Update on The Cancer Genome Atlas for the National Cancer Advisory Board

Mark Guyer

Director, Division of Extramural Research, NHGRI

on behalf of the TCGA Project Team

September 7, 2010
The TCGA Project Team

- NCI
  - Ann Barker
  - Joe Vockley
  - Kenna Shaw
  - Laura Dillon
  - Greg Eley
  - Carl Schaeffer
  - Martin Ferguson
  - Peter Fielding

- NHGRI
  - Brad Ozenberger
  - Jacqueline Palchik
  - Jane Peterson
  - Peter Good
  - Elizabeth Thomson
  - Julia Zhang

The Directors: Harold Varmus, Eric Green, Francis Collins
Review the last presentation to NCAB (Sept 2009)

Summarize overall organization of TCGA and the tumors being targeted (plus the ICGC targets)

Update on status of sample accrual and sequencing

Future sample accrual goals

Challenges

Approaches to challenges
Cancer is a disease of genomic alterations – identification of all genomic changes would enable defining cancer subtypes and generate a comprehensive set of alterations that characterize each type and subtype of cancer – potential to transform cancer drug discovery, diagnostics and prevention
Background

- Cancer is a disease of genomic alterations – identification of all genomic changes would enable defining cancer subtypes and generate a comprehensive set of alterations that characterize each type and subtype of cancer – potential to transform cancer drug discovery, diagnostics and prevention.

- Following several workshops and a specific recommendation by the National Cancer Advisory Board, TCGA was launched as a collaboration between the NCI and NHGRI in 2006.
Cancer is a disease of genomic alterations – identification of all genomic changes would enable defining cancer subtypes and generate a comprehensive set of alterations that characterize each type and subtype of cancer – potential to transform cancer drug discovery, diagnostics and prevention.

Following several workshops and a specific recommendation by the National Cancer Advisory Board, TCGA was launched as a collaboration between the NCI and NHGRI in 2006.

TCGA was initiated as a pilot designed to explore the processes needed to perform high-throughput, large scale disease-focused genome characterization, data integration and analysis:

- Biospecimens
- Large-scale genome characterization and sequencing
- Integration of data, laboratories and teams
- Policies (e.g. data standards, data access, informed consent)
High-throughput large-scale comprehensive characterization is feasible and produces useful data.
- High-throughput large-scale comprehensive characterization is feasible and produces useful data
- New cancer genes can be discovered
TCGA Pilot Program Conclusions

- High-throughput large-scale comprehensive characterization is feasible and produces useful data
- New cancer genes can be discovered
- Tumor subtypes can be differentiated based on comprehensive knowledge of genomic alterations
TCGA Pilot Program
Conclusions

- High-throughput large-scale comprehensive characterization is feasible and produces useful data
- New cancer genes can be discovered
- Tumor subtypes can be differentiated based on comprehensive knowledge of genomic alterations
- Integrated teams can be built – and it will take teams to analyze multi-dimensional data
TCGA Pilot Program
Conclusions

- High-throughput large-scale comprehensive characterization is feasible and produces useful data
- New cancer genes can be discovered
- Tumor subtypes can be differentiated based on comprehensive knowledge of genomic alterations
- Integrated teams can be built – and it will take teams to analyze multi-dimensional data
- This comprehensive approach can produce clinically relevant data
TCGA Pilot Program
Conclusions

- High-throughput large-scale comprehensive characterization is feasible and produces useful data
- New cancer genes can be discovered
- Tumor subtypes can be differentiated based on comprehensive knowledge of genomic alterations
- Integrated teams can be built – and it will take teams to analyze multi-dimensional data
- This comprehensive approach can produce clinically relevant data
- At the level of individual genes, cancer genomics is complex. At the level of pathways, more coherence can be observed.
ARRA funds will be employed for 2 years to collect tissues for years 1-5 of TCGA – and scale up the Biospecimen Core Resource capacity
TCGA Phase II: Overview

