Fakulteten för humaniora och samhällsvetenskap

Andreas Tilevik

Evaluation of clustering methods for analyzing drug cytokine profiles

Utvärdering av olika klustringsmetoder för läkemedelsdata

Statistik

C-uppsats

Datum:	2017-09-25
Handledare:	Abdullah Almasri
Examinator:	Jari Appelgren

Karlstads universitet 651 88 Karlstad
Tfn 054-700 10 00 Fax 054-700 14 60
Information@kau.se www.kau.se

Sammanfattning

Huvudsyftet med denna studie var att utvärdera olika klustringsmetoder för att analysera data hämtade från en läkemedelsstudie där cytokinprofiler hade genererats från 23 olika läkemedel. Hierarkisk klustring användes eftersom antal kluster inte var förutbestämt. Olika distansmått och s.k. länkfunktioner utvärderades för hierarkisk klustring. Utvärderingen av de olika distansmåtten visade att Pearsons korrelationskoefficient lämpade sig bäst vid klustring av de olika läkemedlen eftersom likheter i mönster var viktigare än de faktiska mätvärdena. Även fyra länkfunktioner för att slå samman kluster utvärderades. Den länkfunktion som beräknade medelavståndet mellan objektens kluster visade sig vara den optimala metoden baserat på robusthet och korrelation mellan avstånden i dendrogramet och avstånden i distansmatrisen. Genom att använda hierarkisk klustring baserat på Pearsons korrelationskoefficient och medelavstånd så kunde ett antal intressanta läkemedelsgrupper identifieras. De s.k. JAK-inhibitorerna grupperades i ett distinkt kluster medan calcineurin inhibitorerna återfanns i ett robust kluster tillsammans med proteinkinasinhibitorer. Denna studie visar att klustring av läkemedel baserat på cytokinprofiler kan erbjuda viktig information som beslutstöd för framtida projekt inom läkemedelsutveckling, samt att avstånd baserade på Pearsons korrelationskoefficient och att en länkfunktion som beräknar medelavstånd lämpar sig bäst för den här typen av data.

Abstract

The aim of this study was to evaluate different hierarchical clustering techniques for data obtained from a study where cytokine profiles had been generated for 23 different drugs. Both distance metrics and linkage functions were evaluated. The evaluation of the distance metrics showed that the Pearson correlation coefficient was the most appropriate distance metric since similarity in patterns of the profiles was more important than similarity based on the actual values. Out of the four evaluated linkage functions: single, complete, average and Ward's, the average linkage function was the best clustering method based on the cophenetic correlation and the bootstrap probability value. Using the Pearson correlation clustering with the average linkage function, the JAK inhibitors were successfully clustered and the calcineurin inhibitors were found in a robust cluster together with protein kinase inhibitors. This study indicates that cytokine profiles from drugs may provide valuable information where similar drugs can be found in the same clusters. In addition, the study shows that the Pearson correlation coefficient and the average linkage functions were the most appropriate distance metric and linkage function, respectively, for this type of data.

Table of Contents

1. Introduction

Cluster analysis, or simply clustering, is a method where a set of objects are assigned into groups, called clusters. The aim with clustering is to form clusters that include objects that are more similar compared to objects in a different cluster. Clustering is a widely use multivariate statistical technique that allows the investigator to identify groups of objects that have similar characteristics. Clustering is mainly used as a method for exploratory data mining, which can be applied on data on different scales [Jain *et al*. 1999; Eriksson *et al*. 2013]. Cluster analysis was initially developed for biological classification by Sokal and Sneath [1963], but is today a common technique in many fields. Clustering is a type of unsupervised classification, where no predefined class of the objects are known, in comparison to supervised classification. Hence, clustering is mainly used to identify groups when no prior knowledge is known about the class membership. Cluster analysis can be performed by a variety of algorithms that differ in their methodology. Two common clustering methods are hierarchical clustering and k-means clustering. The main difference between these two methods is that k-means clustering algorithms require the number of clusters to be specified in advanced whereas hierarchical clustering does not require such information. Hierarchical clustering produces a number of clusters that can be reduced by a cut-off value that can be applied by the investigator after the clustering has been performed [Yim & Ramdeen, 2015; Jain *et al*. 1999]. Hence, hierarchical clustering is preferred over k-means clustering when no prior information about the number of clusters can be obtained. Hierarchical clustering can be categorized as two types: agglomerative and divisive. Agglomerative hierarchical clustering is a “bottom up” approach that starts by separating each object into a cluster so that the number of the clusters in the first step corresponds to the number of objects [Jain *et al*. 1999]. In the subsequent steps, the clusters are merged based on the similarity where the most similar clusters are first merged into one cluster. This process is repeated until all objects are merged into one single cluster. The divisive hierarchical clustering works in the reverse order, i.e. a “top down” approach, where all objects initially are merged into one cluster. The single cluster is then separated into smaller clusters until each object is in its own cluster.

Since the divisive hierarchical clustering is more computational time consuming compared to the agglomerative hierarchical clustering, the agglomerative approach is more commonly used [Yim & Ramdeen, 2015; Rencher, 2002].

The fundamental feature in clustering is the definition of similarity between the objects. Similarity is usually defined in terms of distance measures between the objects. There exist a number of distance measures that can be applied on continuous and categorical data [Yim & Ramdeen, 2015; Jain, 1999]. For continuous data, the Euclidean distance is the most commonly used metric for measuring distances. Other distance measures take into account the correlation between the objects when the distance is computed. In addition to selecting the most appropriate distance measure, another selection must be made based on how to calculate the distance between clusters that include more than one object. For example, the distance from one object to a cluster including several objects can be evaluated in a number of ways. For this problem, several so called linkage functions have been developed. For example, the complete linkage function utilizes the maximum distance between two objects in a cluster in order to evaluate similarity between clusters including more than one object.

Clustering techniques are widely used in biology, and especially in bioinformatics, as an exploratory method when thousands of objects need to be analyzed. For example, gene expression assays may involve the analysis of about 30,000 genes from different conditions. Clustering techniques can then be used to identify genes that show similar expression (genes that belong to the same cluster) in order to predict which genes that are related. Similarly, clustering is commonly used for proteomic data mining. Cluster analysis have, for example, been used to identify groups of cytokines (molecules produced by the immune system) that are involves in diseases such as chronic lymphocytic leukemia [Yan *et al*. 2011], pathogen infection [Chromy *et al*. 2012] and juvenile idiopathic arthritis [van den Ham *et al*. 2009]. These studies have utilized clustering to identify groups of patients that are related based on their cytokine profile or groups of cytokines that

have a similar expression in the subjects. In the current study, clustering will be used as a tool for decision support during the drug discovery process.

A bottleneck in the drug discovery process is to identify and understand the immune regulating effect of new and safe pharmaceuticals. Immunosuppressive drugs are used to inhibit the immune system in order to treat e.g. allergy, inflammation, autoimmune diseases or to prevent rejection after organ transplants. Immunosuppression involves reduction of the activation and efficiency of the immune systems by suppressing signaling pathways of the immune cells [Wiseman, 2015; O'Shea *et al*. 2013; Noble *et al*. 2004]. Such suppression may cause an alteration in the production and release of cytokines that are central molecules that regulate the immune system [Wiseman, 2015]. Cytokines are usually classified into groups based on their function or because they are produced by the same immune cell. The pro-inflammatory cytokines, which induce inflammation, include interleukine-1 (IL-1), IL-6 and tumor necrosis factor alpha (TNF-α) [Steinke & Borish, 2006; Dinarello, 2000]. Cytokines are also highly involved in the development and activation of the adaptive immunity, and especially in T-cell development. IL-2 stimulates the activation of T-cells [Dinarello, 2000], which develop into either Th1 or Th2 cells. Th1 cells are characterized by their production of IL-2 and IFN-γ whereas Th2 cells produce the cytokines IL-4, IL-5 and IL-13 [Mosmann *et al*. 2005; Grakoui *et al*. 1999]. A range of different drugs have been developed to inhibit the immune system. The three main groups of immunosuppressive drugs include immunophilin-targeting drugs, glucocorticoids and protein kinase inhibitors [Wiseman, 2015; O'Shea *et al*. 2013]. Immunophilin-targeting drugs include the calcineurin inhibitors tacrolimus and cyclosporine A as well as the mammalian targets of rapamycin (mTOR) inhibitors sirolimus and everolimus [Wiseman, 2015]. The glucocorticoids are steroids that bind to the glucocorticoid receptor and thereby up-regulate the expression of several anti-inflammatory proteins, leading to both immunosuppressive and anti-inflammatory effects. Glucocorticoids include the drugs prednisolone, dexamethasone and fluticasone propionate [Wiseman, 2015]. The protein kinase inhibitors block the activity of the protein kinase enzymes, which are central for many signaling

pathways in immune cells. The protein kinase inhibitors include the Janus kinase (JAK) inhibitors tofacitinib and ruxolitinib as well as the tyrosine kinase inhibitors nilotinib and fostamatinib disodium [O'Shea *et al.* 2013; Noble *et al.* 2004]. Given the fact that immunosuppressive drugs have distinct targets it is likely that different groups of drugs will modulate the cytokine production differently. One way to test this prediction is to cluster drugs based on how they alter the cytokine production. This can be achieved by generating drug cytokine profiles by measuring how each drug effects the cytokine production from stimulated immune cells. In the current study, different types of distance measures and linkage functions will be evaluated on the drug cytokine data. The most appropriate method will then be used to identify similarities between drugs. Such information can then be utilized in drug discovery process where clinical candidates with unknown function can be matched to the well-known drugs that is used in this study.

