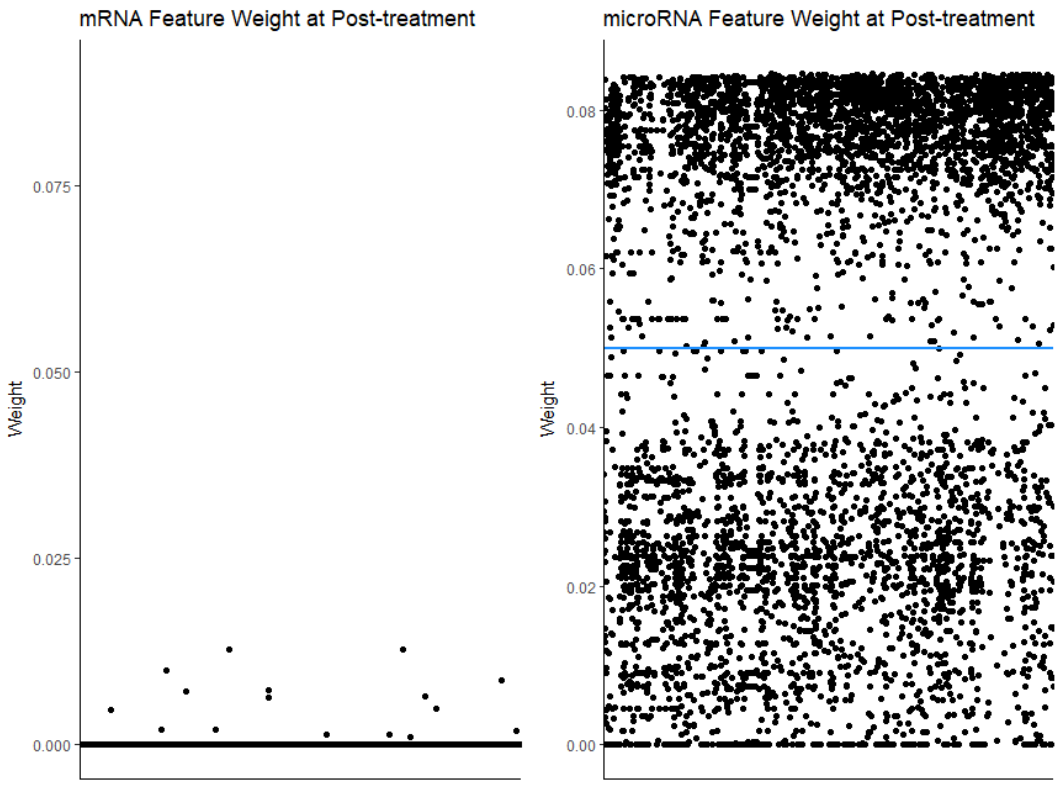
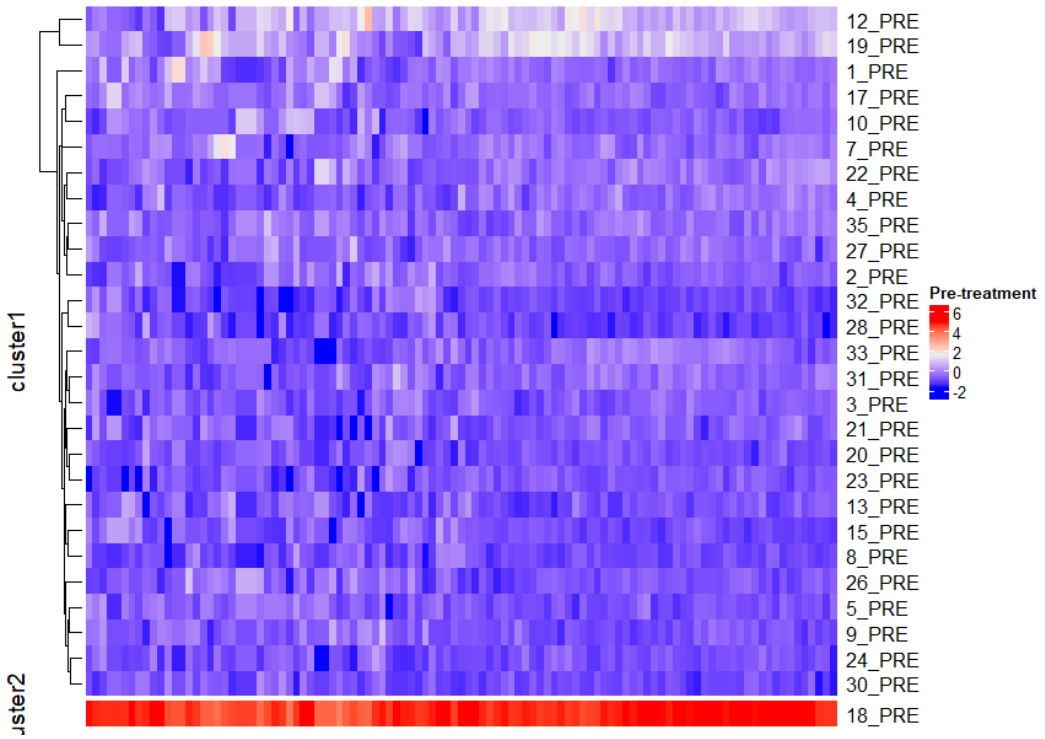
