G-DOC – Enabling Systems Medicine through Innovations in Informatics

Subha Madhavan, Ph.D.
Director, Clinical Research & Biomedical Informatics
Georgetown University

CCSB and ISIS Grand Rounds
Stanford University
Systems Medicine Defined

- The new and emerging field of Systems Medicine, an application of Systems Biology approaches to biomedical problems in the clinical setting, leverages complex computational tools and high dimensional data to derive personalized assessments of disease risk.
- Systems Medicine offers the potential for more effective individualized diagnosis, prognosis, and treatment options.
- Achieving this goal requires the effective use of petabytes of data, which necessitates the development of new types of tools.
Outline

• The Motivation

• Three Pillars That Drive Systems Medicine
  – Informatics Vision, Mission, Team
  – Scalable Technology Platform(s)
  – Methods Development

• Driving Scientific Research In Systems Medicine
Research – Data Analysis Today!!!
(not surprisingly..)
Reality of Personalized (or Systems) Medicine Today

• The breast cancer drug Herceptin is considered the model for the future of medicine tailored to each individual.

• The drug is given only to the 25 percent of breast cancer patients whose tumors have a particular genetic characteristic.
  – 230,000 breast cancer patients per year (US) x 25% = 57,500 patients
  – Response rate 40% x 57,500 patients = 23,000 patients respond to Herceptin treatment
  – Therefore, only 10% of BC patients respond to Herceptin
  – Test is less than 50% accurate

• Many patients suffer from overtreatment or under treatment!
But we live in best of times – an era of Scientific Transformation

EGF Receptor Interactome

638 Genes
But we live in best of times – an era of Scientific Transformation

The Human Metabolome
Yet, we cannot answer many of patient’s individualized questions

- Will my cancer spread?
- Do I need chemotherapy after surgery for my cancer type?
- What are the benefits and side effects of chemotherapy for me?
- Are there any new drugs targeted for my type of cancer?
- Will I survive?
Need improved efforts

- Information continuum (care -> research -> back to care): Connect the dots
- Incorporation of “omics-based evidence” in Clinical Research and in Care settings (EHRs, PHRs)
- Collect data once and use it multiple times – clinical care, secondary use for research
- Connect research platforms to accelerate progress
- Efficiently utilize molecular and clinical information to transform patient care
Georgetown’s Vision For Systems Medicine

Georgetown Ushers in New Era of Systems-Based Medicine

(This is part 1 of 3 of a series of articles excerpted from "Georgetown Ushers in New Era of Systems-Based Medicine" in the Spring/Summer 2009 issue of Georgetown Medicine Magazine)

"The future of medicine will continue to be evidence-based, but it will also be holistic — it must be based in an understanding of systems and networks." — Howard J. Federoff, MD, PhD

Howard J. Federoff, MD, PhD, executive vice president for health sciences at Georgetown University Medical Center, is acutely aware of the hurdles facing the introduction of a sweeping new vision for healthcare, systems medicine. Federoff sees this approach impacting, not only how patient care is delivered, but perhaps even more importantly, how GUMC teaches, conducts research and engages with the community.

He knows that properly conveying the meaning of this approach, to start, and then addressing the various logistical, technological, financial and ethical issues that surround it will be a great challenge. So strongly does he believe in its promise, however, and in Georgetown’s potential to help usher it in, that he is unwavering in his commitment to getting out the word.

"This is the future. It is a question, not really of if it’s going to come, but really when," Federoff said.

Levels of understanding vary widely as to exactly how this concept will play out — not only at Georgetown, but in many forward-looking academic health centers around the country. But throughout the Medical Center, researchers, faculty and administrators are gearing up to implement this approach.

