Plan

• Basic gene expression profile analysis
• Some issues in gene expression analysis
• Batch effect & normalization
• Improving reproducibility
• Improving cross-batch classification
• More advanced analysis
  – Small-sample-size analysis
  – Novel principle for childhood ALL relapse prediction
Basic Gene Expression Profile Analysis
Affymetrix GeneChip Array
Making Affymetrix GeneChip Array

quartz is washed to ensure uniform hydroxylation across its surface and to attach linker molecules.

Exposed linkers become deprotected and are available for nucleotide coupling.

Source: Affymetrix
Gene Expression Measurement by Affymetrix GeneChip Array

Source: Affymetrix
Diagnosis Using Microarray

Image credit: Affymetrix

<table>
<thead>
<tr>
<th>Description</th>
<th>Positive</th>
<th>Negative</th>
<th>Pairs InAct Avg Diff</th>
<th>Abs Call</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFFX-MurF</td>
<td>5</td>
<td>2</td>
<td>19</td>
<td>397.5 A</td>
</tr>
<tr>
<td>AFFX-MurF</td>
<td>3</td>
<td>2</td>
<td>19</td>
<td>564.2 A</td>
</tr>
<tr>
<td>AFFX-MurF</td>
<td>4</td>
<td>2</td>
<td>19</td>
<td>308.6 A</td>
</tr>
<tr>
<td>AFFX-MurF</td>
<td>1</td>
<td>3</td>
<td>19</td>
<td>141 A</td>
</tr>
<tr>
<td>AFFX-EcoE</td>
<td>13</td>
<td>1</td>
<td>19</td>
<td>9340.8 P</td>
</tr>
<tr>
<td>AFFX-EcoE</td>
<td>15</td>
<td>0</td>
<td>19</td>
<td>12882.4 P</td>
</tr>
<tr>
<td>AFFX-EcoE</td>
<td>12</td>
<td>0</td>
<td>19</td>
<td>8716.5 P</td>
</tr>
<tr>
<td>AFFX-EcoE</td>
<td>17</td>
<td>0</td>
<td>19</td>
<td>25942.5 P</td>
</tr>
<tr>
<td>AFFX-EcoE</td>
<td>16</td>
<td>0</td>
<td>20</td>
<td>28833.5 P</td>
</tr>
<tr>
<td>AFFX-EcoE</td>
<td>17</td>
<td>0</td>
<td>19</td>
<td>25765.2 P</td>
</tr>
<tr>
<td>AFFX-EcoE</td>
<td>19</td>
<td>0</td>
<td>20</td>
<td>140113.2 P</td>
</tr>
<tr>
<td>AFFX-CraF</td>
<td>20</td>
<td>0</td>
<td>20</td>
<td>280306.6 P</td>
</tr>
<tr>
<td>AFFX-CraF</td>
<td>20</td>
<td>0</td>
<td>20</td>
<td>401741.8 P</td>
</tr>
<tr>
<td>AFFX-EcoE</td>
<td>7</td>
<td>5</td>
<td>18</td>
<td>-803 A</td>
</tr>
<tr>
<td>AFFX-EcoE</td>
<td>5</td>
<td>4</td>
<td>18</td>
<td>313.7 A</td>
</tr>
<tr>
<td>AFFX-EcoE</td>
<td>7</td>
<td>6</td>
<td>20</td>
<td>-1016.2 A</td>
</tr>
</tbody>
</table>
Application: Disease Subtype Diagnosis

samples

genes

benign

benign

malign

malign

malign

malign

malign

???
Application: Drug Action Detection

Which group of genes are the drug affecting on?
Typical Analysis Workflow

• Gene expression data collection
• DE gene selection by, e.g., t-statistic
• Classifier training based on selected DE genes
• Apply the classifier for diagnosis of future cases

Terminology: DE gene = differentially expressed gene


If you don’t remember this, you should go back to revise your CS2220 notes 😊
You can build a gene expression profile classifier in a simple or in a more complex way.

- **Parallel-multiclass classification scheme**

Next, we take an example to demonstrate the scores used by PCL. A BCR-ABL test sample contained almost all of the top-20 BCR-ABL discriminators. So, a score of 19.6 was assigned to it. Several top-20 OTHERS discriminators together with some beyond the top-20 list were also contained in this test sample. So, another score of 6.97 was assigned. This test sample did not contain any discriminators of E2A-PBX1, Hyperdip>50, or T-ALL. So, the scores are as follows:

<table>
<thead>
<tr>
<th>subtype</th>
<th>BCR-ABL</th>
<th>E2A-PBX1</th>
<th>Hyperdip&gt;50</th>
<th>T-ALL</th>
<th>MLL</th>
<th>TEL-AML1</th>
<th>OTHERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>score</td>
<td>19.63</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.71</td>
<td>2.96</td>
<td>6.97</td>
</tr>
</tbody>
</table>


- **Tree-structured classification scheme**

Childhood ALL Subtype Diagnosis Workflow

A tree-structured diagnostic workflow was recommended by our doctor collaborator.
Hierarchical Clustering

PCA Plots

Image credit: Yeoh et al, Cancer Cell, 1:133-143, 2002
Some Issues in Gene Expression Analysis
Some Headaches

- Natural fluctuations of gene expression in a person

- Noise in experimental protocols
  - Numbers mean diff things in diff batches
  - Numbers mean diff things in data obtained from diff platforms

⇒ Selected genes may not be meaningful
  - Diff genes get selected in diff expts
Natural Fluctuations & Expt Noise

Intrinsic & extrinsic noise

Measurement errors


Sometimes, a gene expression study may involve batches of data collected over a long period of time...

Batch Effects

- Samples from diff batches are grouped together, regardless of subtypes and treatment response

Image credit: Difeng Dong’s PhD dissertation, 2011
Percentage of Overlapping Genes

- Low % of overlapping genes from diff expt in general
  - Prostate cancer
    - Lapointe et al, 2004
    - Singh et al, 2002
  - Lung cancer
    - Garber et al, 2001
    - Bhattacharjee et al, 2001
  - DMD
    - Haslett et al, 2002
    - Pescatori et al, 2007

<table>
<thead>
<tr>
<th>Datasets</th>
<th>DEG</th>
<th>POG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prostate Cancer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top 10</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td>Top 50</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>Top100</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>Lung Cancer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top 10</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Top 50</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>Top100</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td>DMD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top 10</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>Top 50</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td>Top100</td>
<td>0.54</td>
<td></td>
</tr>
</tbody>
</table>

“Most random gene expression signatures are significantly associated with breast cancer outcome”

Batch Effect & Normalization
Approaches to Normalization

- **Aim of normalization:** Reduce variance w/o increasing bias
- **Scaling method**
  - Intensities are scaled so that each array has same ave value
  - E.g., Affymetrix’s
- **Transform data so that distribution of probe intensities is same on all arrays**
  - E.g., \( \frac{x - \mu}{\sigma} \)
- **Quantile normalization**
Quantile Normalization

- Given $n$ arrays of length $p$, form $X$ of size $p \times n$ where each array is a column.
- Sort each column of $X$ to give $X_{\text{sort}}$.
- Take means across rows of $X_{\text{sort}}$ and assign this mean to each elem in the row to get $X'_{\text{sort}}$.
- Get $X_{\text{normalized}}$ by arranging each column of $X'_{\text{sort}}$ to have same ordering as $X$.

- Implemented in some microarray s/w, e.g., EXPANDER.
In such a case, batch effect may be severe... to the extent that you can predict the batch that each sample comes!

Need normalization to correct for batch effect

After quantile normalization

Image credit: Difeng Dong’s PhD dissertation, 2011
Caution: “Over normalize” signals in cancer samples

A gene normalized by quantile normalization (RMA) was detected as down-regulated DE gene, but the original probe intensities in cancer samples were higher than those in normal samples.

A gene was detected as an up-regulated DE gene in the non-normalized data, but was not identified as a DE gene in the quantile normalized data.

