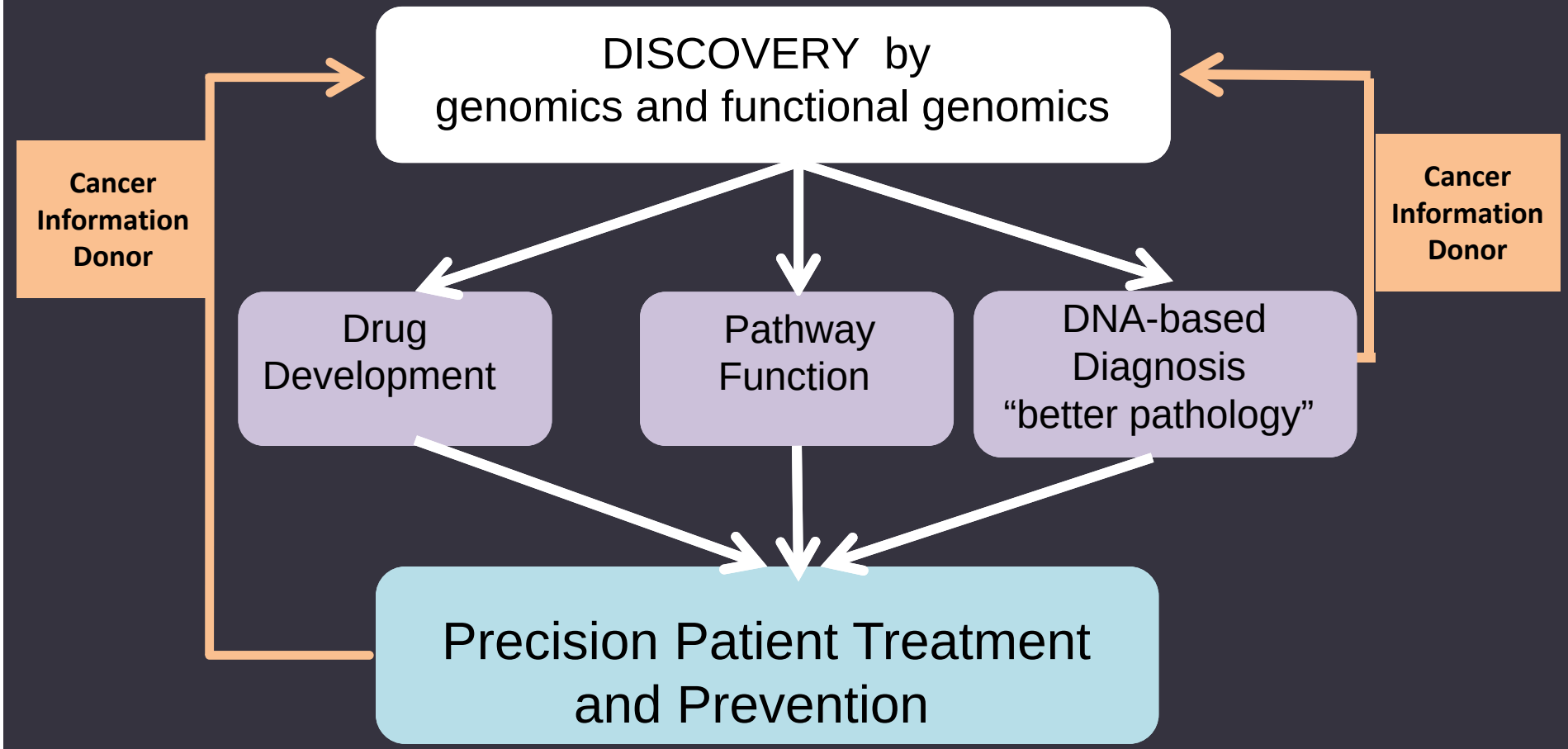
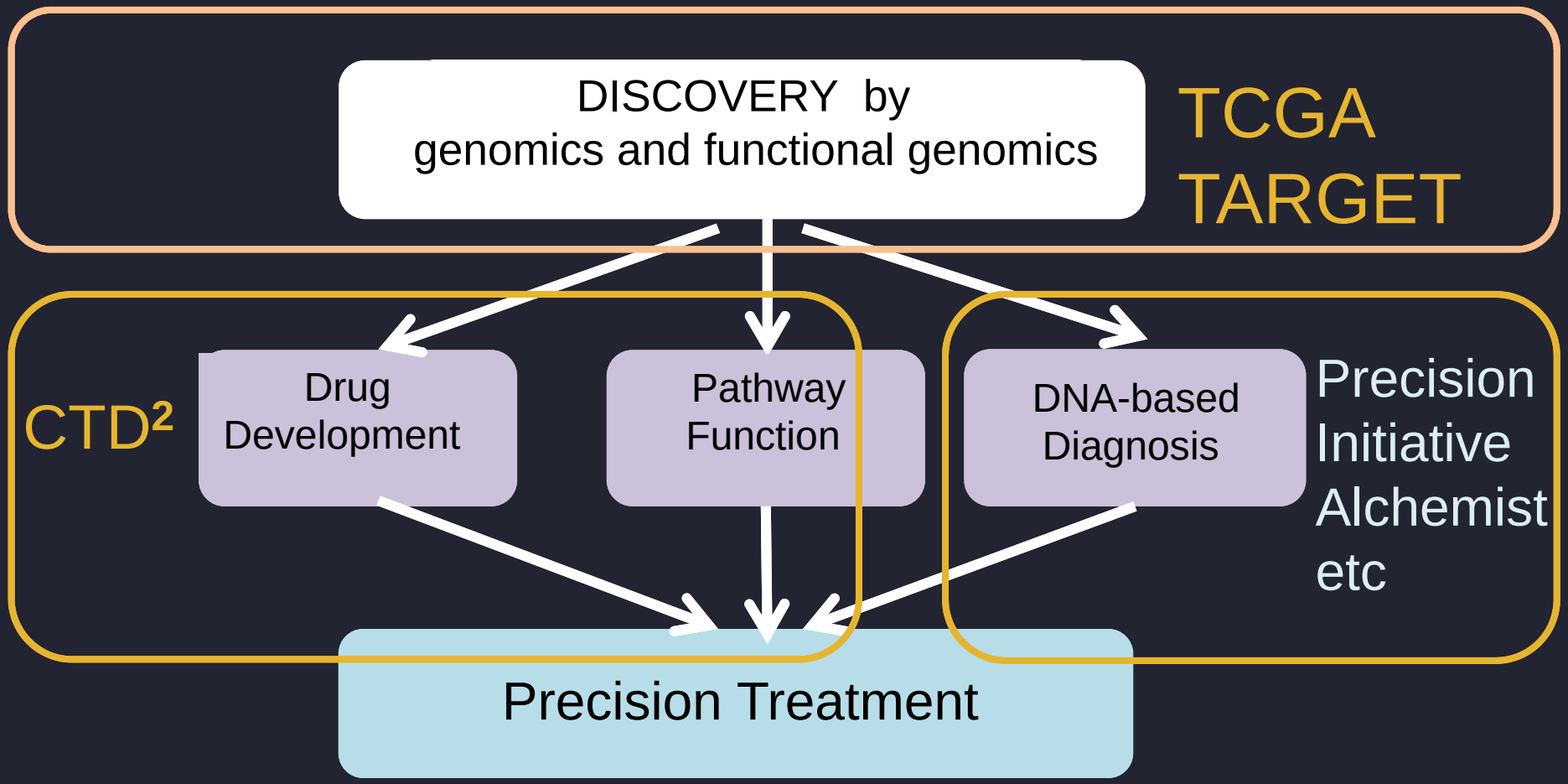
Large Scale Cancer Genomics at NCI
Present and Future



Where we are...

