Large Scale Cancer Genomics at NCI
Present and Future

DISCOVERY by genomics and functional genomics

Drug Development

Pathway Function

DNA-based Diagnosis “better pathology”

Precision Patient Treatment and Prevention

Cancer Information Donor

Cancer Information Donor
Where we are...

DISCOVERY by genomics and functional genomics

TCGA TARGET

CTD²

Drug Development → Pathway Function → DNA-based Diagnosis → Precision Treatment

Precision Initiative Alchemist etc
TCGA = The Cancer Genome Atlas
Adult Cancers
No Prior Treatment

Kenna Shaw PhD

TARGET = Therapeutically Applicable Research to Generate Effective Treatments
Pediatric Cancers
Selected poor outcome tumors

Daniela Gerhard PhD

CTD² = Cancer Target and Drug Discovery

Brad Ozenberger PhD
Major Goals of TCGA and TARGET

Discover “driver” genes; learn frequencies

Discover mutation combinations: pathways, networks

Discover RNA expression, methylation, copy number, LOH
Integrate across data types and tumor types

Mine data to suggest treatment - actionable signatures
Trials follow!

Mine data to focus drug development and other treatments

Develop ever-better methods for analysis and make available
Implicit Goals / Questions for TCGA and TARGET

What is the added impact of big “reference data” that are comprehensive, coherent, high quality, widely accessible?

What is the impact of these “Team Science” communities?

Can new TCGA pipeline partner intimately with clinical trials?

With community care?

With RO1 Genomics?

......and vice versa?
Major Goals of CTD$^2$

Translate genomic candidates into treatment targets

Develop and use high throughput screens for target validation

Develop and use computational approaches: pathways, drugs

Identify lead drugs
25* forms of cancer

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

Multiple data types

- Clinical diagnosis
- Treatment history
- Histologic diagnosis
- Pathologic report/images
- Tissue anatomic site
- Surgical history
- Gene expression/RNA sequence
- Chromosomal copy number
- Loss of heterozygosity
- Methylation patterns
- miRNA expression
- DNA sequence
- RPPA (protein)
- Subset for Mass Spec

Biospecimen Core Resource with more than 150 Tissue Source Sites

6 Cancer Genomic Characterization Centers

3 Genome Sequencing Centers

7 Genome Data Analysis Centers

Data Coordinating Center
Robust Pipeline for Comprehensive Genomic Characterization

Tissue Sample

Pathology QC

DNA & RNA Isolation, QC

Sequence Exomes
Whole genomes

RNA Copy number
DNA methylation

Data and Results Storage & QC

Integrative computational Analysis

Comprehensive Characterization

High quality frozen samples costly, rate-limiting

BCR = GSCs = CGCCs = DCC = GDACs
TCGA Adult Tumors
Complete 500 primary tumors per type

- AML
- Breast Ductal*
- Breast Lobular/Breast Other
- Bladder
- Cervical adeno & squamous
- Colorectal*
- Clear cell kidney*
- Diffuse Large B-cell Lymphoma
- Endometrial carcinoma*
- Esophageal adeno & squamous
- Gastric adenocarcinoma
- Glioblastoma multiforme*
- Head and Neck Squamous
- Hepatocellular
- Lower Grade Glioma
- Lung adeno
- Lung squamous
- Melanoma
- Ovarian serous cystadenocarcinoma*
- Papillary kidney
- Pancreas
- Prostate
- Sarcoma (expanding to 10 subtypes)
- Papillary Thyroid*

*Reached 500 tumor goal

9 Research papers published or in preparation

The Cancer Genome Atlas
TCGA Progress by Tumor Type

- Manuscript in preparation, submitted or published
- Analysis underway
- Sample acquisition phase
- Rare tumor project

* Only accepting AA cases
Accrual challenge is great: Outcome range

The figure shows a graph comparing Total Reality and Total Ideal over time from 11/16/2006 to 11/16/2013. The x-axis represents dates, and the y-axis represents the outcome range from 0 to 20,000. The graph indicates that the Total Ideal consistently exceeds the Total Reality, highlighting the challenge in achieving the ideal outcome.
New Rare Tumor Project - Launched 2012
50 -100 tumors per type

- Adrenocortical Carcinoma*
- Adult ALL (B-cell and T-Cell)
- Anaplastic Thyroid
- Cholangiocarcinoma or Gall Bladder
- Chromophobe kidney*
- High Risk MDS (del 5q- cases)
- Mesothelioma*
- MPNST
- Paraganglioma/Pheochromocytoma
- Testicular Germ Cell
- Uterine Carcinosarcoma*
- Thymoma

* - Sample Acquisition Ongoing
Revised Data Access and Publication Policy

- All data are available pre-publication, but users are asked to allow TCGA a first comprehensive publication.

- Before TCGA paper, users may publish on *any* tumor type, any time, as long as only one platform is used.

- After TCGA paper publication, OR 18 months after 100 cases have shipped, any user may use data in any way.

- Users may use data in grant applications, posters at meetings, etc. all prior to any TCGA paper.

- For questions – write [tcga@mail.nih.gov](mailto:tcga@mail.nih.gov)
Raw Sequence Downloads from CGHub

TCGA Data Portal Snapshot: October 2012

- >38,000 archive downloads
- ~350 controlled data; <1% of use is controlled access
- Data use “spikes” after publications
Results: Squamous Cell Lung
“Driver Genes” in diverse combinations
EXOME sequencing
More drivers:
Statistical power issues

10 additional candidates (COSMIC)

> Implications for future study design numbers – how deep is important?

> Meaning of low frequency drivers overall? Meaning in a specific patient?