The report is divided into the following sections; data, method, results, discussion and conclusion. The data section describes how the data were collected and normalized. The data section also includes some descriptive statistics of the drugs. The method section describes the different distance metrics and linkage functions that is commonly used in hierarchical clustering as well as different evaluation techniques for clustering. The result section shows the analysis of the evaluation of the different clustering methods on the drug cytokine data, as well as the clustering of the drugs. The discussion is mainly focused around evaluation of the different clustering methods and how the evaluation relates to other studies where different distance metrics and linkage functions have been evaluated.

2. Data

The raw data has previously been produced by Redoxis AB, a small preclinical company in Lund, in order to generate unique cytokine profiles for different drugs. The company´s aim was to identify similarity in the cytokine profiles between the drugs to improve decision support during the drug discovery process. The data that has been analyzed in this report includes only a fraction of the original data, including cytokine profiles from 23 drugs.

2.1 Data collection

The animal experiments were approved by the local ethical committee (Malmö/Lund, Sweden, M167-12). Nine female Dark Agouti (DA) rats, 8-10 weeks of age, were injected with 500 µl of the adjuvant pristane to induce immune activation. After 2 weeks, rats were sacrificed and single cell suspension was prepared from the spleens. Red blood cells were lysed, and the remaining cells, 4.5 x 10^6 cells/ml, were stimulated with 3 µg/ml of ConA in 96 well plates in the absence (positive control) or presence of a specific drug. Supernatant was harvested after 44 hours of *in vitro* stimulation and the cytokine content was analyzed by a Bio-Plex ProTM rat cytokine assay (Bio-Rad Laboratories). The drug concentrations that were used were determined based on previous studies at Redoxis where the maximum concentration that did not induce cell apoptosis was used. The following compounds and their concentrations were used in the cell cultures: apremilast (Selleck Chemicals; 0.125 µM), apilimod mesylate (Axon MedChem; 125 nM), bardoxolone methyl (Toronto Research Chemicals; 6.25 nM), bortezomib (Selleck Chemicals; 6.25 nM), cyclosporine A (Sigma; 125 nM), dexamethasone (Sigma; 125 nM), dimethyl fumarate (Sigma; 125 nM), everolimus (Selleck Chemicals; 2.5 µM), fostamatinib disodium (Selleck Chemicals; 2.5 µM), fluticasone propionate (mcule; 50 µM), glatiramer Acetate (Toronto Research Chemicals; 50 µM), losmapimod (Selleck Chemicals; 50 µM), mycophenolic acid (Sigma; 50 µM), nilotinib (Selleck Chemicals; 2.5 µM), pilocarpine hydrochloride (Sigma; 125 nM), prednisolone (Sigma; 2.5 µM), rosiglitazone (Cayman; 50 µM), ruxolitinib (Selleck Chemicals; 125 nM), sirolimus (Cayman; 2.5 µM), sotrastaurin acetate (Axon; 2.5 µM), tacrolimus

(Toronto Research Chemicals; 125 nM), tofacitinib (Axon; 125 nM), triptolide (Toronto Research Chemicals; 6.25 nM).

2.2 Data preparation

The raw data obtained from Redoxis was fluorescent intensity (FI) values from the instrument Bio-Plex 200 system. For each rat (n = 9), there is 576 FI values, corresponding of FI values from 24 different cytokines for each of the 23 different drugs and the positive control (stimulation of immune cells in the absence of drug). The cytokine levels for drug treated cells were normalized by dividing by the corresponding cytokine level from the positive control to obtain fold change (FC) values:

$$FC = \frac{Level\ of\ cytokine\ X\ for\ drug\ Y}{Level\ of\ cytokine\ X\ for\ positive\ control} \qquad (2-1)$$

In order to obtain the same distance of up- and downregulated cytokines relative to a FC of 1, a log2 FC was computed before the cluster analysis was performed. Since some distance metrics for clustering are very sensitive to extreme values, the data set was first filtered by removing outliers as detected by boxplots (data points outside 1.5 interquartile range above the upper quartile or below the lower quartile).

2.3 Descriptive statistics

The cytokine profiles are shown for all drugs as boxplots in the appendix (Fig. A1). A scatter plot matrix based on a sample of seven randomly chosen drugs was generated as a representative illustration of the relationship between the drugs (Fig. A2). The scatter plot matrix shows data points of the filtered FC data whereas the boxplots show the raw FC values.

3. Method

In order to deal with the large data set, a multivariate technique must be utilized. For example, principal component analysis and clustering techniques can be used to identify observations with similar patterns and to facilitate the interpretation of multivariate data. However, to constrain the analysis I have here focused on evaluating different clustering techniques and then applied the "best" clustering method on the drug cytokine data.

3.1 Clustering

Clustering is a method where a set of objects are assigned into clusters. When no prior knowledge about the number of clusters is available, hierarchical clustering is an appropriate method to use for creating clusters [Yim & Ramdeen, 2015; Jain *et al*. 1999]. Clustering utilize different types of measures to evaluate the similarity between the objects. In addition, agglomerative hierarchical clustering also involves different types of linkage functions to join the clusters.

3.2 Measures of similarity

A number of different measures of similarities have been developed for different purposes. However, to constrain the analysis, only a few distance measures will be described in the following section.

3.2.1 Euclidean distance

A common similarity metric for continuous variables is the Euclidean distance [Gower, 1982], which is defined as:

$$d(x,y) = \sqrt{(x-y)'(x-y)} = \sqrt{\sum_{k=1}^{n}(x_k - y_k)^2}\ , \qquad (3-1)$$

where x and y are two vectors: $x = (x_1,x_2,\ldots,x_n)$ and $y = (y_1,y_2,\ldots,y_n)$. Hence, the distance between two points in the Euclidean space is given by the Pythagorean formula. Thus, the Euclidean distance evaluates the similarity between two objects

based on the squared distances between the data points. The Euclidean distance can be squared, then called the squared Euclidean distance, where more weight is put on objects that are far apart.

3.2.2 Mahalanobis distance

To account for covariance between the different variables, the Mahalanobis distance [Mahalanobis, 1936] is defined by:

$$d(x,y) = \sqrt{(x-y)'S^{-1}(x-y)}\,, \qquad (3-2)$$

where S is the sample covariance. This method weights the distances based on the underlying correlation between the variables. Hence, two objects are considered to be more similar if they correlate in addition to the distances between the data points. If there is no correlation between the objects, the Mahalanobis distance will essentially produce the same clusters as generated by the Euclidean distance.

3.2.3 Minkowski distance

The Minkowski metric can be seen as an extension of the Euclidian distance, where the parameter r can be adjusted by the user. The Minkowski distance is defined as:

$$d(x,y) = \left[\sum_{k=1}^{n} |x_k - y_k|^r\right]^{1/r}. \qquad (3-3)$$

If the parameter r equals 2, the Minkowski distance is equivalent to the Euclidean distance and if r equals 1, the distance metric is called Manhattan or city block distance [Hassan *et al.* 2014].

3.2.4 Distances based on correlation

Another common method to identify similarities between objects is to first compute pairwise correlation between all objects and then use 1 minus the correlation coefficient as a measure of similarity [Gibbons and Roth, 2002]. Distances based on correlation can be defined by:

$$d(x,y) = 1 - corr(x,y) \qquad 0 < d(x,y) < 2\,, \qquad (3-4)$$

Hence, two objects with a strong positive correlation coefficient will have a distance close to zero, which defines the similarity between the objects. For example, the distance measure based on the sample Pearson correlation coefficient can be defined as:

$$d(x,y) = 1 - \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \qquad 0 < d(x,y) < 2. \qquad (3-5)$$

3.3 Linkage functions

Agglomerative hierarchical clustering starts by linking single objects into a cluster. However, once the tree grows, an object will be linked to an existing cluster and the distance between the object and the members of the cluster must be determined by a linkage rule. Many linkage methods exist, but to limit the analysis only four common linkage functions [Ferreira & Hitchcock, 2009] will be evaluated.

3.3.1 Single-linkage function

The single-linkage clustering, also called the nearest neighbor clustering, uses the minimum distance between two objects within two clusters as a distance measure [Sokal & Sneath, 1963], defined by:

$$D(X,Y) = \min_{x\in X, y\in Y} d(x,y), \qquad (3-6)$$

where $d(x,y)$ is the distance between the object x and y in cluster X and Y. Hence, clusters are evaluated and joined based only on their nearest neighbors. The single linkage function tends to generate long thin clusters since objects further away in a cluster are not evaluated in the clustering process [Ferreira & Hitchcock, 2009].

3.3.2 Complete-linkage function

Complete-linkage clustering uses the maximum distance between objects within two clusters as a measure [Sokal & Sneath, 1963], defined by:

$$D(X,Y) = \max_{x \in X, y \in Y} d(x,y). \qquad (3-7)$$

The method is also called farthest neighbor clustering since only the distance between the objects that are farthest away from each other in two clusters are evaluated. The shortest distance of these maximum distances between the clusters is then selected to determine which clusters that should be joined. In comparison to the single-linkage method, the complete-linkage function tends to form compact clusters of equal size.

