Closing Remarks

Elaine Mardis
1st Annual TCGA Scientific Symposium
Challenges from the 2005 White Paper

- **Overarching Goal**
  - Obtain a comprehensive description of the genetic basis of human cancer

- **Overarching Assessment of Success**
  - Short-term
    - Milestones such as samples collected, analyzed, made accessible to the community
  - Long-term
    - Impact on the lives of patients

- **Strategy**
  - Create a large collection of samples from all major types of cancer ensuring:
    - Patient consent
    - **Clinical annotation**
    - Sample quality
    - **Sample quantity**
    - Availability of matched DNA
    - **Ethnic diversity**
TCGA Data Production Pipeline

Massive Scale of raw Sequence Data Production, Processing, and Storage

CGHub sequence data repository at San Diego Supercomputer Center (SDSC) ~5 PB BAMs for 10K tumor cases

Petascale data with massive compute and storage cost

360° Integration and Data Mining

Mutation calling
Structural variation
DNA vs RNAseq
Gene expression
Transcript
Methylation
Pathway analysis
Clinical correlations

DCC data sets (Terascale)

Issues: Data-intensive computing with large-scale raw data processing, massive transfer and storage, controlled access, privacy, integration with metadata

Issues: Data quality with collection, validation, provenance

The Cancer Genome Atlas
TCGA: The Pipeline for Comprehensive Characterization

Tissue Sample → Pathology QC → DNA & RNA Isolation, QC → Sequencing → Expression, CNA & LOH, Epigenetics → Data and Results Storage & QC → Analysis → Integrate Analysis & Publication
2012: The Year of TCGA Post-Pilot Publications

- Colorectal Cancer
- Acute Myeloid Leukemia
- Breast Cancer
- Endometrial Cancer
- Kidney Clear Cell
- Lung Adeno/Squamous
- Head and Neck Cancer
- Etc.
## TCGA: Whole Genome Sequencing

<table>
<thead>
<tr>
<th>Tumor Type</th>
<th>In Progress</th>
<th>Completed</th>
</tr>
</thead>
<tbody>
<tr>
<td>GBM</td>
<td>-</td>
<td>22</td>
</tr>
<tr>
<td>Colorectal</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>Renal</td>
<td>-</td>
<td>10</td>
</tr>
<tr>
<td>Breast (triple negative)</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>AML</td>
<td>-</td>
<td>49</td>
</tr>
<tr>
<td>Ovarian</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>Endometrial (serous type)</td>
<td>28</td>
<td>2</td>
</tr>
<tr>
<td>LUSC</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>LUAD</td>
<td>14</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>66</td>
<td>146</td>
</tr>
</tbody>
</table>

Updated Nov 1, 2011
TCGA: Expanding the Enterprise

- Pilot project: FFPE-preserved tissues
- Pilot project: Mouse models of human cancers
- Projects to study rare tumor types (smaller numbers yet comprehensive focus of assays and analysis)
- Integration efforts: International Cancer Genomics Consortium
- Integration efforts: interface between TCGA (genomics of cancer samples) and CPTAC (proteomics of cancer samples)
mTCGA Committee Members

- Elaine Mardis (Chair)
- Tyler Jacks
- Monica Justice
- David Threadgill
- Allan Balmain
- Glenn Merlino
- Kenna Shaw (NCI)
mTCGA Pilot Projects

- **Prostate cancer: Cory Abate-Shen**
  - Models are based on a tamoxifen-inducible conditional allele, Nkx3.1CreErT2 (Wang et al., Nature 2009) crossed with a Pten floxed allele (Hong Wu) crossed with an activateable K-ras allele (Tyler Jacks).
  - Tumors are induced by administration of Tamoxifen in adult mice.
  - Phenotype: The mutant mice develop preinvasive lesions (called PIN) that progress to invasive cancer and ultimately metastatic disease. By 4 months of age, these mice display 100% penetrance of adenocarcinoma as well as 30% incidence of distant metastases. Tumors are epithelial in origin.

- **Melanoma: Glenn Merlino**
  - Melanoma model is driven by constitutive c-MET signaling and neonatal UV irradiation, and accelerated by loss of one Ink4a allele.
  - Mean latency is about 3-4 months to melanoma. Premalignant "spots" can be seen earlier. Metastasis can be seen in about 20% of the mice.
mTCGA Pilot Projects

- **NSCLC: Christopher Kemp**
  - A single injection of urethane is given to pre-weanling mice and tumors resembling non small cell adenomas and adenocarcinomas arise after a long latency.
  - >75% of all induced tumors contain activating mutations in Kras.

- **Breast cancer: Muller, Green and Sharpless models**
  - **Green**: Overexpression of PyMT in mammary epithelium results in highly aggressive tumor formation and metastases.
  - **Sharpless**: Classical MMTV-Neu GEMs (in 100% FVB/n) treated for 2-8 months with lapatinib and now resistant. Both resistant and sensitive tumors in-hand.
  - **Muller**: MMTV/Activated ErbB2 IRES Cre transgenic where ErbB2 expression is coupled to Cre mediated excision of any conditional allele.
Clinical Proteomic Tumor Analysis Consortium

- 5 teams funded to perform proteomic work in relationship to genomic discoveries from programs like TCGA

- First projects to look at genome-proteome correlations in breast (Matthew Ellis), Ovarian (Dan Chan) and Colorectal (Dan Liebler)

- Will receive portions of TCGA-characterized cases from BCRs to perform proteomic analysis (shotgun or targeted)

- Identifying validation samples from similar cohorts, blood samples for next phases
Next Steps

- Thank you for attending!

- We value your feedback to improve the meeting…

- Please look for announcements on the 2nd Annual TCGA Scientific Symposium!!