

Machine Learning

Katie Pollard

BMI 206

<http://docpollard.org/bmi206>

October 19, 2016

Clustering vs. Classification

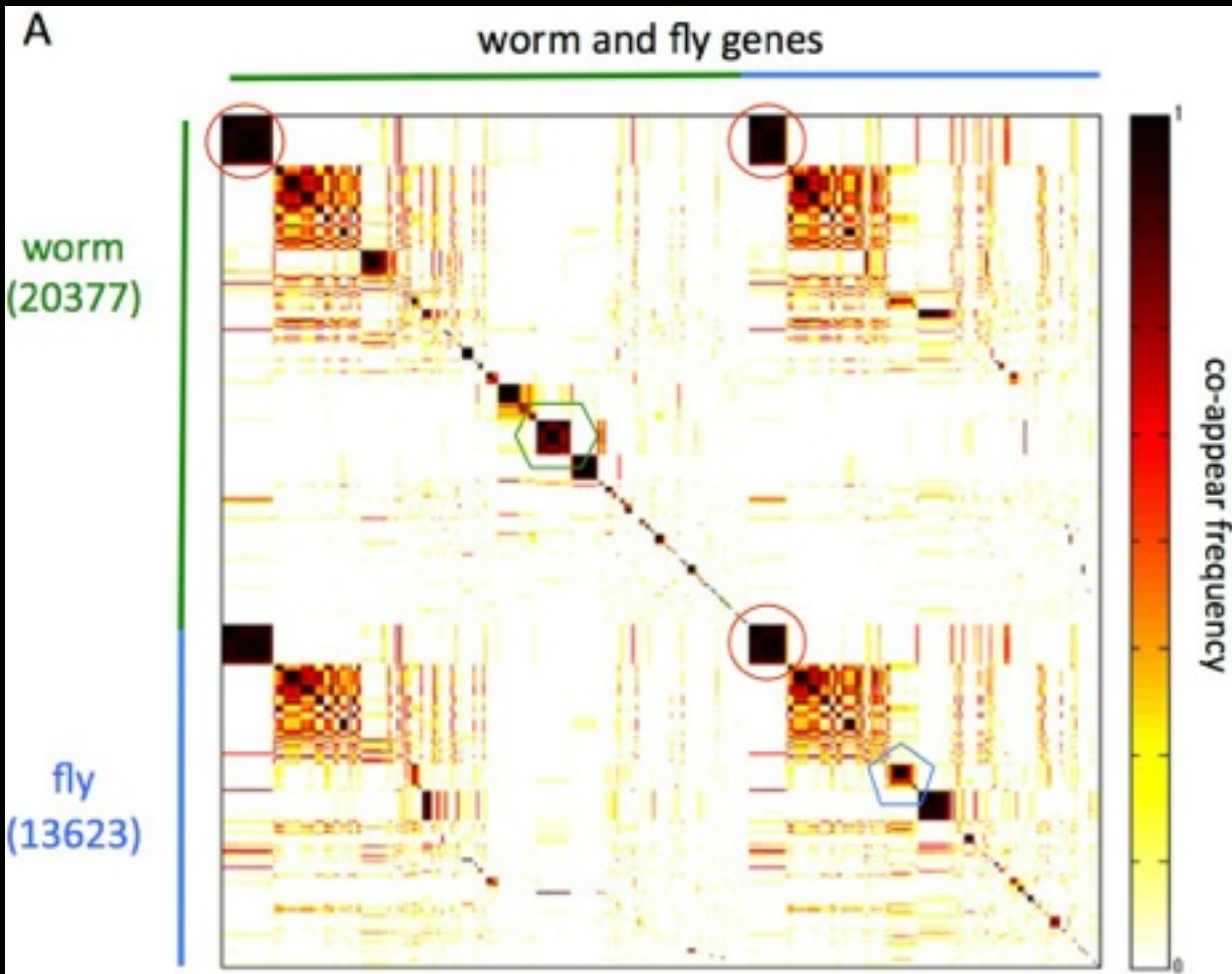
Clustering = unsupervised learning

- Classes: unknown a priori
- Goal: discover classes from the data

Classification = supervised learning

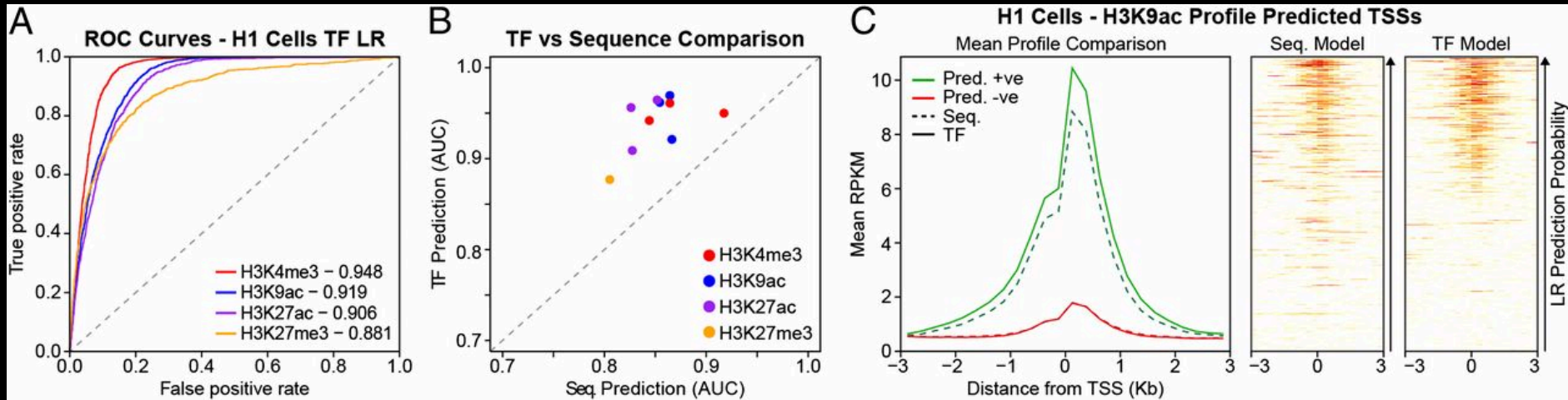
- Classes: known/predefined
- Goals: understand the basis for the classes and build a predictor (classify new data)
- Prediction with a categorical outcome variable

Clustering in Bioinformatics



Worm orthologs of co-expressed genes in fly are some times also co-expressed.

Classification in Bioinformatics



Presence of histone modifications at promoters can be predicted from transcription factor binding and (to a lesser degree) sequence motifs.

DISTANCES

Distances

Machine learning algorithms require a notion of **pairwise distance** between objects.

- **Dissimilarity**

 - Non-negative: $d(x,y) \geq 0$

 - Symmetric: $d(x,y) = d(y,x)$

 - Monotone: $d(x,y) > d(x,z)$ if z more similar to x

- **Metric** (additional conditions)

 - Definite: $d(x,y) = 0$ iff $x=y$

 - Triangle inequality: $d(x,y) + d(y,z) \geq d(x,z)$

Distance Metrics

- Manhattan distance

$$d(x, y) = \sum_i |x_i - y_i| \in (0, \infty)$$

- Euclidean distance

$$d(x, y) = \sqrt{\sum_i (x_i - y_i)^2} \in (0, \infty)$$

Examples of the Minkowski metric

Correlation Distances

$$d(x, y) = 1 - r(x, y) \in (0, 2)$$

Sample correlation measures $r(x, y)$:

- Pearson
- Uncentered (cosine-angle distance)
- Spearman
- Kendall's Tau
- Maximal Information Coefficient

(Correlation measure from Reshef et al. Science 2011 – not used for clustering yet!)

