

Precision Medicine and Cancer Informatics

Warren Kibbe, PhD

NCI Center for Biomedical Informatics

Outline

■ Background

- Attributes require for scalable precision medicine informatics
- How we got here: Lessons from TCGA
- Where we are: Role of the NCI Cancer Genomic Data Commons (GDC) & Cloud Pilots
- Where we are going: Information problems we intend to solve with the Precision Medicine Initiative for Oncology

Some Basic Ingredients for Precision Medicine Big Data

- **Open Science.** Supporting **Open Access, Open Data, Open Source,** and **Data Liquidity** for the cancer community
- **Standardization** through CDEs and Case Report Forms
- **Interoperability** by exposing existing knowledge through **appropriate integration of ontologies, vocabularies and taxonomies**
- **Sustainable models** for informatics infrastructure, services, data, metadata, curation

The Cancer Genome Atlas

A comprehensive effort to accelerate our understanding of the molecular basis of cancer

TCGA: The Cancer Genome Atlas

- Launched in 2006 by NCI & NHGRI
- Complete characterization of ~35 adult cancers
 - ~20 common cancers at 500 cases each
 - ~15 rare cancers at 50-150 cases each
- Copy Number, Gene Expression, Methylation, DNA Sequencing (WGS/WXS), Clinical data
 - ~11,000 cases
- Project ending in 2016
 - Future projects to use the TCGA infrastructure
 - Exceptional Responders, ALChEMIST, Clinical Trial Sequencing Program (CTSP), Cancer Driver Discovery Program (CDDP)
- <http://cancergenome.nih.gov/>