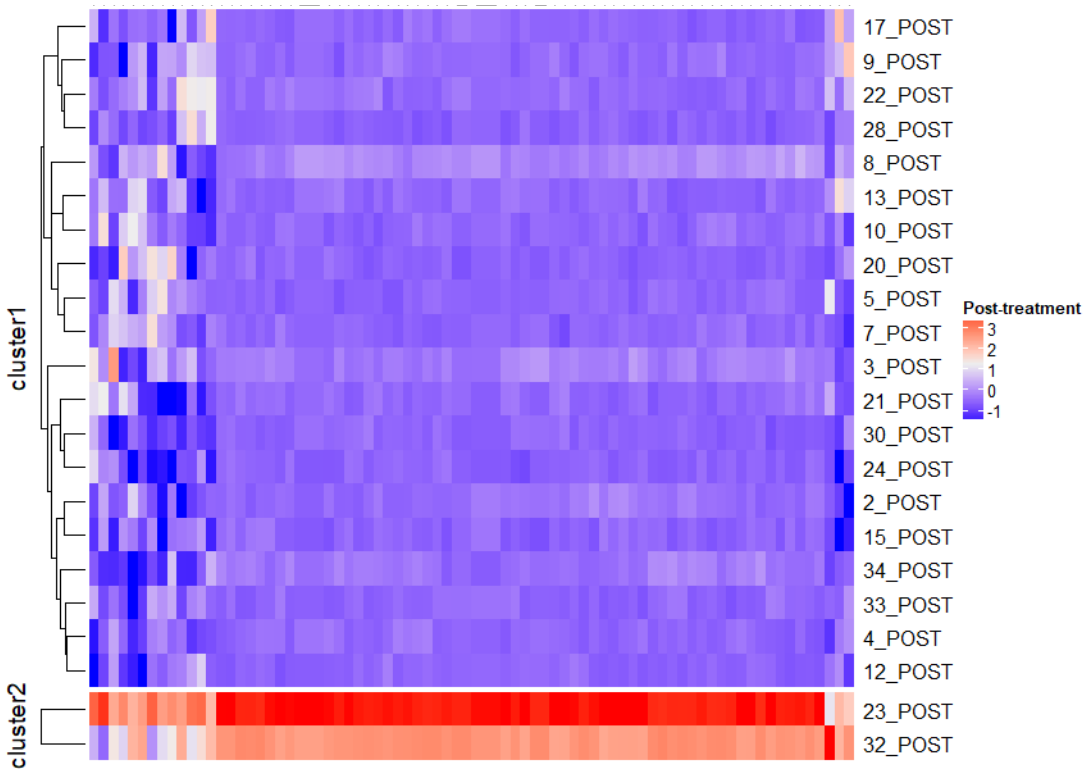


