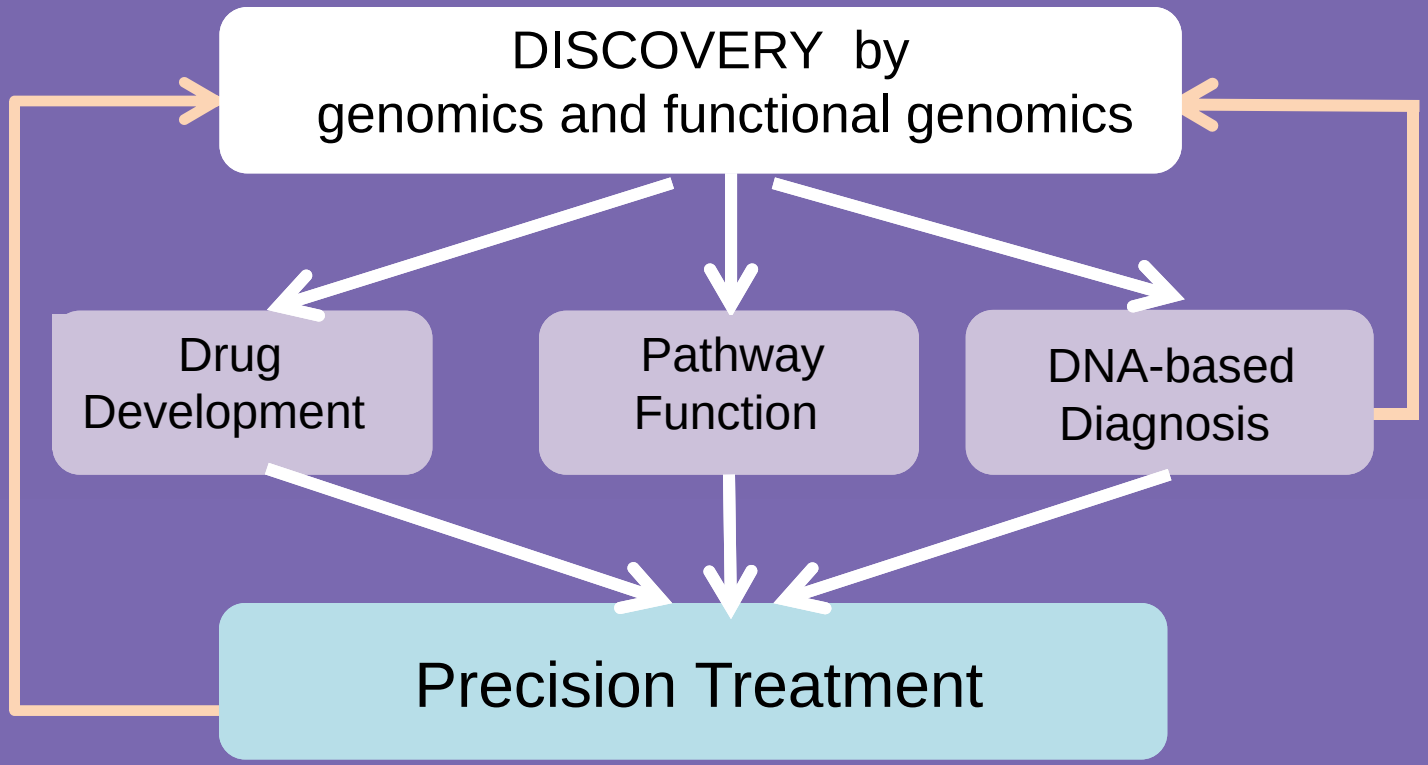
National Cancer Institute

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

NCI Center for Cancer Genomics

Our Mission is to develop and apply cutting edge genome science to better treat cancer patients



2012 and beyond NCI genomics

Now is the time to ask:

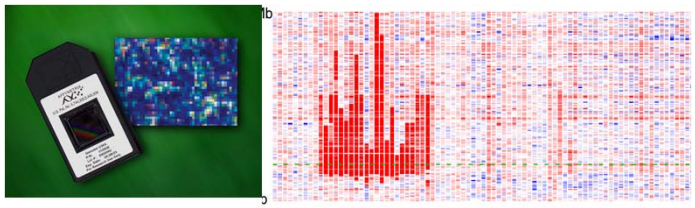
- What are the key science opportunities?
- What is the best and speediest path to the clinic?
- What are the bottlenecks: discovery -> trial design -> SOC?
- How will we capture clinical sequencing to drive further discovery?

CONTEXT

Early Genome Era (2001-2005)

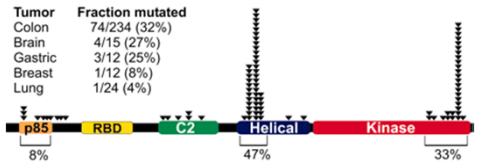
Global views in normal and disease

Microarrays (DNA, RNA),
DNA Re-sequencing
RNAi



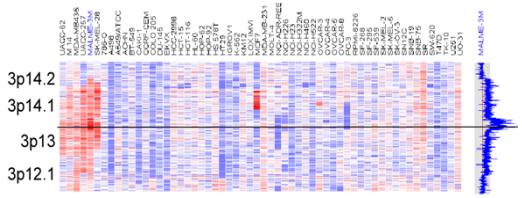
Systematic sequencing of pathways, gene classes

BRAF	Sanger	(Davies et al. 2002)
PIK3CA	Hopkins	(Samuels et al. 2004)
EGFR	Boston	(Paez et al. 2004, Lynch et al. 2004)



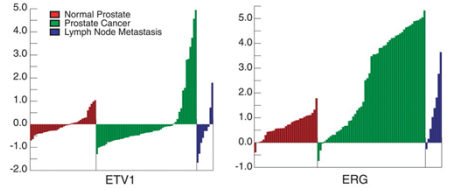
Systematic microarray studies and integrative genomics

MITF/melanoma (Garraway et al. 2005)



Systematic translocation discovery from genomic data

ETS/prostate Michigan (Tomlins et al. 2005)



Eric Lander: TCGA Symposium 11/11

Recommendation for a Human Cancer Genome Project

National Cancer Advisory Board
Working Group on Biomedical Technology
(Feb 2005)

CO-CHAIRS

Eric Lander (Broad Inst of MIT and Harvard)

Lee Hartwell (Fred Hutchinson)

MEMBERS

David Baltimore (Caltech)

Anna Barker (NCI)

Joan Brugge (Harvard)

Brian Druker (Oregon)

Geoffrey Duyk (TPG Ventures)

Chris Logothetis (M.D. Anderson)

John Niederhuber (Wisconsin)

Edward Penhoet (Moore Foundation)

Kathleen Schlom (NCI)

Bennett Shapiro (formerly Merck)

Ellen Sigal (Friends of Cancer Research)

Dinah Singer (NCI)

Margaret Spitz (M.D. Anderson)

Bruce Stillman (Cold Spring Harbor)

Harold Varmus (Memorial Sloan-Kettering)

Bert Vogelstein (Johns Hopkins)

Ralph Weissleder (Massachusetts General)

(Archival slide from 2005 courtesy of Eric Lander)

NCAB Working Group Report (2005)

