

Determining Method of Action in Drug Discovery Using Affymetrix Microarray Data

Max Kuhn

max.kuhn@pfizer.com

Pfizer Global R&D
Research Statistics
Groton, CT

Method of Action

As the level of drug resistance increases, the need for antibiotics with novel method of action (MOA) has also increased.

An important part of drug discovery is solidifying the MOA of promising anti-infective compounds. This can increase the odds of the compound becoming a successful drug.

Discovery scientists would like to use data on existing compounds with known MOA to predict or rule out specific MOA for new compounds. They would also like to know what predictors have an influence of method of action.

Gene Expression

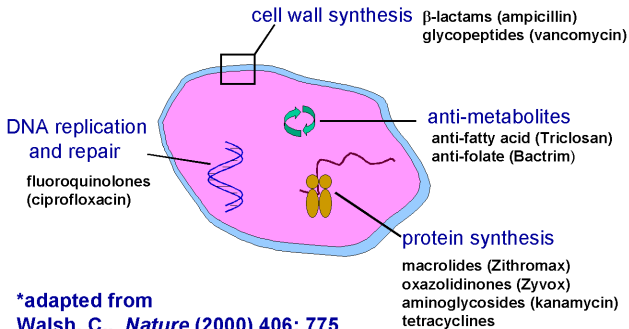
Several publications have linked gene transcript profiles to method of action and we assume that gene expression in bacteria contains relevant information.

Gene expression profiles for a set of existing compounds/drugs with known MOA were generated and used to develop a predictive model for defining the MOA in new compounds. In some cases, it is enough to rule out several mechanisms.

staph. aureus RN4220 samples were treated with 27 antibiotics and noxious agents.

Their RNA was harvested, QC'ed and converted to cDNA. The cDNA was assayed using a custom Affy gene chip with 7775 probes for *staph. aureus* bacteria was developed to represent the genomes of several clinical isolates.

Mechanism of Action for Common Antibiotics*



*adapted from
Walsh, C., *Nature* (2000) 406: 775.

Sample Allocation

There were 114 staph samples across 9 MOAs. They were partitioned into training sets and test sets using a roughly 80/20 split:

MOA	Label	Training	Test
RNA synthesis inhibitors	A	8	2
DNA synthesis inhibitors	B	12	2
Protein Synthesis Inhibitors (30S)	C	13	3
Protein Synthesis Inhibitors (50S)	D	12	3
Cell Wall Synthesis Inhibitors	E	22	5
Anti-metabolites	F	9	2
Fatty Acid Biosynthesis Inhibitors	G	6	1
PMF Uncouplers	H	6	1
Noxious Agents	I	6	1
Total		94	20

Data Processing

Typically, we would run `rma` on samples. However, this is not a good solution for this project since parts of `rma` are batch-oriented:

- 1 Background correction happens within sample (i.e. batch independent)
- 2 Normalization is batch dependent as it takes the “average” distribution over samples and normalizes all samples to this average. For example, the average quantiles are determined across samples and this is the reference distribution that all samples are coerced to.
- 3 Expression value calculation by default uses the `median polish` to fit a model with effects for probes and samples and thus is batch dependent.

(Given the number publications using Affy data to classify samples, it's surprising that this issue is not discussed more)

Data Processing

Another algorithm, `mas5`, is not batch oriented, but performance using this technique was abysmal (shown later).

Instead, an `rma`-like technique was evaluated:

- 1 Same background correction
- 2 Same normalization procedure, but all samples are normalized to the reference distribution of the training set
- 3 Expression is calculated using a 10% trimmed mean instead of a median polish.

Performance was evaluated for this method, `rma` and `mas5` (results shown later).

Classification Model

Random forests was used to predict MOA, generate class probabilities and calculate variable importance.

The tuning parameter, the random subset size, was determined by finding the optimal bootstrap accuracy across a grid of 5 candidate values.

For calculating variable importance: “For each tree, the prediction accuracy on the out-of-bag portion of the data is recorded. Then the same is done after permuting each predictor variable. The difference between the two accuracies are then averaged over all trees, and normalized by the standard error. ” (Andy Liaw in *Rnews*, 2002)

MOA-specific importance measures were calculated for each probe

Selection Bias

Selecting features is tricky and can quickly lead to over-fitting.

A common approach: measure “importance” for each predictor from the training data. Remove the least important features and re-fit the model. Measured performance usually improves.

This is a circular argument. Features are important for these training samples and may not generalize well.

With $p \gg n$, the problem of finding a model that classifies perfectly is not difficult. For example, the odds that a non-informative factor will randomly show a group effect goes up as $p \rightarrow \text{large}$.

Will resampling solve this problem?

Selection Bias and Resampling

Resampling *can* solve this problem, but it must be done correctly.

We usually think of cross-validation or bootstrapping to select model parameters (e.g. the number of PLS components etc)

It is important to realize that feature selection is part of the model building process and must also be cross-validated.

“External” cross-validation encompasses feature selection and model tuning.