As a concept, systems medicine is at once complex and straightforward. Federoff, who marked two years at
G-DOC – Systems Medicine Vision of the Cancer Center

APPLICATIONS
- Utility
- Informatics
  - Molecular Data
  - Patient Data
- Institutions

Georgetown Database of Cancer
  - IT, Biostatistics, Bioinformatics

Prevention
Prognosis
Prediction

Biomarkers
Drug Development

Gene Expression
DNA and Gene Analysis
Proteome Metabolome Kinome

Clinical
Behavioral
Epidemiologic
Tissue Banking

Lombardi
GUMC
External Collaborations
Three Pillars Driving Systems Medicine

Team
- Informatics Vision
- Informatics Mission
- Cross-Disciplinary Team

Scalable Technology Platform(s)
- Data Integration & Analysis
- Electronic Data Capture
- Reporting and BI tools

Method Development
- Multi-Omics Capabilities
- EHR Integration
- Biomarker Discovery & Disease Classification
Who We Are

Commitment to driving biomedical research
Effective use of technology to solve real problems

End to end view of projects
Project management and planning

Data analysis expertise
Extracting knowledge from data

World-class multi-disciplinary team facilitation
Information strategy and metadata management

Agile and broad knowledge of technologies
Excellence in software tools for biomedical informatics

Our values
• Delivery focus
• Team work
• Bold ideas
• Innovation
• Pride in our work
• Excellence
Who We Are (contd)

Focus on web-based user experience
User-targeted delivery on software projects

State-of-the-art database design
Attention to process detail and quality

Systems management expertise
Information security

Scientific data management
Genotype-Phenotype Associations

Great collaborators

Not listed: Grants/Finance Office, RAs, Students, Other Faculty
G-DOC Suite Of Tools

Pathway Studio
- Systems biology analysis
- Literature mining

Cytoscape
- Visualization networks (pathways, interactions)

Gene Pattern
- Genomic/Proteomic Analysis
- Visualizations

Clinical Research
- REDCap

JMol/Marvin
- 3-D Structure and Molecule Visualization

JBrowse
- Genome Visualization

Heatmap Viewer
- Visualization of copy number data

EHR
- ARIA
- AMALGA
- CENTRICITY

catissue
- Sample Tracking
G-DOC Supports A Variety of Data Types

- Clinical
- Chromosomal Instability Index
- CNA
- 3D Structures
- Curated Findings
- Metabolites
- Chem
- miRNA
- mRNA Microarray
The Georgetown Database of Cancer (G-DOC) is a cutting-edge data integration platform and knowledge discovery system for the oncology and translational research communities. G-DOC users can access public and proprietary clinical and -omics data aggregated from across the Medical Center, along with a comprehensive set of advanced analysis and visualization tools, to generate and test hypotheses across biomedical disciplines.

<table>
<thead>
<tr>
<th>Disease</th>
<th>Study Count</th>
<th>Patient Count</th>
<th>Biospecimen Count</th>
<th>Available Data Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>BREAST CANCER</td>
<td>19</td>
<td>3319</td>
<td>3690</td>
<td></td>
</tr>
<tr>
<td>Disease</td>
<td>Study Count</td>
<td>Patient Count</td>
<td>Biospecimen Count</td>
<td>Available Data Types</td>
</tr>
<tr>
<td>------------------</td>
<td>-------------</td>
<td>---------------</td>
<td>-------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>BREAST CANCER</td>
<td>19</td>
<td>3319</td>
<td>3690</td>
<td></td>
</tr>
<tr>
<td>COLON CANCER</td>
<td>9</td>
<td>662</td>
<td>1125</td>
<td></td>
</tr>
<tr>
<td>LIVER CANCER</td>
<td>3</td>
<td>280</td>
<td>468</td>
<td></td>
</tr>
<tr>
<td>PANCREATIC CANCER</td>
<td>1</td>
<td>52</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>STOMACH CANCER</td>
<td>1</td>
<td>197</td>
<td>165</td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>33</strong></td>
<td><strong>4510</strong></td>
<td><strong>5499</strong></td>
<td></td>
</tr>
</tbody>
</table>

**FINDINGS**

**NEWS**
List of genes that can be used to classify G2 breast tumors into G2a (low grade) and G2b (high grade) subsets, which are similar in survival outcome to C1 and C3 tumors, respectively; validated in 2 separate cohorts based on Ivshina, et al. 2006 - loaded on: Wed Aug 17, 2011

**PUBLICATIONS**
List of genes which are associated with predicting metastatic colon cancer patient respondents to Cetuximab (anti-EGFR) based on Khambata-Ford, et al. 2007 - loaded
Welcome back, your last login was Fri Sep 16, 2011. You can check if you have been granted access to new lists or analyses since your last login.

(enter published findings: genes, proteins, cancer type, studies, investigators, authors ..)