Cancer Genome: Premises

(1) Cancer is a genetic disease

(2) Cancer is a highly heterogeneous disease

(3) Cancer is an understandable disease

(4) Systematic understanding would have major implications for

- Identification of cellular pathways that underlie cancer
- Improved selection of therapeutic targets
- Resolution of cancer into more homogeneous groups
- Faster and more efficient clinical trials
- Improved applications of drugs
- Design of epidemiological studies
- Identification of markers for early detection

(5) We are still ignorant about many key aspects of the genetic basis of cancer

(6) Systematic understanding of the cancer genome is

- technologically feasible within the next decade
- reasonable cost in context (requires 3% budget increase)

Archival slide from 2005 Courtesy of Eric Lander

NCAB Working Group Report (2005)

Human Cancer Genome Project: Goal

Comprehensive description of genetic basis of all major cancer types

Identify all genomic alterations significantly associated with all major cancer types

(1) creating large collection of appropriate, clinically annotated samples from all major types of cancer

(2) completely characterizing each sample in terms of:

- all regions of genomic loss or amplification
- all chromosomal rearrangements
- all mutations in coding regions of all human genes
- all regions of aberrant methylation
- complete gene expression profile, and other appropriate technologies.

NCAB Working Group Report (2005)

Human Cancer Genome Project: Goal

Comprehensive description of genetic basis of all major cancer types

Identify all genomic alterations significantly associated with all major cancer types

(1) creating large collection of appropriate, clinically annotated samples from all major types of cancer

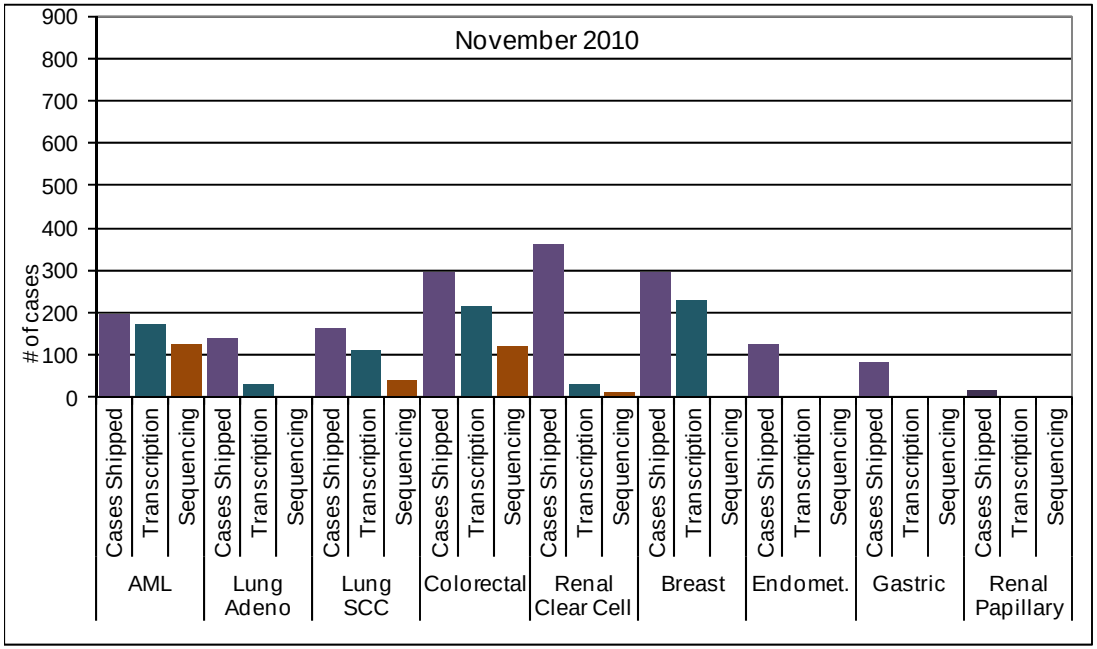
(2) completely characterizing each sample in terms of:

- all regions of genomic loss or amplification
- all chromosomal rearrangements
- all mutations in coding regions of all human genes
- all regions of aberrant methylation
- complete gene expression profile, and other appropriate technologies.

Outside concerns:

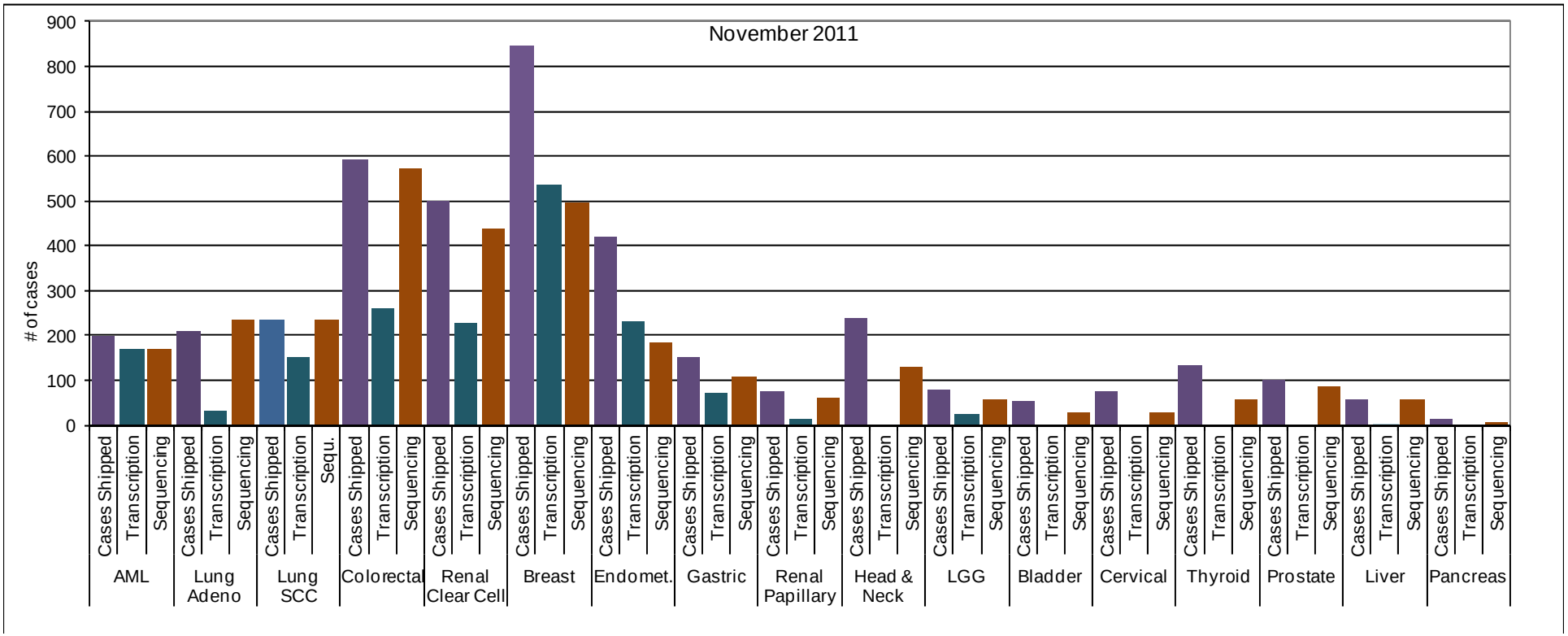
- ***Cancer genes all known***
- ***Cancer so complicated, never know genes***
- ***Cost too high***