3.3.3 Average-linkage function

The average linkage function, also called the unweighted pair group method (UPGMA), is a method that calculates the average distance of all possible pairs of objects between the two clusters [Sokal & Sneath, 1963]. The function is defined as:

$$D(X,Y) = \frac{1}{N_X N_Y} \sum_{x \in X} \sum_{y \in Y} d(x,y), \qquad (3-8)$$

where N_x and N_y is the size of each cluster. Thus, clusters are joined based on the shortest average of all distances between pairs of objects in any two clusters. Since it is an unweighted method, all distances contribute equally to the average that is computed.

3.3.4 Ward's method

Ward's linkage method is a recursive algorithm that minimize the total within-cluster variance [Ward, 1963]. Instead of distance metrics, Ward's method cluster based on the variance. The sum of squared errors (SSE) for cluster *K* is the within cluster error sum of squares defined by:

$$SSE = \sum_{i=1}^{K} \sum_{j=1}^{n} (y_{ij} - \bar{y}_i)'(y_{ij} - \bar{y}_i), \qquad (3-9)$$

where y_{ij} is the *j*th object in cluster *i* and *n* is the number of objects in cluster *K*. At the initial step, all objects represent their own clusters. Then n-1 clusters are formed,

which means that only two objects form one cluster. The within group sum of squared errors is then evaluated for all possible n-1 clusters. The cluster, which generate the lowest within group sum of squared errors is selected. Ward's method tend to produce clusters with equal number of objects and is very sensitive to outliers [Murtagh & Legendre, 2014].

3.4 Dendrogram

The results of a hierarchical cluster analysis can be illustrated in a tree diagram, called a dendrogram. A dendrogram comprises of lines that connect clusters. The bottom row of a dendrogram consists of leaves, which represent the individual objects. Hence, the number of leaves, or leaf nodes, in a dendrogram corresponds to the number of objects included in the cluster analysis. A cluster is usually defined as two of more leaves that are joined by a common node and a branch. The height in a dendrogram represents the distance between two objects or clusters (Fig. 1).

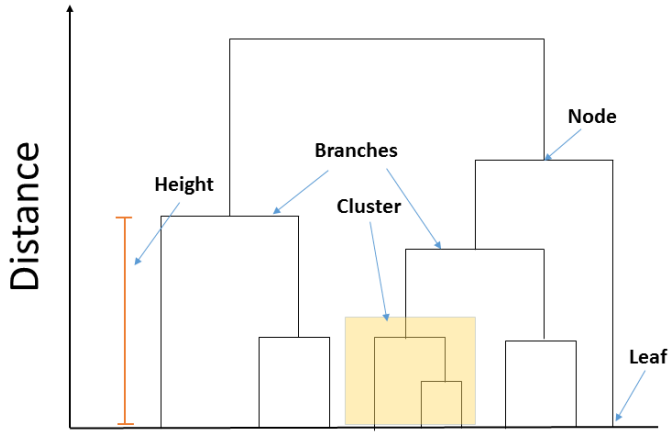


Figure 1. An illustration of a dendrogram. The vertical lines show the distance between the clusters whereas the horizontal lines (branches) connect the clusters. Leaves represent the objects, which are included in the clusters.

3.5 Bootstrapping cluster analysis

To assess the uncertainty and robustness of a cluster, bootstrapping cluster analysis has been developed [Kerr & Churchill, 2001]. Bootstrapping is a method that involves random sampling with replacement. Robustness of a cluster can be evaluated by taking thousands of bootstrap samples of the original data. For each bootstrap a dendrogram is generated. A bootstrap probability (bp) value can then be calculated for each cluster. The bp value indicates how often the original clusters are identified from the same cluster analysis based on the re-sampled data. Hence, the bp value can be used as a measurement of robustness where a high bp value indicates that the cluster is maintained even though a smaller sample is taken from the original data to produce the clusters.

3.6 Cophenetic correlation

The cophenetic distance is simply the height of the dendrogram where two branches merge into a single branch. The cophenetic correlation for a cluster tree is defined as the correlation between the cophenetic distance and the original distances used to generate the tree [Sokal & Rohlf, 1962]. This correlation is a measure of how well the dendrogram represents the actual distances. The cophenetic correlation coefficient is defined as:

$$c = \frac{\sum_{i<j}(Y_{ij} - \bar{y})(Z_{ij} - \bar{z})}{\sqrt{\sum_{i<j}(Y_{ij} - \bar{y})^2 \sum_{i<j}(Z_{ij} - \bar{z})^2}}, \qquad (3-10)$$

where Y_{ij} is the distance between objects i and j in Y, and Z_{ij} is the cophenetic distance between objects i and j.

4. Results

All statistical analyses and figure preparations were performed using R, version 3.3.2. R code for evaluation of the clusters can be found in the appendix. The results are divided into three sections; section 4.1 and 4.2 show the results from evaluating different distance metrics and linkage functions, respectively. The last section, 4.3, shows the clustering analysis of the drugs based on the most appropriate method identified from the evaluation.

4.1 Evaluation of different distance measures

In order to evaluate the most appropriate distance measure to use for the drug cytokine data, three simple scenarios were generated (Fig. 2). The three scenarios were defined so that the absolute distance between the drugs were equivalent, such as:

$$d(drug_1, drug_2)_A = d(drug_1, drug_2)_B = d(drug_1, drug_2)_C = \sum_{i=1}^{5} |x_{i,1} - x_{i,2}|. \qquad (4-1)$$

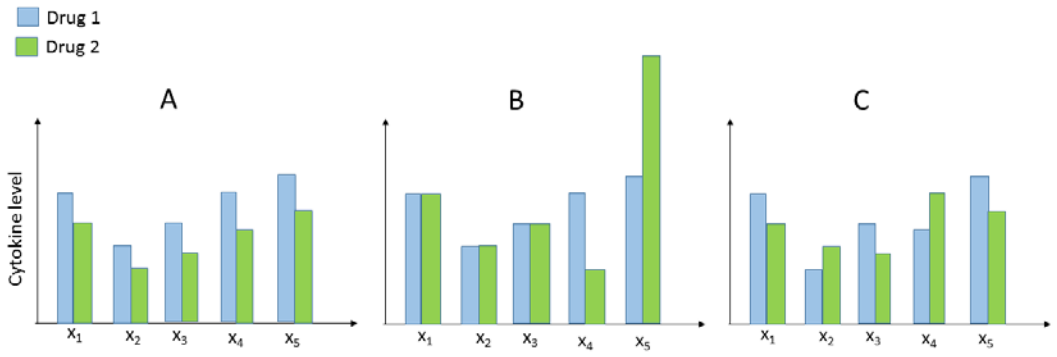


Figure 2. The three scenarios (A, B and C) that were evaluated for appropriate distance measures.

In scenario A, the two drugs have a similar relationship where drug 2 has induced an overall lower cytokine level compared to drug 1. For scenario B, the two drugs have equivalent cytokine levels of the first three variables but includes large variation between variables 4 and 5. Scenario C is similar to scenario A, but does not involve any consistency in cytokine levels between the two drugs. To evaluate

the different distance measures of the three scenarios, a simple simulations study was performed. For each scenario, a set of initial values (Table A1, appendix) were set based on the patterns seen in figure 2. The values of drug 1 were also multiplied with the factors 0.5 and 2 to obtain three different sets of data for each scenario. For example, the Euclidean distance showed the shortest distance in scenario A with the factor 0.5, whereas scenario C was identified as the shortest distance for factor 2 (Table 1). By using factor 1, scenario A and C generated identical distances for the Euclidean distance. Hence, the Euclidean distance is not favorable for scenario B and does not discriminate between scenario A and C. Generally, the results show that only the distance based on the Pearson correlation coefficient generates the shortest distance for scenario A for all factors. In addition, the correlation distance is independent on the factors since it is a standardized measure of association.

Table 1. Calculated distances for each scenario.

	Scenario		
Distances	**A**	**B**	**C**
Euclidean distance	**3.12**; 2.23; 12.73	6.14; 3.61; 11.40	4.5; 2.23; **10.95**
Mahalanobis distance	**3.87**; **1.96**; 7.89	6.66; 2.76; **6.18**	6.36; 2.24; 7.75
Minkowski distance (r=1)	**6.5**; 5; 28	11.5; 5; 24	9.5; 5; **22**
Minkowski distance (r=3)	**2.51**; 1.71, 9.88	5.34; 3.27; 9.25	3.62; 1.71; **8.93**
Pearson correlation distance	**0, 0, 0**	0.35; 0.35; 0.35	0.33, 0.33, 0.33

* The calculated distances were based on default values as in table A1 (appendix), where the values for drug 1 have been multiplied with the factors 0.5, 1, and 2 to simulate a variety of values. The shortest distance for each scenario is highlighted with bold numbers, whereas ties are underlined.

The aim of the cluster analysis is to cluster drugs with similar cytokine profile. Since drugs have different effect depending on their concentrations, the similarity measure must also account for the different drug concentrations. Hence, to account for the fact that two drugs might have identical profiles if their concentration have been optimized to have the same efficiency, only scenario A may account for this. In addition, experts at the company Redoxis agreed that scenario A is favorable for measuring similarity between their drugs. Hence, the Pearson correlation distance, which has the shortest distances for scenario A, is the only distance metric that is

based on the pattern rather than actual values. As such, the correlation distance seems to be the most appropriate distance measure for this type of data since the correlation coefficient measures the extent to which two variables tend to change together and is independent on the actual values of the two drugs.