Must do experiments.....
Driver Mutation Pathways
Mutations aggregate in pathways and networks “Actionable” fraction

PI(3)K/RTK/RAS signalling
69% altered

PTEN 15%
PIK3CA 16%

STK11 2%
AKT1 <1%
AKT2 4%
AKT3 16%

AMPK
MTOR

EGFR 9%
ERBB2 4%
ERBB3 2%
FGFR1 7%
FGFR2 3%
FGFR3 2%

KRAS 3%
HRAS 3%
NRAS <1%

BRAF 4%

Cases (%)

Alteration pattern
RTK 26%
RAS 24%
PI(3)K 47%

Proliferation, cell survival, translation
Activation
Inhibition
Cross-Tumor Integration
Similarities among tumor subsets suggested by Somatic Copy Number data

Andrew Cherniak, Matthew Meyerson
Broad Institute
Specific (numerically rare) subset of Gliomas display “ride along” deletions of ENO1

This renders them sensitive to ENO2 inhibition
ENO1 “Passenger” deletion creates druggable ENO2 vulnerability – small and specific subset of GBMs
Pediatric Cancer Genomics

Emphasize tumors with poor outcomes to current treatment
TARGET: Pediatric Cancer Genomics

@ 100-200 cases per tumor type

- Acute lymphoblastic leukemia (ALL), including relapse
- Acute myeloid leukemia (AML), including relapse
- Neuroblastoma (stage 4)
- Osteosarcoma
- Wilms tumor (relapsed patients and anaplasia)
### Summary: TARGET Sequencing completed - August 24, 2012

<table>
<thead>
<tr>
<th>Disease</th>
<th>WGS Cases (CGI)</th>
<th>Trios (T)</th>
<th>WGS D Cases (Illumina)</th>
<th>WES cases</th>
<th>mRNA-seq</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALL</td>
<td>114</td>
<td>50</td>
<td>2</td>
<td>21 T</td>
<td>12 D</td>
</tr>
<tr>
<td>AML</td>
<td>112</td>
<td>52</td>
<td>NA</td>
<td>20 T + 2 D</td>
<td>~100</td>
</tr>
<tr>
<td>NBL</td>
<td>10</td>
<td>NA</td>
<td>10</td>
<td>254 D</td>
<td>~35 D</td>
</tr>
<tr>
<td>OS</td>
<td>19</td>
<td>NA</td>
<td>12</td>
<td>54 D</td>
<td>54 D</td>
</tr>
<tr>
<td>WT</td>
<td>48</td>
<td>NA</td>
<td>NA</td>
<td>28 (T and D)</td>
<td>NA</td>
</tr>
<tr>
<td>Total</td>
<td>303*</td>
<td>24</td>
<td>379</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TARGET Whole Genome Sequences show AML “quiet” genome vs Osteosarcoma “agitated” genome.
Clusters of mutations close together surround rearrangements – implications for mechanism

Slide adapted from Paul Meltzer, TARGET Osteo Group
CLUSTERS: 17 Osteosarcoma Whole Genomes

• 114 TOTAL CLUSTERS (MEDIAN 7; RANGE 1-20)

• 72% SHOW STRAND COORDINATION - NEARLY ALL AT G-C bp

• 4.4% (1538) OF ALL SOMATIC SNV’S ARE IN STRAND COORDINATED CLUSTERS. (MEDIAN 1.9%; RANGE 0.28%-5.6%)

• 71 OVERLAP REFSEQ EXONS

......In pursuit of mechanistic implications

Slide adapted from Paul Meltzer for TARGET Osteo
CTD² Result: siRNA target gene evaluation
ID4 in ovarian tumors

Ren et al  Science Translational Med 4, 147, 2012
CTD\(^2\) Result: siRNA target gene evaluation ID4 in ovarian tumors

Human Xenograft test of nanoparticle ID4 siRNA efficacy

Ren et al  Science Translational Med 4, 147, 2012
Update FFPE: Formalin Fixed Paraffin Embedded

Critical path to trials and all clinical samples

State of the art

> DNA FFPE ready for many uses

(samples in 5-10 year range; buffered formalin superior)

> Becoming strong for RNA alone

> Promising new TCGA protocol for joint DNA/RNA**

**Scott Morris and Erik Zmuda TCGA BCRs
DNA Sequence coverage: Frozen vs. FFPE
Exome data

Frozen (89.0%)

FFPE (82.5%)

BCR + Baylor TCGA
Major Cancer Genomics Opportunity: Genomics of Progression, Resistance, Metastasis
Path forward - Partner CCG pipelines with new trials: e.g. Alchemist, “Exceptional Cases” ......
Major Cancer Genomics Opportunities 2013 cont...

- Tumor heterogeneity and microenvironment

- Epigenomics broadly defined – Cancer “ENCODE”?
  Provides framework for deep individual projects

- Germline genomics

- Interface with Systems Biology, predictive modeling
Whatever problems top your list, you will need

Informatics and Analysis: Toward a Cancer Genome Commons

- Joint mining of genomics data and EHRs
- Data aggregation and access
  - CGHub is new, working, but will not scale 10X, 100X, etc
  - dbGAP will have serious scaling issues

Guidelines and Bake-offs wet and dry
  - Example = mutation calling series
Future Cancer Genomics at NCI
Make the Cancer Information Donor real: Multiple Steps

1. Partner in trials; answer key questions, fill Library core
2. Pilot RO1 data – a separate Commons Library Branch?
3. Pilot Library branch for true clinical patient donated information
Now Leading CCG

Dr. Louis Staudt  Dr. Stephen Chanock

Joint NCI NHGRI workshop on the future of Cancer Genomics
November 30, 2012