Work Performed by

• Ochs Lab
  – Andrew Kossenkov (Wistar Institute)
  – Ghislain Bidaut (CRC-INSERM Marseille)

• GIST work in collaboration with Andrew Godwin (Fox Chase Cancer Center)
  – Lori Rink, Chi Tarn
Outline

• Motivation
  – Cell signaling and cancer therapeutics
  – Biomarkers of signaling changes

• Transcriptional Signatures
  – The response to a transcriptional regulator
  – Bayesian Decomposition

• Gastrointestinal Stromal Tumors
  – cKIT signaling and Gleevec
Cell Signaling

Extracellular Signal

Signal Transduction

Metabolic Changes

Transcription

RAS

Sebti and Der, Nat Rev Cancer, 3, 945, 2003

Proliferation

Apoptosis
Targeted Therapeutics

Personalized Medicine
Measuring Signaling

- **Protein Levels**
  - very hard in vivo for signaling proteins
  - efforts using 2D gel have been successful, but expensive and not scalable
  - not a global measure, only what you seek

- **Transcript Levels**
  - easily measured from biopsies/tumors
  - but …
Expression and Signaling

• Concordance of mRNA and Protein
  – correlation coefficient of 0.36 (yeast)
  – human prostate cancer (60% concordant meaning *only* same direction of change)

• Signaling
  – driven by post-translational modifications
  – signaling proteins have low expression
Targeted Therapeutics

Personalized Medicine
Transcriptional Signatures

[Diagram showing a flowchart of transcriptional pathways involving genes and signaling molecules like Ste7, Fus3, and Hog1 under different conditions such as pheromone, starvation, and high osmolarity.]
Biomarkers

• “Signaling Biomarker”
  – Set of genes regulated by downstream effectors
  – Isolate behavior of each gene due to regulator
  – Total expression of a gene is not strongly linked to signaling phenotype

• “Traditional” Biomarker
  – Strong association of gene expression level with phenotype
  – Strong association of relative expression levels of multiple genes with phenotype
Outline

• Motivation
  – Cell signaling and cancer therapeutics
  – Biomarkers of signaling changes

• Transcriptional Signatures
  – The response to a transcriptional regulator
  – Bayesian Decomposition

• Gastrointestinal Stromal Tumors
  – cKIT signaling and Gleevec
Issues for Analysis

• Multiple Regulation
  – Genes are regulated by multiple pathways and processes
  – Multiple processes active at all times

• Statistical Power
  – Grouping genes increases statistical power
  – But knowledge is incomplete and prone to error
Bayesian Decomposition

\( D: \) Data vs Mock

\( A: \) Distribution of Genes in Patterns

\( P: \) Patterns of Behavior

Determining \( k \) Still an Issue


Moloshok et al, *Bioinformatics*, 18, 566, 2002
Markov Chain Monte Carlo

Find $A$ and $P$
Simultaneously
We cannot always solve the problem analytically (unlike PCA)
We can only estimate relative probabilities of possible solutions

Markov Chain Monte Carlo is used to explore the possible solutions (Gibbs sampler with simulated annealing)
MCMC and Posterior

\[ p(A, P | D) = \]

Posterior Distribution

Prior

Minimization of Structure
Positivity of mRNA levels
Known TF regulation

Oncology Bioinformatics
Outline

• Motivation
  – Cell signaling and cancer therapeutics
  – Biomarkers of signaling changes

• Transcriptional Signatures
  – The response to a transcriptional regulator
  – Bayesian Decomposition

• Gastrointestinal Stromal Tumors
  – cKIT signaling and Gleevec
Signaling in GIST
GIST Cell Line Study

- Treatment of GIST-T1 IM Sensitive Line*
  - 0, 2, 4, 6, 9, 12, 18, 24, 48 hr 10 μM IM
  - Grown in Triplicate
  - Triplicate Agilent HGUG4112a Arrays

- Preprocessing in R
  - Limma: Loess Normalization, Cross-slide Normalization
  - Agilent and R QC/QA

Preprocessing

• Replicates by UniGene
  – Combine all probes for same ID
  – Estimate Mean and SDOM

• TRANSFAC
  – Map TRANSFAC to UniGene using ASAP*
  – Retain All TFs with more than 5 targets