4.2 Evaluation of different linkage functions

In order to evaluate different types of linkage functions, bootstrapping cluster analysis was performed in order to assess the robustness of the clusters generated by Pearson correlation clustering. Cluster bootstrapping was performed using the Pvclust package [Suzuki & Shimodaira, 2015]. The single, complete, average and Ward's linkage functions were evaluated by performing 10.000 bootstraps. The mean bp value for each linkage function was computed (Table 2). The linkage functions: average, single and Ward showed similar robustness with a mean bp value of 0.60. The complete linkage functions generated the lowest robustness (bp = 0.54) out of the four linkage methods that were evaluated. The linkage functions were also evaluated based on their cophenetic correlation coefficient, which is a measure for how well the dendrogram preserves the pairwise distances between the original data points. The average linkage method generated the strongest cophenetic correlation, indicating that this method best preserves the pairwise correlations of the drugs (Table 2).

Table 2. Evaluation of different linkage functions

Linkage function	Mean bp value	Cophenetic correlation coefficient (*c*)
Average	0.60	0.77
Single	0.60	0.73
Complete	0.54	0.73
Ward	0.60	0.67

Based on the robustness and cophenetic correlation, the average linkage method was selected as the most appropriate method for this analysis. This was also supported by inspecting the dendrogram for each linkage function, where the average linkage function produced clusters including drugs that have similar function or target (Fig. A3, appendix)

4.3 Cluster analysis of the drugs

To identify groups of drugs with similar cytokine profile, correlation hierarchical clustering was performed by using the average linkage function. Out of the 253 pairwise correlations, 26 were found to have a correlation coefficient greater than 0.8. Table A2 (appendix) shows the top 20 drug correlations identified in this study. By using one minus the Pearson correlation coefficient as a distance measure, the clustering method generates a dendrogram where drugs with strong positive correlation, based on the cytokine profiles, are grouped together (Fig. 3). To evaluate the robustness of the clusters, bootstrapping (n=10,000) was performed where the bootstrap probability (bp) value is calculated for each cluster. The bp value shows the fraction of times the original cluster was generated from the clustering of the re-sampled data. Based on the bp values, it is possible to identify two very robust clusters. The first cluster includes the JAK inhibitors ruxolitinib and tofacitinib (bp = 99), whereas the second cluster includes the calcineurin inhibitors cyclosporine A and tacrolimus, and the protein kinase inhibitors fostamatinib disodium and sotrastaurin acetate (bp = 97).

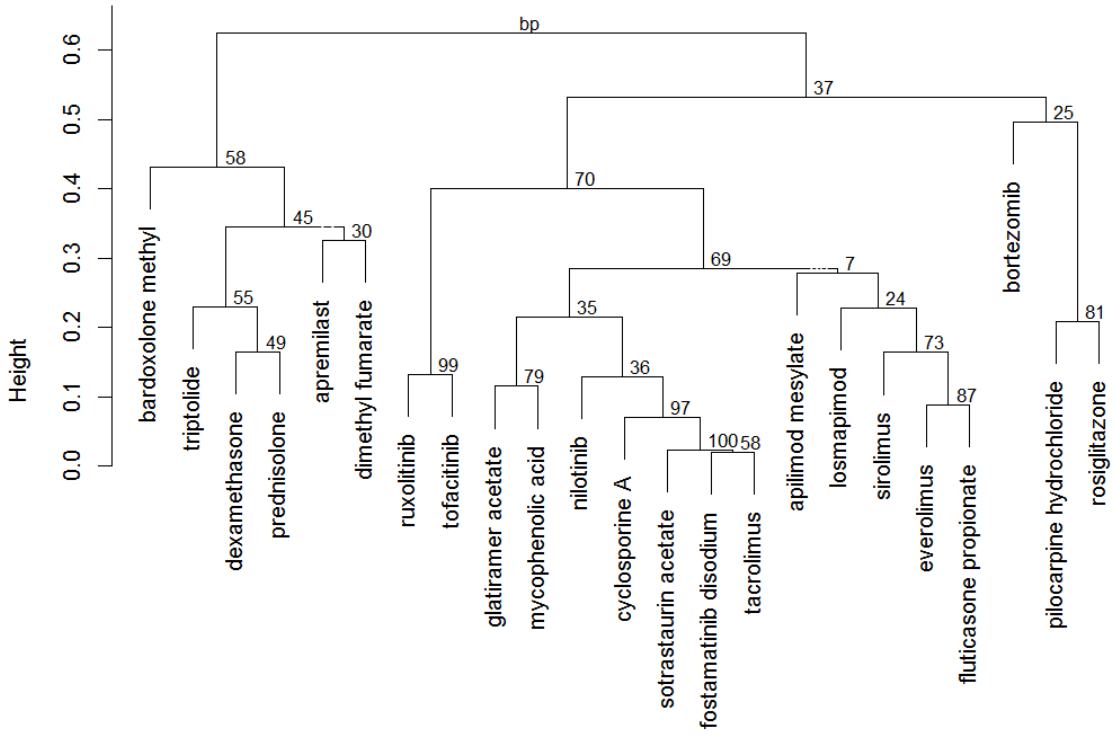


Figure 3. Cluster analysis of the drug cytokine profiles. Agglomerative hierarchical clustering was performed based on one minus the Pearson correlation coefficient, using the average linkage method. Clusters in the lower part of the dendrogram include drugs with a strong positive correlation. The bootstrap probability values are shown at the tree nodes and indicate how often the original cluster is generated from re-sampled data (n = 10.000).

In order to relate the clusters to the cytokine profiles, a heatmap was produced (Fig. 4). The heatmap shows the log 2 fold change values as colors, where a dark blue color corresponds to a FC of -5 to -4, which indicates that the drug efficiently suppresses the production of the cytokine, whereas the dark red color represents a FC value between 2 and 3, which indicates that the drug increases the production of the cytokine relative to the control. The rows and column variables are sorted based on the correlation hierarchical cluster analysis using the average linkage method. The pro-inflammatory cytokines IL-1α, IL-1β, IL-6 and TNF-α were found in the same cluster, whereas the cytokines IL-2 and IL-17 were grouped together with MIP-3α. The cytokines GM-CSF, IL-18, IL-13, IL-12, IL-7, EPO and IL-5 are barely affected by the drugs. Hence, these cytokines have a very weak impact on the cluster analysis. The JAK inhibitors ruxolitinib and tofacitinib show distinct cytokine profiles as these are the only drugs that induce an increased expression of IL-2 and MIP-3α, and at the same time cause a strong inhibition of IFN-γ. The previously identified robust cluster, which included the drugs cyclosporine A, tacrolimus, fostamatinib disodium and sotrastaurin acetate shows an overall inhibitory effect on the cytokines. The drugs in the cluster on the bottom rows in figure 4 (from dimethyl fumarate to bardoxolone methyl) have a very weak effect on the cytokines compared to the other drugs.

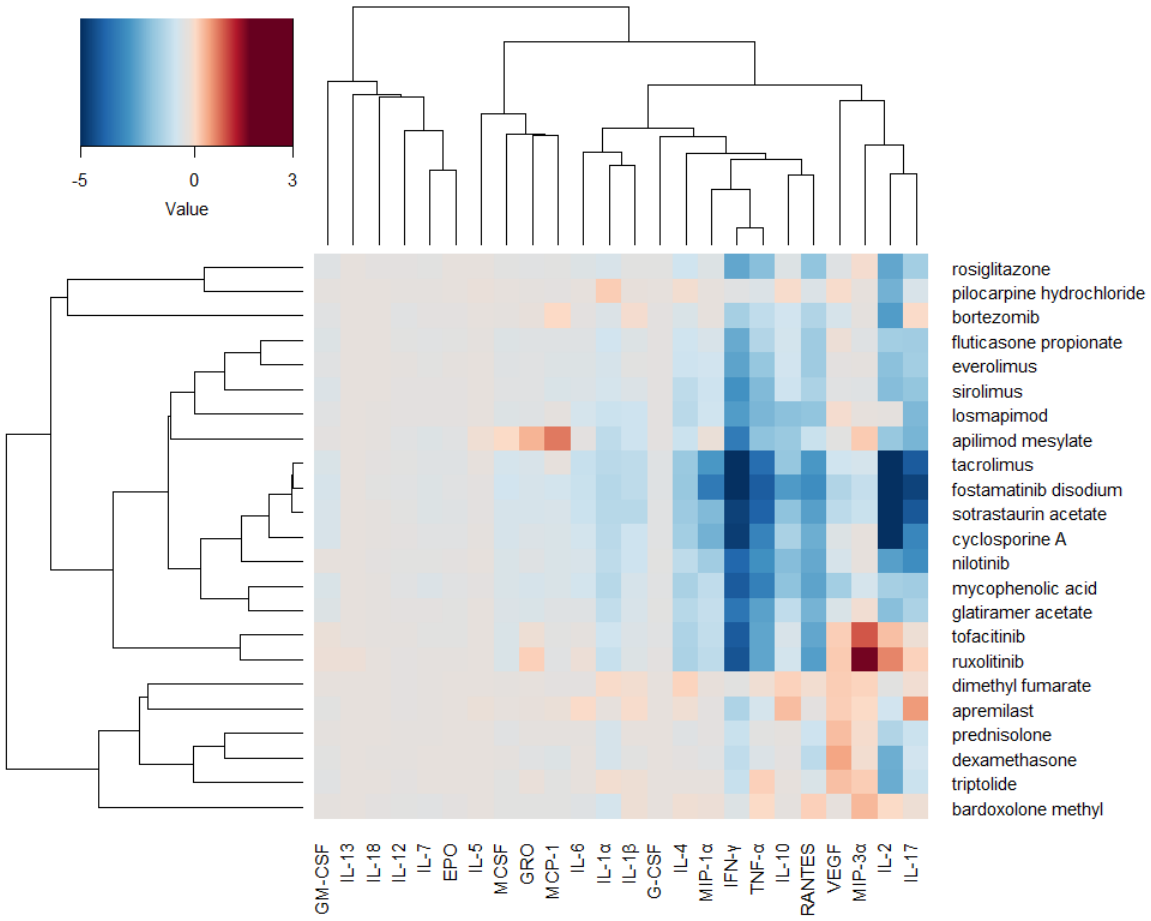


Figure 4. Heatmap of the log2 FC values. The FC was calculated by dividing the cytokine level from ConA stimulated splenocytes in the presence of drugs by the positive control (ConA in the absence of drug). Red color indicates in an increased cytokine production compared to the control, whereas a blue color indicates that the drugs suppress the production of the cytokine.