- ARRA funds will be employed for 2 years to collect tissues for years 1-5 of TCGA – and scale up the Biospecimen Core Resource capacity

- During the two years of ARRA funding – plan to complete comprehensive genome characterization of 10 tumor types (at 200 cases/tumor type as a discovery set and more depending on tumor type; 180 exomes; 20 whole genomes/tumor), and partial characterization of 10 more (at 100 cases/tumor type)
  - GCCs will perform expression, copy number, methylation and miRNA characterization
  - Genome Sequencing Centers will use Nex-Gen sequencing technologies – exomes and whole genomes (cost dependent)
ARRA funds will be employed for 2 years to collect tissues for years 1-5 of TCGA – and scale up the Biospecimen Core Resource capacity.

During the two years of ARRA funding – plan to complete comprehensive genome characterization of 10 tumor types (at 200 cases/tumor type as a discovery set and more depending on tumor type; 180 exomes; 20 whole genomes/tumor), and partial characterization of 10 more (at 100 cases/tumor)

- GCCs will perform expression, copy number, methylation and miRNA characterization
- Genome Sequencing Centers will use Nex-Gen sequencing technologies – exomes and whole genomes (cost dependent)

Genome Data Analysis Centers will integrate data from GCCs – GDAC-Bs will further integrate data, create new models and tools to refine and further add value to data for communities.
TCGA Project Pipeline

Supplementary Figure 1

Pathology QC
DNA & RNA Isolation, QC
Analysis
Sequencing
Expression, CNA & LOH, Epigenetics
Analysis
Data and Results Storage & QC
Integrative Analysis
GDAC

Comprehensive Knowledge of a Cancer Genome

= Process
----- = Data
-- = Results

= BCR
= GSCs
= CGCCs
= DCC
= GDACs
TCGA Project Pipeline

Supplementary Figure 1

Tissue Sample Pathology QC

DNA & RNA Isolation, QC

Sequencing

Expression, CNA & LOH, Epigenetics

Data and Results Storage & QC

Analysis

Integrative Analysis

Comprehensive Characterization of a Cancer Genome

GDAC

BCR = GSCs = CGCCs = DCC = GDACs

= Process

= Data

= Results
TCGA Tumor Projects

Projects with comprehensive data available

- Glioblastoma
- Ovarian

Projects in progress (partial data sets available)

- Acute Myeloid Leukemia
- Colon Adenocarcinoma
- Rectal Carcinoma
- Lung Adenocarcinoma
- Lung Squamous Cell Carcinoma
- Renal Clear Cell Carcinoma
TCGA Tumor Projects

Projects recently begun or upcoming

- Breast (multiple types)*
- Bladder
- Cervical
- Head and neck
- Liver
- Lymphoma
- Melanoma
- Multiple myeloma
- Pancreatic
- Prostate
- Sarcoma
- Stomach*
- Thyroid
- Uterine* (endometrial)
ICGC Projects (June, 2010)

- USA
  - TCGA projects
- Canada
  - pancreatic*
  - prostate
- Australia
  - pancreatic*
  - ovarian
- China
  - stomach
- EU/FRance
  - renal carcinomas
- EU/United Kingdom
  - breast cancers*
- France
  - breast cancers*
  - hepatic (alcohol-associated)
- Germany
  - pediatric brain cancers
- India
  - oral
- Italy
  - rare pancreatic types
- Japan
  - hepatic (virus-associated)
- Spain
  - chronic lymphocytic leukemia
- United Kingdom
  - breast cancers*
Goals for ARRA Period

DNA In/Data Out (Cumulative)

Number of Cases

Cumulative samples needed to enter GSCs
Cumulative samples sequenced-Goal
Cumulative samples received-Actual
Cumulative samples sequenced-Actual
Planned Data Submission
TCGA Sample Accrual (Sept 2010)

Samples Received/Qualified

[Bar chart showing the number of samples received and qualified for different types of cancer.]