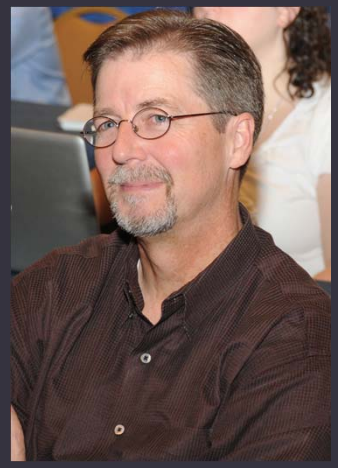

DISCOVERY by
genomics and functional genomics
TCGA
TARGET
CTD²
Drug
Development
Pathway
Function
DNA-based
Diagnosis
Precision
Initiative
Alchemist
etc
Precision Treatment

Kenna Shaw PhD

TCGA = The Cancer Genome Atlas

Adult Cancers

No Prior Treatment



Brad Ozenberger PhD

Daniela Gerhard PhD

TARGET = Therapeutically Applicable Research to Generate Effective Treatments

Pediatric Cancers

Selected poor outcome tumors

CTD^2 = Cancer Target and Drug Discovery

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

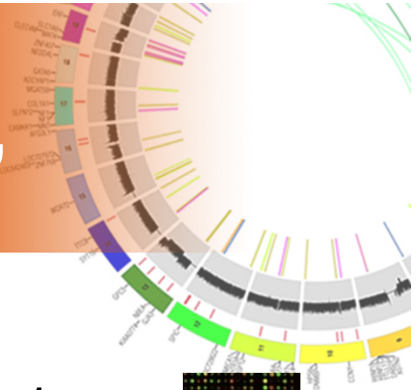
Translate genomic candidates into treatment targets

Develop and use high throughput screens for target validation

Develop and use computational approaches: pathways, drugs

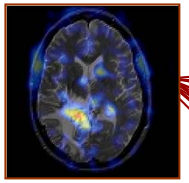
Identify lead drugs

TCGA Design: No Platform Left Behind
Distributed "Team Science"

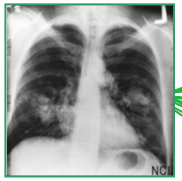


25* forms of cancer

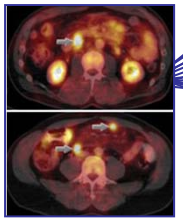
glioblastoma multiforme (brain)



squamous carcinoma (lung)



serous cystadenocarcinoma (ovarian)



Etc. Etc. Etc.

Biospecimen Core Resource with more than 150 Tissue Source Sites

6 Cancer Genomic Characterization Centers

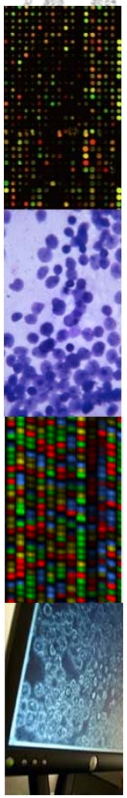
3 Genome Sequencing Centers

7 Genome Data Analysis Centers

Data Coordinating Center

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



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

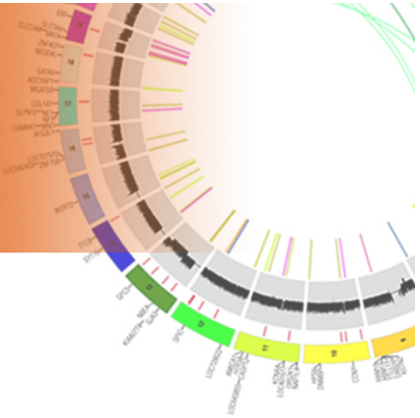
Comprehensive Characterization

High quality frozen samples
costly, rate-limiting

= BCR = GSCs = CGCCs = DCC = GDACs

The Cancer Genome Atlas

TCGA Adult Tumors
Complete 500 primary tumors per type

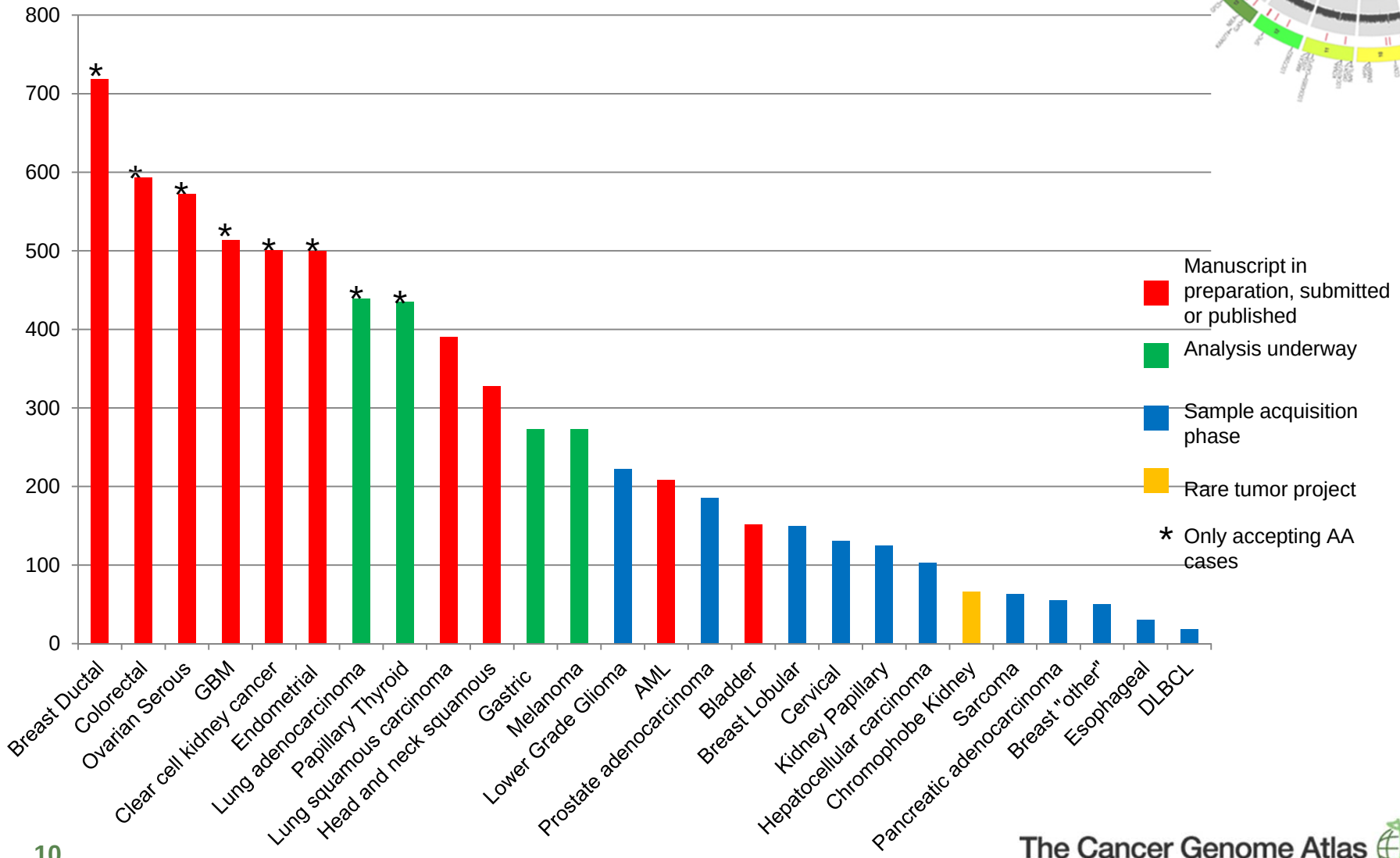


- AML
- Breast Ductal*
- Breast Lobular/Breast Other
- Bladder
- Cervical adeno & squamous
- Colorectal*
- Clear cell kidney*
- Diffuse Large B-cell Lymphoma
- Endometrial carcinoma*
- Esophageal adeno & squamous
- Gastric adenocarcinoma
- Glioblastoma multiforme*
- Head and Neck Squamous
- Hepatocellular
- 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*