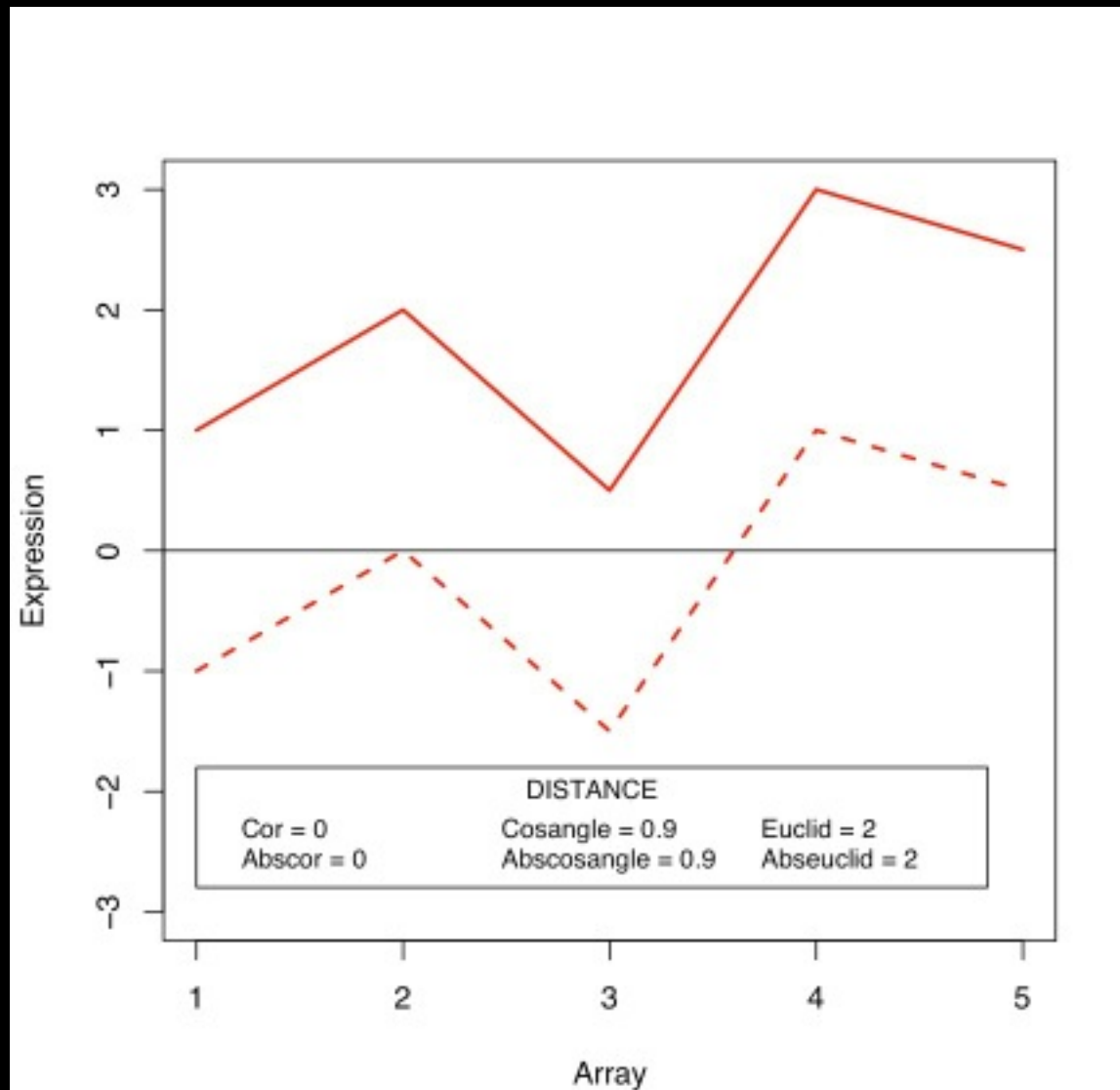
More on Distances

- Minkowski metrics: **magnitude**
- Correlation distances: **pattern** (or both)
- The **absolute value** of any distance can also be used, e.g.

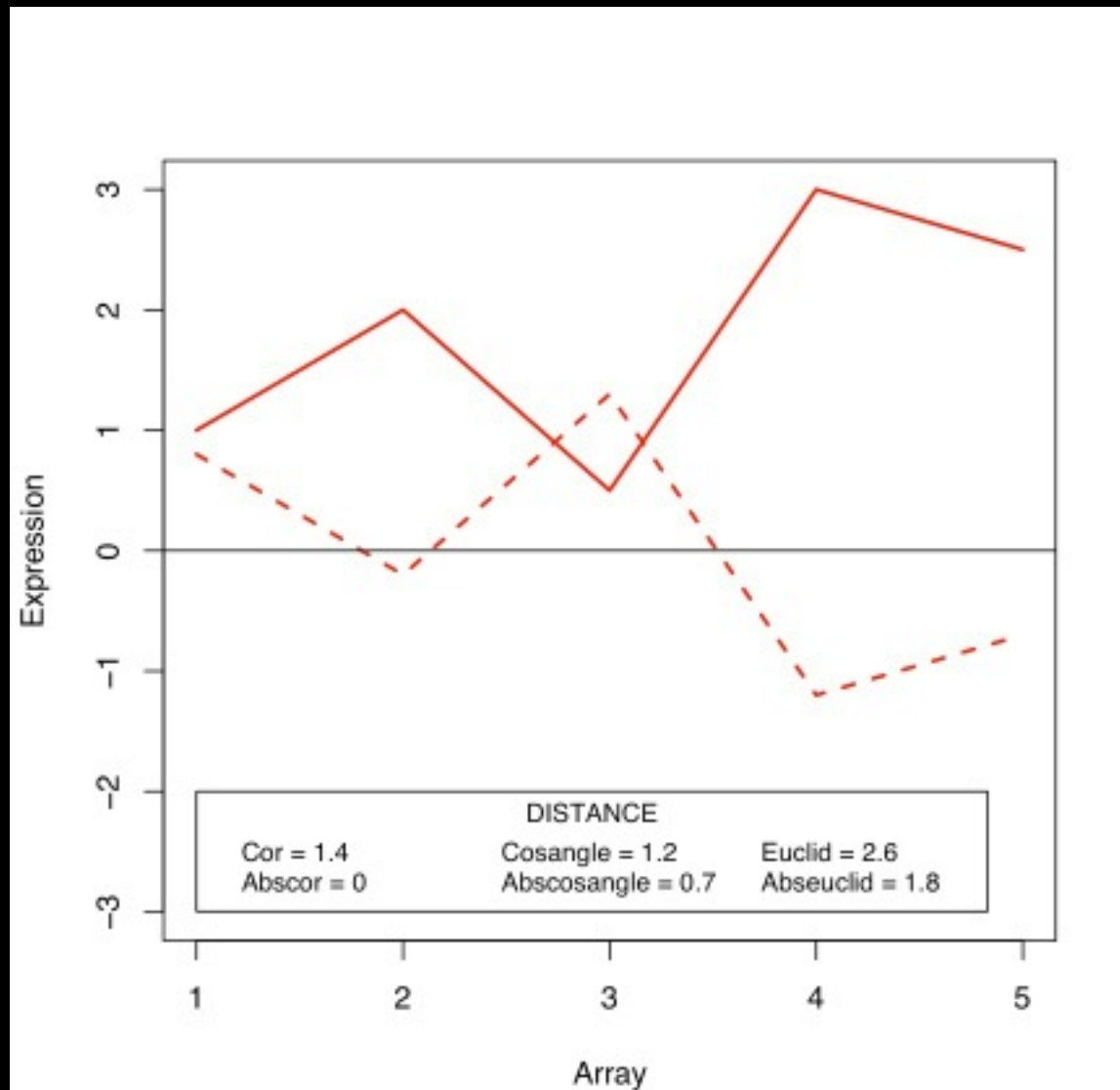
$$d(x,y) = 1 - |r(x,y)| \in (0,1)$$

- Distances between **distributions** are a different concept, e.g., Kullback-Leibler
 $D_{KL}(p(X),q(X)) = - \sum_x p(x) \log\{p(x)/q(x)\}$

