The Future of Low-Dose Radiation Research in the United States

“Models for Coordinated Research: Lessons Learned from Large-Scale Biology Initiatives”

Two Initiatives: TCGA and ISPY 2

Anna (Ann) D. Barker, Ph.D.
Co-Director, Complex Adaptive Systems Initiative
Director, Transformative Healthcare Knowledge Networks
Director, National Biomarker Development Alliance
Professor, School of Life Sciences
Arizona State University
Large Scale Coordinated (Collaborative) Biology Program Models

- Driver(s) for a specific initiative, rationale for organization and funding sources
- Setting goals and agendas – infrastructure, operations, communications and “products”
- Assessing progress, insuring outcomes and impact
- Lessons learned and potential applications in the low-dose radiation research
The Cancer Genome Atlas (TCGA)

“A large-scale coordinated/collaborative cancer genome sequencing initiative to identify all of the significant genomic alterations in the major types of cancer”

REALITIES THAT “SET THE STAGE” FOR TCGA
In 2005 (When TCGA was launched), there were an estimated 1,372,910 of new cancer diagnoses and 570,280 deaths from cancer in the United States.

In 2019, it is projected that there will be an estimated 1,762,450 new cancer diagnoses and 606,880 deaths from cancer in the U.S.

Cancer risk increases significantly beyond age 50, and half of all cancers occur at age 66 and above (10,000 baby boomers turn 65 every day).

Estimated cost of cancer care in the U.S. for 2017 was $147.3 billion in direct medical costs.

Worldwide cancer predicted to increase by 70% by 2030. There were 14 million cases and 8.2 million deaths in 2012; 21.7 million cases and 13 million deaths/year projected by 2030.

National Cancer Institute (NCI), American Cancer Society (ACS), Facts and Figures, 2019
World Health Organization (WHO)
The “Omics Revolution” Became Real in 2003-04 with Completion of the Human Genome Sequence

- Since 2003: Sequencing costs fell from ~$1.0 million in 2001 to a promised $100 per genome
- Estimated well over a million human genomes have been sequenced to date (difficult to estimate) – could approach 1000 PB per year at scale
How Did We Prepare for TCGA?

- Planning began in 2003 – Two large workshops for input – April, 2004; July 2005 – numerous presentations to NCI’s Boards (Major input from the affected Scientific communities)
- September, 2003 – Ad Hoc Group of the NCI’s National Cancer Advisory Board (NCAB) undertook broad study to determine what areas of science/technology would be critical in accelerating progress against cancer
- NCAB Ad Hoc Group recommended TCGA as a critical strategic project in its report of February, 2005 (Hartwell-Lander Report)
- NCI-NHGRI Program Work Group formed in February 2005
- Built on prior NHGRI/NCI experience with large scale initiatives – HGP, CGAP, MGC, HapMap, ENCODE, KOMP, TSP, etc.
- RFAs and RFPs issued in 2005 – Awards made and TCGA launched in late 2006
- Three-year $100 million dollar pilot
Goals of the TCGA Pilot

- To develop, deploy and connect a high quality biospecimen resource with genome characterization, sequencing and bioinformatics centers into a network with unprecedented capacity to sequence cancer genomes, collect and integrate the data and make it publicly available.

- To define all relevant genomic changes in three tumors through genome characterization and re-sequencing.

- To create and deploy a pre-competitive, integrated public TCGA database of all of the various genome characterization, sequence and clinical data for the three tumors being studied.
The TCGA Network (Organizational Steps)

- Organized the NCI-NHGRI Project Team – 2005 (Met every week for project duration)
- All components selected late in 2006
- Set up the TCGA Steering Committee (SC) – added physicians contributing biospecimens 2007
- Organized Working Groups (Biospecimens; Production; Informatics; and Analysis; – recently Disease Work Groups, Biospecimen Access Work Groups and Publication Working Groups - 2007
- Regular meeting schedules – SC meetings every 2 weeks
- 2X per year Network Meetings
TCGA: Organized in Two Phases: A Pilot Project Followed by a Large-Scale Project to Sequence Most Cancer Types

**TCGA Pilot – Phase I:** A three-year pilot project of the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI) to identify the major genomic changes associated with three types of cancer

- The Pilot phase tested the feasibility of a systematic large-scale approach to identifying the genomic alterations in glioblastoma, ovarian cancer and lung cancer.