5. Discussion

Cluster analysis involves sorting objects into groups based on some similarity metric. A number of different types of clustering methods were evaluated on the drug cytokine profiles and the most appropriate method was used to cluster the drugs. The results showed that drugs that inhibit the IL-2 production usually also reduce the production of especially IL-17 and IFN-γ. This is not surprising since activated Th cells secrete mainly IL-2 and IFN-γ and that Th17 cells, which produce IL-17, develop in relation to Th1 cells [Damsker *et al*. 2010]. Both Th1 and Th17 cells are highly involved in autoimmune diseases and effective drugs for treating such diseases need to block these types of cells. Exception to the general suppressive drugs are the JAK inhibitors tofacitinib and ruxolitinib, which have stimulatory effect on the cytokines MIP-3α and IL-2 but strong inhibitory effect on IFN-γ. This indicates that cytokine profiles can provide a fingerprint-like identification of a drug. The results show that drugs with similar or identical targets produce similar cytokine profiles and cluster together. However, the calcineurin inhibitors tacrolimus and cyclosporine A clustered together with the protein kinase inhibitors fostamatinib disodium and sotrastaurin acetate. This indicates that drugs with complete different targets may generate similar effect on the cytokine production. However, only about half of the 24 cytokines are affected by the drugs, which provides quite few indicators to separate drugs efficiently. The pristane-treated DA rats are known to prone Th1-mediated autoimmune disease [Beech *et al*. 1997], which explains why the drug cytokine profiles are dominated by the pro-inflammatory cytokines and the cytokines produced by Th1 and Th17 cells. Further studies would benefit from a more diverse immune response where more cytokines are involves as well as including more variables that may help to separate drugs even further.

The main focus of the study was to evaluate different types of clustering methods for a data set involving cytokine profiles from drug treated immune cells. Based on three simple scenarios, the distance based on a correlation coefficient showed to be most appropriate for this type of data. This is due to that the data set was produced based on 23 different treatments (drugs) with different

concentrations and effectiveness. Since the concentration of the drug is believed to effect only the levels of cytokines and not the cytokine pattern, correlation distance efficiently normalize for differences in the concentrations of the drugs. Hence, experiments where pattern similarity is more important than similarity in absolute values, distances based on correlation are more appropriate. There are two main types of correlation coefficients: the parametric Pearson correlation coefficient and the non-parametric Spearman correlation coefficient. In contrast to the Spearman correlation coefficient, Pearson correlation coefficient assumes bivariate normal distribution and is very sensitive to extreme values. However, the significance of the Pearson correlation coefficient has been found to be robust even though the underlying distribution departs from bivariate normality [Edgell & Noon, 1984; Bishara & Hittner, 2012]. In addition, no significance test is computed on the correlation coefficients in the current study since the coefficients are only used as distance measures. Besides, the Pearson correlation coefficient has been found to remain approximately unbiased when the normality assumption is not fulfilled [Puth *et al*. 2014]. Many cytokines were not influenced by the drugs, which generated a distribution with many data points scattered around a FC 0f zero. To minimize the effect of these data points, Pearson correlation is favored over the Spearman correlation, since the Spearman correlation is based on rank values that make no discrimination of small or big values [Puth *et al*. 2014]. On the other hand, the advantage of using the Spearman correlation is that it is robust against extreme values. Hence, the selection of method falls somewhere between the Pearson and Spearman correlation, with a method that put less weight on data points around a FC of zero, but is still robust to extreme values. To account for this, the data set was pre-filtered by removing outliers before the Pearson correlation coefficient was calculated. Thus, extreme values that could be very influential on the correlation coefficient were removed prior to the analysis.

In addition to the distance measure, different types of linkage functions were evaluated by using bootstrapping cluster analysis and by computing the cophenetic correlation. Since the reproducibility of these particular experiments was of high importance, the results from bootstrapping cluster analysis were considered more

important compared to the cophenetic correlation coefficients when the linkage functions were evaluated. Based on the bootstrapping cluster analysis, the complete linkage function was found to be the least robust clustering method, whereas the linkage functions: average, single and Ward's method showed similar robustness. When the linkage functions were evaluated based on the cophenetic correlation, the average method was found to best preserve the pairwise correlations of the drugs. Several previous studies have tested the performance of different linkage functions for hierarchical clustering. Hands and Everitt [1987] evaluated different linkage functions on simulated data by comparing the linkage functions ability to recover the original structure of the clusters. They found that the single linkage function performed worse than the other linkage functions: complete, average, centroid and Ward's method, which all performed similarly. Ferreira and Hitchcock [2009] tested the performance of single, complete, average and Ward's method on simulated functional data. The Rand index was used, which is based on the proportion of correct grouping of objected from the original data and simulated data, to evaluate the linkage functions. They found that the Ward's method performed best when the clusters were of equal size, whereas the average linkage function performed best when the clusters were of different sizes. Ferreira and Hitchcock [2009] recommended that Ward's method should be used when one expects clusters of equal size whereas the average method was recommended when clusters are expected to be of unequal size. In the current study, the clusters were expected to be of different size since unequal number of drugs were selected from the different drug groups. For example, only two JAK inhibitors were selected, which show a unique profile, whereas four immunophilin-targeting drugs were selected. Thus, in agreement with Ferreira and Hitchcock's [2009] analysis, the average method seems to be the most appropriate linkage function for the drug cytokine data set. In addition, Saracli *et al.* [2013] evaluated a range of different linkage functions based on the cophenetic correlation coefficient. They used a number of different distance measures on simulated data from a multivariate normal distribution with or without outliers. Overall, the centroid and the average linkage function were found to generate the strongest cophenetic correlation coefficients. However, the average linkage function generated the strongest

cophenetic correlation coefficients for all simulations when the Pearson correlation coefficient was applied as a distance measure.

6. Conclusion

The goal of this study was to evaluate different distance measures and linkage functions on drug cytokine profiles as well as to test if the clustering method could be used to identify groups of similar drugs. For this particular data set, a distance measure based on the Pearson correlation coefficient was found to be most appropriate since the experiments did not include any normalization of the effect and concentrations of the drugs. The average linkage function was the best clustering method based on the cophenetic correlation and based on the fact that the clusters were expected to be of unequal size. Using the Pearson correlation clustering with the average linkage function, the JAK inhibitors were successfully clustered and the calcineurin inhibitors were found in the same cluster. This study shows that the cytokine profiles from immunomodulatory drugs may provide valuable information during the drug discovery process where cytokine profiles produced by a drug candidate can be matched to the profiles in this study. However, the mixture of drugs with different targets identified in the clusters indicates that too few response variables were included in order to achieve a clear separation between the drugs.

References

Beech JT, Siew LK, Ghoraishian M, Stasiuk LM, Elson CJ, Thompson SJ. CD4+ Th2 cells specific for mycobacterial 65-kilodalton heat shock protein protect against pristane-induced arthritis. *The Journal of Immunology*. 1997, **159**: 3692–3697.

Bishara AJ, Hittner JB. Testing the significance of a correlation with nonnormal data: comparison of Pearson, Spearman, transformation, and resampling approaches. *Psychological methods*, 2012, **17**: 399-417.

Chromy BA, Fodor IK, Montgomery NK, Luciw PA, McCutchen-Maloney SL. Cluster analysis of host cytokine responses to biodefense pathogens in a whole blood ex vivo exposure model (WEEM). *BMC Microbiology*. 2012, **12**:79.

Damsker JM, Hansen AM, Caspi RR. Th1 and Th17 cells: adversaries and collaborators. *Annals of the New York Academy of Sciences*. 2010, **1183**: 211-221.

Dinarello CA. Proinflammatory cytokines. *Chest*. 2000, **118**: 503-508.

Edgell SE, Noon SM. Effect of violation of normality on the t test of the correlation coefficient. *Psychological Bulletin*. 1984, **95**: 576-583.

Eriksson L, Byrne T, Johansson E, Trygg J, Vikström C. Multi-and megavariate data analysis basic principles and applications. 2013. Third edition. Umetrics Academy.

Ferreira L, Hitchcock DB. A comparison of hierarchical methods for clustering functional data. *Communications in Statistics-Simulation and Computation*, 2009, **38**: 1925-1949.