National Cancer Institute

National Human Genome Research Institute



The Cancer Genome Atlas
Data Portal

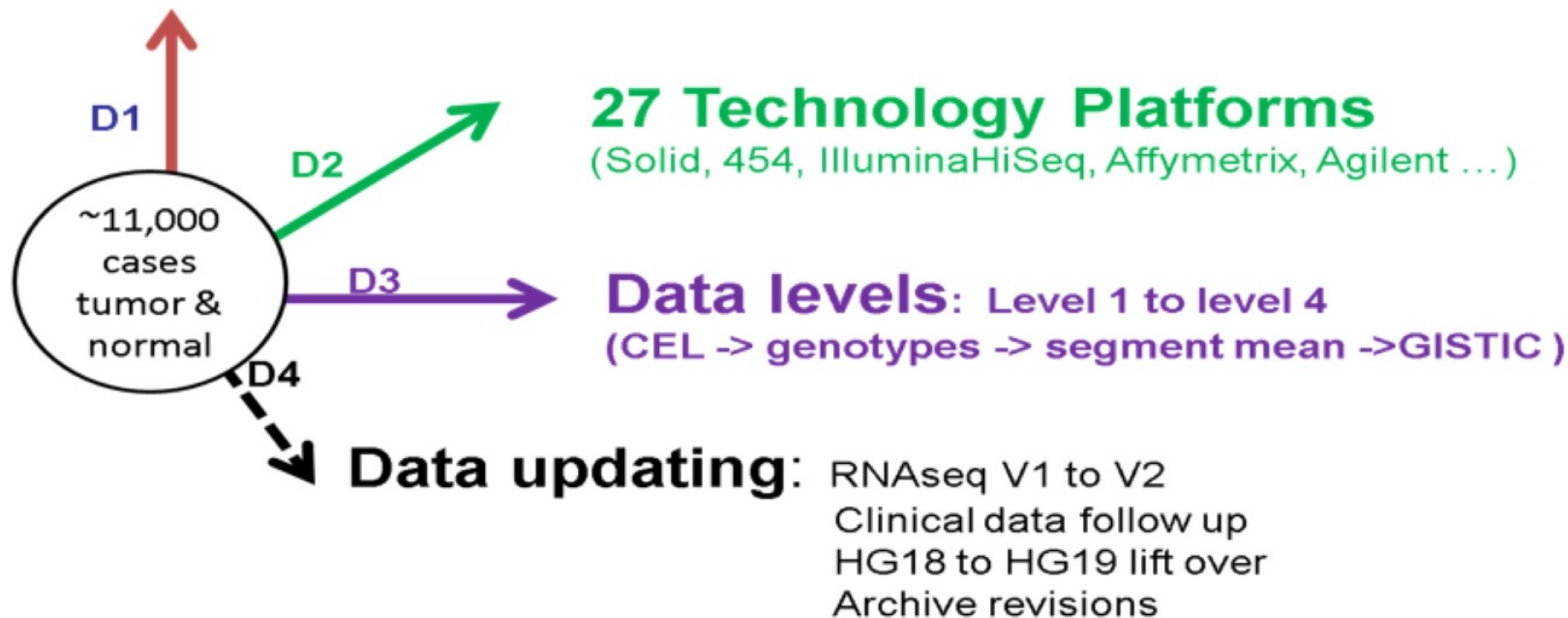


Understanding genomics
to improve cancer care

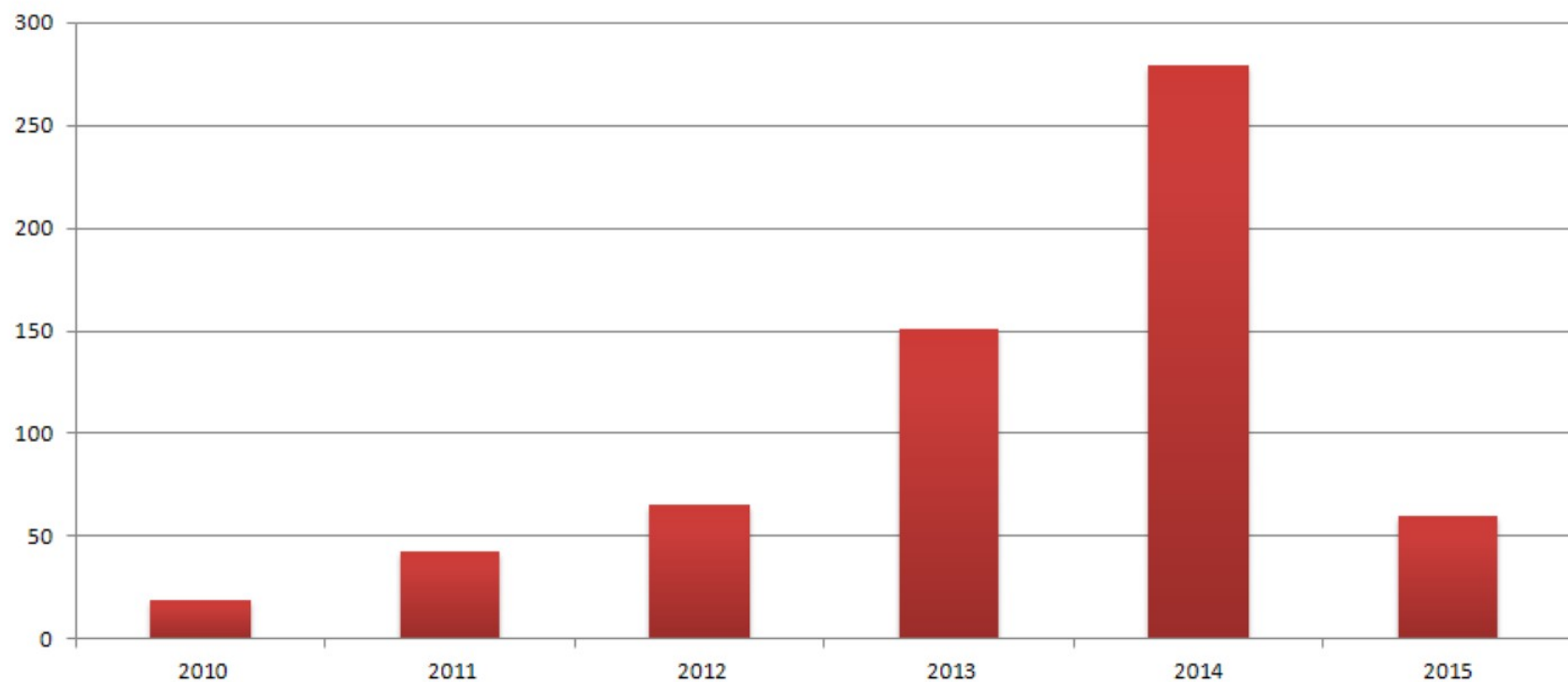
[TCGA Home](#) | [Contact Us](#) | [For the Media](#)