- The Pilot phase established the project infrastructure and organization including the data collections and analytics and established the procedures to ensure that the data was public as soon as possible.

**TCGA – Phase II:** Results from the TCGA Pilot were used to design and scale up the follow-on large scale TCGA project that operated from 2009-2018 – exceeding goals and providing an unprecedented array of cancer specific data sets and a number of seminal findings beyond the original goals.
Tumors Selected for Study in the Pilot Project

- Brain (glioblastoma)
- Lung, squamous cell
- Ovarian, serous

- These three cancers collectively account for more than 210,000 cancer cases each year in the United States.

- Repositories identified containing specimens meeting TCGA’s strict scientific, technical, and ethical requirements.
TCGA Infrastructure and Organization

**TCGA Teams**
- **Data Coordinating Center, DCC**
- Analyses of data
- **Genome Sequencing Centers**
  - High throughput sequencing of genes and genomic regions identified through cancer characterization
- **Cancer Genome Characterization Centers**
  - Identification of expression alternation
  - Detection of DNA fragment copy number changes and LOH
  - Epigenetics
- **Human Cancer Biospecimen Core Resource**
  - Biospecimens-related data storage
  - Histopathology confirmation performed
  - Biomolecules isolated, QC’ed and distributed

**New Analytics Platforms**
- **Tools**
- **Views**

**Technology Development**
- Increased sensitivity of molecular characterization platforms
- Analysis of biomolecules from 1000 cells or less

**Management and Steering Committees**

**NCI and NHGRI Funded/Managed Programs**
- **NCI Funded/Co-Managed**
- **NHGRI Funded/Managed**
- NCI (ARRA Funding)

**NCI and NHGRI Funded/Managed Programs**
- NCI and NHGRI Funded/Managed Programs
- **New Analytics Platforms**
  - Development of New Analyses
  - New Analytics Platforms
  - Technology Development

**New Analytics Platforms**
- Development of New Analyses
  - Tools
  - Views

**NCI and NHGRI TCGA Teams**
- NCI and NHGRI TCGA Teams
- NCI and NHGRI Funded/Managed Programs
- NCI Funded/Co-Managed
- NHGRI Funded/Managed
- NCI (ARRA Funding)
Phase 1: TCGA Pilot Project
TCGA Project “Pipeline”

Tissue Sample

Pathology QC

DNA & RNA Isolation, QC

Analysis

Sequencing

Expression, CNA & LOH, Epigenetics

Data and Results Storage & QC

Integrative Analysis

Comprehensive Integrated Molecular Profile of a Cancer

= Process

= Data

= Results

= Collaborators

BCR = BCR
GSCs = GSCs
CGCCs = CGCCs
DCC = DCC
TCGA: Measured and Integrated All Major Genomic Alterations/Cancer Type

Gene Splicing Alterations

Aneuploidy; Re-arrangement; Translocation

Copy number aberrations

Somatic mutations

Methylation or histone modification

Altered expression

Each Sample Clinical Data

Adapted from Ron DePinho
Three forms of cancer

- glioblastoma multiforme (brain)
- squamous carcinoma (lung)
- serous cystadenocarcinoma (ovarian)

Multiple data types

- Clinical diagnosis
- Treatment history
- Histologic diagnosis
- Pathologic status
- Tissue anatomic site
- Surgical history
- Gene expression
- Chromosomal copy number
- Loss of heterozygosity
- Methylation patterns
- miRNA expression
- DNA sequence

TCGA: Connecting multiple sources, experiments, and data types
Ovarian Cancer Analysis Team

Copy Number
Gaddy Getz
- Adam Olshen
- Barry Taylor
- Chad Creighton
- Devlin Absher
- Henrik Bengtsson
- Jun Li
- Nick Gauthier
- Peter Park
- Ronglai Shen
- Scott Morris

Coordination
Paul Spellman
- Julia Zhang/NCI Staff

Mutation Detection and Significance
Li Ding
- Gaddy Getz
- Kristian Cibulskis
- Larry Donehower
- Rachel Karchin
- Gavin Sherlock
- Jinghui Zhang
- Dave Larson
- Li Ding
- Carrie Sougnez
- Mike Wendl
- Hannah Carter
- Boris Reva
- Anil Sood
- Dan Koboldt