Wang et al. Molecular Biosystems, 8:818-827, 2012
Embracing Noise to Improve Cross-Batch Prediction Accuracy
A comparison of batch effect removal methods for enhancement of prediction performance using MAQC-II microarray gene expression data

- Study how various batch effect removal algorithm influence cross-batch prediction performance
Results

Increased: Difference in MCC with and without batch removal > 0.05
Decreased: Difference in MCC with and without batch removal < -0.05
Unchanged: Difference in MCC with and without batch removal ≤ 0.05 & ≥ -0.05

Figure 10 Percentages of increased, decreased and unchanged cases in prediction performance after applying different batch effect removal methods. The total number of cases explored is 120.
Findings

• Around 10-20% of the times, doing batch effect removal actually reduces prediction power
• Batch removal is not practical in real situations

constructed predictive models to future data sets. It is desirable to have a large sample size or good quality data in each batch, so that the characteristics of each batch can be summarized more accurately and batch effects can be removed more effectively. If the sample sizes of the training and the test set are too small, it is difficult to draw a conclusive inference due to the large uncertainty. In the context of implementing an array-based diagnostic test in a clinical setting, it should be appreciated that batches may, in practice, be composed of a single clinical sample. In this regard, the use of reference samples for the purpose of calibrating batch effects may be of paramount importance.
Batch Effect Approaches

• Typical
  – Attempt to accurately estimate the batch effects
  – Then remove them
  – Therefore large sample sizes are often required for each batch and a balanced class ratio is often desired

• A new approach
  – “Embracing noise”
The “Embracing Noise” Approach

• **Ranking values (c.f. quantile normalization)**
  – Instead of absolute values
  – Inspired by MAQC project
    • “Absolute values may be different (among batches) but relative values are conserved between different platforms”

• **Stochastic sampling with replacement**
  – Bootstrapping suppresses noise
    • Training clones produced are likely to be enriched with more “clean” samples
Ranking Values

• Findings from MAQC project
  – Median coefficient of variations
    • Within Lab: 5-15% for different platforms
    • Inter Lab: 10-20% for different platforms
  – High correlation between the ranks of log ratios between different platforms
    • Absolute values may be different but relative values are conserved between different platforms
  – Conclusion
    • Microarray are still reproducible (ranking values) despite being noisy
Ranking values are stable even when sample size is small
Bootstrapping suppresses noise

Figure 3 Theoretical values of $P_b(x) - P_b(>x)$: Theoretical values of $P_b(x) - P_b(>x)$ for different sample size (i.e., $m$) at varying percentage of "good" samples (i.e., $p$).
Dynamic Bagging

• Integrates bootstrapping with sequential hypothesis testing

• Removes the need to a priori and arbitrarily fixing the number of bootstrap replicates (N)

• N is minimum for each test instance with statistical guarantees on the error rates
  – An error is defined as the difference in decision with the minimum N and infinite N

## Datasets

<table>
<thead>
<tr>
<th>Data set code</th>
<th>Data set description</th>
<th>Training set</th>
<th>Validation set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Number of samples</td>
<td>Positives</td>
</tr>
<tr>
<td>A</td>
<td>Lung tumorigen vs. non-tumorigen (Mouse)</td>
<td>70</td>
<td>26</td>
</tr>
<tr>
<td>D</td>
<td>Breast cancer pre-operative treatment response (pathologic complete response)</td>
<td>130</td>
<td>33</td>
</tr>
<tr>
<td>F</td>
<td>Multiple myeloma overall survival milestone outcome</td>
<td>340</td>
<td>51</td>
</tr>
<tr>
<td>I</td>
<td>Same as data set F but class labels are randomly assigned</td>
<td>340</td>
<td>200</td>
</tr>
</tbody>
</table>

- Two additional data sets of size 25% or 50% of the above (with same class ratio)
- Total of 12 training sets and 12 validation sets
PCA of Datasets
Experiments

• Feature Selection
  – t-Test (Parametric)
  – Wilcoxon Rank Sum Test (Non-Parametric)

• Classification Algorithms
  – C4.5 (Tree)
  – Support Vector Machine (Linear)
  – Nearest Neighbor (Instance-based)

• Performance Metric
  – Area Under Curve
Overall AUC changes in various settings (108)

- A. Rank Values
- B. Bagging (10)
- C. Bagging (100)
- D. Dynamic Bagging

Bar chart indicating:
- Decreased
- Decreased Slightly
- Increased Slightly
- Increased

Percentage of cases (%) distribution for each algorithm.
Influence of algorithms over various subset sizes

AUC Change

Dynamic Bagging.0.25  Dynamic Bagging.0.5  Dynamic Bagging.1.0
Conclusion

• An unconventional yet simple approach
  – Ranking values
  – Dynamic bagging

• Great performance
  – Shows improvements in most cases

• Practically applicable
  – Works on small training data sets
  – Independent of the sample size of the test data
Improving Reproducibility of Gene Expression Profile Analysis
**Individual Genes**

- **Suppose**
  - Each gene has 50% chance to be high
  - You have 3 disease and 3 normal samples

- **How many genes on a microarray are expected to perfectly correlate to these samples?**

- **Prob(a gene is correlated) = 1/2^6**
- **# of genes on array = 25,000**
  \[ \Rightarrow E(# \text{ of correlated genes}) = 390 \]

  \[ \Rightarrow \text{Many false positives} \]

- **These cannot be eliminated based on pure statistics!**
Group of Genes

• Suppose
  – Each gene has 50% chance to be high
  – You have 3 disease and 3 normal samples
• What is the chance of a group of 5 genes being perfectly correlated to these samples?

• Prob(group of genes correlated) = \((1/2^6)^5\)
  – Good, \(< < 1/2^6\)
• # of groups = \(\binom{25000}{5}\)
  \(\Rightarrow E(# \text{ of groups of genes correlated}) = \binom{25000}{5} \times (1/2^6)^5 = 7.58 \times 10^{10}\)

\(\Rightarrow\) Even more false positives?
• Perhaps no need to consider every group
Gene Regulatory Circuits

- Each disease phenotype has some underlying cause
- There is some unifying biological theme for genes that are truly associated with a disease subtype
- Uncertainty in selected genes can be reduced by considering biological processes of the genes
- The unifying biological theme is basis for inferring the underlying cause of disease subtype
Types of Biological Networks

• Natural biological pathways
  – Metabolic pathway
  – Gene regulation network
  – Cell signaling network

• Protein-protein interaction networks
Taming false positives by considering pathways instead of all possible groups

Group of Genes

- Suppose
  - Each gene has 50% chance to be high
  - You have 3 disease and 3 normal samples
- What is the chance of a group of 5 genes being perfectly correlated to these samples?

- # of pathways = 1000
- \( E(\text{# of pathways correlated}) = 1000 \times \left(\frac{1}{2^6}\right)^5 \)
  - Good, \( \ll \frac{1}{2^6} \)
- \( \# \text{ of groups} = \binom{25000}{5} \)
- \( E(\text{# of groups of genes correlated}) = \binom{25000}{5} \times \left(\frac{1}{2^6}\right)^5 = 7.50 \times 10^{10} \)

\( \Rightarrow \) Even more false positives?
- Perhaps no need to consider every group
Towards More Meaningful Genes

- **ORA**
  - Khatri et al
  - *Genomics*, 2002

- **FCS**
  - Pavlidis & Noble
  - PSB 2002

- **GSEA**
  - Subramanian et al
  - *PNAS*, 2005

- **PFSNet**
  - Lim & Wong
  - *Bioinformatics*, 2014

**Overlap Analysis**

**Direct-Group Analysis**

**Network-Based Analysis**
Overlap Analysis: ORA

ORA tests whether a pathway is significant by intersecting the genes in the pathway with a pre-determined list of DE genes (e.g., genes whose t-statistic meets the 5% significance threshold of t-test), and checking the significance of the size of the intersection using the hypergeometric test.