Figure 22: RNA feature weight given by iClusterPlus clustering.



(A) Pre-treatment clustering result



(B) Post-treatment clustering result

Figure 23: Heatmap of k-means clustering using top 1% weight RNA features.

3.4 ASSOCIATION WITH CLINICAL ENDPOINTS

In single level analysis, baseline and post-treatment gave very different result. In general, many times we found subjects were very homogeneous with a few outliers (Table 2). At both timepoints, flow cytometry data group the subjects into two clusters with more even size, but the cluster pattern was not very striking (Figure 7). At baseline, using cytokine data subjects were clustered in two groups. However, in the PCA plot, we could only see one tight cluster (subject 1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 14, 15, 17), subject 7, 13, 16 were outliers (Figure 2). mRNA picked out 3 outliers at baseline while miR picked out 2 outliers post-treatment.

Table 2: Single level clustering results.

Signature	Cluster	Patient ID	Number of Subject
Cytokine (pre)	1	7, 13, 16	3
	2	1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 14, 15, 17	14
Cytokine (post)	1	16	1
	2	1, 2, 3, 4, 6, 7, 8, 9, 10, 12, 13, 14, 15, 17	14
Flow cytometry (pre)	1	1, 3, 4, 9, 10, 11, 12, 14, 15, 16	10
	2	2, 5, 6, 7, 8, 13, 17	7
Flow cytometry (Post)	1	2, 3, 5, 6, 8, 9, 13, 14, 17	9
	2	1, 4, 7, 10, 11, 12, 15, 16	8
mRNA (Pre)	1	12, 18, 19	3
	2	17, 10, 24, 1, 22, 5, 31, 9, 2, 7, 35, 15, 3, 4, 8, 13, 20, 21, 23, 26, 27, 28, 30, 32, 33	25
mRNA (Post)	1	1, 2, 8, 10, 13, 24,	6
	2	3, 4, 5, 7, 9, 12, 15, 17, 20, 22, 23, 28, 30, 32, 33, 34	16
microRNA (Pre)	1	3, 4, 8, 13, 20, 21, 23, 26, 27, 28, 30, 32, 33	13
	2	18, 17, 12, 19, 10, 24, 1, 22, 5, 31, 9, 2, 7, 35, 15	15
microRNA (Post)	1	23, 32	2
	2	2, 3, 4, 5, 7, 8, 9, 10, 12, 13, 15, 17, 20, 23, 24, 28, 30, 32, 33, 34	20

The ad-hoc integration using co-membership matrix gave us slightly different results (Table 3). At both baseline and post-treatment, the flow cytometry data further divided the homogeneous cytokine cluster into two and resulted in 3 subgroups. At post-treatment, one of the three clusters contained only one subject (subject 16). Although this outlier may be an interesting observation clinically, there would be no power for statistical analysis. Only the rest two clusters were used for later association with clinical prognosis. At baseline, the large homogeneous cluster resulted from

mRNA data were split in two with the addition of miR result. Similarly, at post-treatment, the large cluster resulted from miR data were split in two with the addition of mRNA result. As a result, 3 subgroups were formed at both time points.

Table 3: Traditional ad-hoc integrative analysis result.

Signature	Cluster	Patient ID	Number of Subject
Protein I (Pre)	1	1, 3, 4, 9, 10, 11, 12, 14, 15	9
	2	2, 5, 6, 8, 17	5
	3	7, 13, 16	3
Protein I (Post)	1	2, 3, 6, 8, 9, 13, 14, 17	8
	2	1, 4, 7, 10, 12, 15	6
	3	16	1
RNA I (Pre)	1	1, 2, 5, 7, 9, 10, 15, 17, 22, 24, 31, 35	12
	2	3, 4, 8, 13, 20, 21, 23, 26, 27, 28, 30, 32, 33	13
	3	12, 18, 19	3
RNA I (Post)	1	2, 8, 10,13, 21, 24	6
	2	3, 4, 5, 7, 9, 12, 15, 17, 20, 22, 28, 30, 33, 34	14
	3	23, 32	2

Integrative analysis results using sparse k-means and iClusterPlus at both protein and RNA level were driven by one data type with the only exception of post-treatment protein data using sparse k-means clustering. However, the result of post-treatment protein data using sparse k-means clustering is almost consistence with result of baseline protein data using sparse k-means clustering.

Table 4: Integrative analysis result.

Signature	Cluster	Patient ID	Number of Subject
Protein II (Pre) * From integrative analysis sparse k-means clustering	1	7, 13, 16	3
	2	1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 14, 15, 17	14
Protein II (Post) * From integrative analysis sparse k-means clustering	1	6, 7, 13, 16	4
	2	1, 2, 3, 4, 8, 9, 10, 12, 14, 15, 17	11
Protein III (Pre) * From integrative analysis iClusterPlus clustering	1	7, 13, 16	3
	2	1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 14, 15, 17	14
Protein III (Post) * From integrative analysis iClusterPlus clustering	1	1, 2, 3, 4, 6, 7, 8, 9, 10, 12, 13, 14, 15, 17	14
	2	16	1
RNA II (Pre) * From integrative analysis sparse k-means clustering	1	18	1
	2	1, 2, 3, 4, 5, 7, 8, 9,10, 12, 13, 15, 17, 19, 20, 21, 22, 23, 24, 26, 27, 28, 30, 31,32, 33, 35,	27
RNA II (Post) * From integrative analysis sparse k-means clustering	1	23, 32	2
	2	2, 3, 4, 5, 7, 8, 9, 10, 12, 13, 15, 17, 20, 23, 24, 28, 30, 32, 33, 34	20
RNA III (Pre) * From integrative analysis iClusterPlus clustering	1	18	1
	2	1, 2, 3, 4, 5, 7, 8, 9,10, 12, 13, 15, 17, 19, 20, 21, 22, 23, 24, 26, 27, 28, 30, 31,32, 33, 35	27
RNA III (Post) * From integrative analysis iClusterPlus clustering	1	23, 32	2
	2	2, 3, 4, 5, 7, 8, 9, 10, 12, 13, 15, 17, 20, 23, 24, 28, 30, 32, 33, 34	20

Based on the previous clustering results, we further association of the following 5 clustering results with clinical prognosis:

(I) Baseline cytokine data single level analysis.