Methylation
Peter Laird
- Dan Weisenberger
- Mike Lawrence
- Dave Larson
- Xiaoqi Shi
- Houtan Noushmal
- Pierre Neuvial

Pathways
Chris Sander
- Niki Schultz
- Rachel Karchin
- Mike Lawrence
- Li Ding
- Svetlana Tyekucheva
- Yonghong Xiao
- Ethem Ceramic
- Larry Donehower
- Mike Wendl
- Lincoln Stein
- Wendy Winckler
- Chad Creighton

Expression
Roel Verhaak
- Katie Hoadley
- Elizabeth Purdom
- Dan Weisenberger
- Nick Socci
- Hailei Zhang
- Ronglai Shen
- Xiaoqi Shi
- Neil Hayes
- Nick Gauthier
- Quanyan Zhang

miRNAs
Neil Hayes
- Dave Wheeler
- Todd Wylie
- Robert Sheridan
- Doug Levine
- Laura Heiser
- Shaoqi Ming
- Anil Sood
- Dan Koboldt
- Preethi Gunaratne

Whole Genome Analysis
Elaine Mardis
- Ben Raphael
- Kristian Cibulskis
- David Wheeler
- Houtan Noushmal

~ 95-100 Scientists per Team per cancer type
TCGA: Glioblastoma Multiforme (Adult Brain Tumors) Subtypes and Molecular Pathways

Copy Number

- Over 100 regions of focal gain and loss
- Over 20 chromosome arms recurrently gained or lost

TCGA – Ovarian Cancer
TCGA Data Portal Created: Search and Analysis

- Platform for researchers to search, download, and analyze data sets generated by TCGA.

- Contains all TCGA data: Clinical information associated with cancer tumors and human subjects, genomic characterization, and high-throughput sequencing analysis of the tumor genomes.

- Tools for integrative analysis

- New data and tool update alerts sent directly to registered users.
TCGA Phase I Pilot: Examples of Outcomes

- Built Infrastructure to obtain specimens, sequence genomes, and produce data
- Tested and developed technologies and tools to address the complexity of cancer biology
- Demonstrated proof of concept for large-scale, integrated approach to sequencing and analyzing the cancer genome
- Evolved solutions for obtaining high-quality, heavily annotated biospecimens necessary for genomic research
- Developed the capacity to review each sample for accuracy of pathology and classify
- Created a networked cancer clinical trials system that allows access to biospecimens
- Built systems for large data collection with unprecedented integrative power
- Developed the bioinformatics tools needed to process and analyze the data for onward mining by academia and industry to develop the next molecular targeted therapies
Lessons Learned from TCGA – Phase 1

Biospecimen Core Resource (BCR)

Tissue Source Sites (TSS)
Surgical Pathology Reports Review
Clinical Records Review

Histopathology lab
Frozen section review for % tumor and necrosis
Molecular analyte isolation and QC

~1/3 Fail for pathology QC
~1/3 Fail for MA QC

Data Standardization
Attempt sample rescue
Fail for pathology
Fail for quantity and/or quality of nucleic acids
~1/3 Pass

Clinical & molecular data submitted to DCC
Molecular analytes sent to CGCCs and GSCs for analysis

Path for Qualifying Biospecimens
Path for Failing Biospecimens
Accrual Site
BCR Process
Path Disqualified
Overall Lessons Learned/Perspectives from the Phase I Pilot Phase of TCGA

- This was team science and “big science” requiring integrating 2-3 different cultures—and was not intuitive—it was not easy but we built those teams!
- TCGA was not just about science, there were policy needs - IP, data access, a complex informed consent—and more
- Obtaining the highest quality samples (stringent criteria) was difficult—but
- The depth and breadth of analysis was unprecedented—we needed to build a new generation of analysis teams
- We needed to communicate constantly about TCGA—the project affected nearly all cancer researchers—and everyone had an opinion!
- In terms of science—signal could be differentiated from noise and new cancer genes could be discovered
- Tumor subtypes could be identified based on comprehensive identification and integration of genomic alterations
- High-throughput large-scale identification of molecular alterations for cancer was possible and prerequisite to defining associated biologic effects
TCGA Phase II Organization

Genome Characterization Centers
- mRNA - Univ. North Carolina
- miRNA Univ. British Columbia
- Genome Copy Number, Harvard
- Methylation, Univ. Southern Cali.
- Adv. Genomics; Harvard, Baylor