Disappointing Performance

upregulated in DMD

- DMD gene expression data
  - Pescatori et al., 2007
  - Haslett et al., 2002

- Pathway data
  - PathwayAPI, Soh et al., 2010
Issue #1 with ORA

- Its null hypothesis basically says “Genes in the given pathway behaves *no differently* from randomly chosen gene sets of the same size”

- This null hypothesis is obviously false

  \[ \implies \text{Lots of false positives} \]

- A biological pathway is a series of actions among molecules in a cell that leads to a certain product or a change in a cell. Thus necessarily the behaviour of genes in a pathway is more coordinated than random ones.
Issue #2 with ORA

- It relies on a pre-determined list of DE genes
- This list is sensitive to the test statistic used and to the significance threshold used
- This list is unstable regardless of the threshold used when sample size is small
Issue #3 with ORA

- It tests whether the entire pathway is significantly differentially expressed.

- If only a branch of the pathway is relevant to the phenotypes, the noise from the large irrelevant part of the pathways can dilute the signal from that branch.
Direct-Group Analysis: FCS

\[ \frac{1}{n} \sum_{k=1}^{n} - \log(P_k) \]

Permutation Test

GO Class 1 → Score 1 → Significant Class 1

GO Class 2 → Score 2 → Non Significant Class 2

GO Class N → Score 3 → Significant Class N

- Score of a pathway C = average of log-transformed p-values of genes in C.
- Null distribution to estimate the p-value of the scores above is by repeated sampling of random sets of genes of the same size as C.

Where will FCS be in comparison to ORA below? Why?

DMD gene expression data
- Pescatori et al., 2007
- Haslett et al., 2002

Pathway data
- PathwayAPI, Soh et al., 2010
Direct-Group Analysis: GSEA

Where will GSEA be in comparison to ORA and FCS? Why?

DMD gene expression data
- Pescatori et al., 2007
- Haslett et al., 2002

Pathway data
- PathwayAPI, Soh et al., 2010
PFSNet: Exploiting subnetworks

- Induce subnetworks from pathways by considering only genes highly expressed in majority of patients in any class

\[
\beta_1^*(g_i) = \sum_{p_j \in D} \frac{f_s(e_{g_i,p_j})}{|D|} \quad \quad \beta_2^*(g_i) = \sum_{p_j \in \neg D} \frac{f_s(e_{g_i,p_j})}{|\neg D|}
\]

- For an irrelevant subnetwork S, the two scores above for each patient \(P_k\) should be roughly equal, regardless of his class

- Do a paired t-test to decide whether S is relevant
  - Get null distribution by permuting class labels
Where will PFSNet be in comparison to ORA, FCS, and GSEA? Why?

DMD gene expression data
- Pescatori et al., 2007
- Haslett et al., 2002

Pathway data
- PathwayAPI, Soh et al., 2010
What have we learned?

• Common headaches in gene expression analysis
  – Natural fluctuation, protocol noise, batch effect

• Use of biological background info to tame false positives

• Overlap analysis → direct-group analysis → network-based analysis

• Subnetwork-based methods yield more consistent and larger disease subnetworks
From pathways to models, From static to dynamic:
A couple of quite recent papers that are worth your leisure reading…

• Geistlinger et al. **From sets to graphs: Towards a realistic enrichment analysis of transcriptomic systems.** *Bioinformatics*, 27(13):i366—i373, 2011

• Zampieri et al. **A system-level approach for deciphering the transcriptional response to prion infection.** *Bioinformatics*, 27(24):3407--3414, 2011

Fig. 1. System response inference: a toy genetic network consisting of six genes exemplifies the advantages of using a system-level data comparison (a). Standard statistical tests (i.e. *t*-test) unveil significant fold change in gene expression variations for each transcript individually (b), neglecting the underlying regulatory network. Such statistical test can identify whether the expression level of a transcript is significantly changed with respect to a reference. Putative gene expression changes are reported in panel (c). In this specific example, two genes are identified to be overexpressed (red/+ nodes) and one downregulated (green/- node), while the remaining three do not show any changes (grey nodes). By knowing the corresponding genetic regulatory network (d), we can discriminate the coherent variations from the unexpected ones. As shown in the example, two of the genes that showed a significant expression variations are consistent with model predictions i.e. the expression changes of genes x and y can be explained by the variation of genes z. This is reflected by a skew distribution of discrepancies (i.e. residues), between model predictions and observed data, centered around 0 (f). At the same time, one transcript, w, is not responding coherently to the initial model. The fact that its expression is unchanged, when it should have been increased, might relate to an anomalous direct effect of the pathology, preventing a synergistic response between all the genes in the system. Hence, the list of ‘perturbed genes’ can be sensibly different from the standard DEGs identified from individual fold change analysis (b/e).
Must Read


Good to Read


Improving Cross-Batch Classification
Batch effects

Batch effects are common
Batch effects cannot always be removed using common normalization methods
Gene-feature-based classifiers do badly when there are batch effects, even after normalization.

**Figure 5.8**: Predictive accuracy of gene-feature-based classifiers with and without rank normalization in the DMD/NOR dataset.

Gene selection by t-test, SAM, or rank product. Classifier by naïve Bayes.
Ensemble classifiers can’t always improve results of gene-feature-based classifiers with normalization.
Genes from subnetworks produced by PFSNet/ESSNet can’t help gene-feature-based classifiers

**Figure 5.15:** Predictive accuracy of gene-feature-based classifier using genes extracted from subnetworks in ESSNet; demonstrating that genes in the subnetworks by themselves are not a good discriminator for classification.
So new ideas to better use subnetwork-based features for successful cross-batch classification is needed…
PFSNet-based features

- **PFSNet**
  - Induce subnetworks from pathways by considering only genes highly expressed in majority of patients in any class
  - For each subnetwork $S$ and each patient $P_k$, compute a pair of scores:
    \[
    \beta_1^*(g_i) = \sum_{p_j \in D} \frac{f_s(e_{g_i,p_j})}{|D|} \\
    \beta_2^*(g_i) = \sum_{p_j \in \neg D} \frac{f_s(e_{g_i,p_j})}{|\neg D|}
    \]
    
    \[
    Score_1^p(S) = \sum_{g_i \in S} f_s(e_{g_i,p_k}) \cdot \beta_1^*(g_i) \\
    Score_2^p(S) = \sum_{g_i \in S} f_s(e_{g_i,p_k}) \cdot \beta_2^*(g_i)
    \]

- **Straightforward to use these scores (and their paired difference) as features**
Successfully reducing batch effects

**Figure 5.6:** A figure showing that the batch effects are reduced by PFSNet subnetwork features. The colors red and blue represent different batches.
Successful cross-batch classification

ALL/AML dataset

Figure 5.7: A figure showing that data points are separated by class labels instead of batch when PFSNet features are used. The colors green and orange represent different classes.
Gene expression profile analysis when sample size is extremely small

This part of the lecture is show you an example of more advanced forms of gene expression analysis
Recall the 3 issues of ORA

Issue #2 with ORA

- It relies on a pre-determined list of DE genes
- This list is sensitive to the test statistic used and to the significance threshold used
- This list is unstable regardless of the threshold used when sample size is small

Issue #3 with ORA

- It tests whether the entire pathway is significantly differentially expressed
- If only a branch of the pathway is relevant to the phenotypes, the noise from the large irrelevant part of the pathways can dilute the signal from that branch
Outline

• A new idea for a valid null hypothesis: ORA-Paired

• Paired test on subnetworks: NEA-Paired

• Achieving a quantum leap: ESSNet

Main reference for this work

• Kevin Lim, "Using biological networks and gene-expression profiles for the analysis of diseases", PhD thesis, November 2014, National University of Singapore
ORA-Paired: Paired test and new null hypothesis

- Let \( g_i \) be genes in a given pathway \( P \)
- Let \( p_j \) be a patient
- Let \( q_k \) be a normal

- Let \( \Delta_{i,j,k} = \text{Expr}(g_i,p_j) - \text{Expr}(g_i,q_k) \)

- Test whether \( \Delta_{i,j,k} \) is a distribution with mean 0

- **Issue #1 is solved**
  - Null hypothesis is “Pathway P is irrelevant to the difference between patients and normals, and the genes in P behave similarly in patients and normals”