(II) Traditional ad-hoc result of protein at baseline.

(III) Traditional ad-hoc result of protein at post-treatment excluding subject 16.

(IV) Traditional ad-hoc result of RNA at baseline.

(V) Traditional ad-hoc result of RNA at post-treatment.

- Clinical benefit

Clinical benefit rate is defined as the percent of patients whose best clinical response are either complete response (CR), partial response (PR) or stable disease (SD). Fisher exact tests were used to test the association between each clustering result and clinical benefit. No statistically significant association was found (Table 5).

Table 5: Association between clinical response with cluster memberships.

(A) Association between clinical response with cluster memberships (I)

Cluster Membership (I)	Total N=17	Cluster 1 N=14	Cluster 2 N=3	p-value
Clinical Benefit	12 (70.6%)	11 (78.6%)	1 (33.3%)	0.191
No Clinical Benefit	5 (29.4%)	3 (23.1%)	2 (50.0%)	

(B) Association between clinical response with cluster memberships (II)

Cluster Membership (II)	Total N=17	Cluster 1 N=9	Cluster 2 N=5	Cluster 3 N=3	p-value
Clinical Benefit	12 (70.6%)	8 (88.9%)	3 (60.0%)	1(33.3%)	0.178
No Clinical Benefit	5 (29.4%)	1 (11.1%)	2 (40.0%)	2(40.0%)	

(C) Association between clinical response with cluster memberships (III)

Cluster Membership (III)	Total	Cluster 1	Cluster 2	p-value

Table 5 Continued

	N=14	N=8	N=6	
Clinical Benefit	12 (85.7%)	11 (75.0%)	6 (100%)	0.473
No Clinical Benefit	2 (14.3%)	3 (25.0%)	0 (00.0%)	

(D) Association between clinical response with cluster memberships (IV)

Cluster Membership (IV)	Total N=27	Cluster 1 N=11	Cluster 2 N=13	Cluster 3 N=3	p-value
Clinical Benefit	21 (77.8%)	9 (81.8%)	9 (69.2%)	3(100%)	0.681
No Clinical Benefit	6 (22.2%)	2 (18.2%)	4 (30.8%)	0(0.00%)	

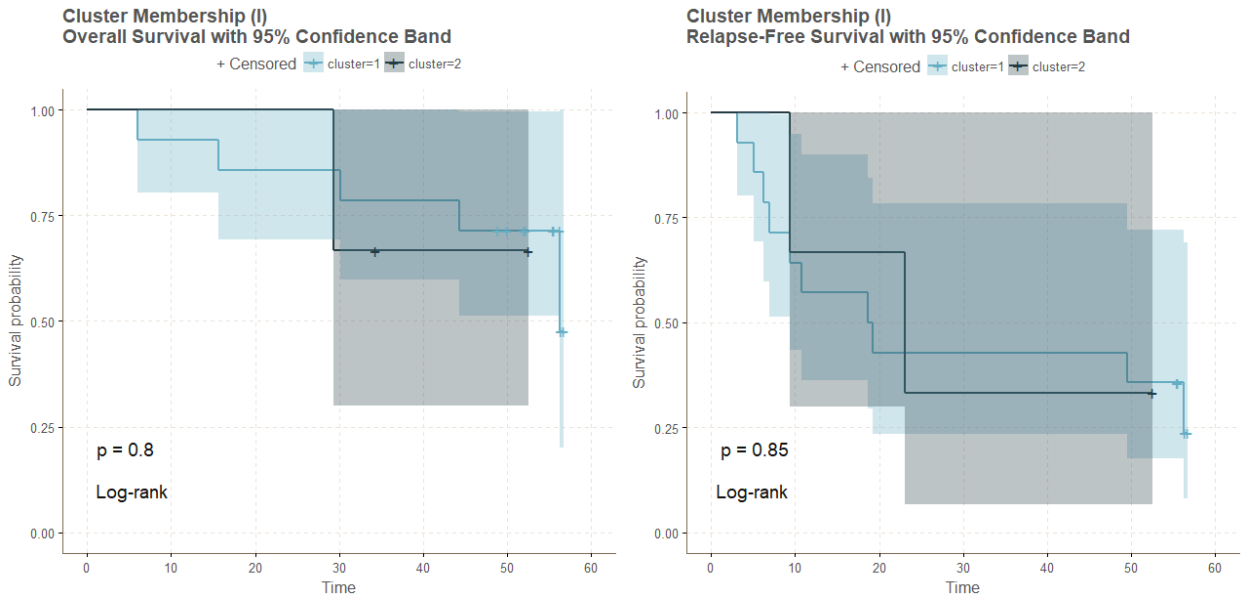
(E) Association between clinical response with cluster memberships (V)