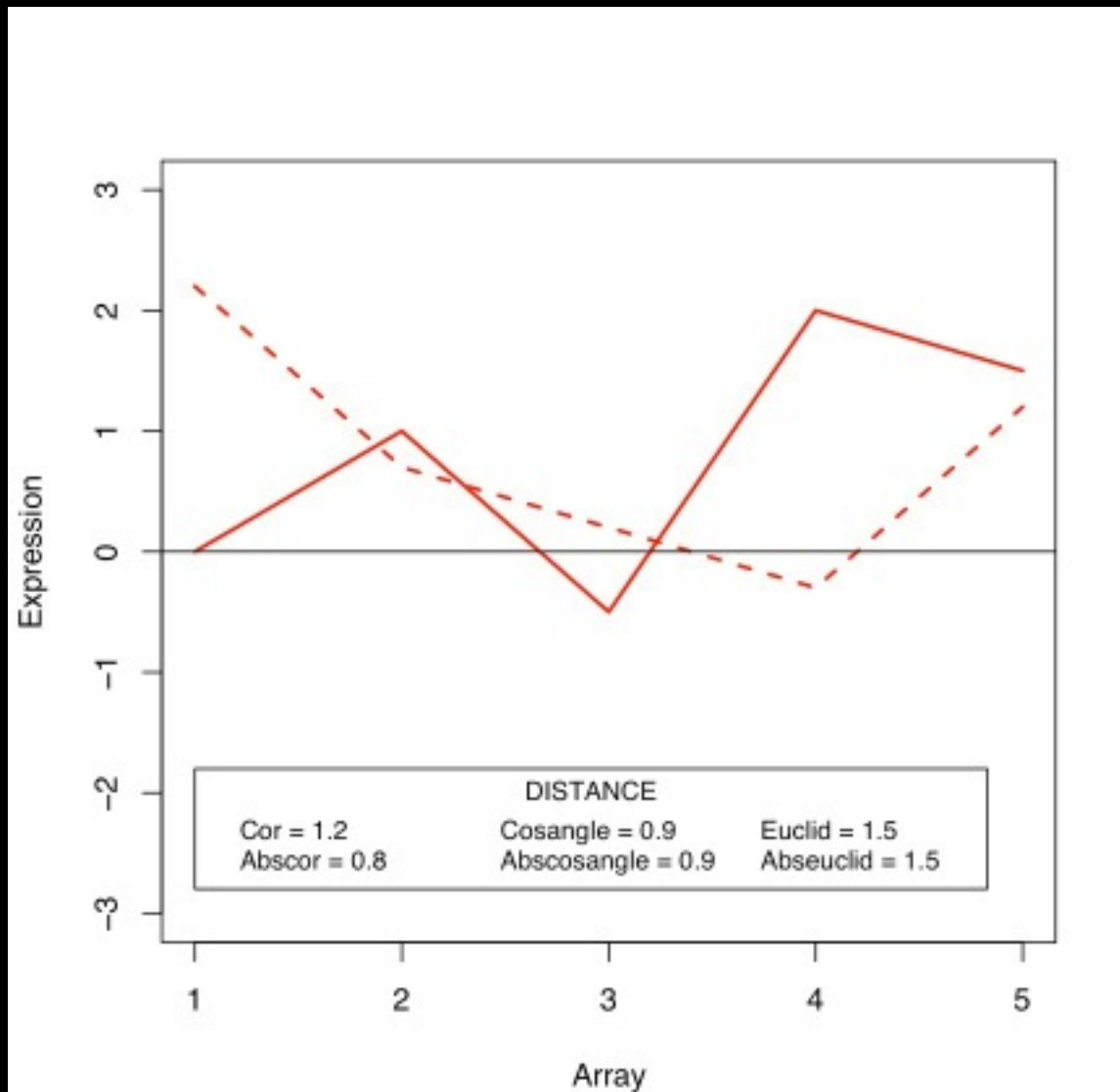
Perfectly Correlated



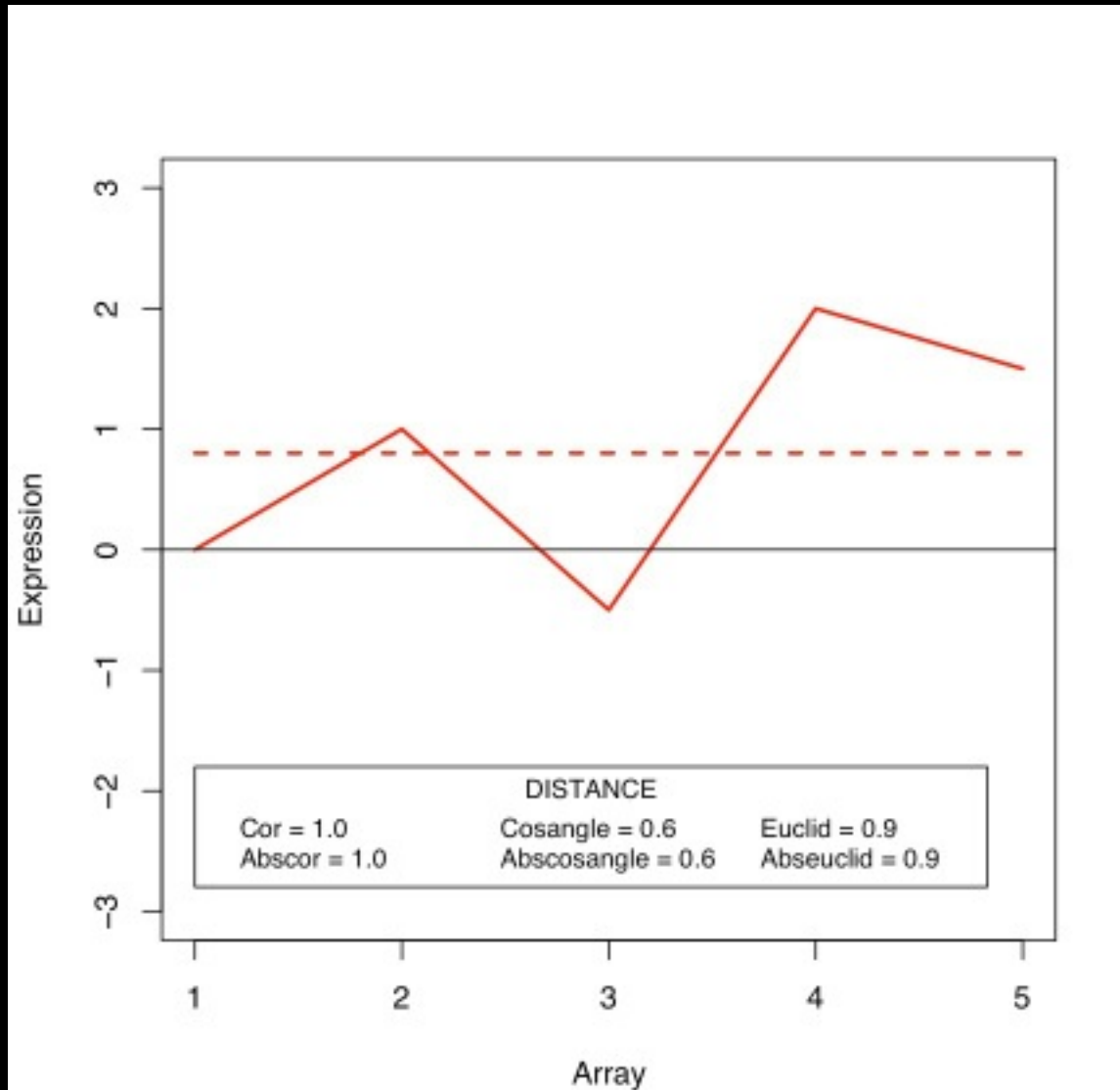
Perfectly Anti-Correlated



Same Mean, Uncorrelated



Same Mean, No Variation



CLUSTERING

Cluster Analysis

Exploratory data analysis methods for:

- Discovering patterns
- Grouping
 - Variables
 - Samples
 - Both simultaneously
- Dimension reduction
 - How many distinct patterns?
- Ordering and organizing

Clustering Methods

Two main components:

1. Distance measure
2. Algorithm

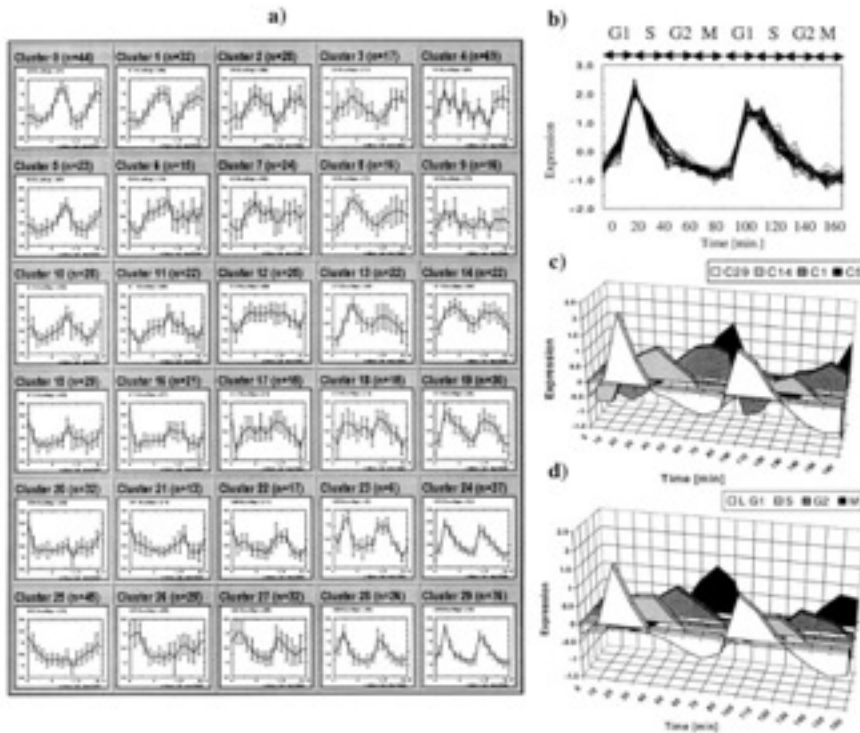
These produce a mapping from data to parameters of interest:

- Cluster labels, sizes, profiles
- Hierarchical tree structure, ordering
- Number of groups

Clustering Algorithms

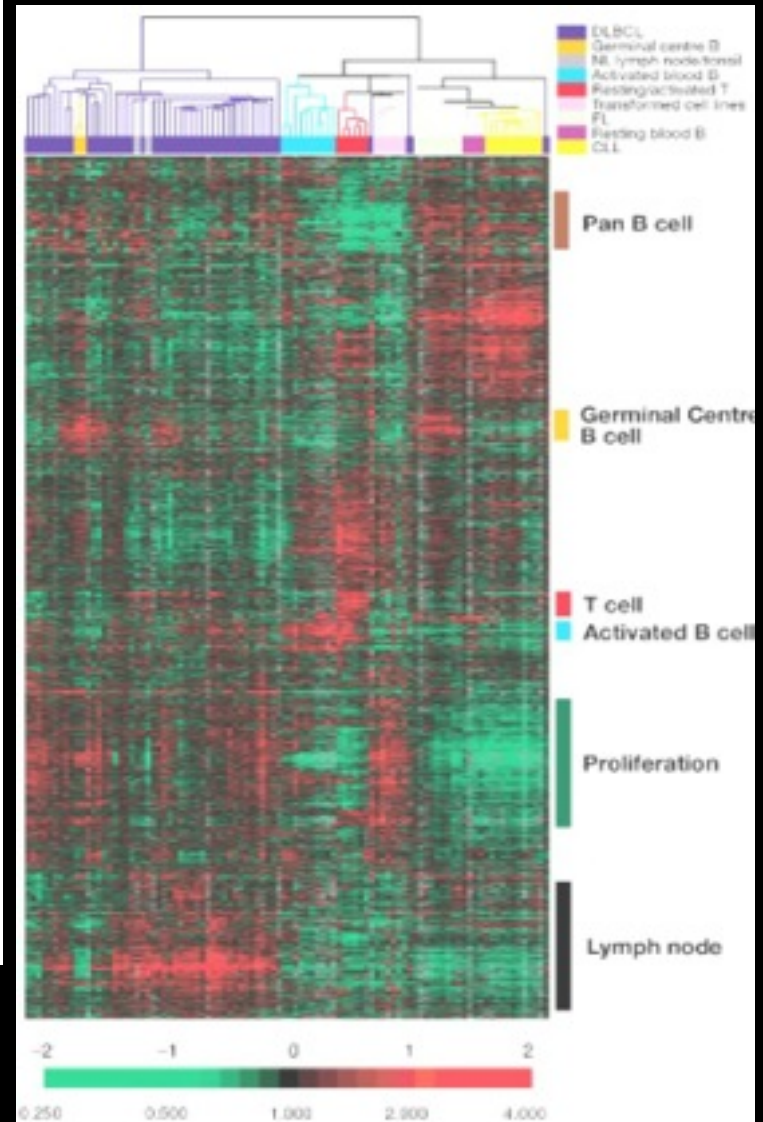
- Model-based (AUTOCLASS, SNOB)
vs. Non-parametric
- Partitioning (SOMs, PAM, KMEANS)
vs. Hierarchical
 - Agglomerative (CLUSTER, AGNES)
 - Linkage: single, complete, average
 - Divisive (SOTA, DIANA, TSVQ)
 - Hybrid (HOPACH, MUTUAL CLUSTERS)

Partitioning vs. Hierarchical



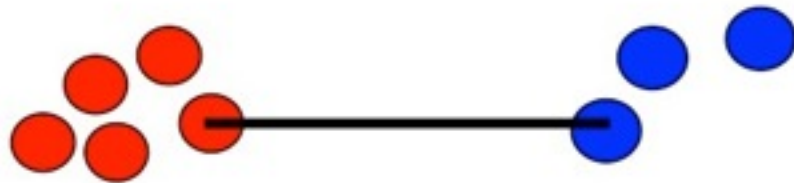
Cho *et al.* (1998) *Molecular Cell*, 2: 65-73

Tamayo *et al.* (1999) *PNAS*, 96: 2907-2912



Alizadeh *et al.* (2000) *Nature*, 403: 503-511

Distance Between Clusters



Complete (minimum)