Progress to Date

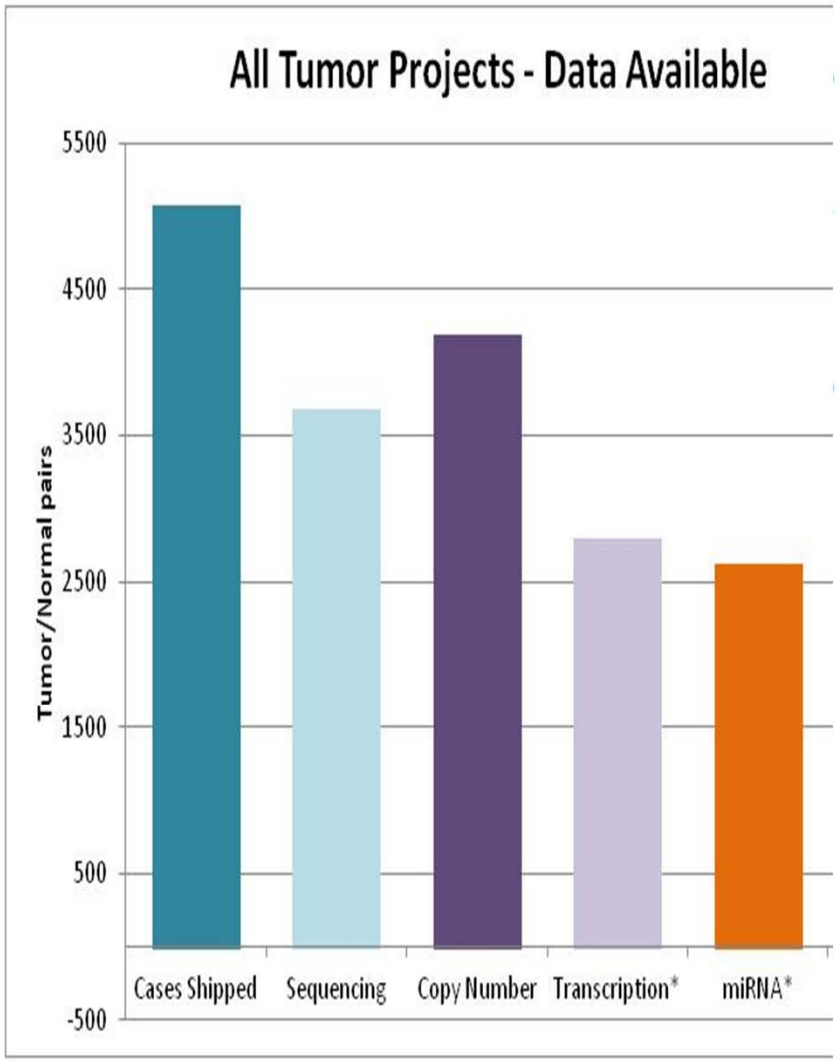


TCGA Project as Example

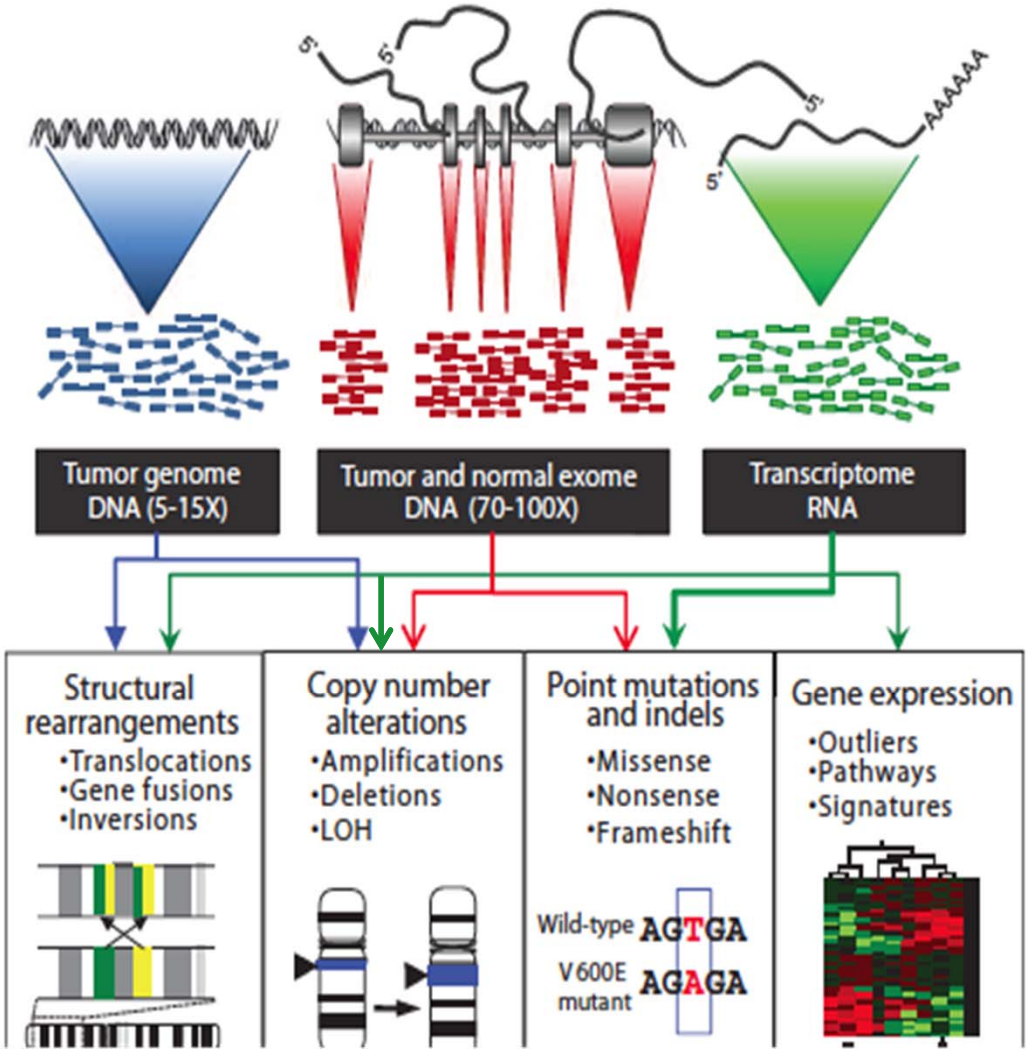
Nov 2010 Status
Vs.
Nov 2011 Status



TCGA data November 2011



Different sequencing data-types identify different genomic lesions: often mutually supportive



Roychowdhury et al, 2011 Dec 1 Science Transl. Med

2012: The Year of TCGA Post-Pilot Publications

- Colorectal Cancer
- Acute Myeloid Leukemia
- Breast Cancer
- Endometrial Cancer
- Kidney Clear Cell
- Lung Adeno/Squamous
- Head and Neck Cancer
- Etc.