Gibbons FD, Roth FP. Judging the quality of gene expression-based clustering methods using gene annotation. *Genome Research*. 2002, **12**:1574-81.

Grakoui A, Donermeyer DL, Kanagawa O, Murphy KM, Allen PM. TCR-independent pathways mediate the effects of antigen dose and altered peptide ligands on Th cell polarization. *The Journal of Immunology*. 1999, **162**: 1923-1930.

Gower JC. Euclidean distance geometry. *The Mathematical Scientist*. 1982, **7**: 1-14

Hands S, Everitt B. A Monte Carlo study of the recovery of cluster structure in binary data by hierarchical clustering techniques. *Multivariate Behavioral Research*, 1987, **22**: 235-243.

Hassan D, Aickelin U, Wagner C. Comparison of Distance metrics for hierarchical data in medical databases. *Neural Networks (IJCNN), 2014 International Joint Conference on. IEEE*, 2014, 3636-3643.

Jain AK, Mutry MN, Flynn PJ. Data Clustering. A Review. *ACM Computing Surveys*. 1999, **31**: 264-232.

Kerr MK, Churchill GA. Bootstrapping cluster analysis: assessing the reliability of conclusions from microarray experiments. *Proceedings of the National Academy of Sciences of the United States of America*. 2001, **98**: 8961-8965.

Mahalanobis, PC. On the Generalised Distance in Statistics. *Proceedings of the National Institute of Sciences of India,* 1936, 2, 49-55.

Mosmann TR, Cherwinski H, Bond MW, Giedlin MA, Coffman RL. Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins. 1986. *The Journal of Immunology*. 2005, **175**: 5-14.

Murtagh F, Legendre P. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *Journal of Classification*: 2014, **31**: 274-295.

Noble ME, Endicott JA, Johnson LN. 2004. Protein kinase inhibitors: insights into drug design from structure. *Science*, 2004, **303**: 1800-1805.

O'Shea JJ, Kontzias A, Yamaoka K, Tanaka Y, Laurence A. Janus kinase inhibitors in autoimmune diseases. *Annals of the rheumatic diseases*, 2013, **72**: ii111-ii115.

Puth MT, Neuhäuser M, Ruxton GD. Effective use of Pearson's product–moment correlation coefficient. *Animal Behaviour*, 2015, **93**: 183-189.

Rencher AC. Methods of Multivariate Analysis. 2002. Second Edition. John Wiley & Sons, Inc.

Saraçli S, Doğan N, Doğan İ. Comparison of hierarchical cluster analysis methods by cophenetic correlation. *Journal of Inequalities and Applications*. 2013, **203**: 1-8.

Sokal RR, Sneath PHA. Principles of numerical taxonomy. 1963. San Francisco: WH Freeman.

Sokal R, Rohlf F. The Comparison of Dendrograms by Objective Methods. *Taxon*. 1962. 11: 33-40.

Steinke JW, Borish L. 3. Cytokines and chemokines. *Journal of Allergy and Clinical Immunology*. 2006, **117**: S441-445.

Suzuki R, Shimodaira H, pvclust: Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling. 2015. Version 2.0-0. https://CRAN.R-project.org/package=pvclust.

van den Ham HJ, de Jager W, Bijlsma JW, Prakken BJ, de Boer RJ. Differential cytokine profiles in juvenile idiopathic arthritis subtypes revealed by cluster analysis. *Rheumatology (Oxford)*. 2009, **48**: 899-905.

Ward JH Jr. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*. 1963, 58:236-244.

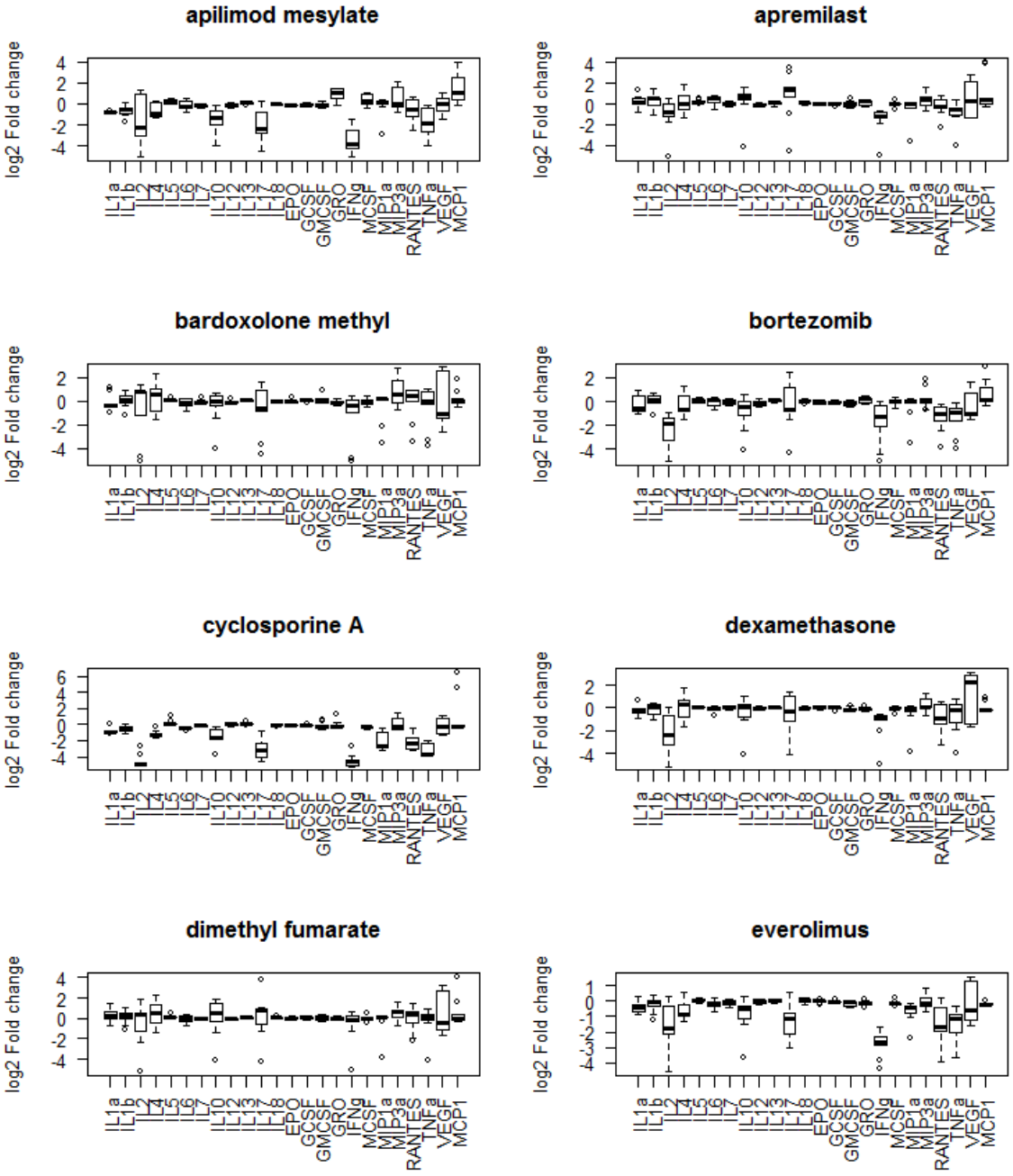
Wiseman AC. Immunosuppressive medications. *Clinical Journal of the American Society of Nephrology*, 2015, **11**: 332-343.

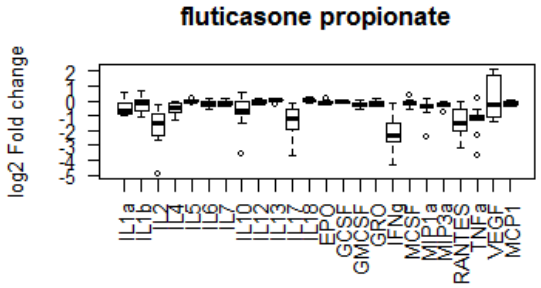
Yan XJ, Dozmorov I, Li W, Yancopoulos S, Sison C, Centola M, Jain P, Allen SL, Kolitz JE, Rai KR, Chiorazzi N, Sherry B. Identification of outcome-correlated cytokine clusters in chronic lymphocytic leukemia. *Blood*. 2011, **118**: 5201-5210.

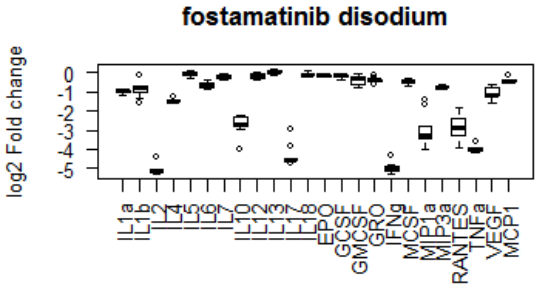
Yim O, Ramdeen KT. Hierarchical Cluster Analysis: Comparison of Three Linkage Measures and Application to Psychological Data. *The quantitative methods of psychology*. 2015; **11**: 8-21.