Getting Started with G-DOC

Most users prefer to start using G-DOC to compare and analyze how groups of subjects within a cancer study differ, either by attributes or via various 'omics' characteristics. The typical process of searching for unique lists before analysis is done for you in the Quick Start feature below.

Quick Start

Let G-DOC organize subjects (patients, cell line, animal models) by cancer type, enabling you to stratify 2 groups by outcome or experimental design (e.g. Relapse, Treated vs. Non-Treated) and quickly take you to the next step of performing an analysis.

Tutorials

Watch step-by-step movies of workflows that are available within the G-DOC application. More instruction is also available in the help section.
<table>
<thead>
<tr>
<th>Study Name</th>
<th>Id</th>
<th>Description</th>
<th>Principal Investigator(s)</th>
<th>Disease</th>
<th>Subject Matter</th>
<th>Point(s) of Contact</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRC_CLARKE_9999_01</td>
<td>222</td>
<td>Breast cancer cell line from Clarke Lab</td>
<td>Robert Clarke, PHD,DSC</td>
<td>BREAST CANCER</td>
<td>CELL LINE</td>
<td>Ayesha Shaibahan</td>
</tr>
<tr>
<td>BRC_CLARKE LIU_9999_01</td>
<td>149</td>
<td>Clarke-Liu Data Set</td>
<td>Minetta Liu, MD</td>
<td>BREAST CANCER</td>
<td>PATIENT</td>
<td>Rebecca Riggins</td>
</tr>
<tr>
<td>BRC_DESMEDT_2007_01</td>
<td>144</td>
<td>Strong Time Dependence of the 76-Cene Prognostic Signature</td>
<td>Public Data Source,</td>
<td>BREAST CANCER</td>
<td>PATIENT</td>
<td>Public Data Source</td>
</tr>
<tr>
<td>BRC_DESMEDT_2009_01</td>
<td>145</td>
<td>CGI: a potential predictor of relapse for endocrine-treated breast cancer patients in the BIC 1-98 trial</td>
<td>Public Data Source,</td>
<td>BREAST CANCER</td>
<td>PATIENT</td>
<td>Public Data Source</td>
</tr>
<tr>
<td>BRC_FINAK_2008_01</td>
<td>167</td>
<td>Tumor-associated stroma derived from primary clinical breast cancer samples</td>
<td>Public Data Source,</td>
<td>BREAST CANCER</td>
<td>PATIENT</td>
<td>Public Data Source</td>
</tr>
<tr>
<td>BRC_FINETTI_2009_01</td>
<td>146</td>
<td>Molecular profiling of ERBB2-amplified breast cancers</td>
<td>Public Data Source,</td>
<td>BREAST CANCER</td>
<td>PATIENT</td>
<td>Public Data Source</td>
</tr>
<tr>
<td>BRC_LIN_2007_01</td>
<td>223</td>
<td>Timecourse of estradiol (10nM) exposure in MCF7 breast cancer cells</td>
<td>Public Data Source,</td>
<td>BREAST CANCER</td>
<td>CELL LINE</td>
<td>Public Data Source</td>
</tr>
<tr>
<td>BRC_LOI_2008_01</td>
<td>141</td>
<td>Molecular profiling with ER and tamoxifen status</td>
<td>Public Data Source,</td>
<td>BREAST CANCER</td>
<td>PATIENT</td>
<td>Public Data Source</td>
</tr>
</tbody>
</table>
BRC_CLARKE_LIU_9999_01 Details

<table>
<thead>
<tr>
<th>Study Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Study Name</strong></td>
</tr>
<tr>
<td><strong>Study Abstract</strong></td>
</tr>
</tbody>
</table>
| **Principal Investigator(s)** | Minetta Liu, MD  
Robert Clarke, PHD,DSC |
| **Disease** | BREAST CANCER |
| **Point(s) of Contact** | Rebecca Riggins |

<table>
<thead>
<tr>
<th>Data Type Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>MICROARRAY</td>
</tr>
<tr>
<td><strong>Data-Type</strong></td>
</tr>
<tr>
<td>Clinical Data</td>
</tr>
<tr>
<td><a href="#">Search</a></td>
</tr>
</tbody>
</table>
**Title:** List of one potential new target for treating a subset of hepatocellular carcinomas based on Chiang, et al. 2008