Cluster Membership (V)	Total N=22	Cluster 1 N=6	Cluster 2 N=14	Cluster 3 N=2	p-value
Clinical Benefit	16 (72.7%)	5 (83.3%)	10 (71.4%)	1(50.0%)	0.799
No Clinical Benefit	6 (27.3%)	1 (16.7%)	4 (28.6%)	1 (50.0%)	

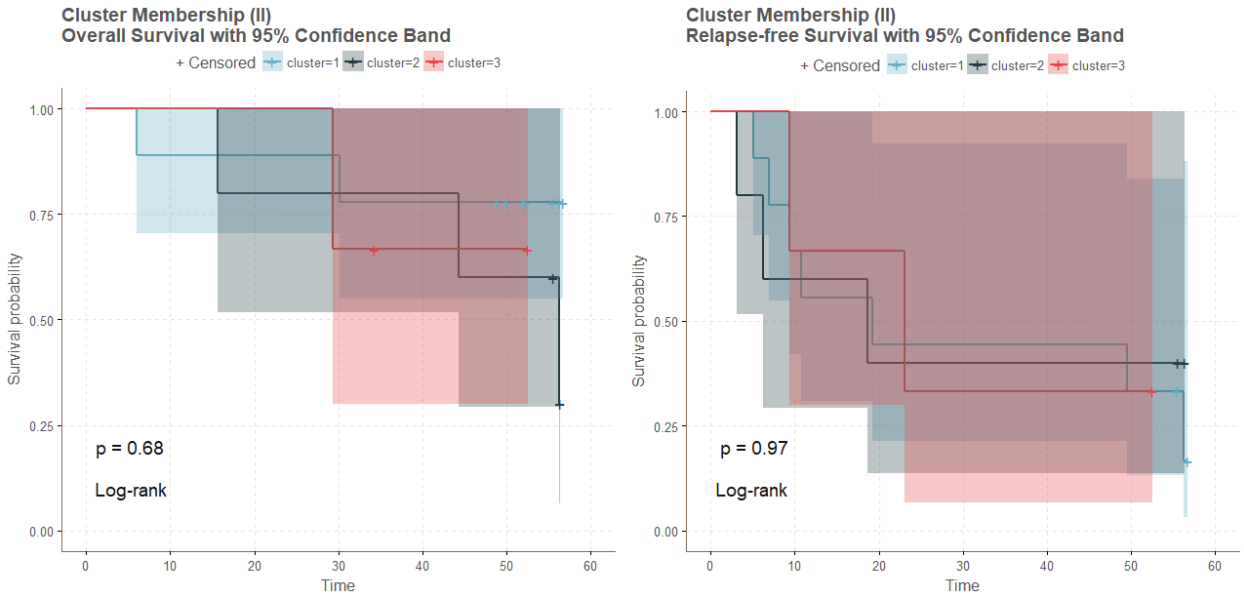
- Overall Survival and Relapse-free Survival

Overall survival is defined as time from the start of treatment to patient's death, which is usually used to measure how well a treatment is. Relapse-free survival is defended as from the start of treatment to patient's death or progression disease.