TCGA: Whole Genome Deep Sequencing

Tumor Type	In Progress	Completed
GBM	-	22
Colorectal	15	5
Renal	-	10
Breast (triple negative)	1	20
AML	-	49
Ovarian	7	13
Endometrial (serous type)	28	2
LUSC	1	19
LUAD	14	6
Total	*66*	*146*

Updated Nov 1, 2011

Novel driver genes have been illuminating: Link between cancer genetics and epigenetics (red)

- *PBRM1* – Renal cell carcinomas
- *EZH2, MEF2B* – Lymphomas
- *KCNJ5* – Adrenal adenomas
- *DNMT3A, SF3B1, SRSF2, U2AF35* – Leukemias
- *MLL2, MLL3* – Medulloblastomas
- *ARID1A, PPP2R1A* – Ovarian cancers
- *DAXX, ATRX* – Pancreatic endocrine tumors
- *BAP1, TTRAP* – Melanomas
- IDH1, 2 – Gliomas
- *CIC, FUBP1* – Oligodendrogliomas
- *MED12* - Leiomyomas

List compiled by B. Vogelstein

Cancer Genomics is valuable and here to stay
Specific projects will sunset

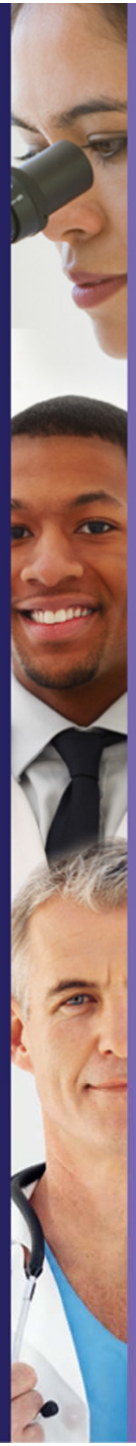
TCGA The Cancer Genome Atlas	FY11	FY12
Appropriated	$31.0	$31.3
ARRA	$118.6	$21.6

TARGET Therapeutically Applicable Research to Generate Effective Treatments	FY11	FY12
Appropriated	$2.3	$1.0
ARRA	$3.0	$9.0

CTDD Cancer Target Discovery Net	FY11	FY12
Appropriated		$10.0

CGCI Cancer Genome Characterization Init.	FY11	FY12
Appropriated	$5.7	$5.7

National Cancer Institute



U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

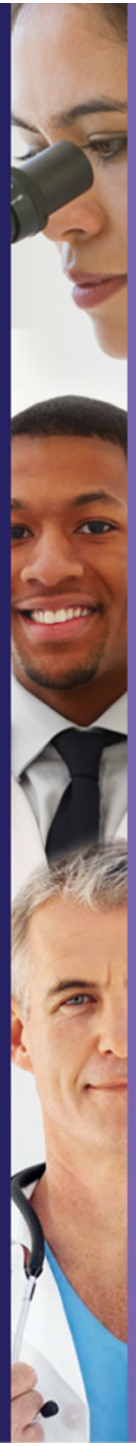
National Institutes of Health

2012 – forward

Immediate science opportunities in 2012:

- Rare tumors - 10 types by 50 samples now launching
- Minority samples – launching with CRCHD
- Key mouse model tumors added (breast , NSCLC, melanoma, prostate)
- FFPE bake-off and hardening ; smaller sample inputs

National Cancer Institute



U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

RNA profiling

Can/should become routine phenotyping in clinical settings

- From arrays to RNA-seq [CLIA]
- From Frozen samples to FFPE
- From big samples to < 100 cells

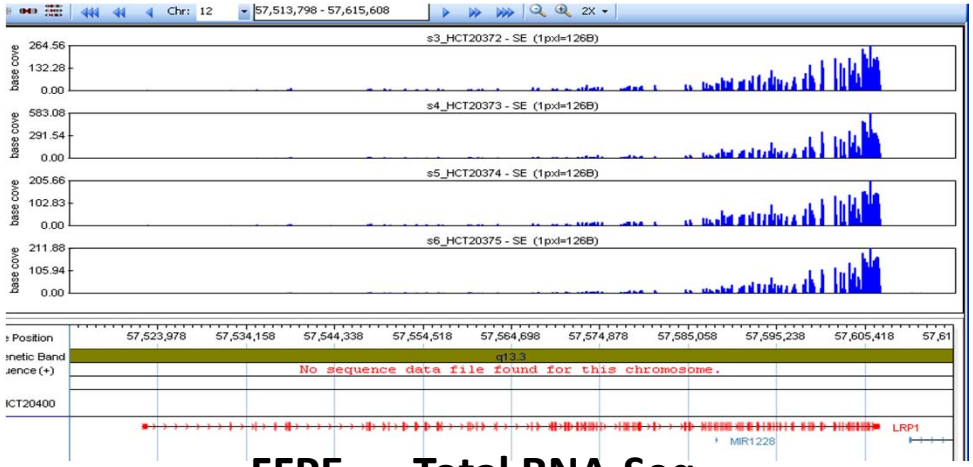
Formaldehyde Fixed Paraffin Embedded samples
Can now produce RNA (and DNA) data of exceedingly high quality

On the path to:

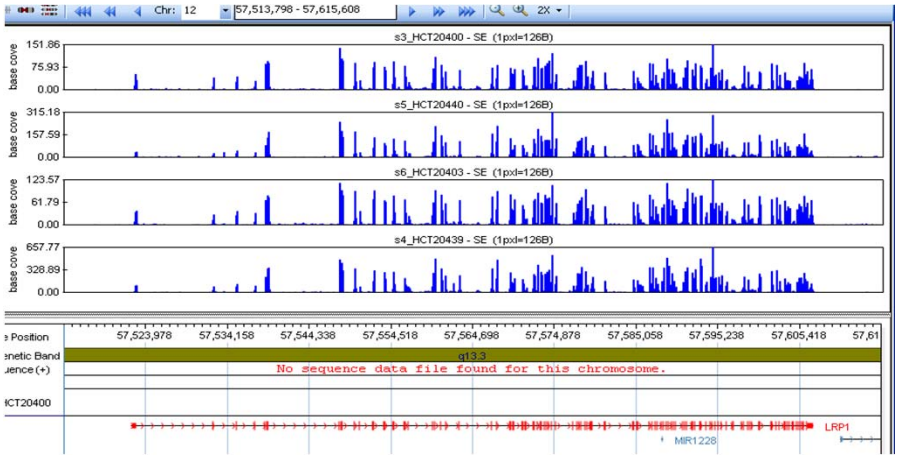
- > rarer tumors
- > minority patient groups
- > tumor progression
- > drug resistance
- > and clinical DNA sequencing