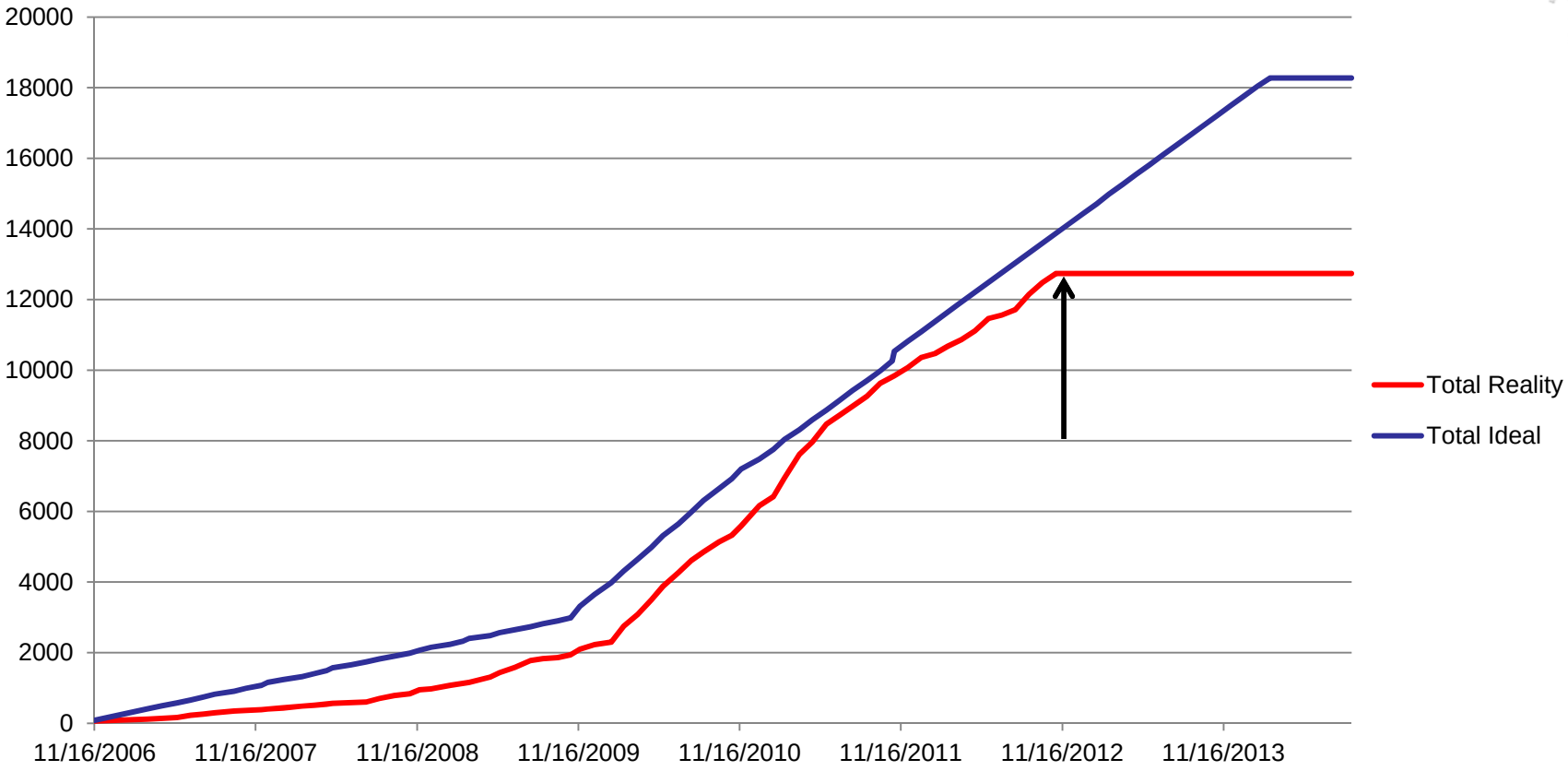
Research papers published or in preparation

TCGA Progress by Tumor Type

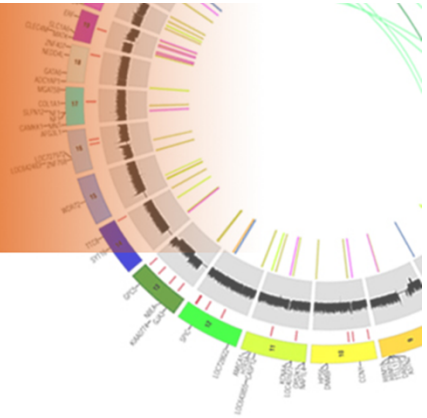


The Cancer Genome Atlas

Accrual challenge is great: Outcome range



New Rare Tumor Project - Launched 2012
50 -100 tumors per type



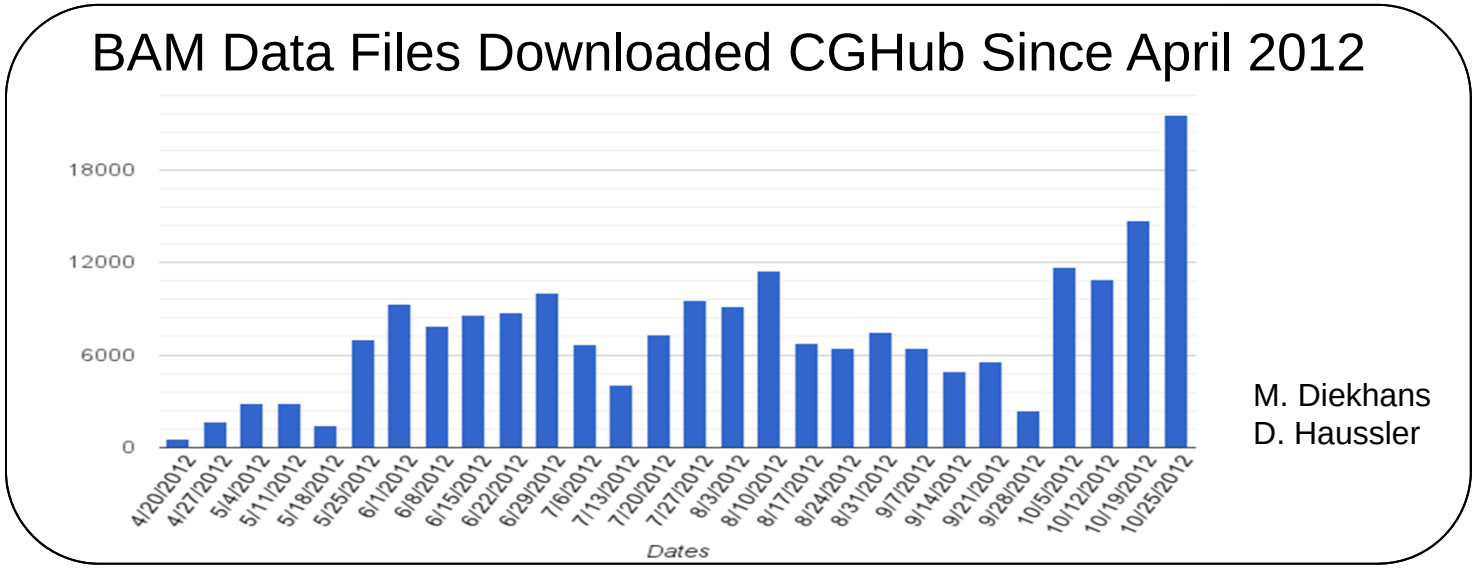
- Adrenocortical Carcinoma*
- Adult ALL (B-cell and T-Cell)
- Anaplastic Thyroid
- Cholangiocarcinoma or Gall Bladder
- Chromophobe kidney*
- High Risk MDS (del 5q- cases)
- Mesothelioma*
- MPNST
- Paraganglioma/Pheochromocytoma
- Testicular Germ Cell
- Uterine Carcinosarcoma*
- Thymoma