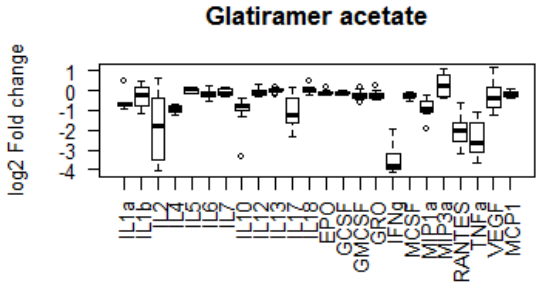
Appendix

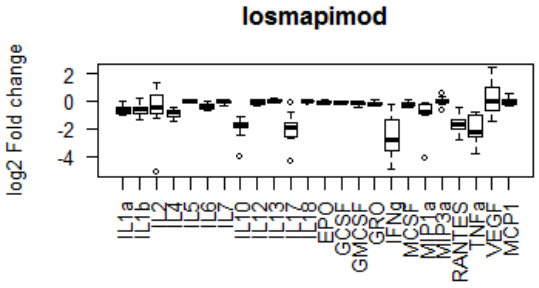
Cytokine profiles illustrated as boxplots for all drugs

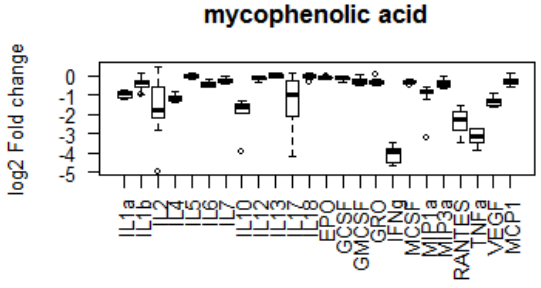


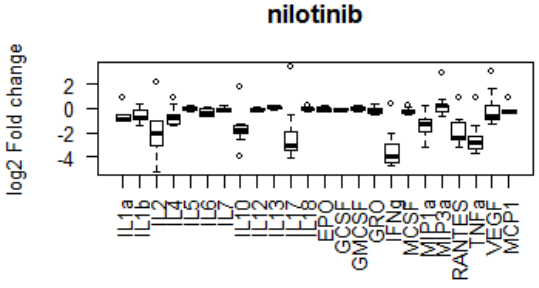

fluticasone propionate
log2 Fold change
IL1a
IL1b
IL2
IL4
IL5
IL6
IL7
IL10
IL12
IL13
IL17
IL18
EPO
GCSF
GMCSF
GRO
IFNg
MCSF
MIP1a
MIP3a
RANTES
TNFa
VEGF
MCP1

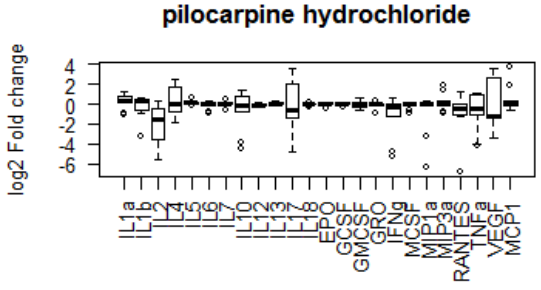

fostamatinib disodium
log2 Fold change
IL1a
IL1b
IL2
IL4
IL5
IL6
IL7
IL10
IL12
IL13
IL17
IL18
EPO
GCSF
GMCSF
GRO
IFNg
MCSF
MIP1a
MIP3a
RANTES
TNFa
VEGF
MCP1

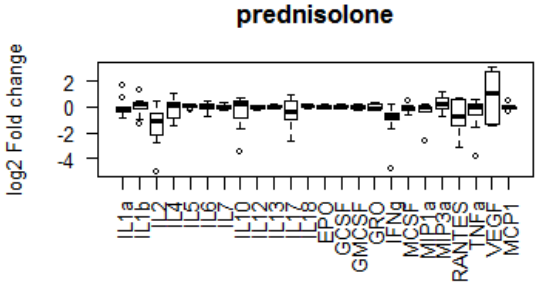

Glatiramer acetate
log2 Fold change
IL1a
IL1b
IL2
IL4
IL5
IL6
IL7
IL10
IL12
IL13
IL17
IL18
EPO
GCSF
GMCSF
GRO
IFNg
MCSF
MIP1a
MIP3a
RANTES
TNFa
VEGF
MCP1


losmapimod
log2 Fold change
IL1a
IL1b
IL2
IL4
IL5
IL6
IL7
IL10
IL12
IL13
IL17
IL18
EPO
GCSF
GMCSF
GRO
IFNg
MCSF
MIP1a
MIP3a
RANTES
TNFa
VEGF
MCP1


mycophenolic acid
log2 Fold change
IL1a
IL1b
IL2
IL4
IL5
IL6
IL7
IL10
IL12
IL13
IL17
IL18
EPO
GCSF
GMCSF
GRO
IFNg
MCSF
MIP1a
MIP3a
RANTES
TNFa
VEGF
MCP1


nilotinib
log2 Fold change
IL1a
IL1b
IL2
IL4
IL5
IL6
IL7
IL10
IL12
IL13
IL17
IL18
EPO
GCSF
GMCSF
GRO
IFNg
MCSF
MIP1a
MIP3a
RANTES
TNFa
VEGF
MCP1


pilocarpine hydrochloride
log2 Fold change
IL1a
IL1b
IL2
IL4
IL5
IL6
IL7
IL10
IL12
IL13
IL17
IL18
EPO
GCSF
GMCSF
GRO
IFNg
MCSF
MIP1a
MIP3a
RANTES
TNFa
VEGF
MCP1


prednisolone
log2 Fold change
IL1a
IL1b
IL2
IL4
IL5
IL6
IL7
IL10
IL12
IL13
IL17
IL18
EPO
GCSF
GMCSF
GRO
IFNg
MCSF
MIP1a
MIP3a
RANTES
TNFa
VEGF
MCP1

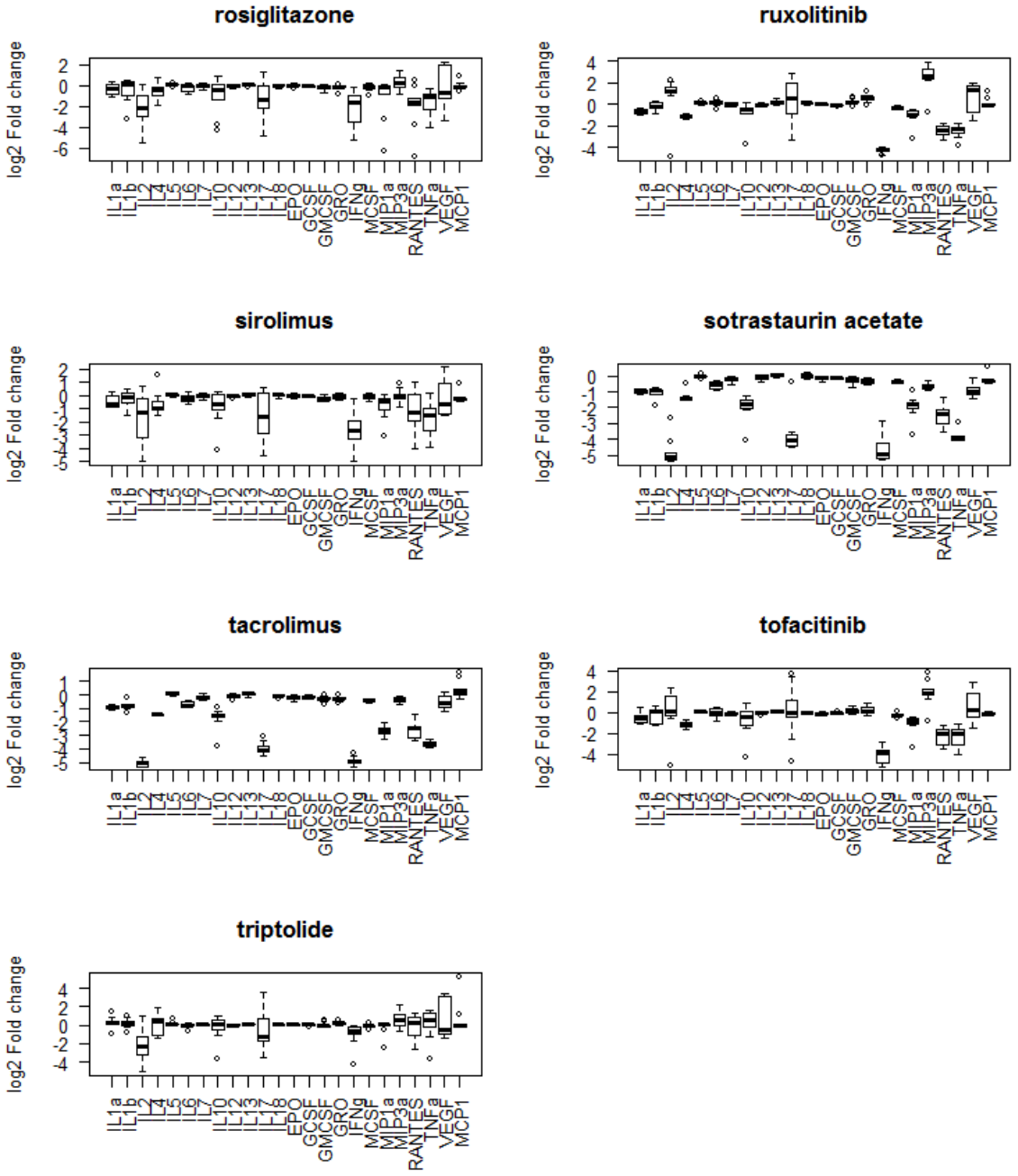


Figure A1. Boxplots of the cytokine profiles. The log2 fold change were calculated for each of the 23 drugs by dividing their FI values by the corresponding positive control (stimulation in the absence of drug). Each box in the figure is based on 9 biological replicates (9 rats).