Probe Selection Procedure

A recursive feature selection (RFE) routine was used to determine the optimal number of probes while **avoiding selection bias**:

```
for Each 10 Fold Cross-Validation Iteration do  
  Separate data based on fold labels  
  Tune/train Random Forests model on 90% of data with all probes  
  Calculate MOA-specific variable importance for each probe  
  for Probe subset size: 900, 450, 225, 108, 54, 27, 18, 9 do  
    Retain most important probes  
    Tune/train Random Forests model on 90% of data  
    Predict the 10% cross-validation samples  
  end  
end  
Calculate cross-validation performance across subset sizes to choose the  
optimal number of probes
```

See Ambroise and McLachlan (PNAS, 2002) for examples demonstrating why this is important.

Filtering Probes

Some MOA were very easy to predict and others were more difficult.

Basic sorting of probes by overall variable importance resulted in poor overall performance since difficult MOAs were not well represented.

A stratified reduction procedure was used to filter probes.

For example, for a probe subset size of 900, the top 100 probes were selected for each of the 9 MOA.

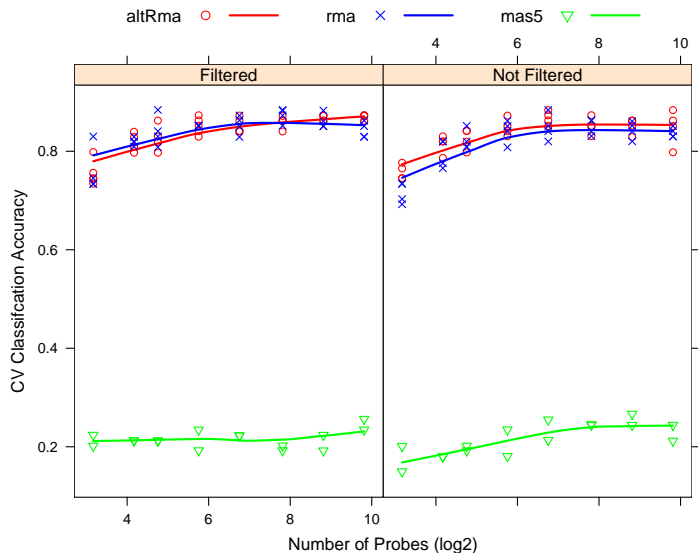
Evaluating the Algorithm

To evaluate the data processing algorithm, the RFE procedure was applied using `rma`, `mas5` and our `rma` alternative.

In Affy experiments, low gene expression signals can also inject significant noise into the results. For each data processing technique, we also dropped the probes whose average expression value fell below the 25th percentile.

For each of these 6 combinations, the cross-validation procedure was repeated 3 times.

RFE Performance



RFE Performance

The performance profiles of `rma` and our alternative are very similar. There was negligible effect of probe filtering based on expression intensity.

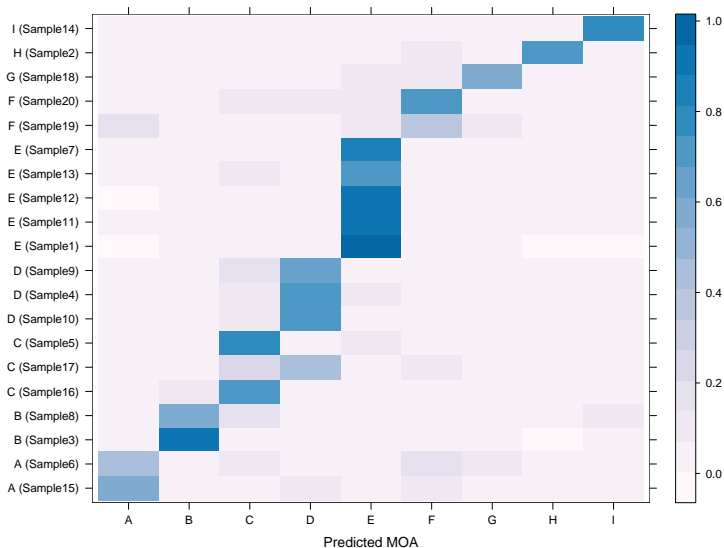
Based on the alternative `rma` procedure, the final model was built using the top 108 probes without the intensity filter. Based on the RFE results, the overall accuracy is estimated to be 85%.

A random forest model was trained using the top 108 probes and the 20 samples in the test set were run using this model. The results are:

Test Set Confusion Matrix

True MOA	Predicted MOA									Sens.	Spec.
	A	B	C	D	E	F	G	H	I		
A	2	0	0	0	0	0	0	0	0	1.00	1.00
B	0	2	0	0	0	0	0	0	0	1.00	1.00
C	0	0	2	1	0	0	0	0	0	0.67	1.00
D	0	0	0	3	0	0	0	0	0	1.00	0.94
E	0	0	0	0	5	0	0	0	0	1.00	1.00
F	0	0	0	0	0	2	0	0	0	1.00	1.00
G	0	0	0	0	0	0	1	0	0	1.00	1.00
H	0	0	0	0	0	0	0	1	0	1.00	1.00
I	0	0	0	0	0	0	0	0	1	1.00	1.00

Test Set Probabilities



Conclusions and Acknowledgements

- Affy gene expression data can be useful in predicting method of action in antibacterials.
- A modified version of the `rma` algorithm can be useful for sequentially processing CEL files.
- There is little effect of a signal intensity filter in this study

Thanks to Alison Jones, Shelley Des Etages, Alita Miller, David Potter ...
... and to Martin for the invitation.