*- *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**

Raw Sequence Downloads from CGHub



M. Diekhans
D. Haussler

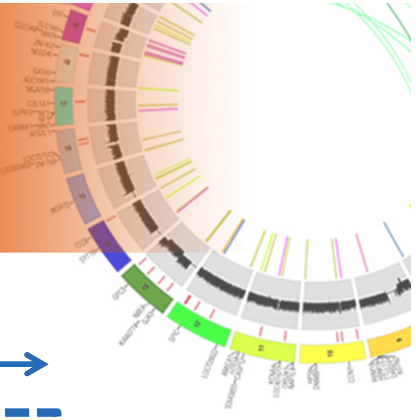
TCGA Data Portal Snapshot: October 2012

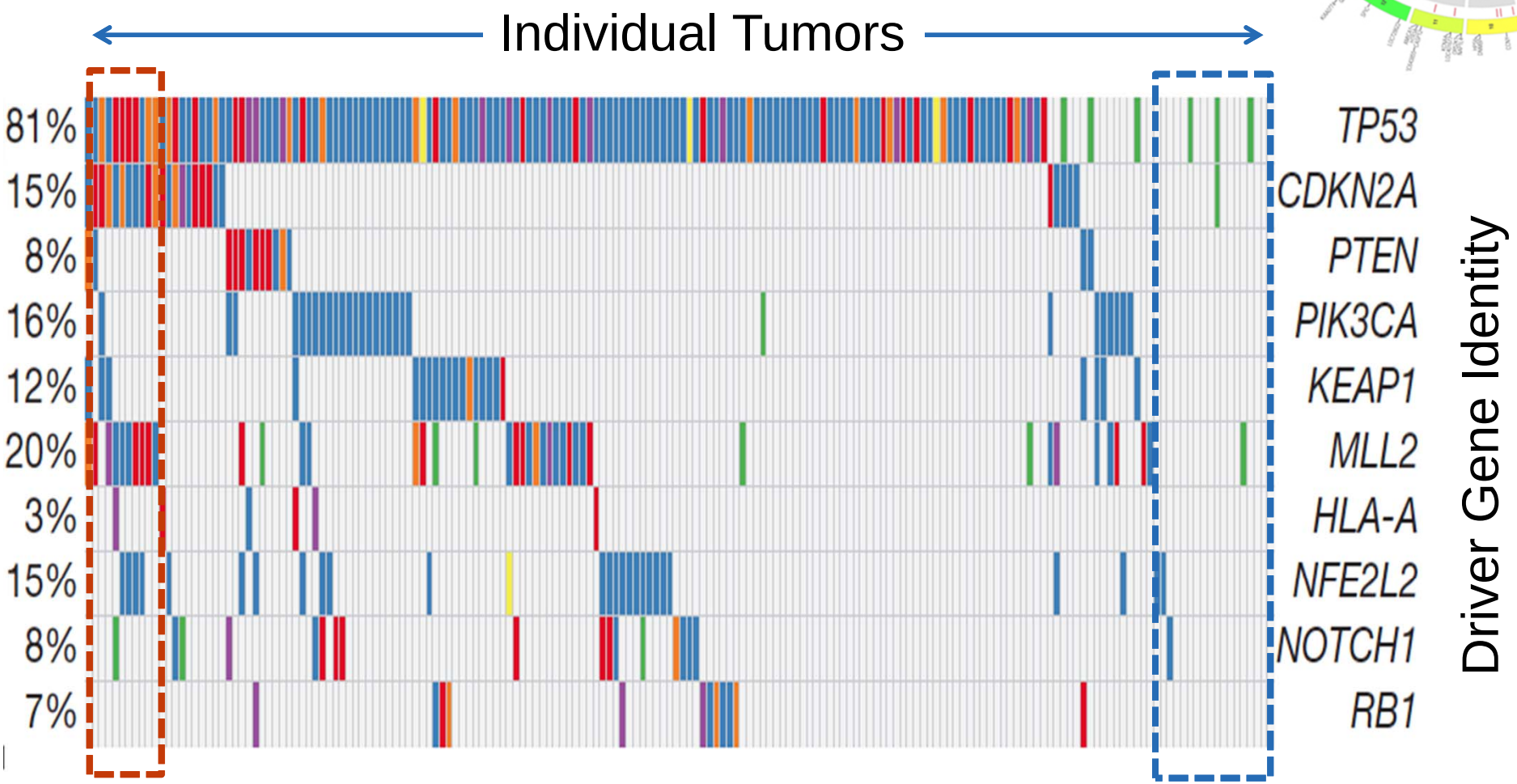
- **>38,000 archive downloads**
- **~350 controlled data; <1% of use is controlled access**
- **Data use "spikes" after publications**

Results: Squamous Cell Lung

"Driver Genes" in diverse combinations

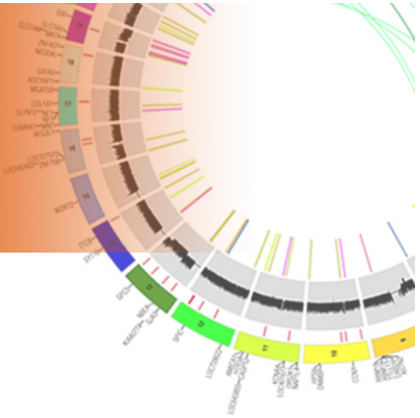
EXOME sequencing





The Cancer Genome Atlas

Results: Squamous Cell Lung



More drivers:
Statistical power issues

10 additional candidates (COSMIC)

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

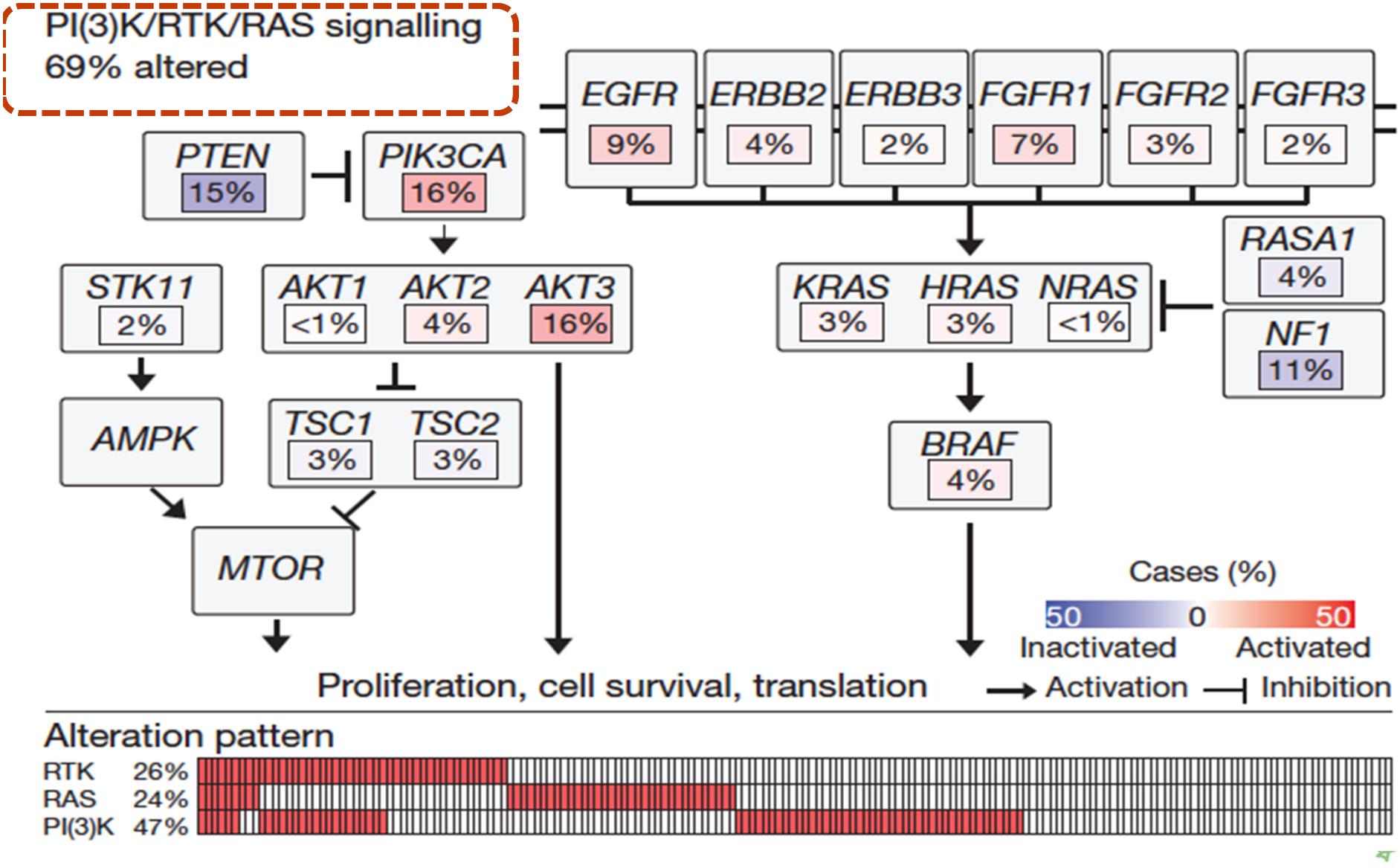
> Meaning of low frequency drivers overall? Meaning in a specific patient?

Must do experiments.....