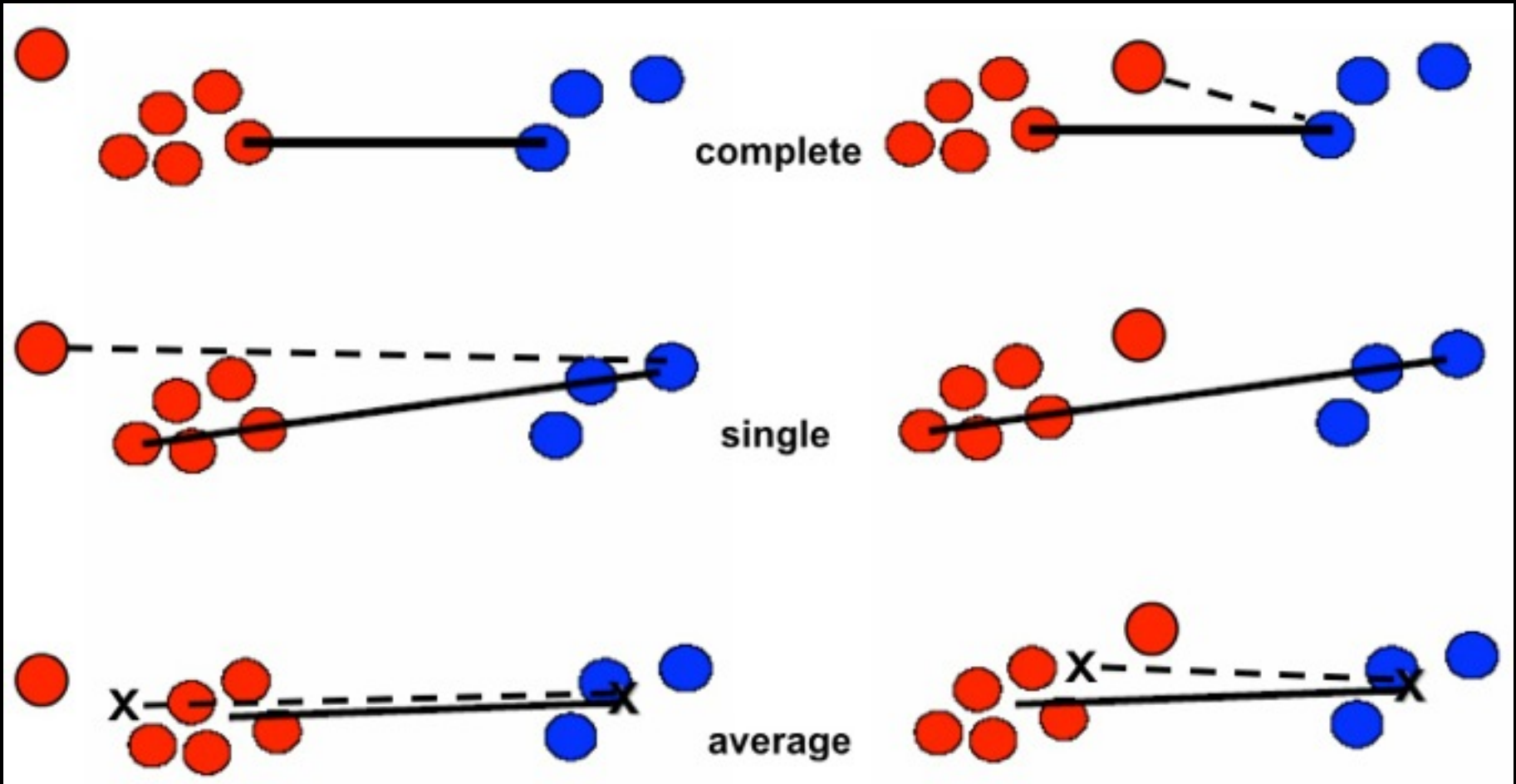


Single (maximum)



Average

Effects of Outliers



How Many Clusters?

- Relevant for both partitioning and hierarchical algorithms (with pruning)
- Level of structure: **global** vs. **detailed**
- Two main approaches:
 - **Direct Methods** (criteria)
e.g. sums of squares, silhouettes
 - **Resampling Methods** (testing)
e.g. Clest, gap statistic, bootstrap

Inference for Clustering

- How **reliable** and **repeatable** are cluster results from a single data set?
- Can view output (e.g. gene cluster labels) as a **parameter estimate**.
- Use resampling methods to estimate the **variability** of this estimator (since no closed form typically).

Example: Bootstrap cluster memberships

Issues in Cluster Analysis

- Results can be very **sensitive** to input (i.e., pre-processing and filtering).
- What method fits your **application**?
 - Distance: capture what “close” means
 - Algorithm: the kind of clusters you seek
- Clustering methods will always return some output, but is it **meaningful**?
 - Evaluate variability
 - Assess biological relevance
 - Confirm hypotheses with experiments

CLASSIFICATION

Classification

- Training: build a **model** for the known class labels of samples as a function of (a subset of) the large set of measured variables.
- Evaluation: quantify how well the model performs on held out or independent data.
- Prediction: model **predicts** unknown class labels of future samples.

Classification Methods

- K nearest neighbors (KNN)
- Discriminant analysis
 - **linear** (LDA)
 - **non-linear** (quadratic, logistic)
- Naïve Bayes
- Support vector machines (SVM)
 - **Multi-Kernel Learning**: linear combination of SVM kernels for subsets of variables.
- Neural networks (CNNs, deep learning)
- Classification trees (CART, random forests)
- Gradient boosting machine (GBM)

Ensemble Classifiers

Build a collection of “weak learners” and combine them for better prediction.

- **Random Forests:** subset of variables.
- **Boosting:** resample and aggregate using weights based on previous misclassifications.
- **Bagging:** classifier for bootstrap samples, aggregate by majority vote.
- **Super Learner:** weighted (via cross validation) combo of different classifiers.

Training Decisions

- **Feature selection** (very important in omics)
 - Correlations with metadata
 - Recursive feature elimination (permutations of out-of-bag observations in random forests)
 - Gradient boosting machine
- **Parameters**
 - KNN: how many neighbors?
 - NNET: how many hidden states?
 - SVM: kernel?
 - Class priors
- Must account for these in **performance evaluation**

Evaluating Performance

- Evaluate **performance** on held out **test data**.
 - Generalizability
 - Over-fitting
- Estimate classification **error rates** by bootstrap (out-of-bag) or cross-validation.
- Must avoid bleeding of information between the training and test sets!
- Prediction error is measured with a **loss function**. Many choices...

Evaluating Performance

		PREDICT	
		1	0
TRUTH	0	FP	TN
	1	TP	FN

Specificity = TNR = $TN / (FP + TN)$

Fall out = FPR = $FP / (FP + TN)$

Sensitivity = TPR = recall = power = $TP / (TP + FN)$

Miss rate = FNR = $FN / (TP + FN)$

Precision = positive predicted value = $TP / (FP + TP)$

False discovery rate = $FP / (FP + TP)$

Negative predicted value = $TN / (TN + FN)$

False omission rate = $FN / (TN + FN)$

Summary Performance Measures

$$\text{Accuracy} = (TP+TN)/(FP+TN+TP+FN)$$

$$\text{F-score} = (1+b^2)(\text{precision} \times \text{recall}) / (b^2 \text{ precision} + \text{recall})$$

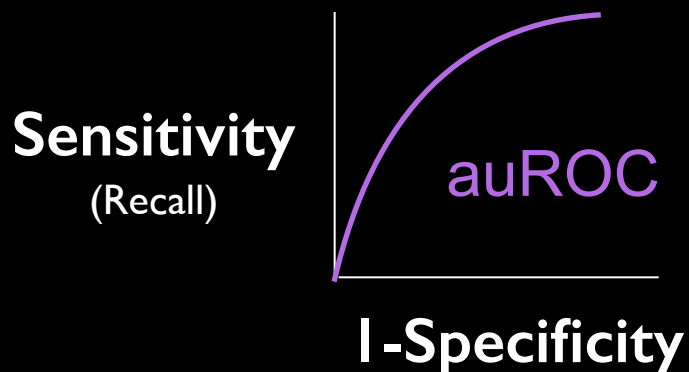
$$F1 = 2TP/(2TP+FP+FN)$$

F1 is harmonic mean of precision and recall

F2 weights recall higher

F0.5 weight precision higher

Area under the curve:



Other Issues...

- Human input for training, not prediction
- Multi-class problems
 - all vs. one and other binary reductions
 - Bagging trees, random forests OK
- Standardization of variables
- Variable importance

EXTRA SLIDES

Comparing Clustering Methods

- Any **distance** measure?
- Reaction to **noise**: robust vs. efficient
- Results **reproducible**?
- Biological **relevance**