**Curator:** David Tanenbaum  
**Date Posted:** 8/17/2011  
[view details...]

**Title:** List of genes that can be used to classify G2 breast tumors into G2a (low grade) and G2b (high grade) subsets, which are similar in survival outcome to G1 and G3 tumors, respectively; validated in 2 separate cohorts based on Ivshina, et al. 2006

**Curator:** David Tanenbaum  
**Date Posted:** 8/17/2011  
[view details...]

**Title:** List of genes which are associated with predicting metastatic colon cancer patient respondents to Cetuximab (anti-EGFR) based on Khambata-Ford, et al. 2007

**Curator:** David Tanenbaum
Input Search | Molecule 'Sketch' Search

Enter name for a gene, protein, molecule: egrf

< molecular weight < 

reset | search

NAME: No name available for the molecule at this time.

This compound is accessible to PUBLIC

Targets: EGFR
(ECFR)

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formula</td>
<td></td>
</tr>
<tr>
<td>Molecular Weight</td>
<td>85.1045</td>
</tr>
<tr>
<td>Refractivity</td>
<td></td>
</tr>
<tr>
<td>Solubility</td>
<td></td>
</tr>
<tr>
<td>PH</td>
<td></td>
</tr>
<tr>
<td>EC50 [nM]</td>
<td></td>
</tr>
<tr>
<td>IC50 [nM]</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Donor Atoms</td>
<td>1</td>
</tr>
<tr>
<td>Acceptor Atoms</td>
<td>1</td>
</tr>
<tr>
<td>Clog P</td>
<td>-0.583</td>
</tr>
<tr>
<td>Rotatable Bonds</td>
<td>0</td>
</tr>
<tr>
<td>ED50 [nM]</td>
<td></td>
</tr>
<tr>
<td>Other Assay</td>
<td></td>
</tr>
<tr>
<td>Chiral</td>
<td></td>
</tr>
</tbody>
</table>

NAME: No name available for the molecule at this time.

This compound is accessible to PUBLIC

Targets: EGFR
(ECFR)
KM Plot Results

Current Study: BRC_SOTIRIOU_2006_01
Gene Expression KM Plot

Reporter: 202704_at, Fold Change: 1

Fold Change: 2
Reporters: 202704_at

Redraw plot

Graph showing disease free survival in days against probability of event.
DNA Copy Number Segments, Chr1
Correlate an abnormality/event with clinical parameters
40 CRC Patients, Stage 2, with >10 years of follow-up
(Samples provided by INDIVUMED Inc., Germany)

- 20 Relapse_Free Patients
  - Tissue DNA: Tumor – 20; Normal – 20
  - Tissue RNA: Tumor – 20; Normal – 20
  - Biofluids microRNA: Serum – 20
  - Biofluids Metabolites: Serum – 20; Urine – 20

- 20 Relapsed Patients
  - Tissue DNA: Tumor – 20; Normal – 20
  - Tissue RNA: Tumor – 20; Normal – 20
  - Biofluids microRNA: Serum – 20
  - Biofluids Metabolites: Serum – 20; Urine – 20

- Clinical Attributes: >100
**Bottom Line:**

40 CRC patients: 20 with relapse vs. 20 relapse-free

What are the molecular correlates of Relapse?

<table>
<thead>
<tr>
<th>Sample Type</th>
<th>Differentially Expressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tumor DNA CNV</td>
<td>37 cytobands</td>
</tr>
<tr>
<td>Tumor RNA Genes</td>
<td>720 reporters</td>
</tr>
<tr>
<td>Tumor RNA microRNA</td>
<td>34 microRNAs</td>
</tr>
<tr>
<td>Serum microRNA</td>
<td>8 microRNAs</td>
</tr>
<tr>
<td>Serum Metabolites</td>
<td>77 peaks</td>
</tr>
<tr>
<td>Urine Metabolites</td>
<td>47 peaks</td>
</tr>
<tr>
<td>DNA Exome-seq (EdgeBio)</td>
<td>Analysis In Progress</td>
</tr>
</tbody>
</table>
Genes in Tumor Samples: Relapse vs Relapse free
PCA based on T-test  p<0.05