- **Issue #2 is solved**
  - No longer need a pre-determined list of DE genes

- **Issue #3 is unsolved**

- **Is sample size now larger?**
  - \(|\text{patients}| \cdot |\text{normals}| \cdot |\text{genes in P}|\)
Testing the null hypothesis

“Pathway P is irrelevant to the difference between patients and normals and so, the genes in P behave similarly in patients and normals”

- **Method #1**
  - T-test w/ a conservative degree of freedom
    - E.g., # normals + # patients

- **Method #2**
  - By the null hypothesis, a dataset and any of its class-label permutations are exchangeable
  => Get null distribution by class-label permutations
    - Only for large-size sample

- **Method #3**
  - Modified null hypothesis
    - “Pathway P induces gene-gene correlations, and genes in P behave according to these gene-gene correlations;
    - P is irrelevant to the diff betw patients and normals and so, genes in P behave similarly in patients and normals”
  => Get null distribution using datasets that conserve gene-gene correlations in the original dataset
    - E.g., array rotation
Array rotation

• QR decomposition

\[ X = X_Q \cdot X_R \]

Where
- \( X \) is gene expression array of \( n \) samples * \( m \) genes
- \( X_Q \) is \( n \) * \( r \) orientation matrix, \( r \) is rank of \( X \)
- \( X_R \) is sufficient statistics of covariance between the \( m \) genes

• Rotation

\[ X' = R_Q \cdot X_Q \cdot X_R \]

Where
- \( R_Q \) is an \( n \) * \( n \) rotation operation

\( \Rightarrow \) \( X' \) is rotation of \( X \)
- preserving gene-gene correlations
- i.e., preserving constraints induced by pathways
Similar to PFSNet, good but not great
NEA-Paired: Paired test on subnetworks

- Given a pathway $P$
- Let each node and its immediate neighbourhood in $P$ be a subnetwork
- Apply ORA-Paired on each subnetwork individually

- Issues #1 & #2 are solved as per ORA-Paired

- Issue #3 is partly solved
  - Testing subnetworks instead of whole pathways
  - But subnetworks derived in a fragmented way
Much better performance

upregulated in DMD

subnetwork agreement

sample size (N)

NEA-Paired
ORA-Paired
PFSNet
GSEA
ORA
ESSNet: Larger subnetworks

- Compute the average rank of a gene based on its expression level in patients in any class.
- Use the top $\alpha\%$ to extract large connected components in pathways.
- Test each component using ORA-Paired.

- Gene rank is very stable.
- Issues #1 - #3 solved.
Fantastic performance

Graph showing the subnetwork agreement for different methods as a function of sample size (N), with the y-axis representing subnetwork agreement and the x-axis representing sample size (N). The methods include ESSNet, NEA-Paired, ORA-Paired, PFSNet, GSEA, and ORA.
More datasets tested
ESSNet is unlikely to report junk

Table 4.2: Average number of subnetworks predicted by ESSNet over the sample sizes ($N$); the first number denotes the number of subnetworks in the numerator of the subnetwork-level agreement and the second number denotes the number of subnetworks in the denominator of the subnetwork-level agreement; cf. equation 4.5.

<table>
<thead>
<tr>
<th>sample size (N)</th>
<th>DMD</th>
<th>ALL</th>
<th>BCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>8.2/13.4</td>
<td>7.0/11.9</td>
<td>4.8/12.6</td>
</tr>
<tr>
<td>3</td>
<td>11.1/15.9</td>
<td>11.3/17.9</td>
<td>5.0/11.7</td>
</tr>
<tr>
<td>4</td>
<td>13.18/16.5</td>
<td>11.9/15.9</td>
<td>6.2/10.4</td>
</tr>
<tr>
<td>5</td>
<td>14.2/16.7</td>
<td>14.6/18.3</td>
<td>7.9/12.7</td>
</tr>
<tr>
<td>6</td>
<td>15.14/17.6</td>
<td>14.9/18.0</td>
<td>11.0/15.7</td>
</tr>
<tr>
<td>7</td>
<td>15.2/17.4</td>
<td>16.1/19.2</td>
<td>12.9/17.5</td>
</tr>
<tr>
<td>8</td>
<td>15.4/17.5</td>
<td>16.2/19.0</td>
<td>15.3/20.4</td>
</tr>
<tr>
<td>9</td>
<td>16.6/18.8</td>
<td>17.0/19.8</td>
<td>15.8/20.8</td>
</tr>
<tr>
<td>10</td>
<td>17.6/19.7</td>
<td>17.3/19.7</td>
<td>16.2/20.8</td>
</tr>
</tbody>
</table>
A negative-control experiment showing that ESSNet does not report junk
ESSNet also dominates when sample size is large

**Table 4.3:** Number of subnetworks predicted by the various methods on a full dataset where the null distribution is computed using array rotation (rot), class-label swapping (cperm) and gene swapping (gswap); the first number denotes the number of subnetworks in the numerator of the subnetwork-level agreement and the second number denotes the number of subnetworks in the denominator of the subnetwork-level agreement; cf. equation 4.5.

<table>
<thead>
<tr>
<th></th>
<th>DMD</th>
<th></th>
<th>ALL</th>
<th></th>
<th></th>
<th>BCR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rot  cperm</td>
<td>rot  cperm</td>
<td>rot  cperm</td>
<td>rot  cperm</td>
<td>rot  cperm</td>
<td></td>
</tr>
<tr>
<td>ESSNet</td>
<td>20/23 13/15</td>
<td>22/24 25/27</td>
<td>24/29 30/32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEA-paired</td>
<td>77/98 91/115</td>
<td>140/163 109/119</td>
<td>176/192 37/43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORA-paired</td>
<td>30/62 30/62</td>
<td>34/74 34/74</td>
<td>53/99 53/99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORA-hypergeo</td>
<td>20/46 41/141</td>
<td>24/60 48/73</td>
<td>4/14 32/166</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSEA</td>
<td>cperm</td>
<td>cperm  gswap</td>
<td>cperm</td>
<td>gswap</td>
<td>cperm</td>
<td>gswap</td>
</tr>
</tbody>
</table>
Do ESSNet results agree on small datasets vs big datasets?

- Use ESSNet’s results on entire datasets as the benchmark to evaluate ESSNet’s results on small subsets of the datasets
- The precision (i.e., agreement) is superb, though some subnetworks are missed when smaller datasets are analysed

<table>
<thead>
<tr>
<th>Sample size (N)</th>
<th>Precision</th>
<th></th>
<th></th>
<th>Recall</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DMD</td>
<td>ALL</td>
<td>BCR</td>
<td>DMD</td>
<td>ALL</td>
<td>BCR</td>
</tr>
<tr>
<td>2</td>
<td>D</td>
<td>0.96</td>
<td>0.87</td>
<td>0.93</td>
<td>0.91</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>0.93</td>
<td>0.86</td>
<td>0.99</td>
<td>0.89</td>
<td>0.90</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
<td>0.88</td>
<td>0.88</td>
<td>0.97</td>
<td>0.92</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>0.89</td>
<td>0.88</td>
<td>0.94</td>
<td>0.90</td>
<td>0.89</td>
</tr>
<tr>
<td>6</td>
<td>D</td>
<td>0.82</td>
<td>0.88</td>
<td>0.93</td>
<td>0.92</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>0.85</td>
<td>0.86</td>
<td>0.95</td>
<td>0.93</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>0.84</td>
<td>0.89</td>
<td>0.97</td>
<td>0.94</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>0.88</td>
<td>0.90</td>
<td>0.94</td>
<td>0.92</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>0.88</td>
<td>0.93</td>
<td>0.97</td>
<td>0.92</td>
<td>0.90</td>
</tr>
</tbody>
</table>
How about cross-batch classification when sample size is small?
ESSNet

- Induce subnetworks using genes highly expressed in majority of samples in any class

  - Let $g_i$ be genes in a given subnetwork $S$
  - Let $p_j$ be patients
  - Let $q_k$ be normals

  - Let $\Delta_{i,j,k} = \text{Expr}(g_i, p_j) - \text{Expr}(g_i, q_k)$

  - Test whether $\Delta_{i,j,k}$ is a distribution with mean 0

ESSNet scores subnetworks but not patients.