Partitioning

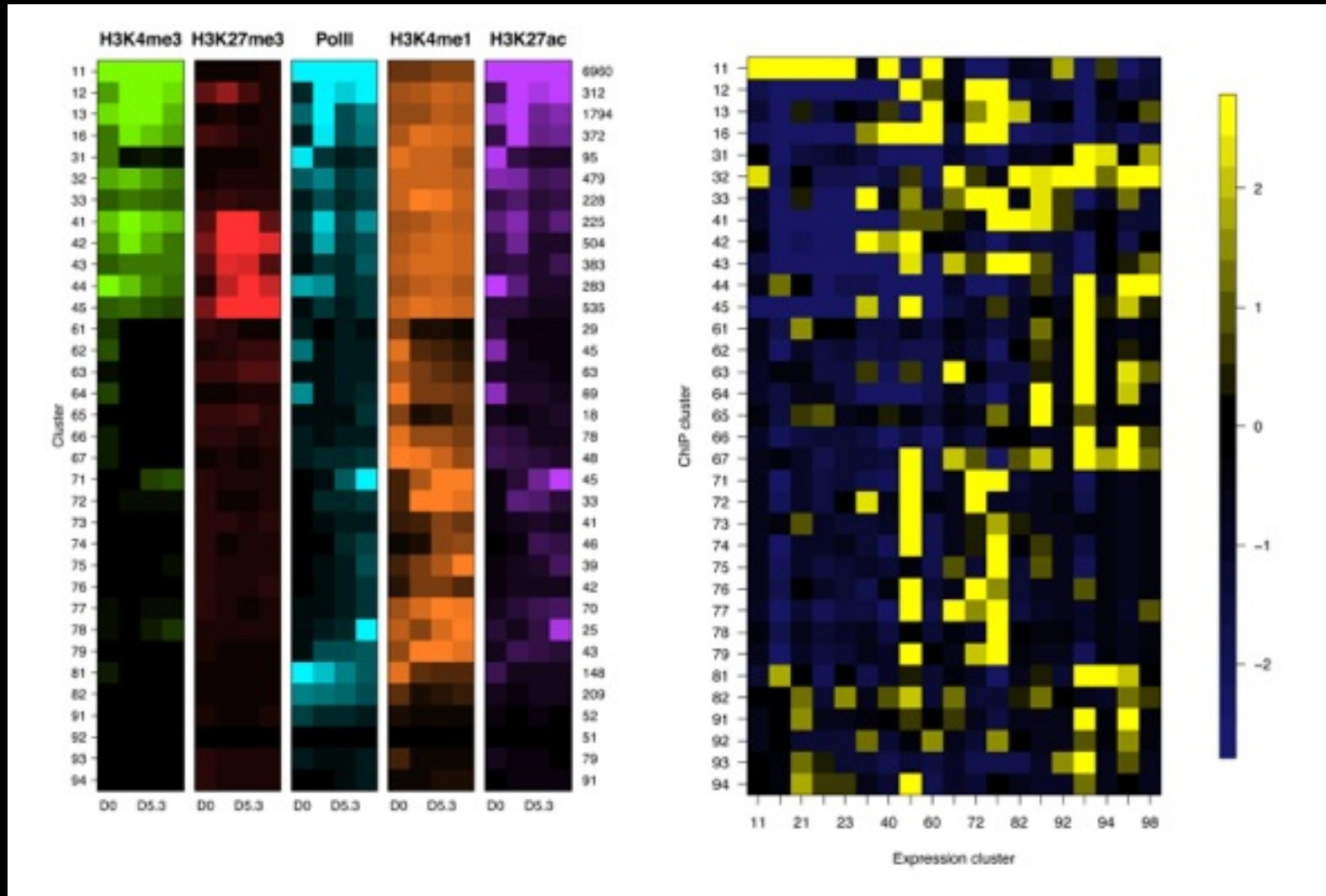
Variety of cluster sizes
Overlapping clusters

Hierarchical

Sensible ordering

Hybrid tree of partitions

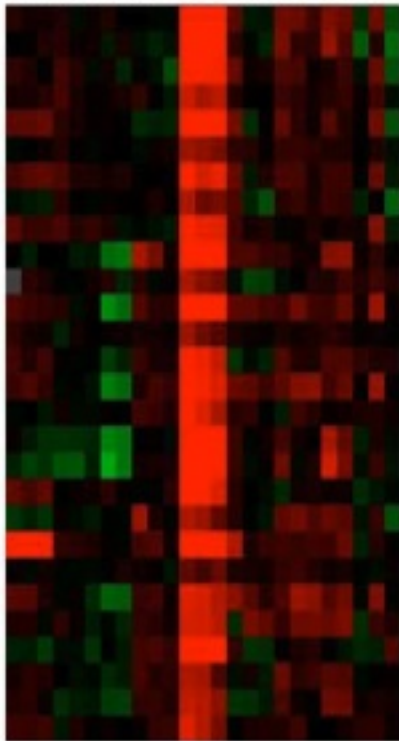
HOPACH: Pruned hierarchical tree produces nested clusters



