Status Report: The Cancer Genome Atlas Pilot Project

National Cancer Advisory Board Meeting
February 6, 2007

Anna D. Barker, Ph.D.
Deputy Director, National Cancer Institute
The Cancer Genome Atlas (TCGA)

is a three-year pilot project of the National Cancer Institute and the National Human Genome Research Institute to increase our comprehensive understanding of the genetic basis of cancer.

It is anticipated that TCGA’s integrated database of molecular and clinical information will provide scientists unprecedented opportunities to discover and develop a new generation of targeted diagnostics, therapies, and preventives for cancer.
Enabling Rationale for TCGA

Achievements:
1. Human Genome Project
2. Gene families and pathways
3. Robust genomic analysis technologies
4. Sanger experience - sequenced known genes (e.g., kinases are druggable)
5. Number of early indications that somatic mutations are important potential targets

Overall Impact:
1. Known human sequence
2. Kinases, phosphatases, transcription factors, hormone responsiveness
3. Copy number changes, expression profiling, potentially epigenomic technologies
4. Survey of known genes that are abnormal prior to sequencing - BRAF
5. **BCR-ABL, EGFR1, ERBB2**
September 2003: NCAB Ad hoc Committee formed

April 2004: NCI-NHGRI Workshop

September 2004: Presentation to EC

February 2005:
- Ad hoc Committee Report to NCAB
- NCI-NHGRI Working Group formed

September 2005:
- NCAB Review
- EC Review
- Meeting for community input

January 2005: Project RFI issued

July 2005:
- NCI Issuance of Human Cancer Biospecimen RFI

December 2005:
- NHGRI Issuance of Sequencing Centers RFA
- Public Launch of Pilot Project

November 2005:
- Presentation to the BSA - Approval
- NCI Issuance of Human Cancer Biospecimen RFI
**TCGA Development Milestones**

**Fiscal Year 2006**

- **March 2006:**
  - NCI Issuance of Cancer Genome Characterization Centers RFA

- **May 2006:**
  - Data Release Workshop

- **September 2006:**
  - Selection of tumor types
  - NCI Funding of Biospecimen Core Resource

- **February 2006:**
  - NCI Issuance of Biospecimen Core Resource RFP

- **July / August 2006:**
  - Site visits to Biorepositories

- **October / November 2006:**
  - Funding of Data Coordinating Center
  - NCI Funding of Cancer Genome Characterization Centers
  - NHGRI Funding of Genome Sequencing Centers

- **December 2006:**
  - First TCGA Steering Committee Meeting – all PIs
February / March 2007:
• TCGA Dry Run

March 2007:
• Biospecimen Core Resource distributes GBM, lung, and ovarian tumor biomolecules to GSCs and CGCCs

March / April 2007:
• GSCs and CGCCs perform sequencing and characterization of tumor biomolecules

May 2007:
• First set of TCGA data is deposited into public databases managed by DCC
TCGA: How it Works
Central to the Success of TCGA Pilot, the BCR is:

- Verifying all biologic and clinical data and performing the pathologic QC of qualified tumors from selected existing collections
- Performing central processing of specimens to provide uniform biomolecules and distributing to both genome characterization and sequencing centers
- Tracking and quality assuring all specimen-related operations (consent, acquisition, transport, processing, QC, distribution)
- Providing “standard” samples for technology platform comparisons
- Developing (with the Office of Biorepositories and Biospecimen Research) and monitor the SOPs for prospective specimen collection
- Serving as a member of TCGA’s Steering Committee
Technology platforms for high-throughput genome characterization:
- Expression profiling
- Copy number changes
- DNA methylation (epigenomics)

Improve existing technologies:
- Epigenomics to meet required throughput rate
- Copy number detection and expression profiling for characterizing small amount of biological samples

Real-time data release into public database
High-throughput Genome Sequencing Centers (NHGRI):

- Sequence large number of targets from three tumor types
- Develop and integrate sequencing technologies
Platform for data collection and management

- Track data produced by components of TCGA
- Ensure that data meets quality standards set for TCGA
- Make TCGA data publicly accessible through databases supported by NCI’s Cancer Biomedical Informatics Grid™ (caBIG™) and the National Library of Medicine’s National Center for Biotechnology Information
- Scientists will have access to TCGA data to generate new insights into causes and potential targets for interventions
- Access to all TCGA data will be provided in a manner that meets the highest standards for protection and respect of the research participants
TCGA Components

- **International Genomics Consortium**
  - Phoenix, Ariz.