- Breast
- Colon
- Kidney clear cell
- Endometrial
- Lung squamous
- Lung adenocarcinoma
- Head and neck
- Thyroid
- Rectal
- Stomach
- Cervical
- Melanoma
- LGG
- Liver
- AML
- Prostate
- Kidney papillary
- Sarcoma
- DLBCL
- Pancreas
GSC Planned Data Submission – 2010
*(does not include GBM, Ovarian)*
<table>
<thead>
<tr>
<th>Tumor Type</th>
<th>GCC assays</th>
<th>Whole Exomes</th>
<th>Whole Genomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ovarian</td>
<td>560</td>
<td>434</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>86 in progress</td>
<td>17 in progress</td>
</tr>
<tr>
<td>AML</td>
<td>162</td>
<td>15</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>39 in progress</td>
<td>135 in progress</td>
<td>29 in progress</td>
</tr>
<tr>
<td>Colon</td>
<td>103</td>
<td>52</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>41 in progress</td>
<td>51 in progress</td>
<td>0</td>
</tr>
<tr>
<td>Rectal</td>
<td>50</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>17 in progress</td>
<td>67 in progress</td>
<td>0</td>
</tr>
<tr>
<td>Breast ductal</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>233 in progress</td>
<td>186 in progress</td>
<td>0</td>
</tr>
<tr>
<td>Lung adeno</td>
<td>21</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>74 in progress</td>
<td>95 in progress</td>
<td>0</td>
</tr>
<tr>
<td>Lung scc</td>
<td>69</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>45 in progress</td>
<td>114 in progress</td>
<td>0</td>
</tr>
<tr>
<td>Endometrial</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>70 in progress</td>
<td>70 in progress</td>
<td>0</td>
</tr>
<tr>
<td>Renal</td>
<td>32</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>32 in progress</td>
<td>32 in progress</td>
<td>0</td>
</tr>
<tr>
<td>Gastric</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>82 in progress</td>
<td>82 in progress</td>
<td>0</td>
</tr>
</tbody>
</table>
http://cancergenome.nih.gov/dataportal
TCGA Sample Criteria (Pilot)

- Primary tumor only
- Snap frozen
- ~ 200 mg
- No more than 20% necrosis; ≥ 80% tumor cells
- Normal tissue: Blood (buffy coat/white cells); adjacent normal tissue or buccal cells; or ≥ 13µg high-quality DNA
- All “Tier One” Clinical Data Elements (15 or more)
- Treatment naïve
Challenges for TCGA Production

- Sample criteria
  - Tumors for which pre-treatment is standard of care
  - Tumors of lower purity
  - Use of adjacent tissue as “germ line” comparison
Challenges for TCGA Production

- Sample criteria
  - Tumors for which pre-treatment is standard of care
  - Tumors of lower purity
  - Use of adjacent tissue as “germ line” comparison

- Multiple concurrent projects

- Project length
GBM and ovarian pilot projects – samples were apportioned among the 3 GSCs
GBM and ovarian pilot projects – samples were apportioned among the 3 GSCs

AML and colon – designated primary and secondary GSCs; 10 exomes and 1 whole genome cases to be done by both centers

- AML – Wash U / Broad Institute
- Colon – Baylor / Wash U
- **GBM and ovarian pilot projects** – samples were apportioned among the 3 GSCs

- **AML and colon** – designated primary and secondary GSCs; 10 exomes and 1 whole genome cases to be done by both centers
  - AML – Wash U / Broad Institute
  - Colon – Baylor / Wash U

- **Subsequent projects** – to be assigned to a single GSC for all data generation, validation, and analysis; 10 exomes to be duplicated in a second GSC for QA
  - Broad Institute: lung adeno, lung squamous, gastric
  - Wash U: breast projects, endometrial
  - Baylor: rectal, kidney