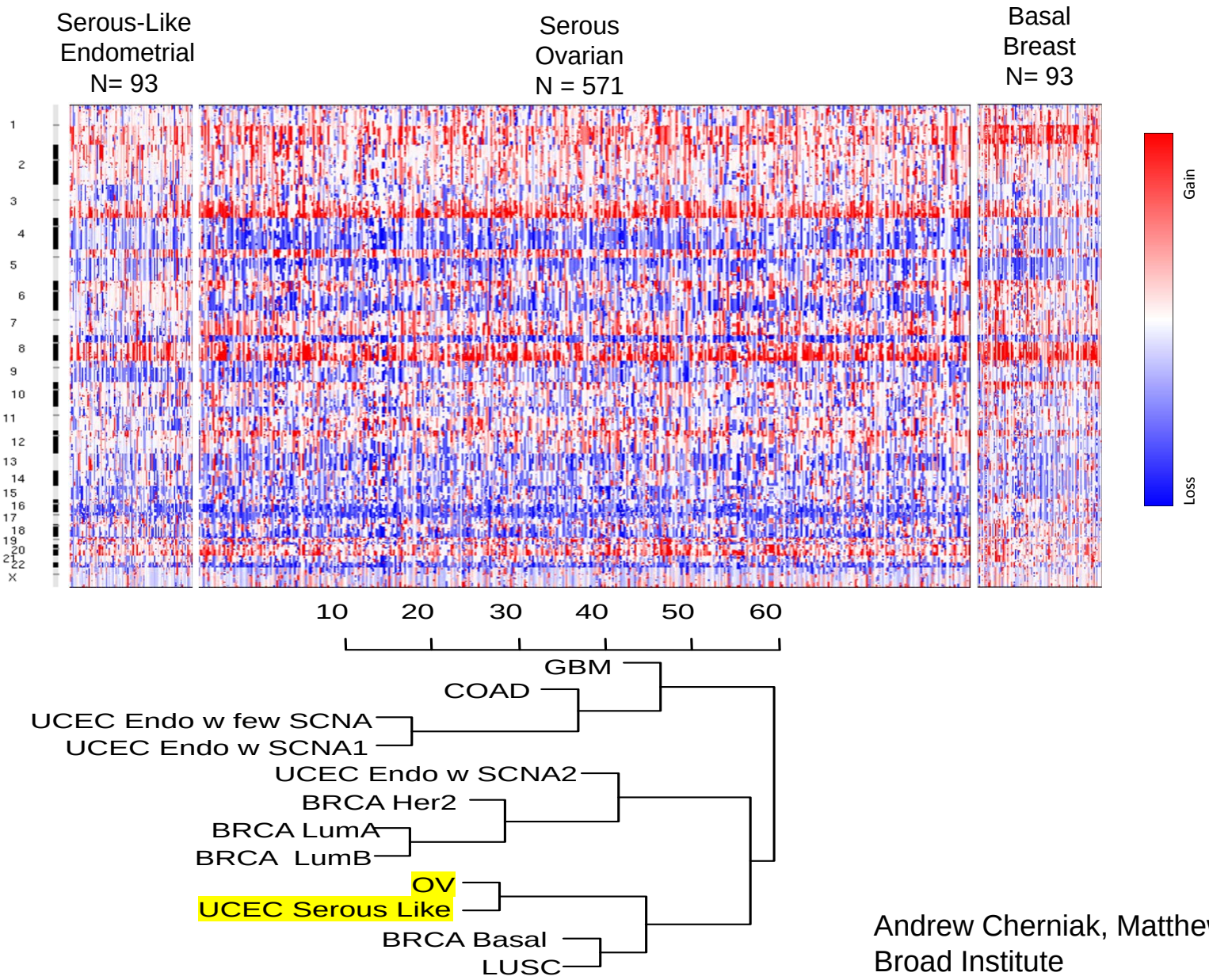
Driver Mutation Pathways

Mutations aggregate in pathways and networks

“Actionable” fraction



Cross-Tumor Integration
Similarities among tumor subsets suggested by Somatic Copy Number data



Andrew Cherniak, Matthew Meyerson
Broad Institute

TCGA reference data mined as starting point for other studies

ARTICLE

doi:10.1038/nature11331

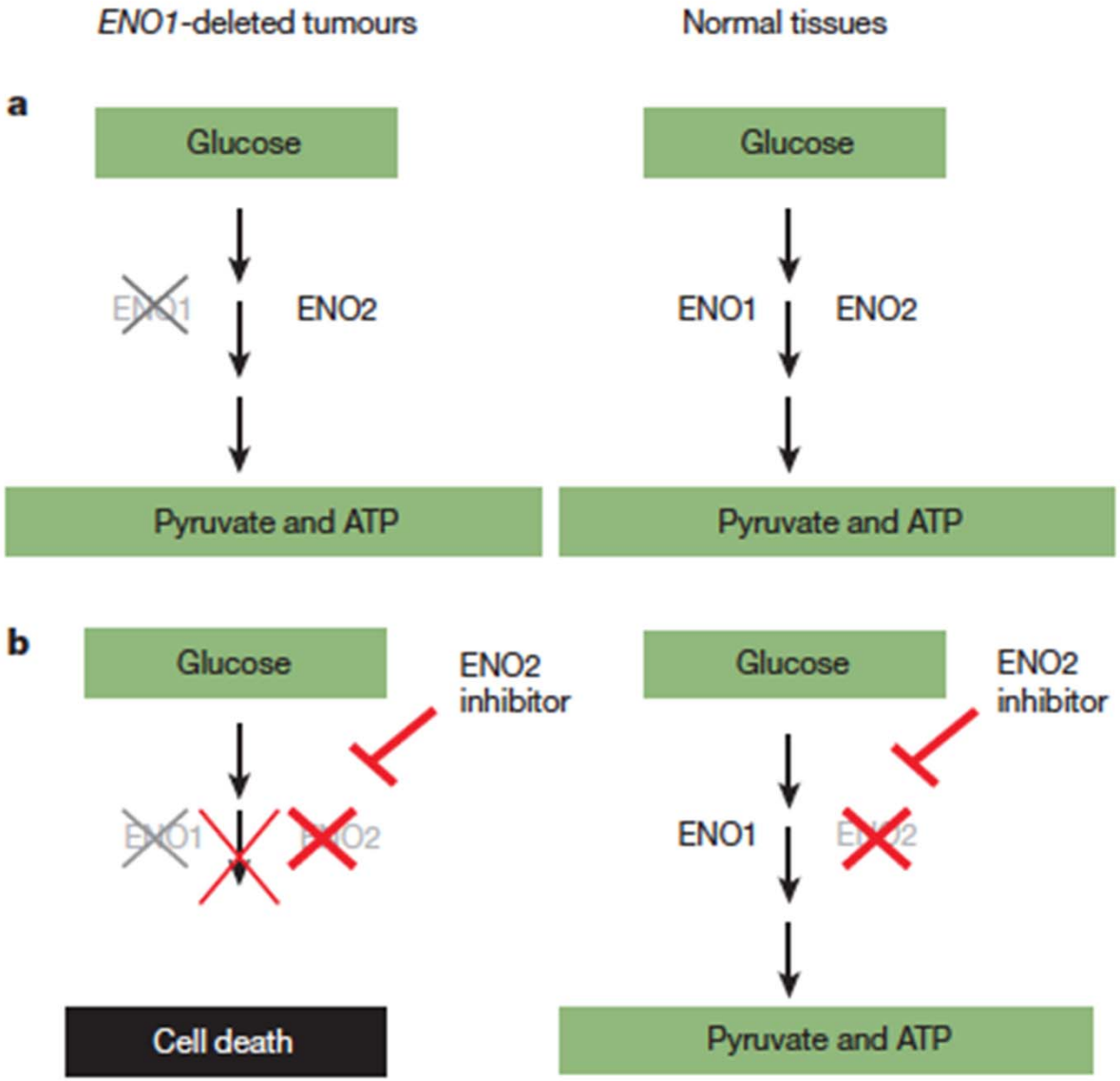
Passenger deletions generate therapeutic vulnerabilities in cancer

Florian L. Muller[1,2,3]*, Simona Colla[1,2,3]*, Elisa Aquilanti[2]*, Veronica E. Manzo[2], Giannicola Genovese[1,2], Jaclyn Lee[2], Daniel Eisenson[2], Rujuta Narurkar[2], Pingna Deng[1,2], Luigi Nezi[1,2], Michelle A. Lee[2,4], Baoli Hu[1,2,5], Jian Hu[1,2,3], Ergun Sahin[2,3], Derrick Ong[1,2,3], Eliot Fletcher-Sananikone[1,2], Dennis Ho[2,3], Lawrence Kwong[1,2], Cameron Brennan[6], Y. Alan Wang[1,2,5], Lynda Chin[1,2,5] & Ronald A. DePinho[2,3,5,7]

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



National Cancer Institute

Pediatric Cancer Genomics

Emphasize tumors with

poor outcomes to current treatment

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
National Institutes of Health

National Cancer Institute

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)