PCA Results:
Complete separation of two groups of patients with one sample on a borderline
Differentially Expressed Genes: Enrichment Analysis
Pathway Studio: Top Signaling Pathways

Gap Junction Regulation Pathway

EGFR/ERBB -> STAT signaling
Proteins Regulating Cell Processes of Inflammatory Response
Proteins Regulating Cell Processes of Immune Response
Gene Expression Findings:

- Strong Expression Pattern of Inflammatory Response:
  - In tumors as well as in normal samples

Possible source: Infiltrating white blood cells

Reference:
microRNA in Tissue: 40 Tumor samples

Tumors Only: Relapse vs Relapse_Free
microRNA in Serum
40 samples, t-test p<0.05 8 miRNAs
microRNA in Serum

PCA 8 microRNAs
DNA Copy Number Analysis
Affy SNP 6.0 arrays

- Raw Data → Probe Level Copy Number: 1.6 million probes
- Probe Level → Segment Level Copy Number: 100K segments
- Segment Level → CIN Index: Whole Chromosome: 22 values
  Individual Cytobands: ~800 values

20 Relapse_Free Patients
Tissue DNA: Tumor – 20; Normal – 20

20 Relapse Patients
Tissue DNA: Tumor – 20; Normal – 20
CIN Index, whole chromosome level: 40 patients
No significant difference between groups

Overall CIN Index  Gains CIN  Losses CIN

---

Cytogenetic analysis showing the distribution of gains and losses in the CIN index across different chromosome regions. The images represent data matrices with color-coded regions indicating the presence of gains or losses.
CIN Index, Cytoband level: 40 patients – significant difference between groups

Chromosome # 4

Overall

Gains

Losses
CIN Index - Cytoband level: Relapse vs Relapse_Free, t-test, P<0.05

Deletion of chromosome 4q predicts outcome in stage II colon cancer patients.

RESULTS: Stage II colon cancers of patients who had relapse of disease showed significantly more losses on chromosomes 4, 5, 15q, 17q and 18q.

In the microsatellite stable (MSS) subgroup (n = 28), only loss of chromosome 4q22.1-4q35.2 was significantly associated with disease relapse.
Metabolites in Biofluids: Relapse vs Relapse_Free

Serum:
Serum Pos – 10 and 30 samples;
Serum Neg – 10 and 30 samples;

Urine:
Urine Pos - 40 samples
Urine Neg – 40 samples
Metabolomics Methods

Sample Preparation

LC-MS Data Preprocessing (MassLynx)
(Filtering, feature extraction, feature matching, retention time correction & handling missing peaks)

<table>
<thead>
<tr>
<th>Sample 1</th>
<th>...</th>
<th>...</th>
<th>Sample N</th>
</tr>
</thead>
<tbody>
<tr>
<td>$RT_1 \cdot m/z_1$</td>
<td>...</td>
<td>Ion Abundance</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$RT_n \cdot m/z_n$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Linear Modeling
“moderated t-statistics”

Feature Identification
(HMDB, KEGG, METLIN, METACYC, LMDB)
& Validation (MS/MS)*

Pathway Analysis
IPA, SMPDB

Network Analysis
(Multi-omics)
INPUT

\[ \text{m/z (+/-) \quad \text{Mass accuracy (ppm)} } \]

\[ \text{INPUT SMALL MOLECULE DATABASES} \]

\[ \text{OUTPUT} \]

\[ \begin{align*}
\text{HMDB} & \quad \text{KEGG} \\
\text{METLIN} & \quad \text{LMDB} \\
\text{METACYC} & 
\end{align*} \]

\[ \text{Metadata} \]

\[ \text{Filtering based on Cross Validation} \]