- **MD Anderson Cancer Center**
  - (glioblastoma tumor biospecimens)
  - Houston, Texas

- **Gynecologic Oncology Group Tissue Bank**
  - Children's Hospital, Ohio State University
  - (ovarian tumor biospecimens)
  - Columbus, Ohio

- **Lung Cancer Tissue Bank of the Cancer and Leukemia Group B (CALGB)**
  - Brigham and Women's Hospital
  - (lung cancer biospecimens)
  - Boston, Mass.

Human Cancer Biospecimen Core Resource (BCR) and Biorepositories Providing Samples
TCGA Components

- Broad Institute of MIT and Harvard
  Cambridge, Mass.
- Dana-Farber Cancer Institute
  Boston, Mass.
- Harvard Medical School
  and Brigham and Women’s Hospital
  Boston, Mass.
- Lung Cancer Tissue Bank of
  the Cancer and Leukemia
  Group B (CALGB),
  Brigham and Women’s Hospital
  Boston, Mass.
- Memorial Sloan-Kettering
  Cancer Center
  New York, N.Y.
- The Sidney Kimmel
  Comprehensive Cancer
  Center at Johns Hopkins
  Baltimore, Md.
- University of North Carolina Lineberger
  Comprehensive Cancer Center
  Chapel Hill, N.C.
A patient donates tumor tissue for study.

Genome Sequencing Centers

Scientists study genetic material in the tumor tissue.

Genome Sequencing Centers

To find the genetic signatures for cancer, scientists study genetic material in many patients' tissues.

Human Cancer Biospecimen Core Resource

Scientists study genetic material in the tumor tissue.

Cancer Genome Characterization Centers

Scientists analyze the data produced by TCGA, and develop a web-based information database.

Data Management, Bioinformatics, and Computational Analysis

Scientists use information in the database to speed research advances.

Research results are translated into new products to help patients.

Technology Development

TCGA Components – Procurement Process Completed in 2006

Patient Community

TGCA Network

Research & Medical Communities
These three cancers collectively account for more than 210,000 cancer cases each year in the United States.
Glioblastoma is Ideal for TCGA Study

- A "homogenous" tumor – or about as good as it gets
  - Single grade (highest grade) of a single histological type of cancer

- Few other cell types, such as stromal cells or inflammatory cells, that might contribute extraneous, non-tumor DNA to the extracted biomolecules
To date, glioblastomas seem to have the greatest number of genetic changes of all astrocyte-based cancers.

A recent study systematically sequenced all tyrosine kinase genes in glioblastomas, confirming the presence of mutations in these and other genes involved in proliferation pathways.

There are many proliferation pathway-targeted drugs in clinical trials, underlining the role of genetic complexity in glioblastoma.

Target Selection

- Identification of known cancer genes
- Integration of all cancer gene data bases
- Selection of a small number of genes (~800-1,000) to begin sequencing in glioblastoma samples
- Meeting with GBM experts to discuss strategy for identification of new genes from TCGA the CGCCs
- Meeting with participants in NCI’s other programs to begin data interrogation processes
Key Issues:

- How detailed should informed consent be – more information vs. less? TCGA’s informed consent is lengthy and detailed.
- Who should have access to data?
- How do we leverage and capitalize on potential for progress against disease and ensure privacy protection? TCGA will provide two levels of data access – one completely open; the other password controlled.
- Solving the issues of data access vs. patient protection will likely require genetic privacy legislation.
TCGA’s Informed Consent Approach

- Permission for detailed genomic research
- Permission for broad future research use of samples and health information
- Permission to place genomic and health information in widely accessible databases – with limitations
- Risks associated with loss of privacy
- Potential benefits for future cancer patients
- Issues related to withdrawal (data and samples)
Some Success Factors for TCGA

Three-Year Time Horizon

- Completion of genomic analysis of three tumors, hopefully leading to identification of new genes involved in these cancers
- Ability to find and identify specific genomic alterations in genes associated with cancer
- Ability to differentiate tumor subtypes based on genomic alterations
- Establishment of a genomics database that scientists can access – new questions – new research
TCGA’s Potential Impact

- Identification of somatic changes in cancer genomes that could establish the molecular basis for each cancer – and inform and enable a new era of molecular oncology

- A molecular taxonomy of cancer

- New molecular targets for diagnostics, therapeutics, and preventives

- Improved ability to stratify patients for clinical trials
Updates on TCGA website:

- Information for patients, scientists, clinicians, policymakers, and the public

Coming Soon:
Sign up at the TCGA website to receive automatic updates and event news

http://cancergenome.nih.gov