Big gene = LRP1: length 14,905 bases

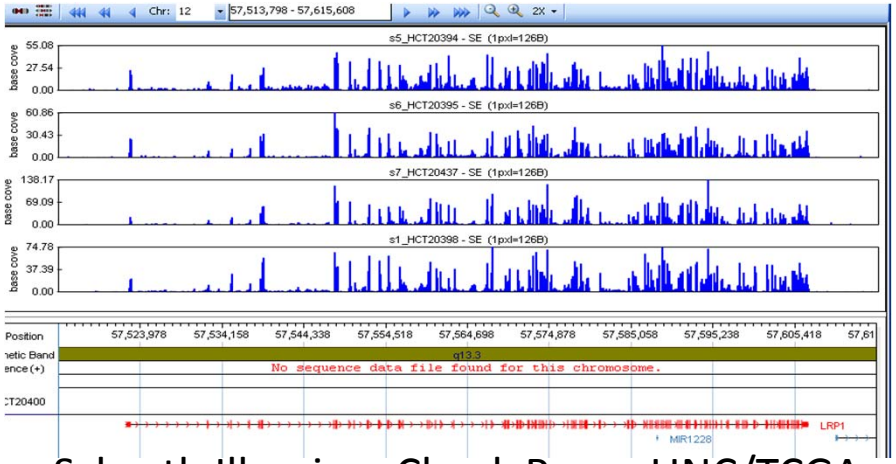
Fresh - mRNA-Seq



Fresh - Total RNA-Seq

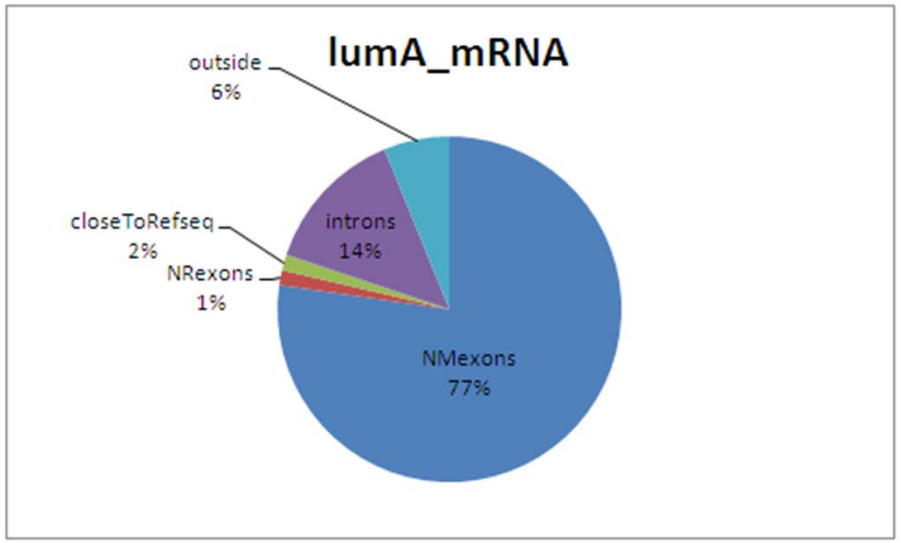


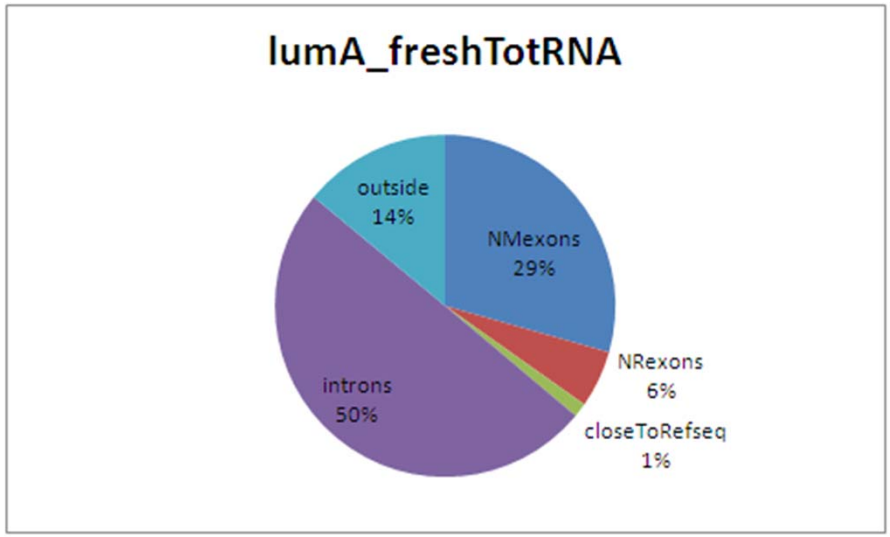
FFPE - Total RNA-Seq

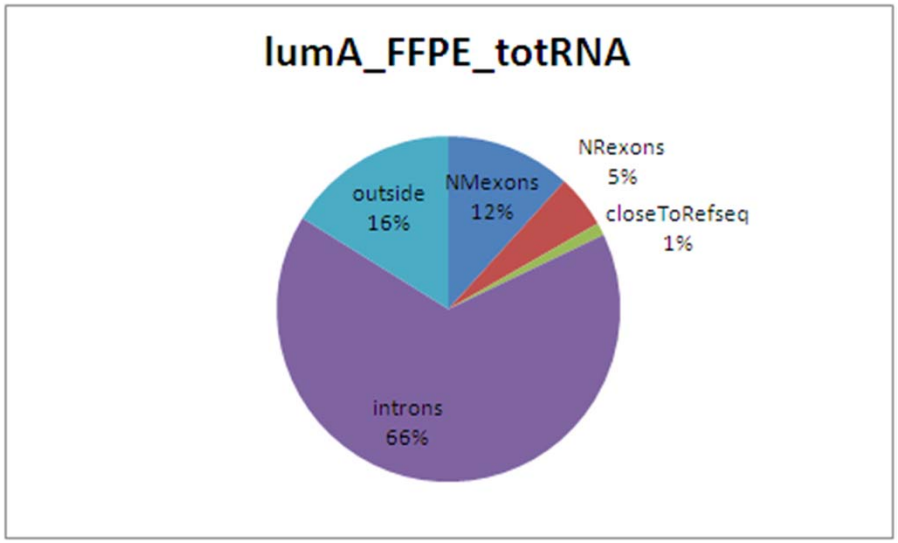


Gary Schroth Illumina; Chuck Perou UNC/TCGA

Total RNA-Seq with DSN Detects Primary Transcripts







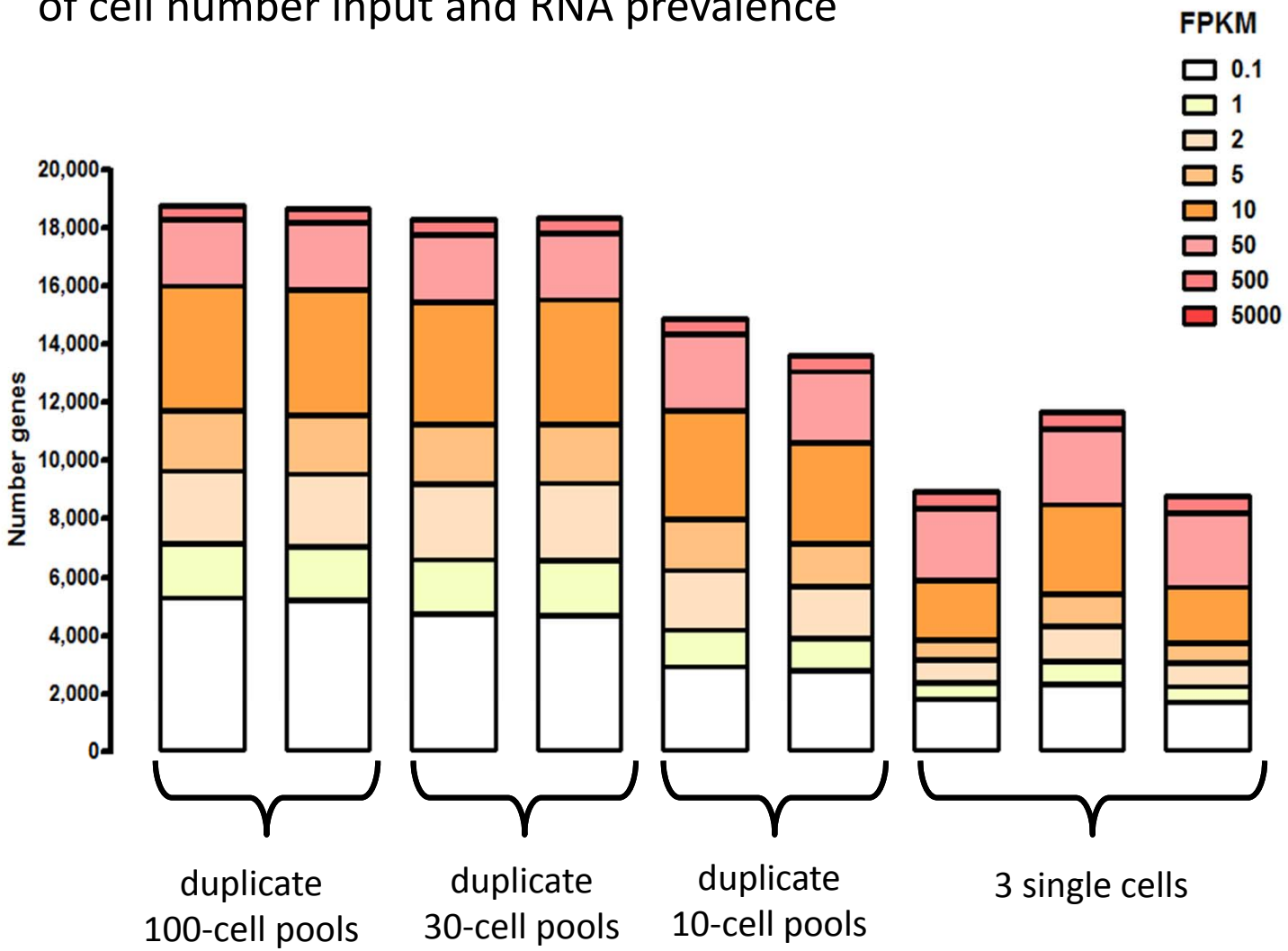