Bootstrap Fuzzy Clustering

boothopach function

PAE05_1500g
PAE06_1500g
PAE07_1500g
PAE08_1500g
PAE09_1500g
PAE10_1500g
PAE11_1500g
PAE12_1500g
PAE13_1500g
PAE14_1500g
PAE15_1500g
PAE16_1500g
PAE17_1500g
PAE18_1500g
PAE19_1500g
PAE20_1500g
PAE21_1500g
PAE22_1500g
PAE23_1500g
PAE24_1500g
PAE25_1500g
PAE26_1500g
PAE27_1500g
PAE28_1500g
PAE29_1500g
PAE30_1500g
PAE31_1500g
PAE32_1500g
PAE33_1500g
PAE34_1500g
PAE35_1500g
PAE36_1500g
PAE37_1500g
PAE38_1500g
PAE39_1500g
PAE40_1500g
PAE41_1500g
PAE42_1500g
PAE43_1500g
PAE44_1500g
PAE45_1500g
PAE46_1500g
PAE47_1500g
PAE48_1500g
PAE49_1500g
PAE50_1500g
PAE51_1500g
PAE52_1500g
PAE53_1500g
PAE54_1500g
PAE55_1500g
PAE56_1500g
PAE57_1500g
PAE58_1500g
PAE59_1500g
PAE60_1500g
PAE61_1500g
PAE62_1500g
PAE63_1500g
PAE64_1500g
PAE65_1500g
PAE66_1500g
PAE67_1500g
PAE68_1500g
PAE69_1500g
PAE70_1500g
PAE71_1500g
PAE72_1500g
PAE73_1500g
PAE74_1500g
PAE75_1500g
PAE76_1500g
PAE77_1500g
PAE78_1500g
PAE79_1500g
PAE80_1500g
PAE81_1500g
PAE82_1500g
PAE83_1500g
PAE84_1500g
PAE85_1500g
PAE86_1500g
PAE87_1500g
PAE88_1500g
PAE89_1500g
PAE90_1500g
PAE91_1500g
PAE92_1500g
PAE93_1500g
PAE94_1500g
PAE95_1500g
PAE96_1500g
PAE97_1500g
PAE98_1500g
PAE99_1500g
PAE100_1500g
PAE101_1500g
PAE102_1500g
PAE103_1500g
PAE104_1500g
PAE105_1500g
PAE106_1500g
PAE107_1500g
PAE108_1500g
PAE109_1500g
PAE110_1500g
PAE111_1500g
PAE112_1500g
PAE113_1500g
PAE114_1500g
PAE115_1500g
PAE116_1500g
PAE117_1500g
PAE118_1500g
PAE119_1500g
PAE120_1500g
PAE121_1500g
PAE122_1500g
PAE123_1500g
PAE124_1500g
PAE125_1500g
PAE126_1500g
PAE127_1500g
PAE128_1500g
PAE129_1500g
PAE130_1500g
PAE131_1500g
PAE132_1500g
PAE133_1500g
PAE134_1500g
PAE135_1500g
PAE136_1500g
PAE137_1500g
PAE138_1500g
PAE139_1500g
PAE140_1500g
PAE141_1500g
PAE142_1500g
PAE143_1500g
PAE144_1500g
PAE145_1500g
PAE146_1500g
PAE147_1500g
PAE148_1500g
PAE149_1500g
PAE150_1500g
PAE151_1500g
PAE152_1500g
PAE153_1500g
PAE154_1500g
PAE155_1500g
PAE156_1500g
PAE157_1500g
PAE158_1500g
PAE159_1500g
PAE160_1500g
PAE161_1500g
PAE162_1500g
PAE163_1500g
PAE164_1500g
PAE165_1500g
PAE166_1500g
PAE167_1500g
PAE168_1500g
PAE169_1500g
PAE170_1500g
PAE171_1500g
PAE172_1500g
PAE173_1500g
PAE174_1500g
PAE175_1500g
PAE176_1500g
PAE177_1500g
PAE178_1500g
PAE179_1500g
PAE180_1500g
PAE181_1500g
PAE182_1500g
PAE183_1500g
PAE184_1500g
PAE185_1500g
PAE186_1500g
PAE187_1500g
PAE188_1500g
PAE189_1500g
PAE190_1500g
PAE191_1500g
PAE192_1500g
PAE193_1500g
PAE194_1500g
PAE195_1500g
PAE196_1500g
PAE197_1500g
PAE198_1500g
PAE199_1500g
PAE200_1500g
PAE201_1500g
PAE202_1500g
PAE203_1500g
PAE204_1500g
PAE205_1500g
PAE206_1500g
PAE207_1500g
PAE208_1500g
PAE209_1500g
PAE210_1500g
PAE211_1500g
PAE212_1500g
PAE213_1500g
PAE214_1500g
PAE215_1500g
PAE216_1500g
PAE217_1500g
PAE218_1500g
PAE219_1500g
PAE220_1500g
PAE221_1500g
PAE222_1500g
PAE223_1500g
PAE224_1500g
PAE225_1500g
PAE226_1500g
PAE227_1500g
PAE228_1500g
PAE229_1500g
PAE230_1500g
PAE231_1500g
PAE232_1500g
PAE233_1500g
PAE234_1500g
PAE235_1500g
PAE236_1500g
PAE237_1500g
PAE238_1500g
PAE239_1500g
PAE240_1500g
PAE241_1500g
PAE242_1500g
PAE243_1500g
PAE244_1500g
PAE245_1500g
PAE246_1500g
PAE247_1500g
PAE248_1500g
PAE249_1500g
PAE250_1500g
PAE251_1500g
PAE252_1500g
PAE253_1500g
PAE254_1500g
PAE255_1500g
PAE256_1500g
PAE257_1500g
PAE258_1500g
PAE259_1500g
PAE260_1500g
PAE261_1500g
PAE262_1500g
PAE263_1500g
PAE264_1500g
PAE265_1500g
PAE266_1500g
PAE267_1500g
PAE268_1500g
PAE269_1500g
PAE270_1500g
PAE271_1500g
PAE272_1500g
PAE273_1500g
PAE274_1500g
PAE275_1500g
PAE276_1500g
PAE277_1500g
PAE278_1500g
PAE279_1500g
PAE280_1500g
PAE281_1500g
PAE282_1500g
PAE283_1500g
PAE284_1500g
PAE285_1500g
PAE286_1500g
PAE287_1500g
PAE288_1500g
PAE289_1500g
PAE290_1500g
PAE291_1500g
PAE292_1500g
PAE293_1500g
PAE294_1500g
PAE295_1500g
PAE296_1500g
PAE297_1500g
PAE298_1500g
PAE299_1500g
PAE300_1500g
PAE301_1500g
PAE302_1500g
PAE303_1500g
PAE304_1500g
PAE305_1500g
PAE306_1500g
PAE307_1500g
PAE308_1500g
PAE309_1500g
PAE310_1500g
PAE311_1500g
PAE312_1500g
PAE313_1500g
PAE314_1500g
PAE315_1500g
PAE316_1500g
PAE317_1500g
PAE318_1500g
PAE319_1500g
PAE320_1500g
PAE321_1500g
PAE322_1500g
PAE323_1500g
PAE324_1500g
PAE325_1500g
PAE326_1500g
PAE327_1500g
PAE328_1500g
PAE329_1500g
PAE330_1500g
PAE331_1500g
PAE332_1500g
PAE333_1500g
PAE334_1500g
PAE335_1500g
PAE336_1500g
PAE337_1500g
PAE338_1500g
PAE339_1500g
PAE340_1500g
PAE341_1500g
PAE342_1500g
PAE343_1500g
PAE344_1500g
PAE345_1500g
PAE346_1500g
PAE347_1500g
PAE348_1500g
PAE349_1500g
PAE350_1500g
PAE351_1500g
PAE352_1500g
PAE353_1500g
PAE354_1500g
PAE355_1500g
PAE356_1500g
PAE357_1500g
PAE358_1500g
PAE359_1500g
PAE360_1500g
PAE361_1500g
PAE362_1500g
PAE363_1500g
PAE364_1500g
PAE365_1500g
PAE366_1500g
PAE367_1500g
PAE368_1500g
PAE369_1500g
PAE370_1500g
PAE371_1500g
PAE372_1500g
PAE373_1500g
PAE374_1500g
PAE375_1500g
PAE376_1500g
PAE377_1500g
PAE378_1500g
PAE379_1500g
PAE380_1500g
PAE381_1500g
PAE382_1500g
PAE383_1500g
PAE384_1500g
PAE385_1500g
PAE386_1500g
PAE387_1500g
PAE388_1500g
PAE389_1500g
PAE390_1500g
PAE391_1500g
PAE392_1500g
PAE393_1500g
PAE394_1500g
PAE395_1500g
PAE396_1500g
PAE397_1500g
PAE398_1500g
PAE399_1500g
PAE400_1500g
PAE401_1500g
PAE402_1500g
PAE403_1500g
PAE404_1500g
PAE405_1500g
PAE406_1500g
PAE407_1500g
PAE408_1500g
PAE409_1500g
PAE410_1500g
PAE411_1500g
PAE412_1500g
PAE413_1500g
PAE414_1500g
PAE415_1500g
PAE416_1500g
PAE417_1500g
PAE418_1500g
PAE419_1500g
PAE420_1500g
PAE421_1500g
PAE422_1500g
PAE423_1500g
PAE424_1500g
PAE425_1500g
PAE426_1500g
PAE427_1500g
PAE428_1500g
PAE429_1500g
PAE430_1500g
PAE431_1500g
PAE432_1500g
PAE433_1500g
PAE434_1500g
PAE435_1500g
PAE436_1500g
PAE437_1500g
PAE438_1500g
PAE439_1500g
PAE440_1500g
PAE441_1500g
PAE442_1500g
PAE443_1500g
PAE444_1500g
PAE445_1500g
PAE446_1500g
PAE447_1500g
PAE448_1500g
PAE449_1500g
PAE450_1500g
PAE451_1500g
PAE452_1500g
PAE453_1500g
PAE454_1500g
PAE455_1500g
PAE456_1500g
PAE457_1500g
PAE458_1500g
PAE459_1500g
PAE460_1500g
PAE461_1500g
PAE462_1500g
PAE463_1500g
PAE464_1500g
PAE465_1500g
PAE466_1500g
PAE467_1500g
PAE468_1500g
PAE469_1500g
PAE470_1500g
PAE471_1500g
PAE472_1500g
PAE473_1500g
PAE474_1500g
PAE475_1500g
PAE476_1500g
PAE477_1500g
PAE478_1500g
PAE479_1500g
PAE480_1500g
PAE481_1500g
PAE482_1500g
PAE483_1500g
PAE484_1500g
PAE485_1500g
PAE486_1500g
PAE487_1500g
PAE488_1500g
PAE489_1500g
PAE490_1500g
PAE491_1500g
PAE492_1500g
PAE493_1500g
PAE494_1500g
PAE495_1500g
PAE496_1500g
PAE497_1500g
PAE498_1500g
PAE499_1500g
PAE500_1500g