Genome Sequencing Centers
- Broad Institute
- Washington University
- Baylor University

Data Coordinating Center

Genome Data Analysis Centers
- Broad Institute, Institute for Systems Biology, MD
- Anderson Cancer Center, Lawrence Berkeley Nat’l Lab., Memorial Sloan Kettering Cancer Center, Univ. California, Santa Cruz, Univ. North Carolina

Public Data Portal
http://cancergenome.nih.gov/dataportal/
TCGA: Phase II

20 Institutions
- 2 biospecimen Core Resources
- Cancer Genome Characterization Centers
- Genome Characterization Centers
- Genome Sequencing Centers
- Data Analysis Centers
- Data Coordination Center
~ 50 Principal Investigators
~ 300 total Scientists, technicians and administrative staff
Flow of Data Within TCGA (Enter ARRA)

**Genome Characterization Centers**

**FIREHOSE**
Broad Institute

**Genome Data Analysis Centers**
- Broad Institute, Cambridge, Mass.
- Institute for Systems Biology, Seattle, Wash.
- University of Texas/M.D. Anderson Cancer Center, Houston, Texas
- Lawrence Berkeley National Laboratory, Berkeley, Calif.
- Memorial Sloan-Kettering Cancer Center, New York, N.Y.
- University of California at Santa Cruz, Calif.
- University of North Carolina, Chapel Hill

**Data Coordinating Center**

**Public Data Portal and Browsers**
http://cancergenome.nih.gov/dataportal/
Outcomes (By the Numbers)

- Tumor tissues provided from over 11,000 patients
- Molecularly characterized 20,000 primary cancer and matched normals
- Competed this level of “omics” characterization for 33 tumor types – including 10 rare tumors (original goal was 20 tumor types)
- Generated ~2.5 petabytes of data (including genomic, transcriptomic, proteomic and epigenomic data)
- Generated numerous publications – and will likely become one of the most cited initiative in cancer research publications
- All TCGA publicly available through Genomic Data Commons
- Engaged 20 institutions – and ~ 300 scientists, technicians, administrators, government management team and cancer survivors
Outcomes (How about Patient Benefit)

Increased our fundamental understanding of cancer by identifying and reporting aberrations in DNA, transcriptome, gene expression, MiRNA’s, methylation, protein expression, methylation - first “pan-cancer” “atlas” based on cell of origin and pathway biology – with significant clinical data

- Organized these findings into functional pathways
- Made specific discoveries accounting for clinical observations/effects in patients
- Revolutionized how cancer is classified – i.e., molecular subtypes (dominated by cell of origin)
- Provided high value targets for both biomarker and therapeutics discovery and development
- Provided unprecedented opportunities for clinical research –innovative clinical trials
Outcomes (The Unexpected/Surprises/Intangibles)

**TCGA:**

- Drove and accelerated progress in sequencing technologies
- Was the “vanguard” of the “big data revolution” in cancer
- Created a new generation of hypothesis (RO1)-driven research
- Enabled a number of significant discoveries that benefitted patients
- Established a model of coordinated/collaborative research showing that large-scale and small-scale science not only works – but is necessary to develop precision medicine
- Scientists can put aside egos and individual rewards – and the results of these teams can not be achieved by the lone scientist
- Now everyone “owns” TCGA – designed it – saw that it was funded and fought for its’ initiation and continuity – the definition of success
TCGA Opened a New Chapter in Cancer: An Era of Big DATA

TCGA

• Sanger Institute
• ICGC
• International Sequencing Projects

Data - Data
Data - Data
Data – Data
(Issues: Public and Private – Quality, Access, Consents, Etc.)

Pharma/Biotech
Individual investigators

Hospitals
NCI-Designated Cancer Centers – Academic Medical Centers

ASU Complex Adaptive Systems Initiative
Arizona State University
The ISPY 2 Trial

A large scale academic – government-private sector adaptive platform trial in early high risk breast cancer
Impetus for the ISPY 2 Trial

Time and attrition are both directly related to a lack of understanding of the impact of heterogeneity of cancer from different patients ostensibly with the same disease.
Launched in 2010, ISPY 2 is a transformative adaptive Bayesian driven platform trial (performed under a Master protocol) in early high risk breast cancer. Agents (including drugs, biologics and combinations) are tested in subsets of patients based on biomarker signatures, with pathologic complete response used as the surrogate endpoint. Successful agents (returned to the respective companies) receive a probability estimate of their anticipated success in a small phase 3 trial. The ISPY 2 trial team (~ 100 physicians and support personnel) has successfully tested 23 agents and graduated 7 – and through their efforts, the FDA approved PCR as a surrogate endpoint in 2014.
What are Master Protocols

**Master Protocol**, as defined by the FDA*: “a master protocol is defined as a protocol designed with multiple sub-studies, which may have different objectives and involves coordinated efforts to evaluate one or more investigational drugs in one or more disease subtypes within the overall trial structure”. Which means: “efficiently answering multiple questions faster under single trial structure”.