How to produce feature vectors for patients?
ESSNet-based features

• The idea is to see whether the pairwise differences of genes with a subnetwork between a given sample $p_x$ and the two separate classes (D and $\neg$D) have a distribution around 0

$$\Delta(D)(S, p_x) = \{e_{g_i, p_x} - e_{g_i, p'} \mid g_i \in S \text{ and } p' \in D\}$$

$$\Delta(\neg D)(S, p_x) = \{e_{g_i, p_x} - e_{g_i, p'} \mid g_i \in S \text{ and } p' \in \neg D\}$$

• We expect $\Delta(D)(S, P_x)$ and $\Delta(\neg D)(S, P_x)$ to have +ve or –ve median for patients in one of the classes iff subnetwork $S$ is useful for classification
  – The median and ±2 std dev of $\Delta(D)(S, P_x)$ and $\Delta(\neg D)(S, P_x)$ give 6 features for $P_x$
ESSNet-based features

- We also obtain pairwise differences of genes within a subnetwork among all possible pairs of patients in $D$ and $\neg D$

$$\Delta_{(D - \neg D)}(S) = \left\{ e_{g_i,p'} - e_{g_i,p''} \mid g_i \in S \text{ and } p' \in D \text{ and } p'' \in \neg D \right\}$$

Similarly for $\Delta_{(\neg D - \neg D)}(S)$, $\Delta_{(\neg D - D)}(S)$, $\Delta_{(D - D)}(S)$

- This gives 4 more features

$$ESSNet\text{\_feature}_{7}^{p_x,S} = T\_statistic(\Delta_{(-D)}(S,p_x), \Delta_{(D-\neg D)}(S))$$

$$ESSNet\text{\_feature}_{8}^{p_x,S} = T\_statistic(\Delta_{(-D)}(S,p_x), \Delta_{(-D-\neg D)}(S))$$

$$ESSNet\text{\_feature}_{9}^{p_x,S} = T\_statistic(\Delta_{(D)}(S,p_x), \Delta_{(D-D)}(S))$$

$$ESSNet\text{\_feature}_{10}^{p_x,S} = T\_statistic(\Delta_{(D)}(S,p_x), \Delta_{(-D-D)}(S))$$
ESSNet-based features lead to high cross-batch classification accuracy.
ESSNet-based cross-batch hierarchical clusterings

**Figure 5.17:** A figure depicting hierarchical clustering performed on the patient’s subnetwork scores.
ESSNet-based features retain high cross-batch classification accuracy even when training-sample size is small
A Novel Principle for Childhood ALL Relapse Prediction

This part of the lecture is show you another example of more advanced forms of gene expression analysis
Childhood Acute Lymphoblastic Leukemia

• The most common cancer in children
  – 3,000 new cases in US
  – 2,000 new cases in ASEAN countries

• 80% achieve long-term relapse-free survival, but
  – 20% relapse and eventually die
  – Large fraction of them suffer severe side effects

⇒ Predict relapse early and treat more aggressively
Outline

- Background
- Hypotheses
- Framework
- Methodologies
  - Data Preparation
  - Model Construction
  - Relapse Prediction
- Validation
- Conclusion

Main reference for this work

- Difeng Dong, "Relapse Prediction in Childhood Acute Lymphoblastic Leukemia by Time-Series Gene Expression Profiling", PhD thesis, November 2011, National University of Singapore
Contemporary ALL Treatment Framework

- **Treatment A**: Intensive Treatment
- **Treatment B**: Moderate Treatment

Relapse Prediction

- **High Risk?**
  - Yes: Treatment A → Treatment B → Relapse & Death
  - No: Treatment B → Treatment A → Side Effects
Previous Work

• Correlate GEP to childhood ALL subtypes
  – Identified subtype-based genetic signatures
  – Diagnostic accuracy is >95%, better than routine diagnostic methods

What does intensity of genetic signature means?

Yeoh et al. *Cancer Cell*, 2002
Hypotheses

- **Treatment gradually removes leukemic cells in patient**
- **Diagnostic GEP captures leukemic subtype signature**

**Hypothesis 1:** Time-series GEP captures reduction of leukemic cells during treatment

**Hypothesis 2:** Poor genetic response suggests high risk of relapse
Framework

• **Time-series GEP data preparation (normalization)**

• **H1: Time-series GEP captures reduction of leukemic cells during treatment**
  – Unsupervised hierarchical clustering
  – Signature dissolution analysis
  – Genetic status shifting (GSS) model

• **H2: Poor genetic response → high risk of relapse**
  – Prediction based on GSS distance

• **Validation in independent datasets**
GEP Data Preparation

- 96 patients, 10 relapses vs 86 remissions
- GEP collected on 4 time points, D0, D8, D15, D33, a matrix of >30,000 genes * >300 samples
- Data generated by MAS5.0
GEP Data Normalization

- Scaling factor >20 → Remove
- Noise mainly in low-expression genes
- Genes with > 70% absent calls → Remove
- 4,736 genes remain
- Perform quantile normalization

![Signal (log2 based)](image)

![3D Scatter Plot](image)
Framework

- **Time-series GEP data preparation (normalization)**

- **H1**: Time-series GEP captures reduction of leukemic cells during treatment
  - Unsupervised hierarchical clustering
  - Signature dissolution analysis
  - Genetic status shifting (GSS) model

- **H2**: Poor genetic response $\Rightarrow$ high risk of relapse
  - Prediction based on GSS distance

- **Validation in independent datasets**
Unsupervised Clustering

- Top 10% of genes with largest variance across whole dataset
  - 1,474 genes
  - Noise mainly in low-expression genes

- Unsupervised hierarchical clustering on patients
  - Pearson's correlation
  - Completed linkage
Unsupervised Clustering: Results
Signature Dissolution Analysis

• How do intensity of genetic signatures change during treatment?

Yeoh et al. *Cancer Cell*, 2002
Signature Dissolution Analysis

• Consider the 3 largest subtypes
  – TEL-AML1, n = 26
  – T-ALL, n = 12
  – Hyperdiploid>50, n = 12

• Select genetic signature genes for them
  – Organize diagnostic samples into Subtype A vs ~A
  – Only consider genes highly expressed in A
  – Pick 20 most diff expressed genes by t-test
    • The selected signature genes was validated by testing their prediction accuracy on public data
    • Accuracy achieved: 95%
Signature Dissolution Analysis

- Red is highly expressed
- Green is lowly expressed
Global Genetic Status Shifting (GSS) Model

- **Select drug responsive genes**
  - Diff expressed genes betw D0 & D8 by t-test \((q<0.0001)\)
  - >2 fold change betw D0 and D8
  \[\Rightarrow 461 \text{ up-} \text{ and } 99 \text{ down-regulated genes}\]

- **Apply principal component analysis**
  - Genes are considered as features
  - PCA
  - Each point is a sample
Global Genetic Status Shifting Model

(a)

<table>
<thead>
<tr>
<th>PC</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance</td>
<td>49.08%</td>
<td>7.56%</td>
<td>5.19%</td>
<td>3.64%</td>
<td>2.10%</td>
<td>1.81%</td>
<td>1.61%</td>
<td>1.35%</td>
<td>72.34%</td>
</tr>
</tbody>
</table>
Verifying the Global GSS Model
GO and Pathway Ingenuity Analysis on Drug-Responsive Genes in the Global GSS Model

- **UP**: Reconstruction of immune system and restoration of normal hematogenesis
- **DOWN**: Cell development and DNA synthesis
- **DOWN**: Negative regulation of apoptosis
Framework

• **Time-series GEP data preparation (normalization)**

• **H1: Time-series GEP captures reduction of leukemic cells during treatment**
  – Unsupervised hierarchical clustering
  – Signature dissolution analysis
  – Genetic status shifting (GSS) model

• **H2: Poor genetic response ➔ high risk of relapse**
  – Prediction based on GSS distance