\[ \text{High Confidence Annotations} \]
Metabolomics – Urine Positive
40 samples, 47 peaks p<0.01
<table>
<thead>
<tr>
<th>m/z</th>
<th>Putative Metabolites</th>
<th>KEGG</th>
<th>HMDB</th>
<th>METLIN</th>
<th>LMDB</th>
<th>METACYC</th>
</tr>
</thead>
<tbody>
<tr>
<td>119.0815</td>
<td>L-2,4-diaminobutyric acid</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>121.0318</td>
<td>3-Methylthiopropionic acid</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>130.0495</td>
<td>1-Pyrrole-4-hydroxy-2-carboxylate</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>130.0495</td>
<td>5-Oxo-D-proline</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>130.0495</td>
<td>5-oxoproline</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>130.0495</td>
<td>L-1-Pyrrole-3-hydroxy-5-carboxylate</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>130.0495</td>
<td>Pyroglutamic acid</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>130.0495</td>
<td>pyrroldone-carboxylate</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>130.0495</td>
<td>pyrroline-hydroxy-carboxylate</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>130.0497</td>
<td>1-Pyrrole-4-hydroxy-2-carboxylate</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>130.0497</td>
<td>5-Oxo-D-proline</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>130.0497</td>
<td>5-oxoproline</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>130.0497</td>
<td>L-1-Pyrrole-3-hydroxy-5-carboxylate</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>130.0497</td>
<td>Pyroglutamic acid</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>130.0497</td>
<td>pyrroldone-carboxylate</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>130.0497</td>
<td>pyrroline-hydroxy-carboxylate</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>130.0499</td>
<td>Pyroglutamic acid</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>130.0499</td>
<td>Pyrrolidonecarboxylic acid</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>130.0499</td>
<td>Pyrroline hydroxycarboxylic acid</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>135.0764</td>
<td>L-Canaline</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>135.0803</td>
<td>cinnamyl alcohol</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>135.0803</td>
<td>phenylacetone</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>149.0267</td>
<td>2-Oxo-4-methylthiobutanoic acid</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>153.0655</td>
<td>N1-Methyl-2-pyridone-5-carboxamide</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>153.0655</td>
<td>N1-Methyl-4-pyridone-5-carboxamide</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>153.0655</td>
<td>N-Methyl-2-pyridone-5-carboxamide</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>153.0655</td>
<td>N-Methyl-4-pyridone-5-carboxamide</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>153.0659</td>
<td>N1-Methyl-2-pyridone-5-carboxamide</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>153.0659</td>
<td>N1-Methyl-4-pyridone-5-carboxamide</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>165.0536</td>
<td>2-coumarate</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>165.0536</td>
<td>2-Hydroxycinnamate</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>165.0536</td>
<td>2-Hydroxycinnamic acid</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>165.0536</td>
<td>4-coumarate</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>165.0536</td>
<td>4-Hydroxycinnamic acid</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>165.0536</td>
<td>cis-p-coumarate</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>165.0536</td>
<td>enol-phenylpyruvate</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>165.0536</td>
<td>m-Coumaric acid</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Classification Algorithm: Support Vector Machine (SVM)

Gene Expression Tumors

AUC: 0.8150

microRNA Expression Tumors

AUC: 0.8425
Classification Algorithm: Support Vector Machine (SVM)

ROC Curve for CIN Index/Cytobands - Tumors

AUC: 0.8775
Classification Algorithm: Support Vector Machine (SVM)

AUC = 0.900

Metabolites
Serum_Pos

AUC = 0.7644

Metabolites
Urine_Pos

AUC = 0.9000
## Results of ROC Analysis for SVM classification

<table>
<thead>
<tr>
<th>Data Type/Tissue Type</th>
<th>Classifier</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metabolites/Serum_Pos</td>
<td>SVM</td>
<td>0.7</td>
</tr>
<tr>
<td>Gene Expression/Tumors</td>
<td>SVM</td>
<td>0.81</td>
</tr>
<tr>
<td>miRNA Expression/Tumors</td>
<td>SVM</td>
<td>0.84</td>
</tr>
<tr>
<td>CIN Index/Tumors</td>
<td>SVM</td>
<td>0.87</td>
</tr>
<tr>
<td>Metabolites/Urine_Pos</td>
<td>SVM</td>
<td>0.9</td>
</tr>
</tbody>
</table>
Critical associations involved in CRC Relapse

• Results with 5 molecular data sets plus clinical outcome:
  Gene Expression, microRNA expression, CNV, Metabolites/Serum; Metabolites/Urine

Target: Relapse Status

• Workflow:
  – Combine 5 data matrixes (pre-filtered on significance of differences) –
  – Input to RF-ACE algorithm,
  – Analyze with RF-ACE with 2K permutations,
  – Results of RF-ACE - Upload to Regulome Explorer instance at AWS
  – Visualize results of analysis on Genome map and/or Network viewer
  – Identify genome locations with large number of highly correlated changes
    – i.e. “Hot Spots”
  – filter data based on importance of association with clinical outcome
  – Find interactions between different molecular features that are highly ranked on importance
Genome View of top 40 molecular features/associations
Network View of correlations/associations for top 40 features:
Top 40 most important biomarkers of relapse