Dimensions of TCGA Data

12 Data types: Expression, Methylation, DNaseq, RNAseq ...



TCGA Publications since 2010



Cancer study reveals powerful new system for classifying tumors

One in ten cancers were reclassified in clinically meaningful ways based on molecular subtypes identified by a comprehensive analysis of data from thousands of patients

August 07, 2014

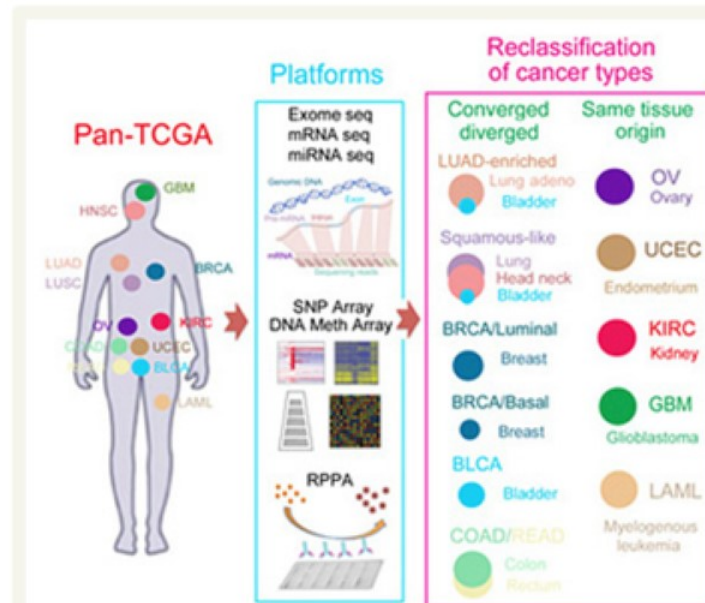
By [Tim Stephens](#)

SHARE THIS STORY: [t](#) [f](#) [g+](#) [in](#) [v](#)

Example publication from TCGA

Cancers are classified primarily on the basis of where in the body the disease originates, as in lung cancer or breast cancer. According to a new study, however, one in ten cancer patients would be classified differently using a new classification system based on molecular subtypes instead of the current tissue-of-origin system. This reclassification could lead to different therapeutic options for those patients, scientists reported in a paper published August 7 in *Cell*.

"It's only ten percent that were classified differently, but it matters a lot if you're one of those patients,"



The Genomic Data Commons

Facilitating the identification of molecular subtypes of cancer and potential drug targets

NCI Cancer Genomic Data Commons (GDC)



Genomic Data Commons (GDC) – Rationale

- TCGA and many other NCI funded cancer genomics projects each currently have their own DCC
 - BAM data and results stored in many different repositories; confusing to users, inefficient, barrier to research
- GDC will be a single repository for all NCI cancer genomics data
 - Will include new, upcoming NCI cancer genomics efforts
 - Store all data including BAMs
 - Harmonize the data as appropriate
 - Realignment to newest human genome standard
 - Recall all variants using a standard calling method
 - Will be the authoritative reference data set