Note reduced signal from Exons in FFPE

Very small numbers of cells for genomic assays

Proof of principle for RNA

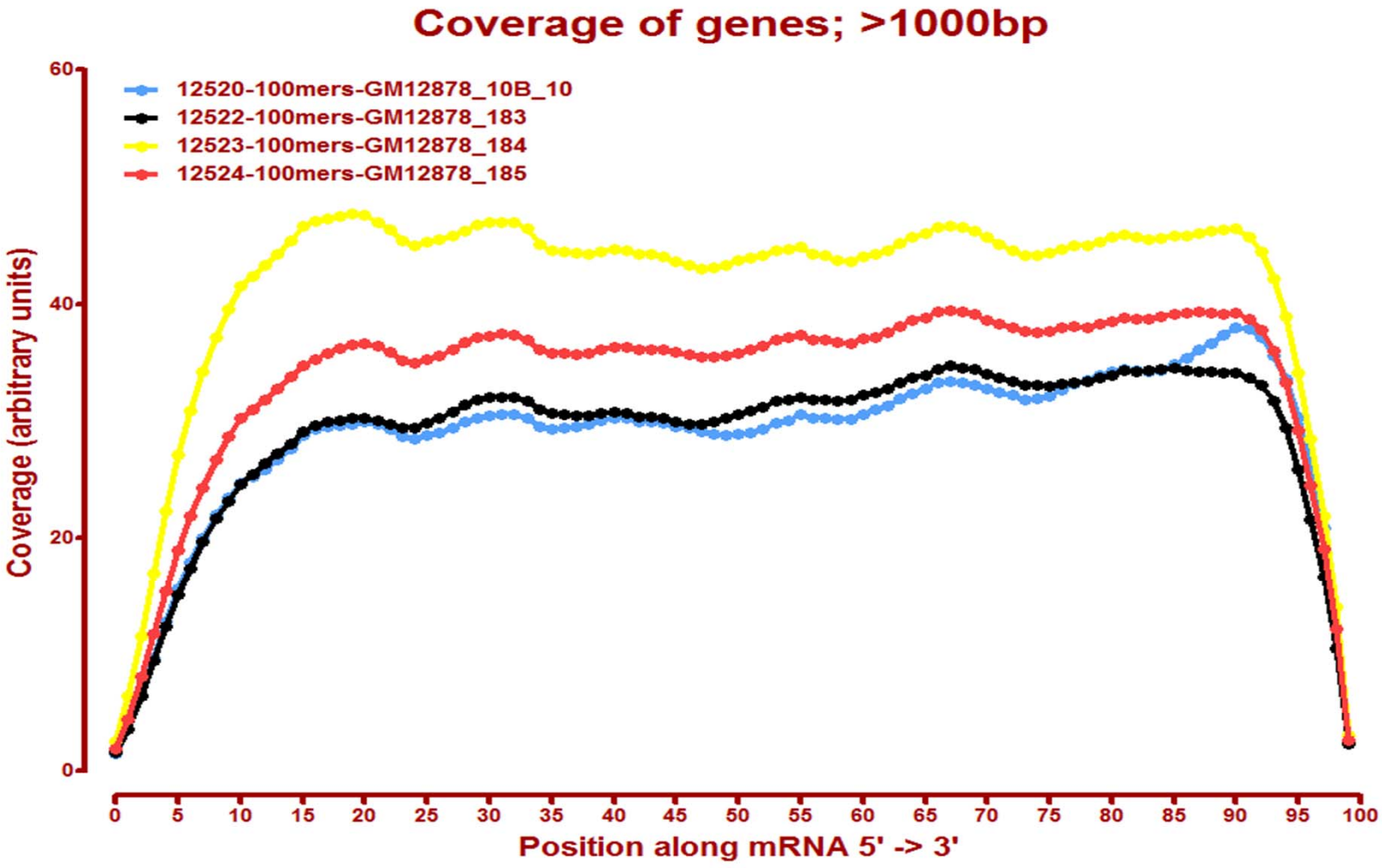
- Relevant for the clinic in early disease
- Relevant for mixed cellularity
- More flexible and general than custom panels, competitive in cost

Genes (Gencode V 7 ~50K) detected by RNA-seq as a function of cell number input and RNA prevalence