Combination of 5 omics data sets - Results of RF-ACE – metabolomics data are included

<table>
<thead>
<tr>
<th>TARGET</th>
<th>PREDICTOR</th>
<th>Log10_P-value</th>
<th>Importance Score</th>
<th>Correlation</th>
<th>Sample#</th>
</tr>
</thead>
<tbody>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:GEXP:UBD:chr6:29523656:29527478:::</td>
<td>-30.0000</td>
<td>0.0040</td>
<td>-0.5655</td>
<td>40</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:METB_Serum:0.3267_440.7302:::</td>
<td>-30.0000</td>
<td>0.0022</td>
<td>0.5380</td>
<td>30</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:METB_Serum:6.1292_226.2166:::</td>
<td>-30.0000</td>
<td>0.0022</td>
<td>-0.6115</td>
<td>30</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:METB_Serum:0.3289_800.4967:::</td>
<td>-11.3782</td>
<td>0.0022</td>
<td>0.4833</td>
<td>30</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:METB_Serum:1.7305_117.0697:::</td>
<td>-10.7396</td>
<td>0.0017</td>
<td>0.4023</td>
<td>30</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:GEXP:IRF6:chr1:209961764:20997475:::</td>
<td>-11.1480</td>
<td>0.0015</td>
<td>0.4570</td>
<td>40</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:GEXP:FITC:chr10:91174555:91178405:::</td>
<td>-11.1209</td>
<td>0.0014</td>
<td>-0.4956</td>
<td>40</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:GEXP:THNS1:chr2:88472669:88485642:::</td>
<td>-11.0729</td>
<td>0.0014</td>
<td>0.5199</td>
<td>40</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:GEXP:CXCL10:chr4:76943082:76944584:::</td>
<td>-11.0759</td>
<td>0.0014</td>
<td>-0.5048</td>
<td>40</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:GEXP:TNIP3:chr4:122053785:122085280:::</td>
<td>-11.4206</td>
<td>0.0013</td>
<td>-0.5002</td>
<td>40</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:METB_Serum:6.1315_156.138:::</td>
<td>-11.0755</td>
<td>0.0012</td>
<td>-0.6166</td>
<td>30</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:METB_Serum:6.1283_301.2485:::</td>
<td>-10.8742</td>
<td>0.0012</td>
<td>-0.5657</td>
<td>30</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:METB_Serum:0.3399_548.6131:::</td>
<td>-11.1661</td>
<td>0.0012</td>
<td>0.5553</td>
<td>30</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:METB_Serum:0.3181_206.8951:::</td>
<td>-30.0000</td>
<td>0.0012</td>
<td>0.5642</td>
<td>30</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:METB_Urine:6.6168_259.1671_Cicutoxin:::</td>
<td>-11.1274</td>
<td>0.0011</td>
<td>0.5012</td>
<td>40</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:METB_Serum:0.2997_242.9263:::</td>
<td>-30.0000</td>
<td>0.0011</td>
<td>-0.6072</td>
<td>30</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:METB_Serum:6.6989_284.2955:::</td>
<td>-11.4763</td>
<td>0.0011</td>
<td>-0.5305</td>
<td>30</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:METB_Serum:5.9522_168.1185:::</td>
<td>-11.0970</td>
<td>0.0011</td>
<td>-0.5849</td>
<td>30</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:METB_Serum:6.1276_366.232:::</td>
<td>-11.2178</td>
<td>0.0011</td>
<td>-0.5686</td>
<td>30</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:MIRN:hsa-mir-147-4373131:chr9:123007256:123007328:::</td>
<td>-10.8805</td>
<td>0.0011</td>
<td>-0.6017</td>
<td>31</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:METB_Serum:6.1373_647.5016:::</td>
<td>-10.8186</td>
<td>0.0011</td>
<td>-0.5989</td>
<td>30</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:METB_Serum:6.1308_441.8824:::</td>
<td>-11.2481</td>
<td>0.0011</td>
<td>-0.5782</td>
<td>30</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:GEXP:CXCL13:chr4:78527019:78532193:::</td>
<td>-11.5340</td>
<td>0.0009</td>
<td>-0.4712</td>
<td>40</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:GEXP:SMAD6:chr15:67074337:67074337:::</td>
<td>-11.5650</td>
<td>0.0009</td>
<td>-0.4331</td>
<td>40</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:GEXP:DHRS12:chr13:52343259:52373795:::</td>
<td>-10.9374</td>
<td>0.0009</td>
<td>0.4473</td>
<td>40</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:GEXP:CCL3:chr4:74902978:74904328:::</td>
<td>-11.0357</td>
<td>0.0009</td>
<td>-0.4396</td>
<td>40</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:GEXP:DUSP22:chr6:292539:350868:::</td>
<td>-10.9272</td>
<td>0.0009</td>
<td>0.4924</td>
<td>40</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:MIRN:hsa-mir-23a-4373074:chr19:13947400:13947473:::</td>
<td>-10.9803</td>
<td>0.0009</td>
<td>-0.5312</td>
<td>34</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:GEXP:FAIM2:chr12:50264286:50297575:::</td>
<td>-10.9929</td>
<td>0.0009</td>
<td>0.4390</td>
<td>40</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:METB_Urine:7.4196_551.3571:::</td>
<td>-10.9869</td>
<td>0.0009</td>
<td>0.5721</td>
<td>40</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:GEXP:SUMO2:chr17:73164433:73178929:::</td>
<td>-11.1583</td>
<td>0.0009</td>
<td>-0.4947</td>
<td>40</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:GEXP:PDE9A:chr21:44073924:44195403:::</td>
<td>-11.0278</td>
<td>0.0008</td>
<td>0.4828</td>
<td>40</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:METB_Urine:5.9726_425.1435:::</td>
<td>-10.9980</td>
<td>0.0008</td>
<td>0.4417</td>
<td>40</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:METB_Urine:7.9934_682.3645:::</td>
<td>-10.9646</td>
<td>0.0008</td>
<td>0.4581</td>
<td>40</td>
</tr>
<tr>
<td>C:CLIN:Relapse:</td>
<td>N:GEXP:TFEC:chr7:115580604:115624495:::</td>
<td>-10.9690</td>
<td>0.0008</td>
<td>-0.5218</td>
<td>40</td>
</tr>
</tbody>
</table>
Top ranked 20 Metabolomics Peaks – network view in Regulome Explorer
Cloud computing refers to the on-demand provision of computational resources via a computer network.