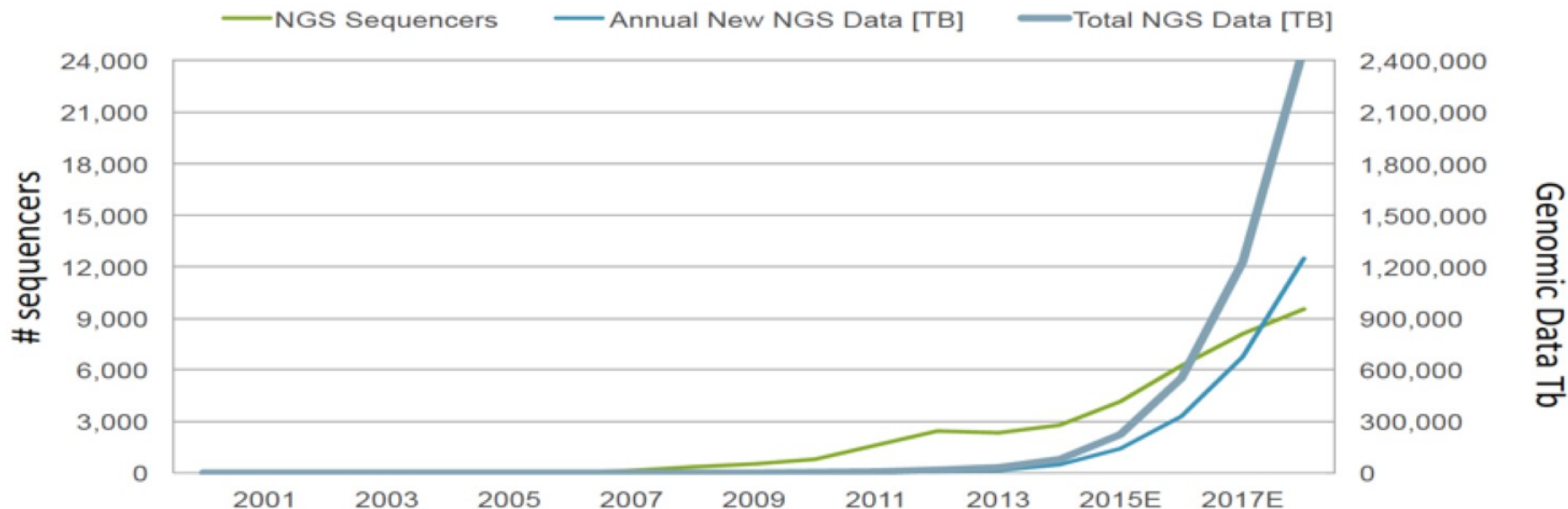
Genomic Data Commons (GDC)

- First step towards development of a knowledge system for cancer
- Foundation for a genomic precision medicine platform
- Project initiated Spring of 2014
 - Contract awarded to University of Chicago
 - PI: Dr. Robert Grossman
 - Go live date: Late Spring 2016
 - Not a commercial cloud
- Data will be freely available for download subject to data access requirements

The NCI Cancer Genomics Cloud Pilots

*Understanding how to meet the research
community's need to analyze large-scale cancer
genomic and clinical data*

Amount of genomic data will exceed available resources



Between 2014-2018 production of new NGS data to exceed **2 Exabytes**

NGS: Next Generation Sequencing

NGS sequencers include machines from Illumina, Life Technologies, and Pacific Biosciences. Human genome data based on estimates of whole human genomes sequenced

Sources: Financial reports of Illumina, Life Technologies, Pacific Biosciences; revenue guidances; JP Morgan; The Economist; Seven Bridges Analysis.