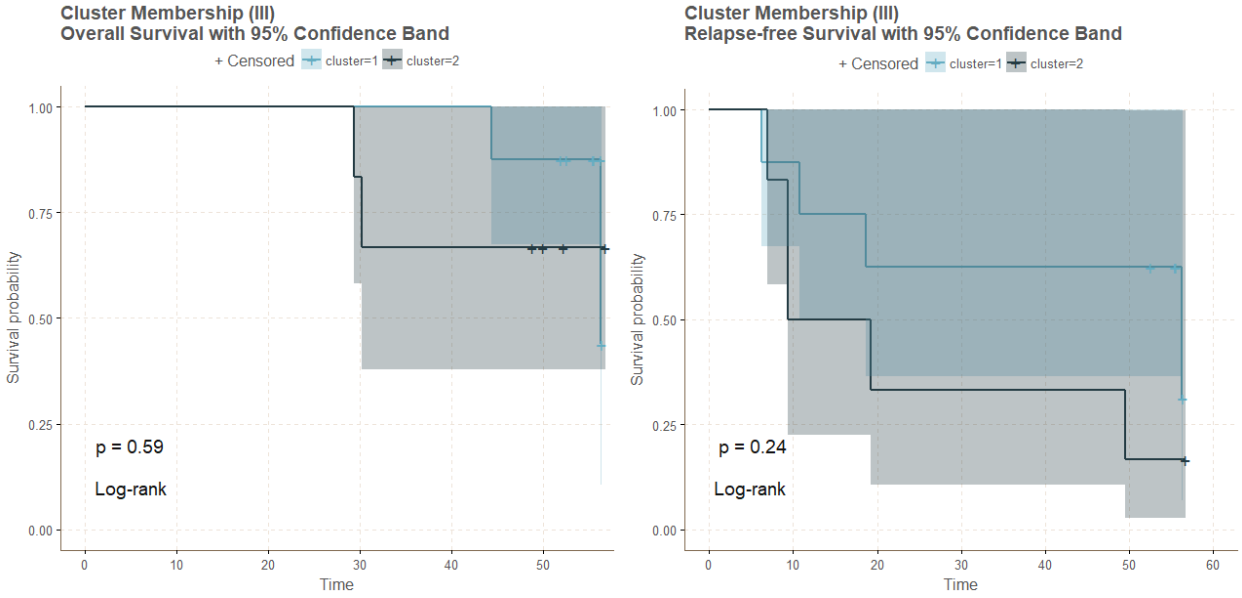
The Kaplan-Meier estimates of overall survival and relapse-free survival were compared using log-rank test between clusters (Figure 24). None the clustering results are significantly associated with either overall survival (OS) or relapse-free survival (RFS). Interestingly, the three outliers identified by the ad-hoc integration of mRNA miR results at baseline seem to have better OS. However, with the small sample size, it is hard to determine if this is a true signal.



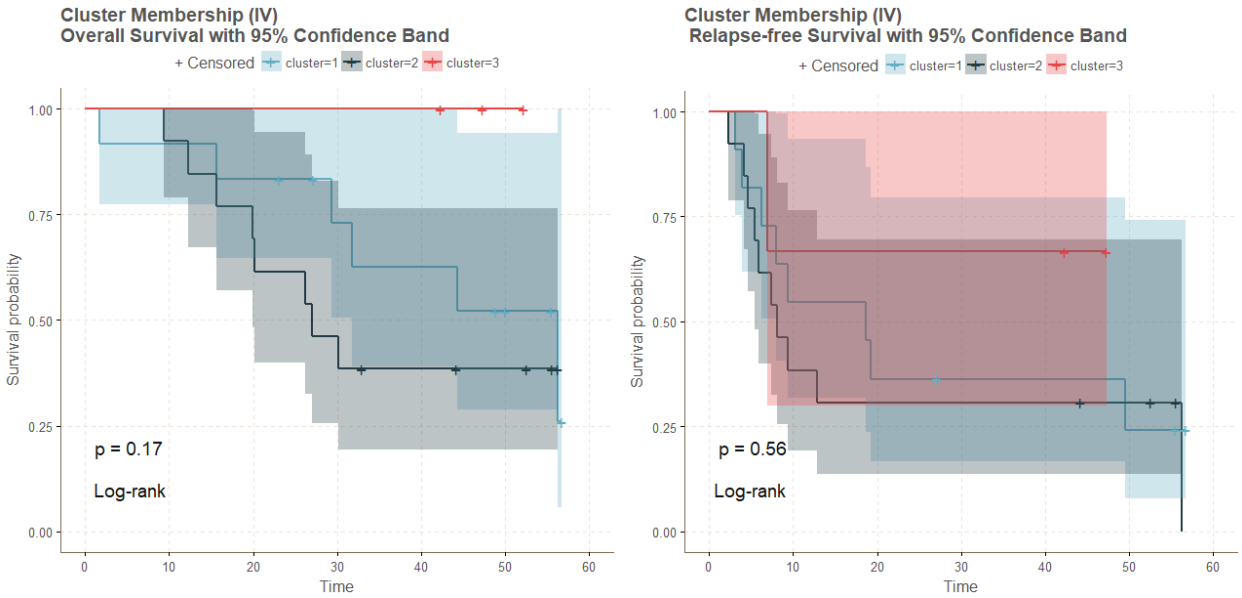
(A) Kaplan-Meier-estimates for cluster membership (I)