Cloud computing is an attractive model for application deployment because it provides the following things:

- Agility
- APIs
- Reduced Cost
- Device independent
- Scalability
- Performance
- Security
Discovery, Clinical Research, Clinical Care

Clinical Research
- ‘omics’ data analysis
- EHR/PHR integration
- Biomarker discovery
- Disease Classification
- Genotype-Phenotype Correlation
- Biospecimen management and analysis
- Novel technology/method development

Clinical Education
- MD/MS Systems Medicine
- Partnership
- Joint Grants
- Joint Appointments
- Contracted Services
- Shared Resources
- Medical Internships
- Joint Clinical Trials
- Shared IP
- Joint publications

Basic/Traslational

Biospecimen

email: sm696@georgetown.edu

G-DOC URL: https://gdoc.georgetown.edu
Vincent T. Lombardi

“Winning isn’t everything, it’s the only thing.”

- Famous professional football coach
- Fell ill at Washington Redskins’ practice on Georgetown University field
- Treated for cancer at Georgetown
- Cancer Center named in his honor
Multidisciplinary Team

- Lombardi Comprehensive Cancer Center:
  - Dr. Louis Weiner’s Lab and Dr Stephen Byers’s Lab;
  - Genomics, Cytogenetics and Metabolomics Shared Resources

- G-DOC Development Team: Michael Harris, Andrew Shinohara, Kevin Rosso, Lavinia Carabet

- Analytical Group: Yuriy Gusev, Krithika Bhuvaneshwar, Robinder Gauba, Lei Song
- Virginia Tech: Joseph Wang

- ISB: Ilya Shmulevich, Hector Rovira, Timo Erkkilä