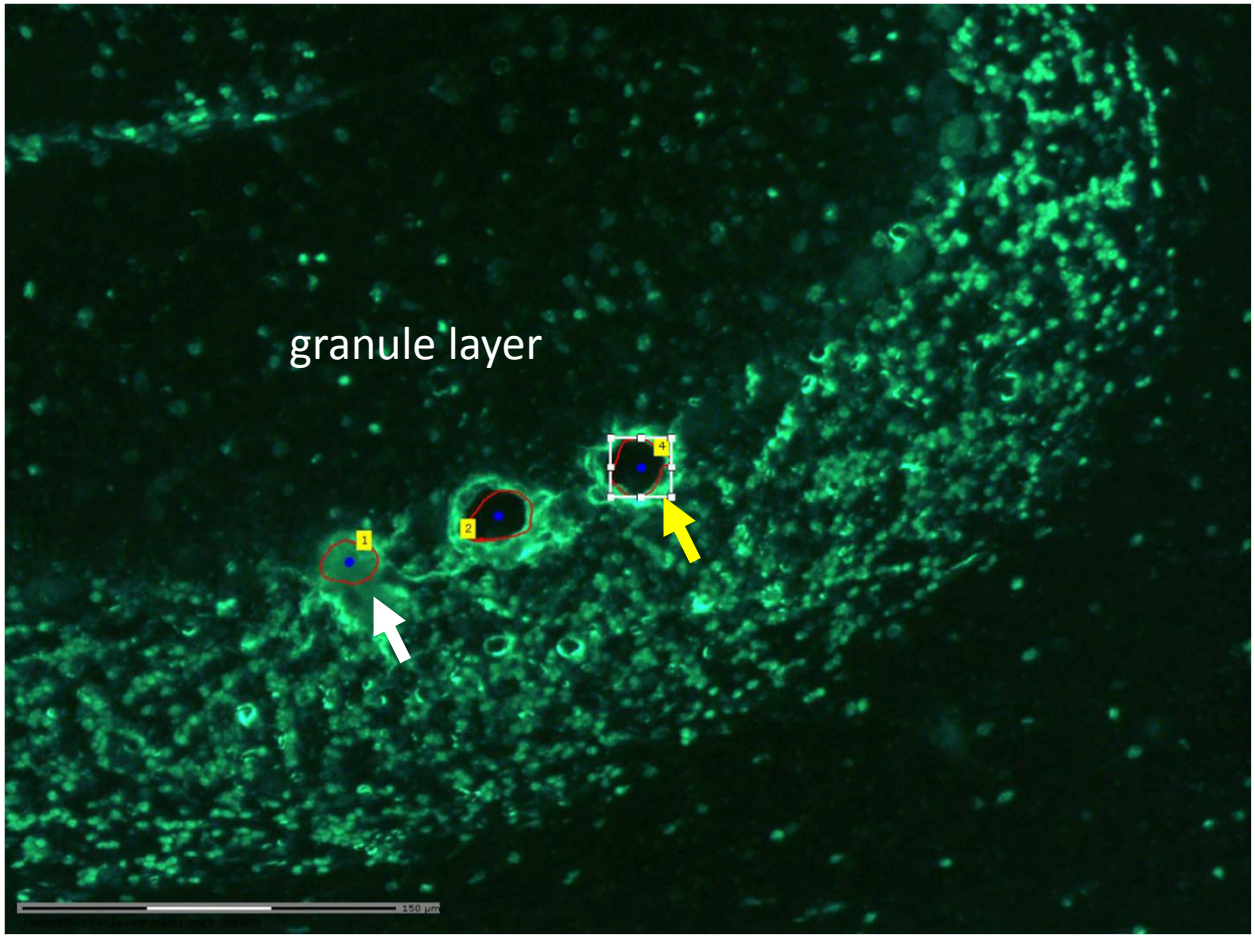
unpublished, B. Williams and G. Marinov, Wold lab Caltech

Transformed B-cells RNA: 10-cell pool and 3 single cell RNA-Seqs mRNA coverage is good across transcript length