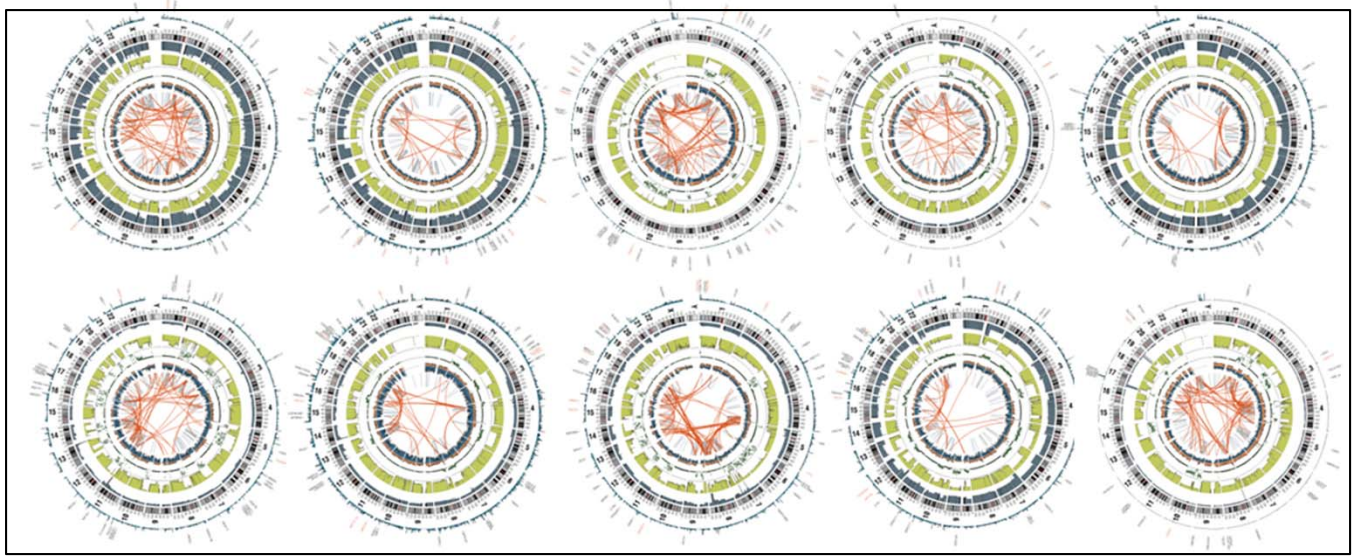
National Cancer Institute

Summary: TARGET Sequencing completed - August 24, 2012

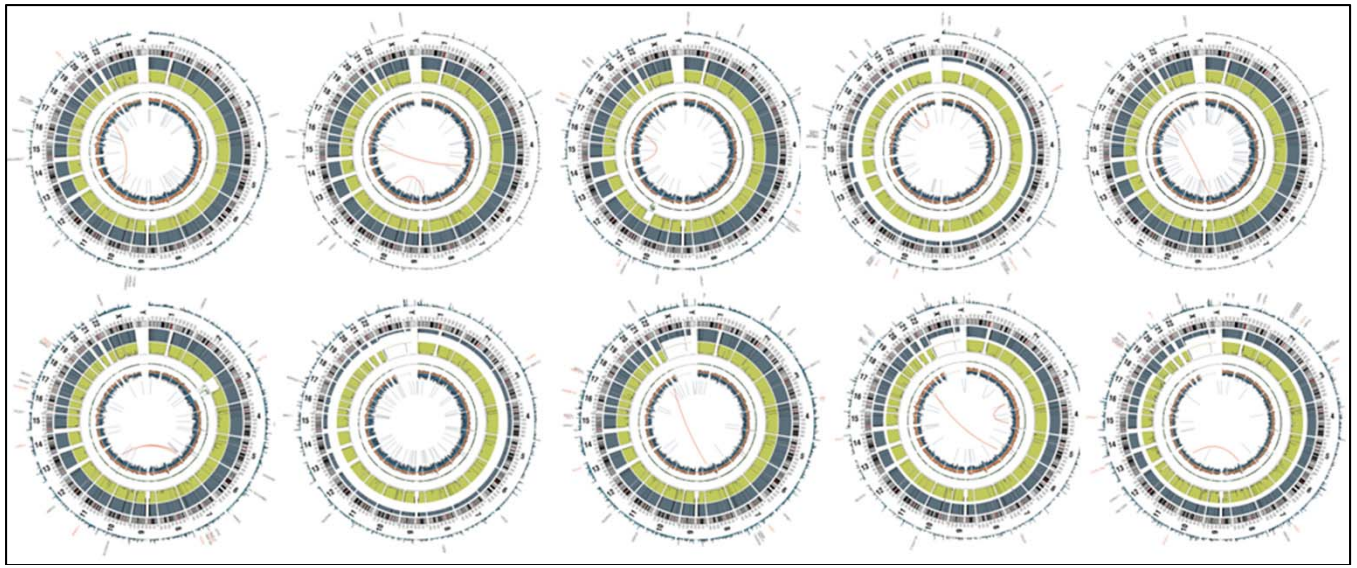
Disease	WGS Cases (CGI)	Trios (T)	WGS D Cases (Illumina)	WES cases	mRNA-seq
ALL	114	50	2	21 T	12 D
AML	112	52	NA	20 T + 2 D	~100
NBL	10	NA	10	254 D	~35 D
OS	19	NA	12	54 D	54 D
WT	48	NA	NA	28 (T and D)	NA
Total	303*		24	379	

National Cancer Institute

TARGET Whole Genome Sequences show AML “quiet” genome vs Osteosarcoma “agitated” genome

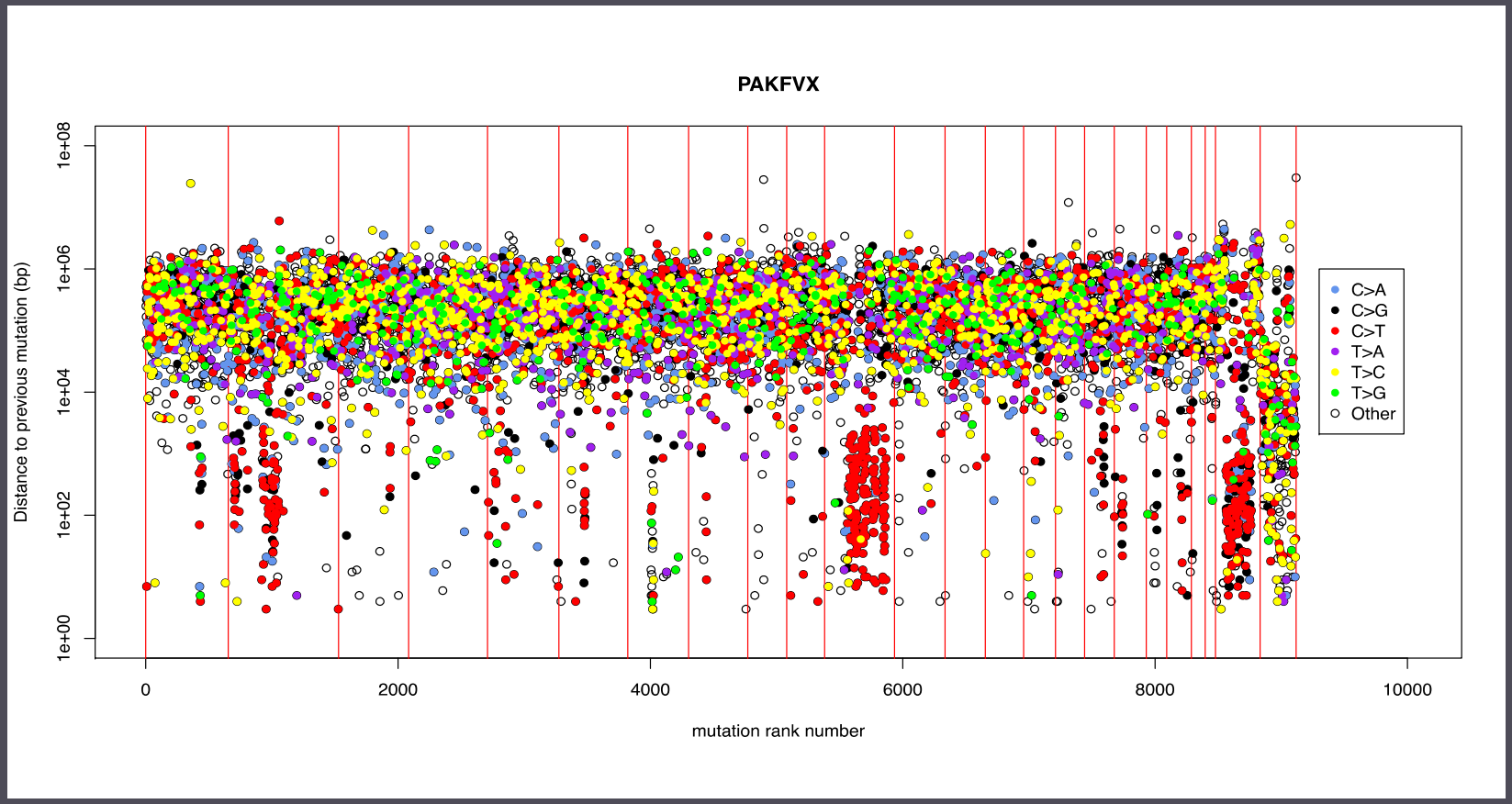


OS



AML

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%)