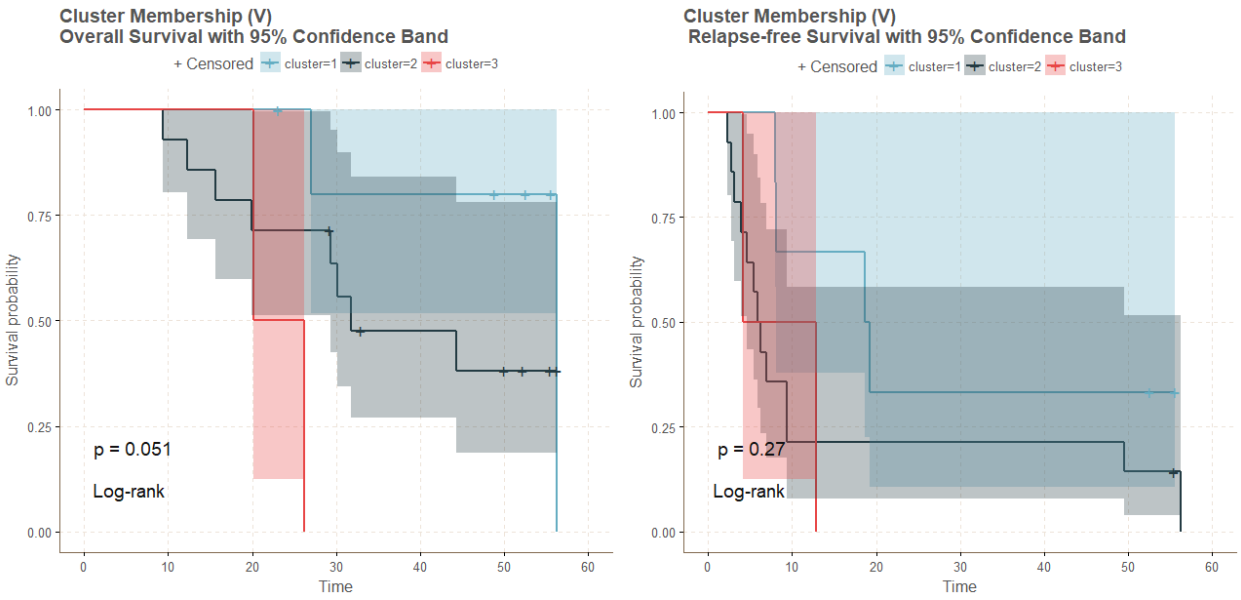
(B) Kaplan-Meier-estimates for cluster membership (II)



(C) Kaplan-Meier-estimates for cluster membership (III)



(D) Kaplan-Meier-estimates for cluster membership (IV)



(E) Kaplan-Meier-estimates for cluster membership (V)

Figure 24: Kaplan-Meier estimates of overall survival and relapse-free survival with 95% confidence band.

- Adverse Event

The worst toxicities in clinical trial UPCI 08-144 included grade 3 diarrhea/colitis, hepatitis, rash and elevated lipase [7]. Grade 3 colitis were the most common adverse event with occurrence among14% patients. Adverse event intensity was graded according to the National Cancer Institute Common Terminology for Adverse Events (version 3.0), where grade 1 represents mild, grade 2 represents moderate, grade 3 and 4 represents severe and very severe. We studied the association of development of severe colitis (grade 2 or 3) with the 5 clustering results. None of the clusters identified was significantly associated with risk of severe colitis (Table 6).

Table 6: Association between adverse event with cluster memberships.

(A) Association between adverse event with cluster membership (I)

Cluster Membership (I)	Total N=17	Cluster 1 N=14	Cluster 2 N=3	p-value
None Severe	10 (58.8%)	10 (71.4%)	0 (0.0%)	0.051
Severe	7 (41.2%)	4 (28.6%)	3 (100.0%)	

(B) Association between adverse event with cluster membership (II)

Cluster Membership (II)	Total N=17	Cluster 1 N=9	Cluster 2 N=5	Cluster 3 N=3	p-value
None Severe	10 (58.8%)	3 (77.8%)	3 (60.0%)	0 (0.00%)	0.266
Severe	7 (41.2%)	6 (22.2%)	2 (40.0%)	3 (100.0%)	

(C) Association between adverse event with cluster membership (III)

Cluster Membership (III)	Total N=14	Cluster 1 N=8	Cluster 2 N=6	p-value
None Severe	8 (57.1%)	5 (62.5%)	3 (50.0%)	1.000
Severe	6 (42.9%)	3 (37.5%)	3 (50.0%)	

Table 6 Continued

(D) Association between clinical response with cluster memberships (IV)

Cluster Membership (IV)	Total N=28	Cluster 1 N=12	Cluster 2 N=13	Cluster 3 N=3	p-value
None Severe	19 (67.9%)	9 (75.0%)	9 (69.2%)	1(33.3%)	0.451
Severe	6 (22.2%)	3 (25.0%)	4 (30.8%)	2(66.7%)	

(E) Association between clinical response with cluster memberships (V)

Cluster Membership (V)	Total N=22	Cluster 1 N=6	Cluster 2 N=14	Cluster 3 N=2	p-value
None Severe	15 (68.2%)	3 (50.0%)	10 (71.4%)	2(100%)	0.507
Severe	7 (31.8%)	3 (50.0%)	4 (28.6%)	0(0.00%)	

4.0 DISCUSSION

We performed multiple single level and integrative analysis for UPCI 08-144 data and observed interesting clustering results. Then we associated the clustering results with clinical prognosis, none of the clusters showed significant association with either clinical efficacy or adverse event risk.

