An Introduction

USC

August 23, 2011
Agenda

» Oncomine History

» Introduction to Compendia Bioscience

» Oncomine Today
  » The Oncomine Platform
  » Oncomine Web Applications
  » Oncomine Services
Oncomine History

» Developed in Chinnaiyan Lab at University of Michigan
  » Led by Compendia Co-founder, Dan Rhodes, Ph.D.

» Led to several major discoveries
  » AMACR, SPINK1, AGTR1
  » Prostate Cancer Gene Fusions

» Rapidly adopted by academia
  » Currently serves more than 20,000 academic users

» Rapidly adopted by pharma
  » 15 of the top 20 pharma
Oncomine Impact

» Significant discoveries
  » ETS gene fusions in prostate cancer (Science, 2005)
  » SPINK1 in ETS-negative prostate cancer (Cancer Cell, 2008)
  » Cancer regulatory programs (Nature Genetics, 2005)
  » Prostate cancer progression (Nature Genetics, 2007)
  » AGTR1 as a target in breast cancer (PNAS, 2009)
  » KRAS addiction signature (Cancer Cell, 2009)
  » Signatures of pathway activation & drug response (unpublished)

» 500+ Oncomine literature references

» 15 of top 20 oncology biotech / pharma as customers

» 20,000+ academics
Gene Fusions in Prostate Cancer

- Novel analysis method - Outlier
  - Marked over-expression in subsets
- Identified ERG and ETV1 in Prostate Cancer
- Validated across 9 cohorts
- Over-expression caused by gene fusions with TMPRSS2
- May explain efficacy of androgen treatment
- Urine-based diagnostic under development

Oncomine Impact

SPINK1
- Biomarker
- Therapeutic target
- Diagnostic
Introduction to Compendia

» Founded in 2006
» Exclusive license to Oncomine
» Rapidly established Oncomine as a standard for pharma/biotech research
» Funding from
  » NCI SBIR Grants – granted for innovation
  » Michigan 21st Century Fund
  » Customer Revenue
Today

Web Applications

Oncomine Platform

Scientific Services
The Oncomine Platform

Data
Published cancer genomic data gathered from sources worldwide

Expert Curation
Dedicated team reviewing each sample, each annotation and each publication and mapping the data to The Compendia Ontology

The Compendia Ontology
Multi-threaded hierarchy of terms and synonyms to describe the data

Standardized Analysis
Leveraging The Compendia Ontology, standardized analyses are performed on every Oncomine dataset

The “Computable” Global Collection of Cancer Genomic Data
Oncomine Platform: Data Collection

» Dedicated team:
  » Collaborate with academic partners
    » University of Michigan
    » Harvard University
    » University of Utah
    » others
  » Monitor major repositories worldwide
  » Constantly review the literature
  » Identify studies and correspond directly with leading authors
  » Gather supplemental data from authors and other supplemental sources
Oncomine Platform: The Data

- World’s largest collection of curated cancer genomics data
  - 587 Datasets (and growing)
  - 55,000+ Samples
- Samples with key properties for powerful analyses
  - 14,000+ samples with known clinical outcomes
  - 7,000+ samples with known mutation status

<table>
<thead>
<tr>
<th>Cancer Type</th>
<th>Datasets</th>
<th>Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bladder</td>
<td>16</td>
<td>1,414</td>
</tr>
<tr>
<td>Brain/CNS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breast</td>
<td>53</td>
<td>3,426</td>
</tr>
<tr>
<td>Cervical</td>
<td>13</td>
<td>472</td>
</tr>
<tr>
<td>Colorectal</td>
<td>63</td>
<td>3,665</td>
</tr>
<tr>
<td>Esophageal</td>
<td>17</td>
<td>417</td>
</tr>
<tr>
<td>Gastric</td>
<td>16</td>
<td>771</td>
</tr>
<tr>
<td>Head/Neck</td>
<td>27</td>
<td>900</td>
</tr>
<tr>
<td>Kidney</td>
<td>29</td>
<td>1,653</td>
</tr>
<tr>
<td>Leukemia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liver</td>
<td>18</td>
<td>1,060</td>
</tr>
<tr>
<td>Lung</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lymphoma</td>
<td>45</td>
<td>2,682</td>
</tr>
<tr>
<td>Melanoma</td>
<td>44</td>
<td>1,369</td>
</tr>
<tr>
<td>Myeloma</td>
<td>27</td>
<td>2,662</td>
</tr>
<tr>
<td>Ovarian</td>
<td>32</td>
<td>2,858</td>
</tr>
<tr>
<td>Pancreatic</td>
<td>24</td>
<td>442</td>
</tr>
<tr>
<td>Prostate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sarcoma</td>
<td>37</td>
<td>1,518</td>
</tr>
</tbody>
</table>