• **Validation in independent datasets**
Genetic Status Shifting Distance

DC: Disease Centroid
NC: NBM Centroid

ASD = s3  (Absolute shifting distance)
ESD = s2  (Effective shifting distance)
ESR = s2/s1  (Effective shifting ratio)
## Relapse Prediction by ESD

<table>
<thead>
<tr>
<th>RANK</th>
<th>SAMPLE</th>
<th>ESD-D8</th>
<th>RANK</th>
<th>SAMPLE</th>
<th>ESD-D8</th>
<th>RANK</th>
<th>SAMPLE</th>
<th>ESD-D8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>56_KL464</td>
<td>-2.21</td>
<td>30</td>
<td>29_KL439</td>
<td>26.20</td>
<td>59</td>
<td>05_KL354</td>
<td>52.56</td>
</tr>
<tr>
<td>2</td>
<td>77_KL401</td>
<td>-1.51</td>
<td>31</td>
<td>49_R432</td>
<td>26.77</td>
<td>60</td>
<td>80_KL423</td>
<td>53.33</td>
</tr>
<tr>
<td>3</td>
<td>33_R247</td>
<td>-0.46</td>
<td>32</td>
<td>83_KL457</td>
<td>29.13</td>
<td>61</td>
<td>75_KL385</td>
<td>54.31</td>
</tr>
<tr>
<td>4</td>
<td>19_KL205</td>
<td>-0.21</td>
<td>33</td>
<td>04_KL322</td>
<td>30.64</td>
<td>62</td>
<td>69_KL313</td>
<td>54.40</td>
</tr>
<tr>
<td>5</td>
<td>45_R194</td>
<td>2.13</td>
<td>34</td>
<td>74_KL383</td>
<td>31.52</td>
<td>63</td>
<td>64_KKH29</td>
<td>55.12</td>
</tr>
<tr>
<td>6</td>
<td>19_R281</td>
<td>2.50</td>
<td>35</td>
<td>60_KKH30</td>
<td>32.88</td>
<td>64</td>
<td>93_R337</td>
<td>55.35</td>
</tr>
<tr>
<td>7</td>
<td>97_R208</td>
<td>2.94</td>
<td>36</td>
<td>38_KL218</td>
<td>33.00</td>
<td>65</td>
<td>26_KL369</td>
<td>56.83</td>
</tr>
<tr>
<td>8</td>
<td>67_KL287</td>
<td>4.63</td>
<td>37</td>
<td>10_R257</td>
<td>33.39</td>
<td>66</td>
<td>14_KKH19</td>
<td>57.79</td>
</tr>
<tr>
<td>9</td>
<td>96_R202</td>
<td>5.14</td>
<td>38</td>
<td>40_KL430</td>
<td>33.55</td>
<td>67</td>
<td>28_KL375</td>
<td>58.21</td>
</tr>
<tr>
<td>10</td>
<td>11_R280</td>
<td>5.90</td>
<td>39</td>
<td>86_KL509</td>
<td>34.48</td>
<td>68</td>
<td>62_KKH22</td>
<td>60.55</td>
</tr>
<tr>
<td>11</td>
<td>27_KL374</td>
<td>6.31</td>
<td>40</td>
<td>51_KL461</td>
<td>34.83</td>
<td>69</td>
<td>21_KL300</td>
<td>60.80</td>
</tr>
<tr>
<td>12</td>
<td>70_KL320</td>
<td>6.64</td>
<td>41</td>
<td>79_KL421</td>
<td>35.43</td>
<td>70</td>
<td>01_KKH18</td>
<td>62.07</td>
</tr>
<tr>
<td>13</td>
<td>41_KL441</td>
<td>7.86</td>
<td>42</td>
<td>94_R354</td>
<td>35.66</td>
<td>71</td>
<td>32_R233</td>
<td>67.35</td>
</tr>
<tr>
<td>14</td>
<td>20_KL274</td>
<td>8.91</td>
<td>43</td>
<td>24_KL328</td>
<td>35.78</td>
<td>72</td>
<td>34_R256</td>
<td>67.95</td>
</tr>
<tr>
<td>15</td>
<td>82_KL454</td>
<td>9.53</td>
<td>44</td>
<td>47_R334</td>
<td>37.35</td>
<td>73</td>
<td>68_KL304</td>
<td>70.06</td>
</tr>
<tr>
<td>16</td>
<td>13_R410</td>
<td>11.45</td>
<td>45</td>
<td>30_KL444</td>
<td>38.26</td>
<td>74</td>
<td>18_KKH28</td>
<td>70.41</td>
</tr>
<tr>
<td>17</td>
<td>39_KL395</td>
<td>11.49</td>
<td>46</td>
<td>57_KL535</td>
<td>39.40</td>
<td>75</td>
<td>36_R343</td>
<td>71.95</td>
</tr>
<tr>
<td>18</td>
<td>17_KKH27</td>
<td>12.64</td>
<td>47</td>
<td>73_KL381</td>
<td>39.51</td>
<td>76</td>
<td>58_KL543</td>
<td>71.96</td>
</tr>
<tr>
<td>19</td>
<td>35_R313</td>
<td>13.76</td>
<td>48</td>
<td>66_KL247</td>
<td>39.69</td>
<td>77</td>
<td>08_KL456</td>
<td>72.39</td>
</tr>
<tr>
<td>20</td>
<td>55_KL419</td>
<td>18.60</td>
<td>49</td>
<td>16_KKH21</td>
<td>40.05</td>
<td>78</td>
<td>12_R297</td>
<td>74.21</td>
</tr>
<tr>
<td>21</td>
<td>99_KL416</td>
<td>19.83</td>
<td>50</td>
<td>84_KL458</td>
<td>40.12</td>
<td>79</td>
<td>88_KL544</td>
<td>75.25</td>
</tr>
<tr>
<td>22</td>
<td>23_KL321</td>
<td>20.37</td>
<td>51</td>
<td>37_R355</td>
<td>41.68</td>
<td>80</td>
<td>72_KL377</td>
<td>76.58</td>
</tr>
<tr>
<td>23</td>
<td>50_KL360</td>
<td>20.80</td>
<td>52</td>
<td>25_KL357</td>
<td>42.14</td>
<td>81</td>
<td>44_KL541</td>
<td>79.45</td>
</tr>
<tr>
<td>24</td>
<td>90_R253</td>
<td>21.40</td>
<td>53</td>
<td>15_KKH20</td>
<td>43.35</td>
<td>82</td>
<td>87_KL522</td>
<td>79.91</td>
</tr>
<tr>
<td>25</td>
<td>61_KKH13</td>
<td>21.67</td>
<td>54</td>
<td>52_R252</td>
<td>45.31</td>
<td>83</td>
<td>89_R245</td>
<td>85.37</td>
</tr>
<tr>
<td>26</td>
<td>43_KL536</td>
<td>23.00</td>
<td>55</td>
<td>85_KL485</td>
<td>45.82</td>
<td>84</td>
<td>63_KKH25</td>
<td>85.67</td>
</tr>
<tr>
<td>27</td>
<td>65_KL224</td>
<td>23.30</td>
<td>56</td>
<td>92_R332</td>
<td>49.15</td>
<td>85</td>
<td>78_KL412</td>
<td>92.38</td>
</tr>
<tr>
<td>28</td>
<td>07_KL417</td>
<td>24.00</td>
<td>57</td>
<td>98_KL387</td>
<td>49.17</td>
<td>86</td>
<td>95_R431</td>
<td>104.57</td>
</tr>
<tr>
<td>29</td>
<td>42_KL507</td>
<td>24.00</td>
<td>58</td>
<td>48_R339</td>
<td>49.26</td>
<td>87</td>
<td>94_KH16</td>
<td>104.89</td>
</tr>
</tbody>
</table>

Relapse Prediction by ESD
## Comparison with Other Clinical & Computational Protocols