**Platform Trial**: Multiple therapies evaluated in a specific disease – depending on the design, agents enter and leave (graduate) the trial based on a specific algorithms and rules.

**Adaptive Platform Trials** are designed (generally based on biomarker-differentiated subsets) to evaluate patient outcomes on an ongoing basis and modify trial parameters (including the protocol) in response to the emerging data. The ability to adapt within the trial enables dropping unsuccessful therapies, graduating successful therapies and adding new agents to the trial. Many of these trials are designed using Bayesian statistical approaches.
Adaptive Platform Trials

Patient Population

Stratify Based on Biomarkers

Adaptive Randomization

Control Arm

Experimental Arm A

Experimental Arm B

Experimental Arm C

Experimental Arm D

Interim Analysis

Determine Efficacy of Treatments in Different Biomarker Signatures

Continue Trial, Adjust Adaptive Randomization So Patients Are More Likely to Be Assigned to Effective Arms
ISPY 2: An Adaptive Platform Trial that Collects RWD to Enrich Trial Data

ISPY2 PLATFORM TRIAL
Biomarker-rich protocol

SUBTYPE

SCREENING

ADAPTIVE RANDOMIZATION

RCB

MRIBiopsyBlood Draw

Paclitaxel* (control arm)
12 weekly cycles

Invest. Agent† 1 ± Paclitaxel*
12 weekly cycles

Invest. Agent† 5 ± Paclitaxel*
12 weekly cycles

Anthacycline (AC)
4 cycles

Anthacycline (AC)
4 cycles

Anthacycline (AC)
4 cycles

T0HR+, HER+, MP+

T1

T2

T3

T5

Surgery

pCR PRIMARY ENDPOINT

* Patients who are HER2+ may also receive tustuzumab (Herceptin)
† An investigational combination of one or more agents may be used to replace all or some of the standard therapy

Attribution, Laura Essermam
Confidential 8.5.2017
I-SPY 2 TRIAL (Organized via a Large Team (FNIH Biomarker Consortium and the FDA) – Launched in 2009

FNIH Original Management Structure
(Now Managed by Quantum Leap Healthcare Collaborative)
Current Status: The I-SPY Network

- Accrues and rapidly tests new agents from the private through a rapid review and activation process (4 months)
- Significant focus on biomarker discovery, qualification, refinement as part of the trial
- Maintains a robust pre-competitive consortium
- ISPY 2 is an unparalleled clinical evidence generation system
- A learning system - over 1000 patients treated in 7 years
- Over 20 agents tested to date – 7 have graduated to qualify for a small phase 3 trial
- The trial is evolving standard of care (e.g., lowering toxicities) – which is translated to patients
- The trial is an incubator for tools that integrate research and care
- The network (ecosystem) is a culture of innovation focused on accelerating knowledge turns

Slide from L. Esserman
ISPY 2: Lessons Learned

- Start with a “coalition of the willing” – need to “crowdsource needed knowledge”
- Platform trials are potentially transformative
- All phases of the planning, design and operations requires large teams – give credit
- Whatever your operating plan, ensure that it can comprehensively manage the construct
- Plan for change – change in science – change in standard of care – etc
- Build models that can address multiple goals (e.g., validate biomarkers)
- It’s a learning system – learn how to use it and open it up
- Engage and convince the industry – form precompetitive coalitions
We have lots of data “tombs” in oncology – driven by “protectionism” – doesn’t seem to be the case in low-dose radiation research – there is power in connecting
If the low-dose radiation research communities believe that we need more research, a lot more and better data, more or fewer regulations, more or less enforcement- is there something that could/should be done – together?

If so, then create a national or international coordinated/collaborative network to define goals – create or mimic a model – maybe start with a pilot – but it really matters - just --

START