*This bar chart represents datasets. The yellow denotes an increase in datasets and/or samples. Some samples are represented in multiple cancer types. Oncomine also has 6,140 normal samples.
## Oncomine Platform: The Data

New studies for July 2011 - new data is constantly being added

<table>
<thead>
<tr>
<th>Includes treatment response or drug sensitivity data</th>
<th>Includes additional data not in GEO</th>
<th>Oncomine Study Name</th>
<th>Tissue</th>
<th>Title</th>
<th>Journal</th>
<th>PubMed ID</th>
<th>Year Published</th>
<th>Sample Total</th>
<th>Data Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>x Biewenga Cervix</td>
<td>Cervix</td>
<td>Gene expression in early stage cervical cancer.</td>
<td>Gynecol Oncol</td>
<td>18191186</td>
<td>2008</td>
<td>45</td>
<td>mRNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x Bogunovic Melanoma</td>
<td>Melanoma</td>
<td>Immune profile and mitotic index of metastatic melanoma lesions enhance clinical staging in predicting patient survival.</td>
<td>Proc Natl Acad Sci</td>
<td>19915147</td>
<td>2009</td>
<td>44</td>
<td>mRNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x Chapman Myeloma</td>
<td>Myeloma</td>
<td>Initial genome sequencing and analysis of multiple myeloma.</td>
<td>Nature</td>
<td>21430775</td>
<td>2011</td>
<td>304</td>
<td>mRNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x Chapman Myeloma 2</td>
<td>Myeloma</td>
<td>Initial genome sequencing and analysis of multiple myeloma.</td>
<td>Nature</td>
<td>21430775</td>
<td>2011</td>
<td>254</td>
<td>DNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x Collison CellLine</td>
<td>Pancreas</td>
<td>Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy.</td>
<td>Nat Med</td>
<td>21460848</td>
<td>2011</td>
<td>20</td>
<td>mRNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x Collison Pancreas</td>
<td>Pancreas</td>
<td>Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy.</td>
<td>Nat Med</td>
<td>21460848</td>
<td>2011</td>
<td>27</td>
<td>mRNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x Estilo Head-Neck</td>
<td>Head-Neck</td>
<td>Oral tongue cancer gene expression profiling: Identification of novel potential prognosticators by oligonucleotide microarray analysis.</td>
<td>BMC Cancer</td>
<td>19138406</td>
<td>2009</td>
<td>58</td>
<td>mRNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x Giordano Thyroid</td>
<td>Thyroid</td>
<td>Delineation, functional validation, and bioinformatic evaluation of gene expression in thyroid follicular carcinomas with the PAX8-PPARG translocation.</td>
<td>Clin Cancer Res</td>
<td>16609007</td>
<td>2006</td>
<td>99</td>
<td>mRNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x Harlin Melanoma</td>
<td>Melanoma</td>
<td>Chemokine expression in melanoma metastases associated with CD8+ T-cell recruitment.</td>
<td>Cancer Res</td>
<td>19293190</td>
<td>2009</td>
<td>52</td>
<td>mRNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x Hegde CellLine</td>
<td>Breast</td>
<td>Delineation of molecular mechanisms of sensitivity to lapatinib in breast cancer cell lines using global gene expression profiles.</td>
<td>Mol Cancer Ther</td>
<td>17513611</td>
<td>2007</td>
<td>36</td>
<td>mRNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x Hu Esophagus</td>
<td>Esophagus</td>
<td>Genome-wide analysis of DNA copy number neutral loss of heterozygosity (CNNLOH) and its relation to gene expression in esophageal squamous cell carcinoma.</td>
<td>BMC Genomics</td>
<td>20955586</td>
<td>2010</td>
<td>34</td>
<td>mRNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x Imadore Cervix</td>
<td>Cervix</td>
<td>Subtypes of cervical adenosquamous carcinomas classified by EpCAM expression related to radiosensitivity.</td>
<td>Cancer Biol Ther</td>
<td>20855995</td>
<td>2010</td>
<td>160</td>
<td>mRNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x Kim Esophagus</td>
<td>Esophagus</td>
<td>Prognostic biomarkers for esophageal adenocarcinoma identified by analysis of tumor transcriptome.</td>
<td>PLoS One</td>
<td>21152079</td>
<td>2010</td>
<td>118</td>
<td>mRNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x Laurent Melanoma</td>
<td>Melanoma</td>
<td>High PTP4A3 phosphatase expression correlates with metastatic risk in ureal melanoma patients.</td>
<td>Cancer Res</td>
<td>21135111</td>
<td>2011</td>
<td>63</td>
<td>mRNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x Rickman Head-Neck</td>
<td>Head-Neck</td>
<td>Prediction of future metastasis and molecular characterization of head and neck squamous-cell carcinoma based on transcriptome and genome analysis by microarrays.</td>
<td>Oncogene</td>
<td>18679425</td>
<td>2008</td>
<td>81</td>
<td>mRNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sengupta Head-Neck</td>
<td>Head-Neck</td>
<td>Genome-wide expression profiling reveals EBV-associated inhibition of MHC class I expression in nasopharyngeal carcinoma.</td>
<td>Cancer Res</td>
<td>16912175</td>
<td>2006</td>
<td>41</td>
<td>mRNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x Takeno Gastric</td>
<td>Gastric</td>
<td>Gene expression profile prospectively predicts peritoneal relapse after curative surgery of gastric cancer.</td>
<td>Ann Surg Oncol</td>
<td>20012301</td>
<td>2010</td>
<td>14</td>
<td>mRNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCGA Head-Neck</td>
<td>Head-Neck</td>
<td>The Cancer Genome Atlas - Head and Neck DNA Copy Number data</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>DNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCGA Liver</td>
<td>Liver</td>
<td>The Cancer Genome Atlas - Liver DNA Copy Number data</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>88</td>
<td>DNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCGA Prostate</td>
<td>Prostate</td>
<td>The Cancer Genome Atlas - Prostate DNA Copy Number data</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>120</td>
<td>DNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x Waddell Breast</td>
<td>Breast</td>
<td>Gene expression profiling of formalin-fixed, paraffin-embedded familial Breast tumours using the whole genome-DASL assay.</td>
<td>J Pathol</td>
<td>20593485</td>
<td>2010</td>
<td>85</td>
<td>mRNA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Oncomine Platform: Expert Curation