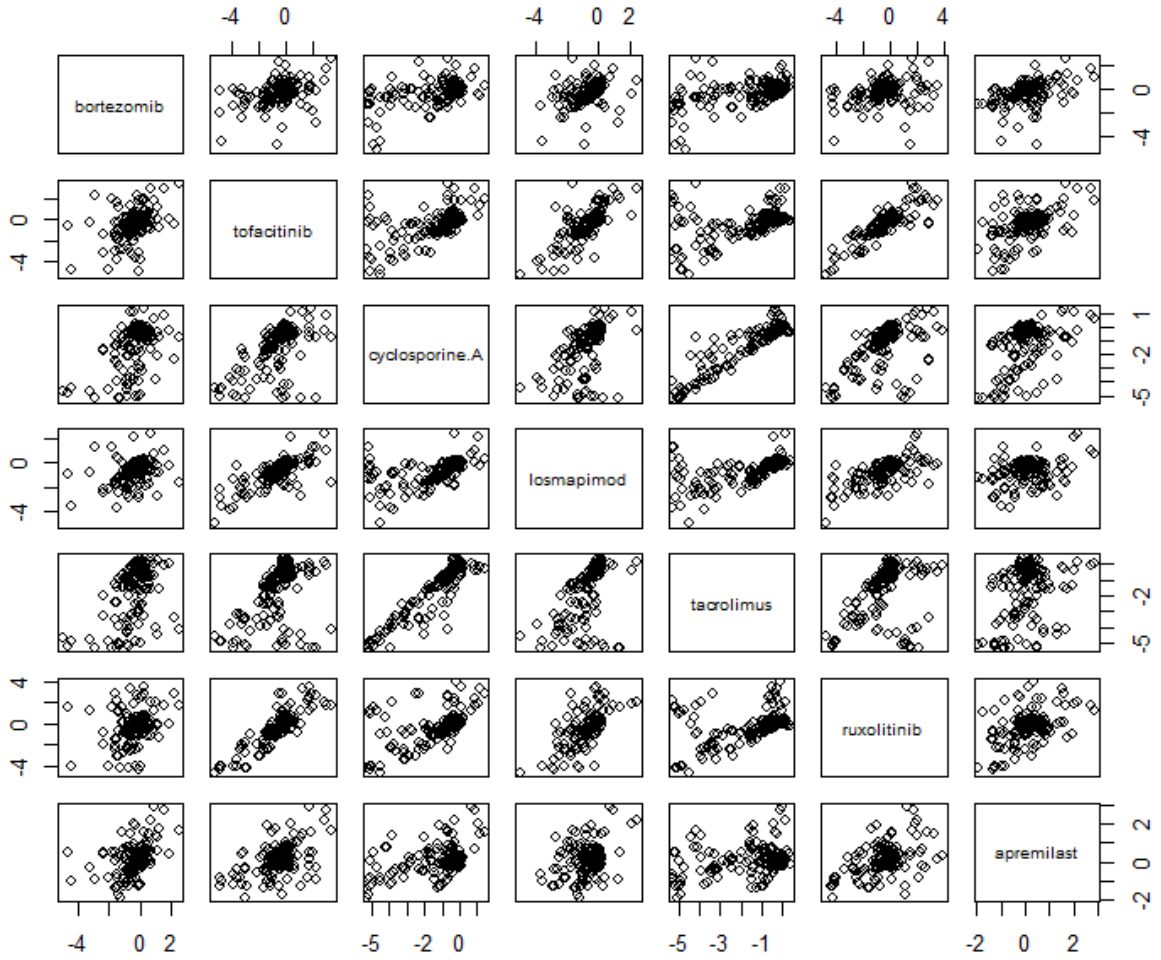


Figure A2. Scatter plot matrix of the log2 fold change values. A sample of seven randomly chosen drugs were drawn out of the 23 drugs in order illustrate a representative figure for the relationship between the drugs. The data set has been pre-filtered by removing outliers as detected by boxplots (data points outside 1.5 interquartile range above the upper quartile or below the lower quartile).

Table A1. The default values used for each scenario.

Scenario	x_1	x_2	x_3	x_4	x_5	$d(drug_1, drug_2)$
A	(5 , 4)	(3 , 2)	(4 , 3)	(5 , 4)	(6 , 5)	5
B	(5 , 5)	(3 , 3)	(4 , 4)	(5 , 2)	(6 , 8)	5
C	(5 , 4)	(2 , 3)	(4 , 3)	(4 , 5)	(6 , 5)	5

* The values for drug 1 and drug 2 for each variable (x_i) used in the simulations for scenarios A, B and C.

Table A2. Top 20 significant Pearson correlation coefficients

Drug 1	***Drug 2***	***r***
tacrolimus	fostamatinib disodium	0.98
tacrolimus	sotrastaurin acetate	0.98
sotrastaurin acetate	fostamatinib disodium	0.98
fostamatinib disodium	cyclosporine A	0.93
tacrolimus	cyclosporine A	0.93
sotrastaurin acetate	cyclosporine A	0.92
fluticasone propionate	everolimus	0.91
nilotinib	fostamatinib disodium	0.90
sotrastaurin acetate	nilotinib	0.89
mycophenolic acid	glatiramer acetate	0.88
nilotinib	losmapimod	0.87
sirolimus	everolimus	0.87
tofacitinib	ruxolitinib	0.87
nilotinib	mycophenolic acid	0.87
tacrolimus	nilotinib	0.86
nilotinib	everolimus	0.84
glatiramer acetate	everolimus	0.84
nilotinib	glatiramer acetate	0.84
prednisolone	dexamethasone	0.83
triptolide	dexamethasone	0.83

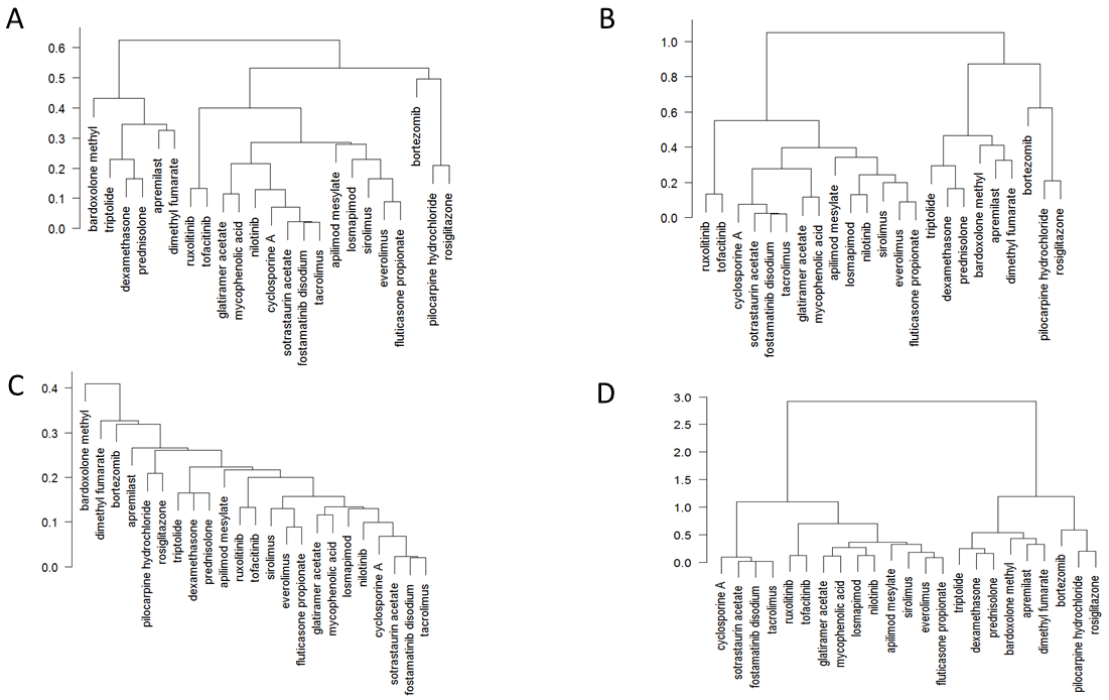


Figure A3. Dendrograms generated by different linkage functions. The dendrograms were generated by agglomerative hierarchical clustering by using the linkage functions: average (A), complete (B), single (C) and Ward (D).

R code for evaluating linkage functions

Calculate the mean bp value of different linkage functions

```
library(pvclust)
# Evaluated linkage functions: "average", "ward.D", "single", "complete"
fit <- pvclust(data, method.hclust="average",method.dist="cor",
use.cor="pairwise.complete.obs", nboot=10000)
mean(fit$edges$bp) # Calculate the mean bp value
# Generate a dendrogram with the bp values
plot(fit,labels=h1, print.pv=TRUE,print.num=FALSE,col.pv=c("white","black"))
```

Calculate cophenetic correlation coefficient

```
library(Hmisc) # Missing values are deleted in pairs
pears=rcorr(as.matrix(data), type="pearson")
dissimilarity <- 1 - pears$r
d1 <- as.dist(dissimilarity)
# Evaluate different linkage functions: "average", "single", "complete", "ward.D"
```

```
klust=hclust(d1,method="average")
d2 <- cophenetic(klust)
cor(d1, d2)
```