NCI Cloud Pilots

The Broad

PI: Gad Getz

Institute for Systems Biology

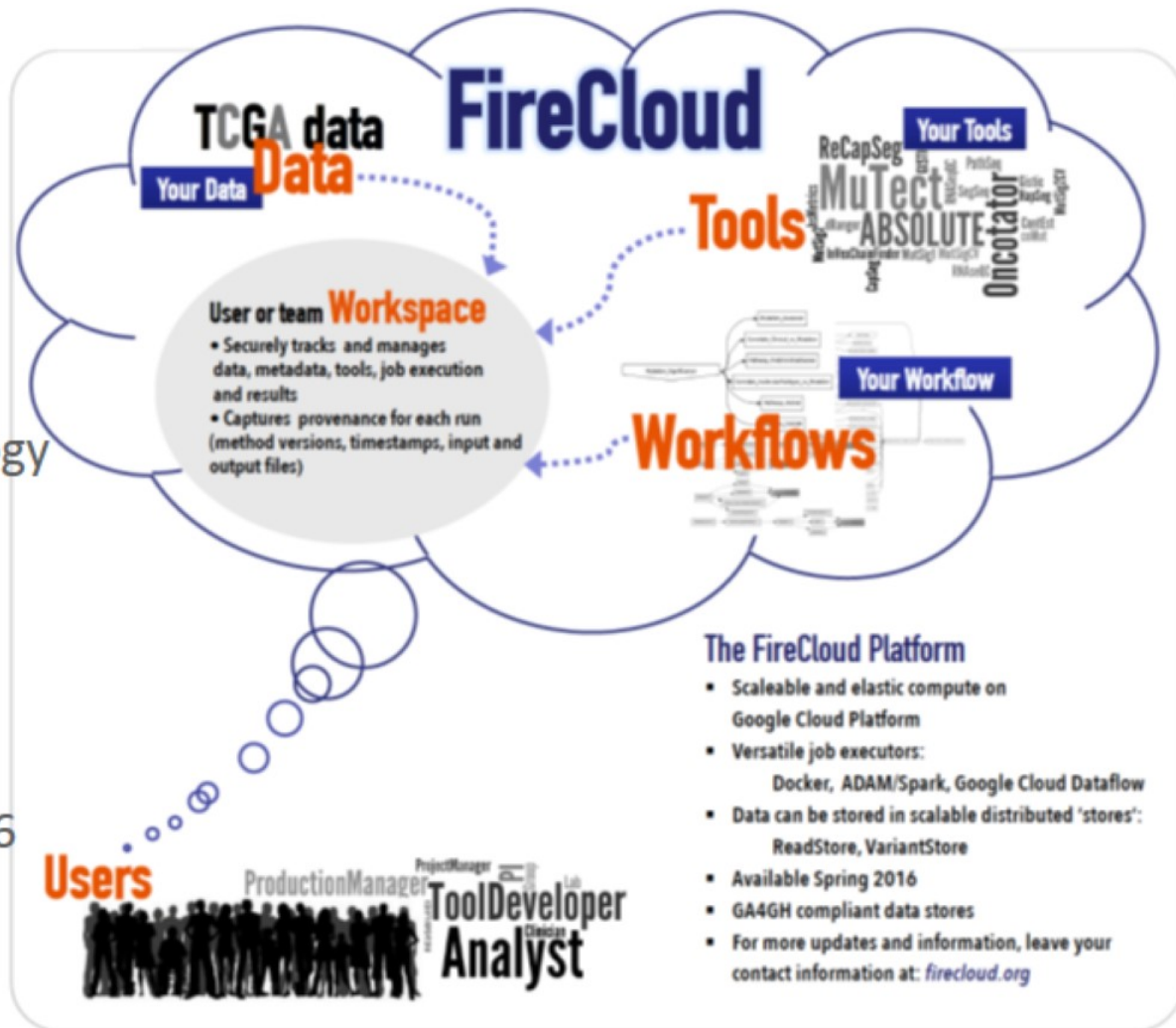
PI: Ilya Shmulevich

Seven Bridges Genomics

PI: Deniz Kural

Period of performance:

Sept 2014 – Sept 2016



NCI GDC and the Cloud Pilots

- Working together to build **common APIs**
- Working with the Global Alliance for Genomics and Health (**GA4GH**) to **define** the next generation of **secure, flexible, meaningful, interoperable, lightweight interfaces**
- Competing on the **implementation**, collaborating on the **interface**
- Aligned with **BD2K** and serving as a part of the **NIH Commons** and working toward shared goals of **FAIR** (Findable, Accessible, Interoperable, Reusable)
- Exploring and defining **sustainable precision medicine information infrastructure**

Information problem(s) we intend to solve with the Precision Medicine Initiative for Oncology

- **Establish** a sustainable infrastructure for cancer genomic data – through the **GDC**
- **Provide** a data integration platform to allow multiple data types, multi-scalar data, temporal data from cancer models and patients
 - Under evaluation, but it is likely to include the GDC, TCIA, Cloud Pilots, tools from the ITCR program, and activities underway at the Global Alliance for Genomics and Health
- **Support** precision medicine-focused clinical research

NCI Precision Medicine Informatics Activities

- As we receive additional funding for Precision Medicine, we plan to:
 - **Expand** the GDC to handle additional data types
 - **Include** the learning from the Cloud Pilots into the GDC
 - **Scale** the GDC from 10PB to hundreds of petabytes
 - Include imaging by interoperating between the GDC and the **Quantitative Imaging Network TCIA** repository
 - **Expand** clinical trials tooling from NCI-MATCH to NCI-MATCH Plus
 - **Strengthen** the ITCR grant program to explicitly include precision medicine-relevant proposals

Bridging Cancer Research and Cancer Care

- Making clinical research relevant in the clinic
- Supporting the virtuous cycle of clinical research informing care, and back again
- Providing decision support tools for precision medicine

But how?

Precision Medicine informatics community engagement

- Ongoing

- Cancer Informatics for Cancer Centers Clinical Genomics Workshops (Nov '13, May '14, Nov '14, March '15) <http://ci4cc.org>
- Global Alliance for Genomics and Health Data and Clinical Working Groups (March '14, Oct '14, June '15) <http://genomicsandhealth.org/>

- Planned

- Convene a community informatics workshop Fall/Winter 2015

Thank you



Warren Kibbe
warren.kibbe@nih.gov

Thanks to content contributors:

Sherri de Coronado, Gilberto Fragoso, Mark Jensen, Warren Kibbe, Juli Klemm, Tony Kerlavage, JC Zenklusen, Elizabeth Gillanders and others.



**NATIONAL
CANCER
INSTITUTE**

www.cancer.gov

www.cancer.gov/espanol

Cancer Genomics Project Teams

CGC Pilot Team Principal Investigators

- **Gad Getz, Ph.D** - Broad Institute - <http://firecloud.org>
- **Ilya Shmulevich, Ph.D** - ISB - <http://cgc.systemsbiology.net/>
- **Deniz Kural, Ph.D** - Seven Bridges – <http://www.cancergenomicscloud.org>

NCI Project Officer & CORs

- Anthony Kerlavage, Ph.D –Project Officer
- Juli Klemm, Ph.D – COR, Broad Institute
- Tanja Davidsen, Ph.D – COR, Institute for Systems Biology
- Ishwar Chandramouliswaran, MS, MBA – COR, Seven Bridges Genomics

GDC Principal Investigator

- Robert Grossman, Ph.D - University of Chicago

Center for Cancer Genomics Partners

- JC Zenklusen, Ph.D
- Daniela Gerhard, Ph.D
- Zhining Wang, Ph.D
- Liming Yang, Ph.D
- Martin Ferguson, Ph.D

NCI Leadership Team

- Warren Kibbe, Ph.D
- Lou Staudt, M.D.
- Steven Chanock, Ph. D
- George Komatsoulis, Ph.D