• BD Analysis
  – All genes that are targets in TRANSFAC

*Kossenkov et al, Bioinformatics, 19, 675, 2003
Data Set

- 9 Time Points
  - Means and SDOM
- 1062 Genes (TRANSFAC code <= 3)
  - UniGene cluster as “gene”
  - Each gene mapped to known TF regulator
- 204 TFs with 5 or more Genes
Measures from BD

• Pattern Matrix
  – Normalized expression at each time point
  – Also, uncertainties on these values (not used here)

• Amplitude Matrix
  – Strength of each gene in each pattern
  – Uncertainty (std dev) on this value

\[ D = A P \]
Patterns

- Rising
- Transient
- Declining
- Undulating
- Flat
Statistic

• For Each Pattern
  – Generate Z-score for all genes
  – Get average Z-score for TF (below left)
  – Generate average Z-score for permutations with same number of genes (below right)

• Statistic
  \[
  \frac{1}{R} \sum_{r \in G} Z_{rp} \quad \text{vs.} \quad \frac{1}{R} \sum_{r \in A} Z_{rp}
  \]
Decreasing with IM

<table>
<thead>
<tr>
<th>TF</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elk.1</td>
<td>0.002</td>
</tr>
<tr>
<td>SRF</td>
<td>0.002</td>
</tr>
<tr>
<td>SRF.L</td>
<td>0.002</td>
</tr>
<tr>
<td>c.Myc</td>
<td>0.01</td>
</tr>
<tr>
<td>MTF.1</td>
<td>0.016</td>
</tr>
<tr>
<td>p53</td>
<td>0.016</td>
</tr>
<tr>
<td>Smad3</td>
<td>0.018</td>
</tr>
<tr>
<td>HIF.1</td>
<td>0.022</td>
</tr>
<tr>
<td>HSF1</td>
<td>0.026</td>
</tr>
<tr>
<td>HES.1</td>
<td>0.028</td>
</tr>
<tr>
<td>p63alpha</td>
<td>0.03</td>
</tr>
<tr>
<td>Sp1</td>
<td>0.03</td>
</tr>
<tr>
<td>TEF.3</td>
<td>0.03</td>
</tr>
<tr>
<td>Erm</td>
<td>0.034</td>
</tr>
<tr>
<td>JunD</td>
<td>0.042</td>
</tr>
<tr>
<td>SREBP.1</td>
<td>0.044</td>
</tr>
<tr>
<td>SREBP.1a</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Signaling and Transcription

• Focus on Signaling Pathways
  – KIT mutation key in GIST
  – Significant TFs also suggest pathways

• TF Measure
  – Treat p-Value as indicator of TF activity
  – If p approaches 1, p-Value of nonactivity approaches 0 because from permutation test
Network Activity

Oncology Bioinformatics
p53: DNA Damage?

- GIST cells without treatment.
- GIST cells treated with imatinib for 48 hours.
- GIST cells treated with 4Gy of irradiation followed by 2 hours recovery at 37°C.
- GIST cells treated with imatinib for 48 hours, followed by irradiation and recovery.
Summary of Comet Assays

Average Tail Moment

Arbitrary #

- 1.5
IM 9.1
IR 7.9
IM+IR 12.2

GIST T1
Summary

- Overlapping Regulation
  - BD can separate strongly overlapping sets of genes

- Signaling Activity
  - Regulated genes to TF to signaling pathway can be used to estimate signaling activity

- TF Activity in Tumors
Acknowledgements

JHU/Fox Chase

Systems Development
Jeffrey Grant
Yue Zhang
Elizabeth Goralczyk
Olga Tchuvatkina
Liat Shimoni
Luke Somers (UPenn)

Bioinformatics Support
Tom Moloshok
Sinoula Apostolou
Michael Slifker

Graduate Students
Ghislain Bidaut (CRC Marseille)
Andrew Kossenkov (Wistar)
Vladimir Minayev (MPEI)

Postdocs
Alexander Favorov
Garo Toby (Biotech)
Yan Zhou (FCCC)
Aidan Petersen (HHMI-UMinn)
Guoli Wang (FCCC)

Summer Students
Bill Speier (Johns Hopkins)
Daniel Chung (Columbia)
DJ Datta (UCSF)
Ryan Goldstein (UPenn)

Collaborators
A. Godwin (FCCC)
D. Geman/Laurent Younes (JHU)
G. Parmigiani (JHU)
O. Favorova (RMSU)
J.-M. Claverie (CNRS)

Oncology Bioinformatics