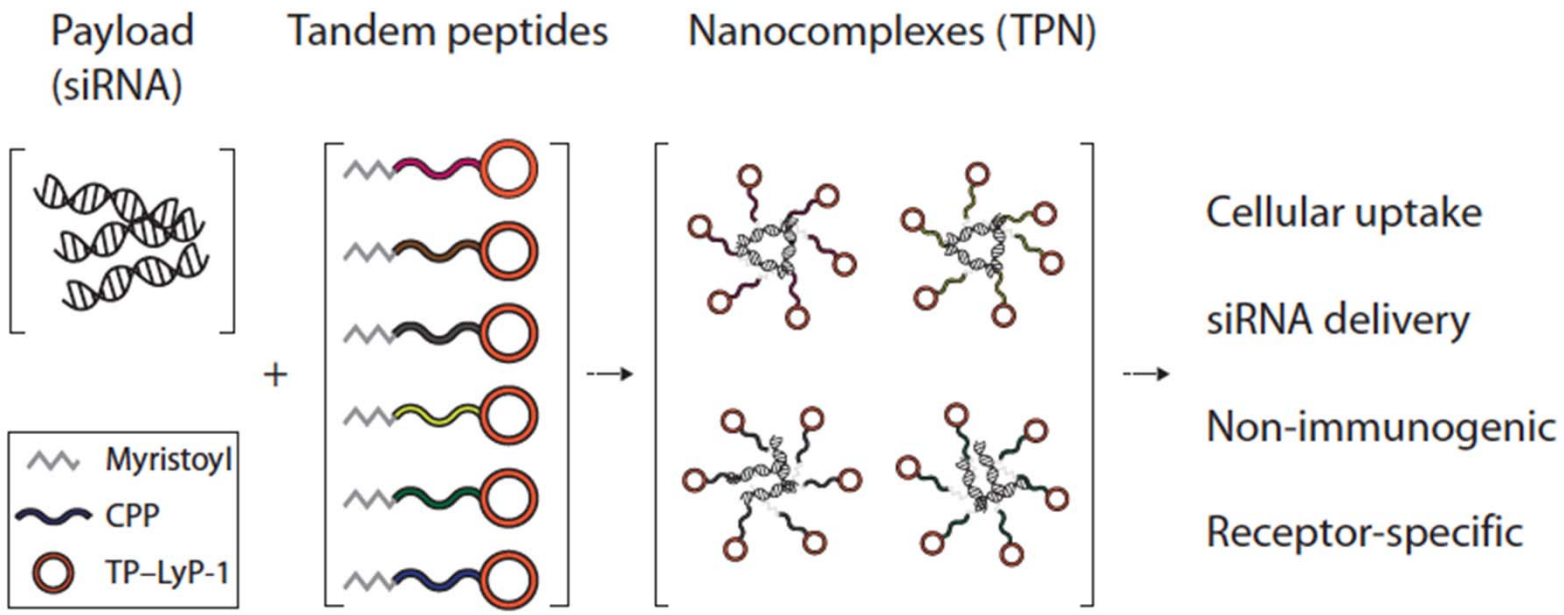
• 71 OVERLAP REFSEQ EXONS

.....In pursuit of mechanistic implications

Slide adapted from Paul Meltzer for TARGET Osteo

National Cancer Institute

CTD^2 Result: siRNA target gene evaluation ID4 in ovarian tumors

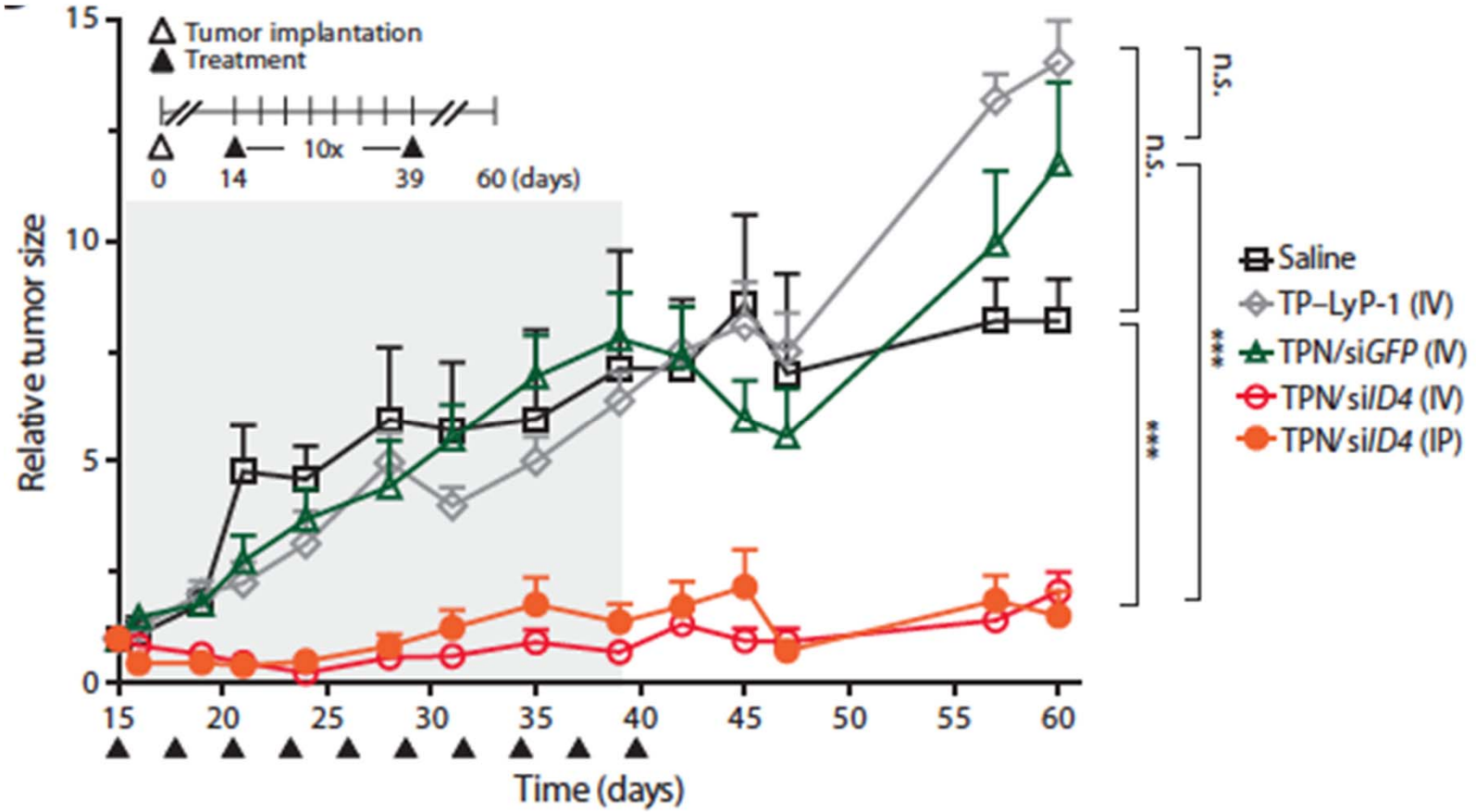


Ren et al Science Translational Med 4, 147, 2012

National Cancer Institute

CTD² 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

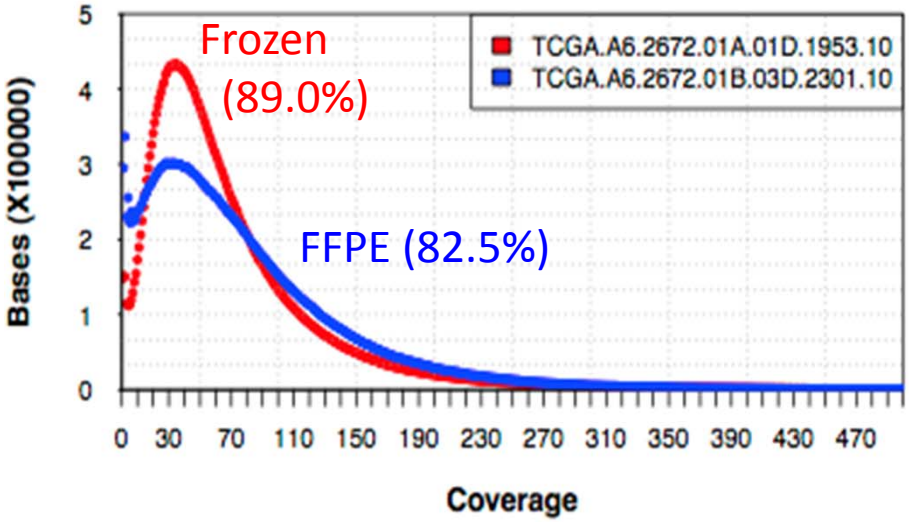
(samples in 5-10 year range; buffered formalin superior)