» Uses a standardized “pipeline” to process incoming data

» Works directly with study authors to gather all available sample facts

» Reviews each sample fact for clarity and consistency across the dataset and across Oncomine

» Applies The Compendia Ontology to each sample for representation in application filter tree and as properties for analyses
Oncomine Platform: The Compendia Ontology

» A new and comprehensive way to characterize Oncomine data, metadata and analyses

» Visually represented to application users in the search and browse interface

» Drives consistent downstream analyses
Oncomine Platform: The Analyses

- Pre-computed analyses
  - Differential Expression
  - Coexpression
  - Outlier Analysis and...
  - Oncomine Concept Generation
- Logical, consistent analysis rules applied per dataset across the database
- Meta-analysis to compare independent results
  - Meta-Differential Expression
  - Meta-Outlier
Oncomine Research Premium Edition Features

- **Multi-Gene Search** allows you to compare the expression of specified, multiple genes at once to understand the relationships between these genes in a cancer type or subtype.
- **Custom Concepts Upload** allows you to create and analyze your own gene sets or signatures against all Oncomine concepts to better understand your experimental results.
- Manage your experimental data by adding descriptions to custom concepts during upload.
- **Expanded Analysis Types** help you to further refine your questions to include cancer subtype, drug sensitivity, and clinical outcome analyses.
- **Export Functionalities to Excel and PowerPoint** facilitate additional, in-depth data analyses and sharing information with colleagues.
- **Export edges and nodes data** from concept associations to build network graphs in third party applications.
- Visualize primary relationships between primary concept and associated concepts, and secondary relationships among associated concepts. **Export visualizations in Oncomine as scalable vector graphics (SVG)**.
- **Enhanced Support** guarantees 24-hour response to support questions and follow-up via telephone as needed to resolve your support questions.