The integrative clustering results are mostly driven by one data type. We also noticed that in our case, we often picked out outliers in the analysis. Several outliers we picked appear to have longer OS than the rest, however, given the extreme small number, it is impossible to distinguish a chance finding from the true biological cluster.

iClusterPlus did not show any advantage comparing with simply applied sparse k-means on merged data set. This is expected given the small sample size. With a larger data set, iClusterPlus may perform better. However, it is reassuring that iClusterPlus identified the same clusters as the sparse k-means, which seems to indicate the use of the latent class model in small studies. Thus, this approach could still be inspiring for future study for that. iClusterPlus can handle data of mixed type, i.e. both categorical and continuous. It will be interesting to compare the performance of this method under different sample sizes. One possibility is to random sample from a large dataset and evaluate the performance of the clustering results.

The small number of samples also prohibit us from performing stability. Exploration using simulation data would also be a good approach in future study to test the performance of integrative analysis verse traditional analysis method.

One last note on the strategy in sample distribution. In most studies, the amount of tissue is limited. When multiple assays were performed, sample selection it is often focused on the quality

of available tissues to ensure the success of the experiment. Thus, very small number of tissues overlapped among different platforms as we observed in our study. We suggest that more emphasis should also be put in the coordination among different assays, so that we could perform integrative data analysis to borrow information from different types of biomarkers.

BIBLIOGRAPHY

1. AM Noone, et al. *SEER Cancer Statistics Review, 1975-2015, National Cancer Institute*. [cited 10 May 2018]; Available from: https://seer.cancer.gov/csr/1975_2015/.
2. A.H. Shain, and B.C. Bastian, *From melanocytes to melanomas.* Nature Reviews Cancer, 2016. **16**: p. 345.
3. American Joint Committee on Cancer, *Melanoma of the skin.* AJCC Cancer Staging Manual, 2010.
4. N.P. Restifo, M.E. Dudley, and S.A. Rosenberg, *Adoptive immunotherapy for cancer: harnessing the T cell response.* Nature Reviews Immunology, 2012. **12**: p. 269.
5. N.L. Syn, et al., *De-novo and acquired resistance to immune checkpoint targeting.* The Lancet Oncology, 2017. **18**(12): p. e731-e741.
6. C.E. Rudd, T. Alison, and S. Helga, *CD28 and CTLA‐4 coreceptor expression and signal transduction.* Immunological Reviews, 2009. **229**(1): p. 12-26.
7. A.A. Tarhini, et al., *Immune Monitoring of the Circulation and the Tumor Microenvironment in Patients with Regionally Advanced Melanoma Receiving Neoadjuvant Ipilimumab.* PLoS ONE, 2014. **9**(2): p. e87705.
8. J.R. Schoenborn and C.B. Wilson, *Regulation of Interferon‐γ During Innate and Adaptive Immune Responses*, in *Advances in Immunology*. 2007, Academic Press. p. 41-101.
9. P.G. Coulie, et al., *Tumour antigens recognized by T lymphocytes: at the core of cancer immunotherapy.* Nature Reviews Cancer, 2014. **14**: p. 135.
10. J.S. Parker , et al., *Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes.* Journal of Clinical Oncology, 2009. **27**(8): p. 1160-1167.
11. R. Shen, A.B. Olshen, and M. Ladanyi, *Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis.* Bioinformatics, 2009. **25**(22): p. 2906-2912.
12. J.A. Hartigan, and M.A. Wong, *Algorithm AS 136: A K-Means Clustering Algorithm.* Journal of the Royal Statistical Society. Series C (Applied Statistics), 1979. **28**(1): p. 100-108.
13. L. Scrucca, et al., *mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models.* The R journal, 2016. **8**(1): p. 289-317.
14. C. Fraley, and A.E. Raftery, *Bayesian Regularization for Normal Mixture Estimation and Model-Based Clustering.* Journal of Classification, 2007. **24**(2): p. 155-181.
15. D.M. Witten, and A.A. Tibshirani, *A Framework for Feature Selection in Clustering.* Journal of the American Statistical Association, 2010. **105**(490): p. 713-726.