<table>
<thead>
<tr>
<th>Method</th>
<th>Prognostic Feature</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Holleman-DT</td>
<td>D0 GEP</td>
<td>60.00%</td>
<td>69.51%</td>
<td>64.76%</td>
</tr>
<tr>
<td>Holleman-NB</td>
<td>D0 GEP</td>
<td>60.00%</td>
<td>69.51%</td>
<td>64.76%</td>
</tr>
<tr>
<td>Holleman-SVM</td>
<td>D0 GEP</td>
<td>60.00%</td>
<td>69.51%</td>
<td>64.76%</td>
</tr>
<tr>
<td>Bhojwani</td>
<td>D0 GEP</td>
<td>20.00%</td>
<td>79.27%</td>
<td>49.63%</td>
</tr>
<tr>
<td>NCI</td>
<td>D0 GEP</td>
<td>80.00%</td>
<td>58.14%</td>
<td>69.07%</td>
</tr>
<tr>
<td>Cytogenetics</td>
<td>Diagnostic Cytogenetics</td>
<td>30.00%</td>
<td><strong>94.19%</strong></td>
<td>62.09%</td>
</tr>
<tr>
<td>MRD-D33</td>
<td>D33 MRD</td>
<td>77.78%</td>
<td>54.12%</td>
<td>65.95%</td>
</tr>
<tr>
<td>D8 Response</td>
<td>D8 Blast Count</td>
<td>30.00%</td>
<td>85.53%</td>
<td>57.76%</td>
</tr>
<tr>
<td>ASD-D8</td>
<td>D0 and D8 GEP</td>
<td>90.00%</td>
<td>73.68%</td>
<td>81.84%</td>
</tr>
<tr>
<td>ESD-D8</td>
<td>D0 and D8 GEP</td>
<td><strong>100.00%</strong></td>
<td>75.00%</td>
<td><strong>87.50%</strong></td>
</tr>
<tr>
<td>ESR-D8</td>
<td>D0 and D8 GEP</td>
<td>90.00%</td>
<td>73.68%</td>
<td>81.84%</td>
</tr>
</tbody>
</table>
Framework

• **Time-series GEP data preparation (normalization)**

• **H1: Time-series GEP captures reduction of leukemic cells during treatment**
  – Unsupervised hierarchical clustering
  – Signature dissolution analysis
  – Genetic status shifting (GSS) model

• **H2: Poor genetic response $\Rightarrow$ high risk of relapse**
  – Prediction based on GSS distance

• **Validation in independent datasets**
Validation on an Independent Dataset

• 8 childhood ALL patients from Europe

• GEP on D0, D8, and D15

• Standard data preprocessing

• Use the same drug-responsive genes as the global GSS model

• Apply PCA and use ESD to make prediction
Recall... so threshold is ESD = 24

<table>
<thead>
<tr>
<th>RANK</th>
<th>SAMPLE</th>
<th>ESD-D8</th>
<th>RANK</th>
<th>SAMPLE</th>
<th>ESD-D8</th>
<th>RANK</th>
<th>SAMPLE</th>
<th>ESD-D8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>56_KL464</td>
<td>-2.21</td>
<td>30</td>
<td>29_KL439</td>
<td>26.20</td>
<td>59</td>
<td>05_KL354</td>
<td>52.56</td>
</tr>
<tr>
<td>2</td>
<td>77_KL401</td>
<td>-1.51</td>
<td>31</td>
<td>49_R432</td>
<td>26.77</td>
<td>60</td>
<td>80_KL423</td>
<td>53.33</td>
</tr>
<tr>
<td>3</td>
<td>33_R247</td>
<td>-0.46</td>
<td>32</td>
<td>83_KL457</td>
<td>29.13</td>
<td>61</td>
<td>75_KL385</td>
<td>54.31</td>
</tr>
<tr>
<td>4</td>
<td>19_KL205</td>
<td>-0.21</td>
<td>33</td>
<td>04_KL322</td>
<td>30.64</td>
<td>62</td>
<td>69_KL313</td>
<td>54.40</td>
</tr>
<tr>
<td>5</td>
<td>45_R194</td>
<td>2.13</td>
<td>34</td>
<td>74_KL383</td>
<td>31.52</td>
<td>63</td>
<td>64_KKH29</td>
<td>55.12</td>
</tr>
<tr>
<td>6</td>
<td>29_R281</td>
<td>2.57</td>
<td>35</td>
<td>60_KKH30</td>
<td>32.88</td>
<td>64</td>
<td>93_R337</td>
<td>55.35</td>
</tr>
<tr>
<td>7</td>
<td>97_R208</td>
<td>2.94</td>
<td>36</td>
<td>38_KL218</td>
<td>33.00</td>
<td>65</td>
<td>26_KL369</td>
<td>56.83</td>
</tr>
<tr>
<td>8</td>
<td>67_KL287</td>
<td>4.63</td>
<td>37</td>
<td>10_R257</td>
<td>33.39</td>
<td>66</td>
<td>14_KKH19</td>
<td>57.79</td>
</tr>
<tr>
<td>9</td>
<td>96_R202</td>
<td>5.14</td>
<td>38</td>
<td>40_KL430</td>
<td>33.55</td>
<td>67</td>
<td>28_KL375</td>
<td>58.21</td>
</tr>
<tr>
<td>10</td>
<td>11_R280</td>
<td>5.90</td>
<td>39</td>
<td>86_KL509</td>
<td>34.48</td>
<td>68</td>
<td>62_KKH22</td>
<td>60.55</td>
</tr>
<tr>
<td>11</td>
<td>27_KL374</td>
<td>6.31</td>
<td>40</td>
<td>51_KL461</td>
<td>34.83</td>
<td>69</td>
<td>21_KL300</td>
<td>60.80</td>
</tr>
<tr>
<td>12</td>
<td>70_KL320</td>
<td>6.64</td>
<td>41</td>
<td>79_KL421</td>
<td>35.43</td>
<td>70</td>
<td>01_KKH18</td>
<td>62.07</td>
</tr>
<tr>
<td>13</td>
<td>41_KL441</td>
<td>7.86</td>
<td>42</td>
<td>94_R354</td>
<td>35.66</td>
<td>71</td>
<td>32_R233</td>
<td>67.35</td>
</tr>
<tr>
<td>14</td>
<td>20_KL274</td>
<td>8.91</td>
<td>43</td>
<td>24_KL328</td>
<td>35.78</td>
<td>72</td>
<td>34_R256</td>
<td>67.95</td>
</tr>
<tr>
<td>15</td>
<td>82_KL454</td>
<td>9.53</td>
<td>44</td>
<td>47_R334</td>
<td>37.35</td>
<td>73</td>
<td>68_KL304</td>
<td>70.06</td>
</tr>
<tr>
<td>16</td>
<td>13_R410</td>
<td>11.45</td>
<td>45</td>
<td>30_KL444</td>
<td>38.26</td>
<td>74</td>
<td>18_KKH28</td>
<td>70.41</td>
</tr>
<tr>
<td>17</td>
<td>39_KL395</td>
<td>11.49</td>
<td>46</td>
<td>57_KL535</td>
<td>39.40</td>
<td>75</td>
<td>36_R343</td>
<td>71.95</td>
</tr>
<tr>
<td>18</td>
<td>17_KKH27</td>
<td>12.64</td>
<td>47</td>
<td>73_KL381</td>
<td>39.51</td>
<td>76</td>
<td>58_KL543</td>
<td>71.96</td>
</tr>
<tr>
<td>19</td>
<td>35_R313</td>
<td>13.76</td>
<td>48</td>
<td>66_KL247</td>
<td>39.69</td>
<td>77</td>
<td>08_KL456</td>
<td>72.39</td>
</tr>
<tr>
<td>20</td>
<td>55_KL419</td>
<td>18.60</td>
<td>49</td>
<td>16_KKH21</td>
<td>40.05</td>
<td>78</td>
<td>12_R297</td>
<td>74.21</td>
</tr>
<tr>
<td>21</td>
<td>99_KL416</td>
<td>19.83</td>
<td>50</td>
<td>84_KL458</td>
<td>40.12</td>
<td>79</td>
<td>88_KL544</td>
<td>75.25</td>
</tr>
<tr>
<td>22</td>
<td>23_KL321</td>
<td>20.37</td>
<td>51</td>
<td>37_R355</td>
<td>41.68</td>
<td>80</td>
<td>72_KL377</td>
<td>76.58</td>
</tr>
<tr>
<td>23</td>
<td>50_KL360</td>
<td>20.80</td>
<td>52</td>
<td>25_KL357</td>
<td>42.14</td>
<td>81</td>
<td>44_KL541</td>
<td>79.45</td>
</tr>
<tr>
<td>24</td>
<td>90_R253</td>
<td>21.40</td>
<td>53</td>
<td>15_KKH20</td>
<td>43.35</td>
<td>82</td>
<td>87_KL522</td>
<td>79.91</td>
</tr>
<tr>
<td>25</td>
<td>61_KKH13</td>
<td>21.67</td>
<td>54</td>
<td>52_R252</td>
<td>45.31</td>
<td>83</td>
<td>89_R245</td>
<td>85.37</td>
</tr>
<tr>
<td>26</td>
<td>43_KL536</td>
<td>23.00</td>
<td>55</td>
<td>85_KL485</td>
<td>45.82</td>
<td>84</td>
<td>63_KKH25</td>
<td>85.67</td>
</tr>
<tr>
<td>27</td>
<td>65_KL224</td>
<td>23.30</td>
<td>56</td>
<td>92_R332</td>
<td>49.15</td>
<td>85</td>
<td>78_KL412</td>
<td>92.38</td>
</tr>
<tr>
<td>28</td>
<td>07_KL417</td>
<td>24.00</td>
<td>57</td>
<td>98_KL387</td>
<td>49.17</td>
<td>86</td>
<td>95_R431</td>
<td>104.57</td>
</tr>
<tr>
<td>29</td>
<td>42_KL507</td>
<td>24.00</td>
<td>58</td>
<td>48_R339</td>
<td>49.26</td>
<td>87</td>
<td>64_KL423</td>
<td>24.00</td>
</tr>
</tbody>
</table>
## Result on the Independent Dataset