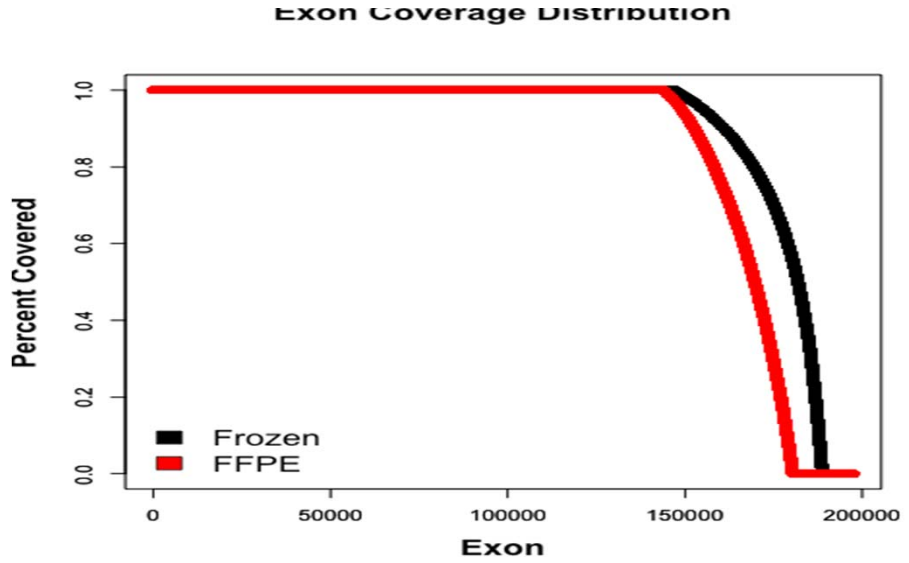
> Becoming strong for RNA alone

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

**Scott Morris and Erik Zmuda TCGA BCRs

DNA Sequence coverage: Frozen vs. FFPE Exome data

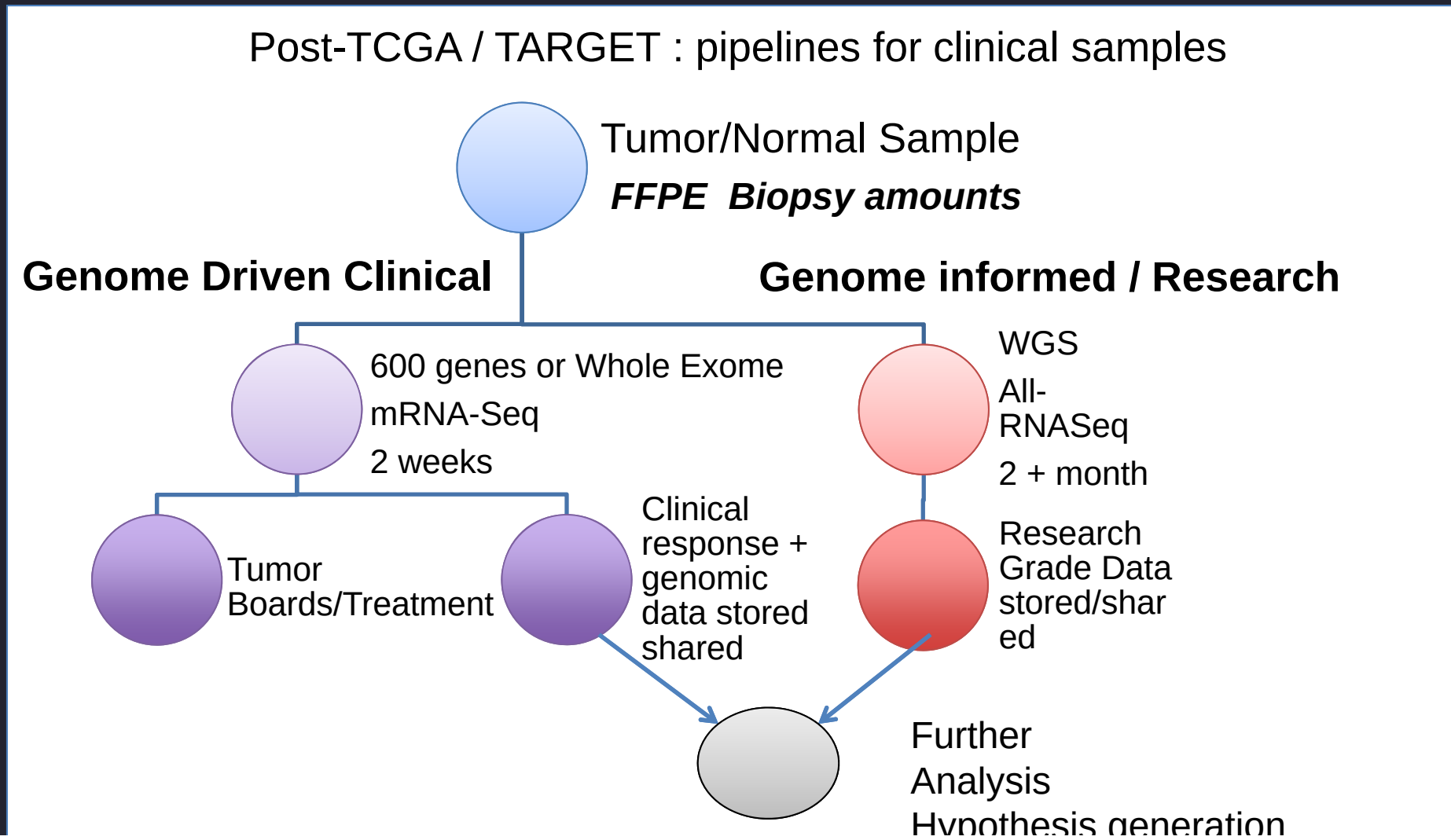




BCR + Baylor TCGA

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

Genome Analysis Access and Standards 2013

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

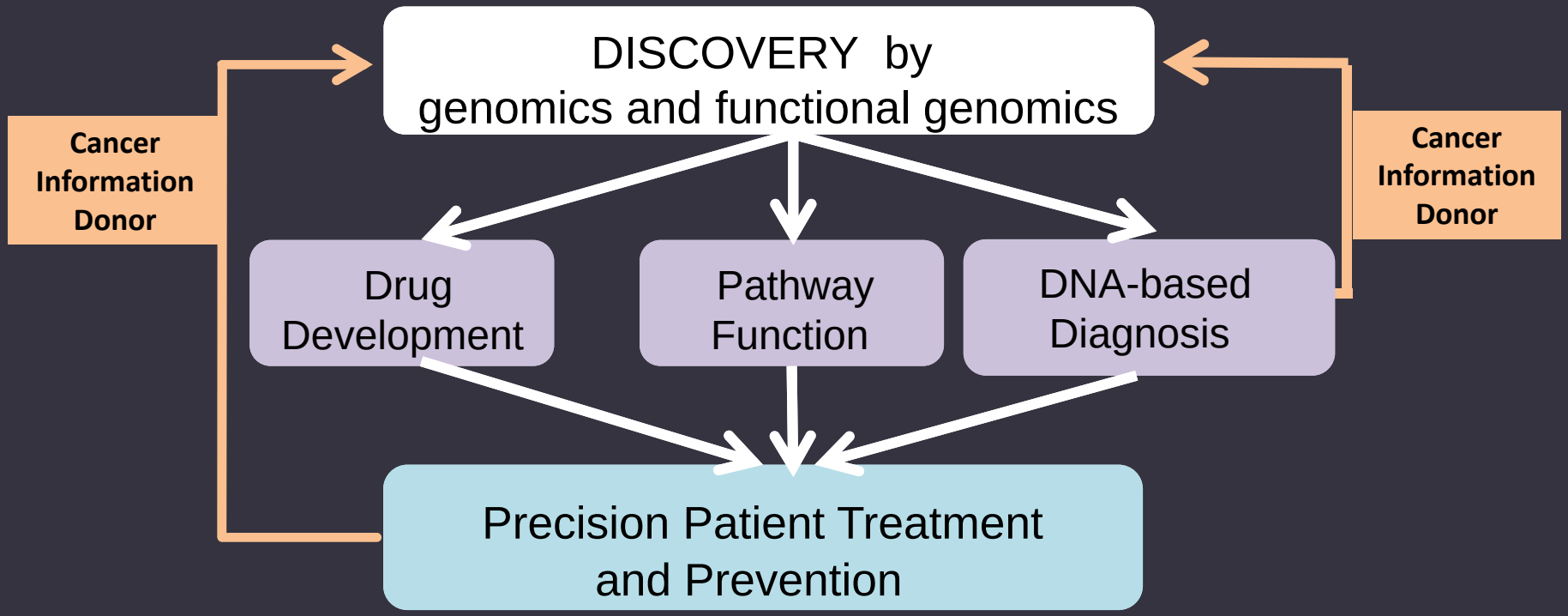
dbGAP will have serious scaling issues

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

photo: Josh Lewandowski