MembersHIP	Name
1.0	PAE2648 C,C : PROD: pyruvate dehydrogenase E2 (dihydrolipoamide acetyltransferase), CAT: Energy metabolism,Glycolysis/gluconeogenesis; HMM-ID: 2-oxoacid_dh
0.999	PAE2695 R : PROD: conserved hypothetical protein, CAT: Hypothetical,Conserved; HMM-ID: Glyoxalase
0.998	PAE2201 S : PROD: conserved hypothetical protein, CAT: Hypothetical,Conserved; HMM-ID: NA
0.993	PAE1408 NA : PROD: hypothetical protein, CAT: Hypothetical,Hypothetical; HMM-ID: NA
0.974	PAE2687 ER,R,R : PROD: alcohol dehydrogenase (Zinc), CAT: Energy metabolism,Fermentation; HMM-ID: adh_zinc
0.97	PAE0956 NA : PROD: amidotransferase (hisH), CAT: Amino acid biosynthesis,Histidine family; HMM-ID: GATase
0.963	PAE0724 NA : Between features PAE1365 and PAE1366 Length: 139 bp
0.948	PAE3562 K : PROD: DNA-directed RNA polymerase subunit E' (poE1), CAT: Transcription,DNA-dependent RNA polymerase; HMM-ID: S1
0.945	PAE1102 NA : PROD: tricom protease, CAT: Protein fate,Degradation of proteins, peptides, and glycopeptides ; HMM-ID: NA
0.945	PAE2284 S : PROD: conserved hypothetical protein, CAT: Hypothetical,Conserved; HMM-ID: filament
0.943	PAE2458 E,E : PROD: tryptophan synthase alpha subunit; CAT: Amino acid biosynthesis,Aromatic amino acid family; HMM-ID: tp_synth
0.94	PAE0152 NA : PROD: conserved hypothetical protein, CAT: Hypothetical,Conserved; HMM-ID: NA
0.939	PAE1035 NA : Between features PAE1931 and PAE1932 Length: 170 bp
0.931	PAE2258 K : PROD: DNA-directed RNA polymerase subunit, CAT: Transcription,DNA-dependent RNA polymerase; HMM-ID: NA
0.922	PAE2564 NA : PROD: adenylosulfate reductase alpha subunit part 2, authentic; frameshift, CAT: Central intermediary metabolism,Sulfur metabolism; HMM-ID: NA
0.917	PAE0958 E,E,E : PROD: histidinol-phosphate aminotransferase (HisC) ; CAT: Amino acid biosynthesis,Histidine family; HMM-ID: aminotran_2,aminotran_1
0.911	PAE2612 C,C : PROD: nitrate reductase beta subunit (narH), CAT: Energy metabolism,Electron transport; HMM-ID: narH
0.909	PAE3584 NA : PROD: citrate synthase part 1, authentic; frameshift, CAT: Energy metabolism,TCA cycle; HMM-ID: citrate_synth
0.893	PAE0325 NA : PROD: possible protein, CAT: may not code; HMM-ID: NA
0.873	PAE3596 NA : PROD: RNA 3'-terminal phosphate cyclase ; CAT: Transcription,RNA processing; HMM-ID: RCT
0.866	PAE1956 NA : PROD: hypothetical protein, CAT: Hypothetical,Hypothetical; HMM-ID: NA
0.856	PAE3488 M : PROD: UDP-N-acetylglucosamine-dolichyl-phosphate N-acetylglucosaminylphosphotransferase; CAT: Cell envelope,Biosynthesis and degradation of surface polysaccharides
0.848	PAE2597 NA : PROD: sulfite reductase beta subunit, CAT: Central intermediary metabolism,Sulfur metabolism; HMM-ID: fer4
0.829	PAE2296 H : PROD: riboflavin synthase, CAT: Biosynthesis of cofactors, prosthetic groups, and carriers;Riboflavin, FMN, and FAD; HMM-ID: DMRL_synthase
0.823	PAE0263 NA : Between features PAE0499 and PAE0500 Length: 193 bp
0.812	PAE2222b NA : PROD: hypothetical protein, CAT: Hypothetical,Hypothetical; HMM-ID: NA
0.809	PAE2254 NA : PROD: hypothetical protein, CAT: Hypothetical,Hypothetical; HMM-ID: bZP
0.808	PAE2034 NA : PROD: hypothetical protein, CAT: Hypothetical,Hypothetical; HMM-ID: NA

Silhouette-based Criteria

- The **silhouette** for j 'th object (e.g., gene):

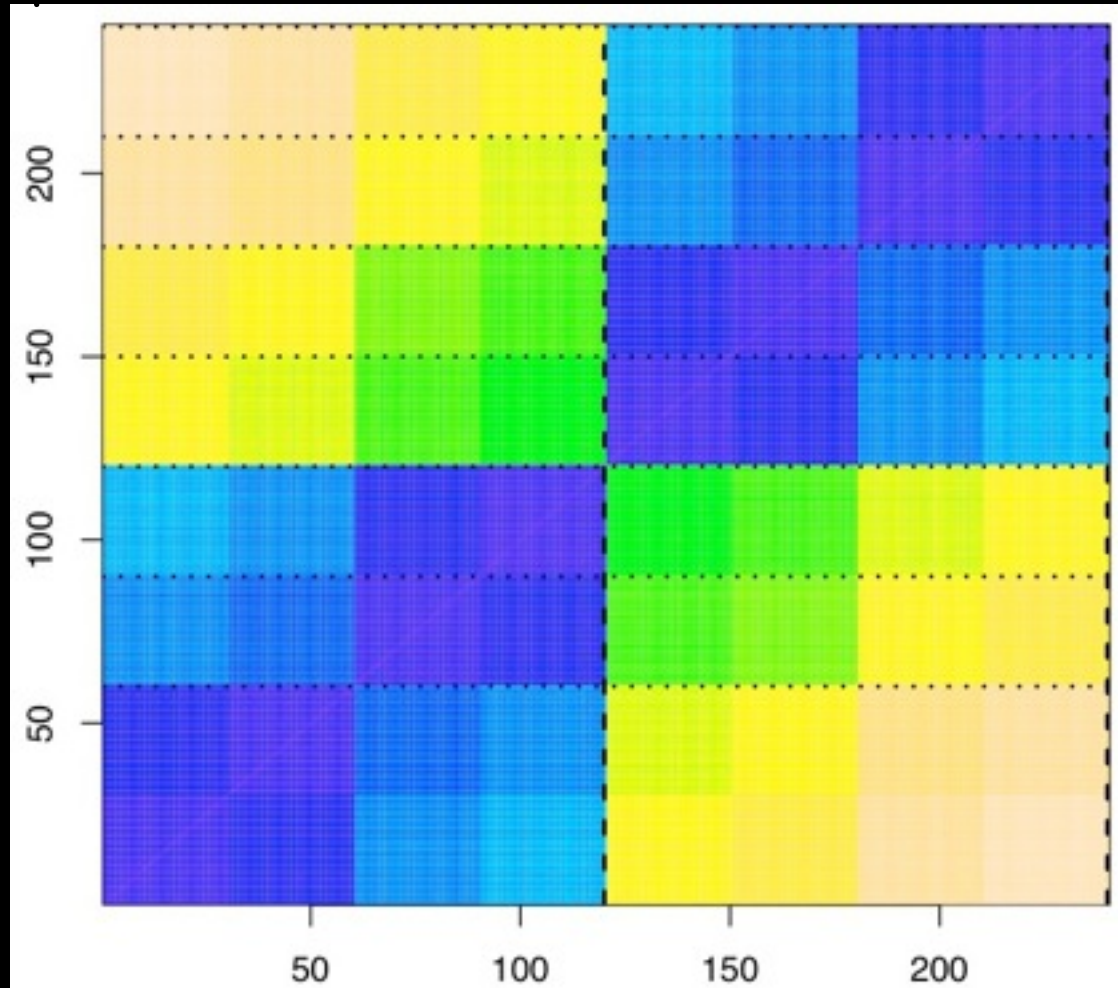
$$S_j = \frac{a_j - b_j}{\max(a_j, b_j)}$$

own cluster \swarrow $a_j - b_j$ \nwarrow next closest cluster

- Average silhouette (**AS**): overall mean
- Median split silhouette (**MSS**): split each cluster and see if silhouettes get smaller

MSS vs. AS

**MSS
(k=7)**



AS (k=2)

PAM clustering
Euclidean distance

Distances in R

Function	Package	Distances
dist	stats	Euclidean, Manhattan, Canberra, max, binary
daisy	cluster bioDist	Euclidean, Manhattan
distancematrix distancevector	hopach	Euclidean, cor, cosine-angle (abs versions)

Clustering Algorithms in R

Package	Functions	Type
stats	kmeans hclust	partitioning divisive
class	SOM	partitioning
cluster	pam agnes diana	partitioning agglomerative divisive
hopach	hopach	hybrid

Other packages: **cclust**, **e1071**, **flexmix**, **fpc**, **mclust**

Classification Methods in R

Package	Functions
<code>e1071</code>	<code>naiveBayes</code> , <code>svm</code>
<code>glmnet</code>	<code>glmnet</code> , <code>cv.glmnet</code>
<code>ipred</code>	<code>lda</code> , <code>bagging</code>
<code>nnet</code>	<code>nnet</code>
<code>rpart</code>	<code>rpart</code>
<code>gbm</code>	<code>gbm</code> , <code>summary</code>
<code>randomForest</code>	<code>randomForest</code>
<code>pamr</code>	<code>knn</code> , <code>cv</code>

`caret`, `MLInterfaces`: frameworks to interface with methods