<table>
<thead>
<tr>
<th>ID</th>
<th>Type</th>
<th>Subtype</th>
<th>ESD</th>
<th>Chance of adverse event</th>
<th>Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>861-M</td>
<td>ALL</td>
<td>BCR-ABL</td>
<td>-4.67</td>
<td>Very high risk</td>
<td>100%</td>
</tr>
<tr>
<td>1021-M</td>
<td>ALL</td>
<td>BCR-ABL</td>
<td>1.86</td>
<td>Very high risk</td>
<td>90%</td>
</tr>
<tr>
<td>961-M</td>
<td>ALL</td>
<td>HYPERDIP</td>
<td>6.26</td>
<td>High risk</td>
<td>60%</td>
</tr>
<tr>
<td>906-M</td>
<td>ALL</td>
<td>OTHERS</td>
<td>17.16</td>
<td>High risk</td>
<td>30%</td>
</tr>
<tr>
<td>980-M</td>
<td>ALL</td>
<td>OTHERS</td>
<td>30.84</td>
<td>Intermediate to low risk</td>
<td>0%</td>
</tr>
<tr>
<td>824-M</td>
<td>ALL</td>
<td>TEL-AML1</td>
<td>39.16</td>
<td>Intermediate to low risk</td>
<td>0%</td>
</tr>
<tr>
<td>1003-M</td>
<td>ALL</td>
<td>HYPERDIP</td>
<td>41.48</td>
<td>Intermediate to low risk</td>
<td>0%</td>
</tr>
<tr>
<td>940-M</td>
<td>ALL</td>
<td>TEL-AML1</td>
<td>59.07</td>
<td>Intermediate to low risk</td>
<td>0%</td>
</tr>
</tbody>
</table>
Validation on AML Datasets: How general is the principle of the GSS Model

- 8 acute myeloid leukemia (AML) patients
- Similar treatment philosophy but much lower long-term event-free survival rate (40%)
- Unsynchronized GEPs betw D0 and D60
- Standard data preprocessing
- Select drug-responsive genes by MILE-AML vs MILE-NBM
- PCA and use ASD and ESD to make prediction
Genetic Status Shifting Model of AML

PC1 vs PC2 plot showing different states of AML, with colors and symbols indicating different statuses:
- Red circle: D0 (Remission)
- Red triangle: D0 (Relapse)
- Blue circle: D5~D20 (Remission)
- Blue triangle: D5~D20 (Relapse)
- Green triangle: D25~D40 (Remission)
- Green circle: D25~D40 (Relapse)
- Purple circle: D45~D60 (Remission)

Key points:
- DC
- KL448
- KL505
- KL336
- KL473
- KL343
- R474
- R318
- R318
- K5014
- N0

Legend:
- D0: Remission
- D0: Relapse
- D5~D20: Remission
- D5~D20: Relapse
- D25~D40: Remission
- D25~D40: Relapse
- D45~D60: Remission
## Results on AML Dataset

| Rank | SAMPLE   | ASD | Outcome | | Rank | Sample   | ESD  | Outcome |
|------|----------|-----|---------| | |        |      |         |
| 1    | R318-D5  | 0.28| R       | | 1    | R318-D33| -11.03| R       |
| 2    | KL473-D32| 3.04| R       | | 2    | R318-D5 | 0.04  | R       |
| 3    | KL343-D36| 4.33| R       | | 3    | KL473-D32| 2.83  | R       |
| 4    | KL448-D17| 8.11|         | | 4    | KL343-D36| 3.34  | R       |
| 5    | KL505-D14| 10.61|        | | 5    | KL448-D17| 6.99  |         |
| 6    | R474-D33 | 11.52|         | | 6    | KL505-D14| 10.33 |         |
| 7    | R318-D33 | 20.10| R       | | 7    | R474-D33| 11.31 |         |
| 8    | R474-D60 | 25.67|         | | 8    | R474-D60| 25.62 |         |
| 9    | KL336-D31| 27.14|         | | 9    | KL336-D31| 26.65 |         |
| 10   | KL505-D45| 31.07|         | | 10   | KL505-D45| 31.04 |         |
| 11   | KKH14-D36| 35.61|         | | 11   | KKH14-D36| 35.61 |         |
| 12   | KL448-D51| 39.71|         | | 12   | KL448-D51| 39.67 |         |
Fast forward to 2014…

- GSS is now called “Effective Response Metric” (ERM) and has been tested on 8 new batches of 181 patients

- Difference in the shift of GEP towards normal remains true
Effectiveness of GSS/ERM

- Patients w/ minimal changes in GEP after 8 days of therapy have 4x higher risk of relapse
  - Good ERM → 14.5% relapse
  - Poor ERM → 44.5%

- Better than Day33 MRD
  - Day33 MRD –ve patients
    - Good ERM → 4.4% relapse
    - Poor ERM → 21.7% (5x more)
  - Day33 MRD +ve patients
    - Good ERM → 29.2% relapse
    - Poor ERM → 62.8% (2x more)

- Better than cytogenetic
  - Favourable cytogenetic patients
    - Good ERM → 2.4% relapse
    - Poor ERM → 36.4% (18x more)
ERM vs MRD33

<table>
<thead>
<tr>
<th></th>
<th>MRD</th>
<th>ERM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal</td>
<td>3/60</td>
<td>U/45</td>
</tr>
<tr>
<td>Bad</td>
<td>5/9</td>
<td>21/33</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Actual</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Goal</td>
<td>Rad</td>
<td>Acc</td>
</tr>
<tr>
<td></td>
<td>81</td>
<td>8</td>
<td>68%</td>
</tr>
<tr>
<td></td>
<td>46</td>
<td>32</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Actual</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Goal</td>
<td>Rad</td>
</tr>
<tr>
<td></td>
<td>91</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>26</td>
</tr>
</tbody>
</table>
Acknowledgements

• Much of this lecture is based on the works of my past/current students
  – Koh Chuan Hock (Ah Fu)
  – Donny Soh
  – Dong Difeng
  – Kevin Lim