unpublished, B. Williams and G. Marinov, Wold lab Caltech

Laser capture of individual cells and multi-cell areas

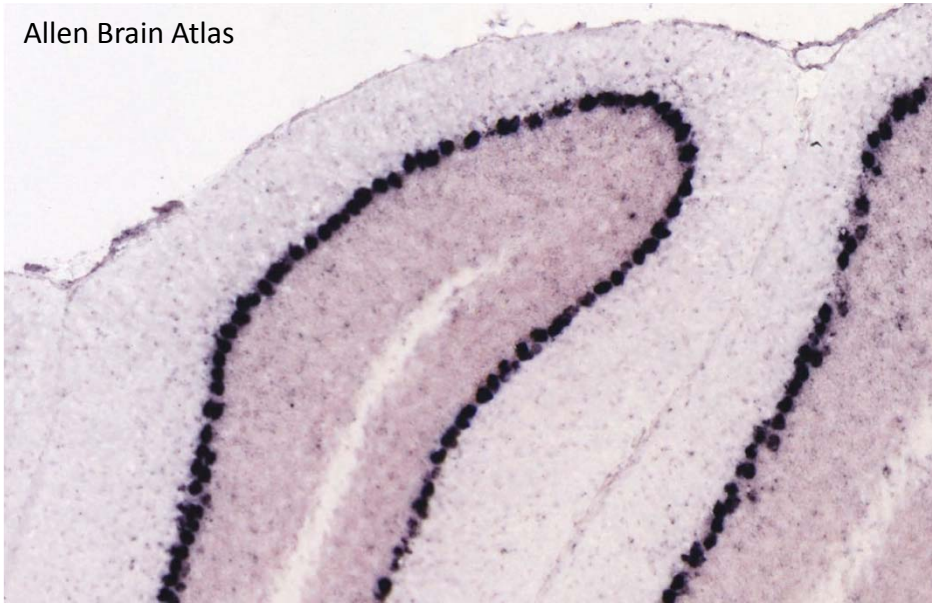


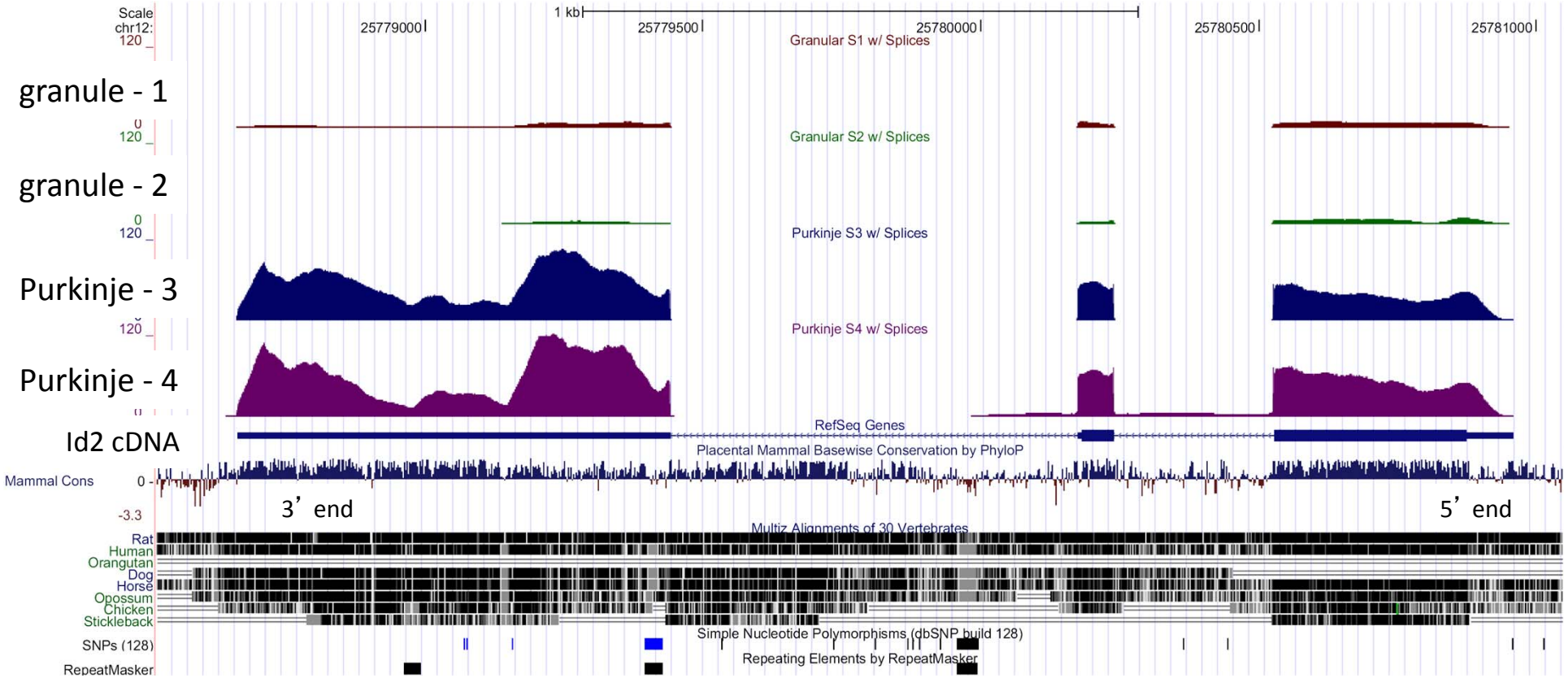
Id2 – Purkinje cell specific

mRNA = 1289 nt

RPKM Purkinje = 1024

RPKM Granule = 64





State-of-the-art

FFPE appears nearly there; 170 tumor test in progress

- Matched pairs with TCGA Frozen data
  - Both DNA and RNA
- Expect results in mid 2012

Small cell numbers, dissected or sorted, fresh or frozen

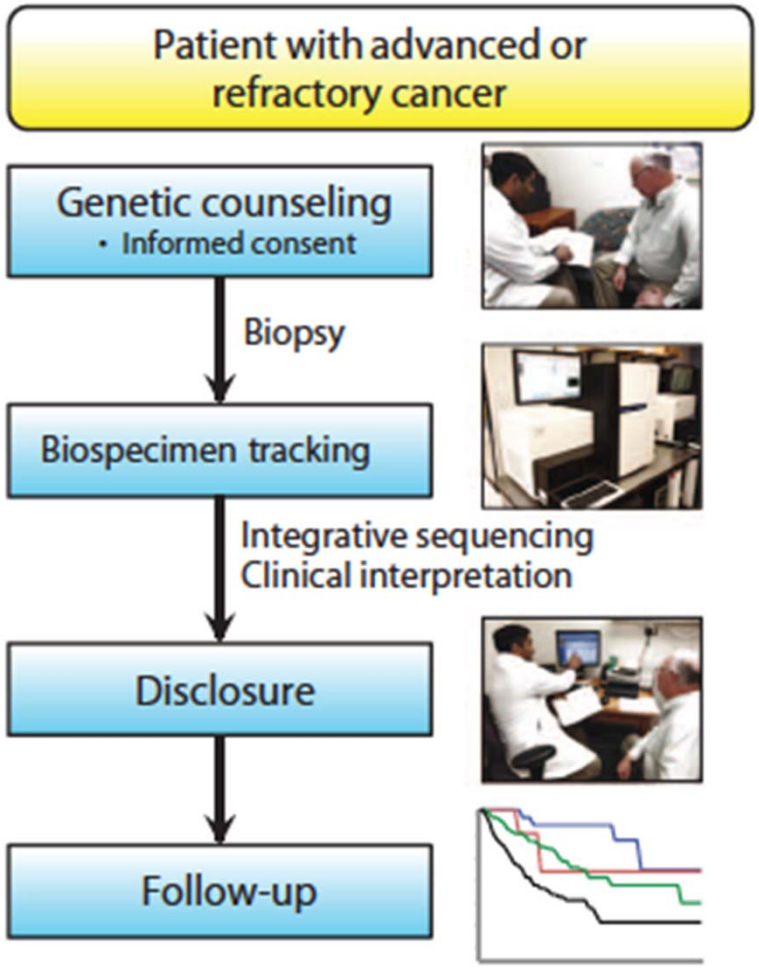
- Ready to generalize in 2012

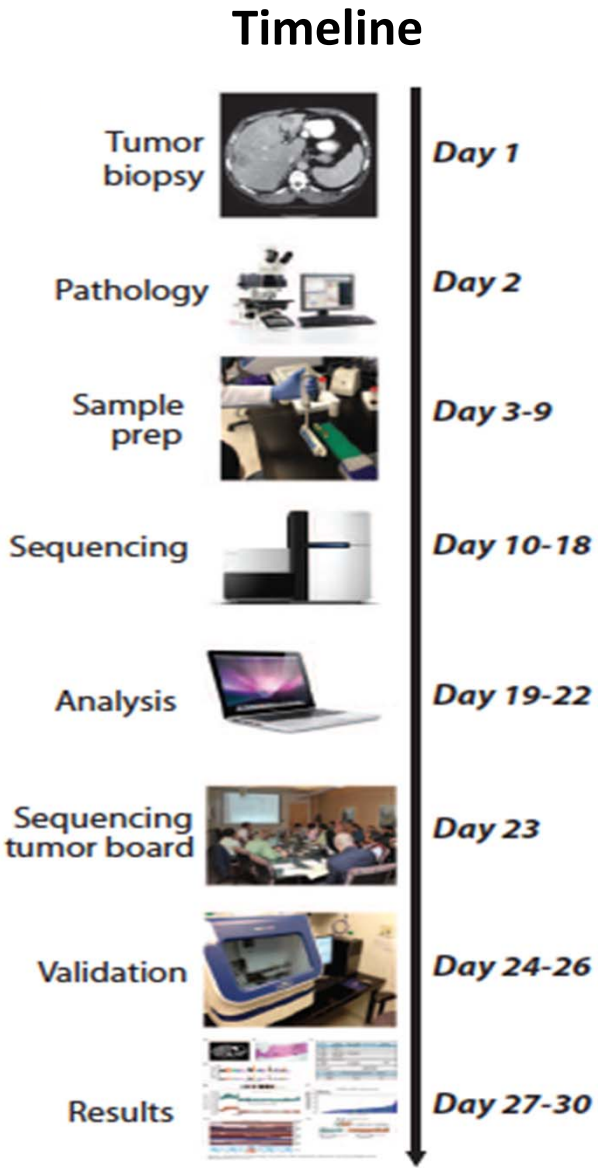
Next:

- Put them together: FFPE on tiny amounts

Path forward to Clinical Sequencing

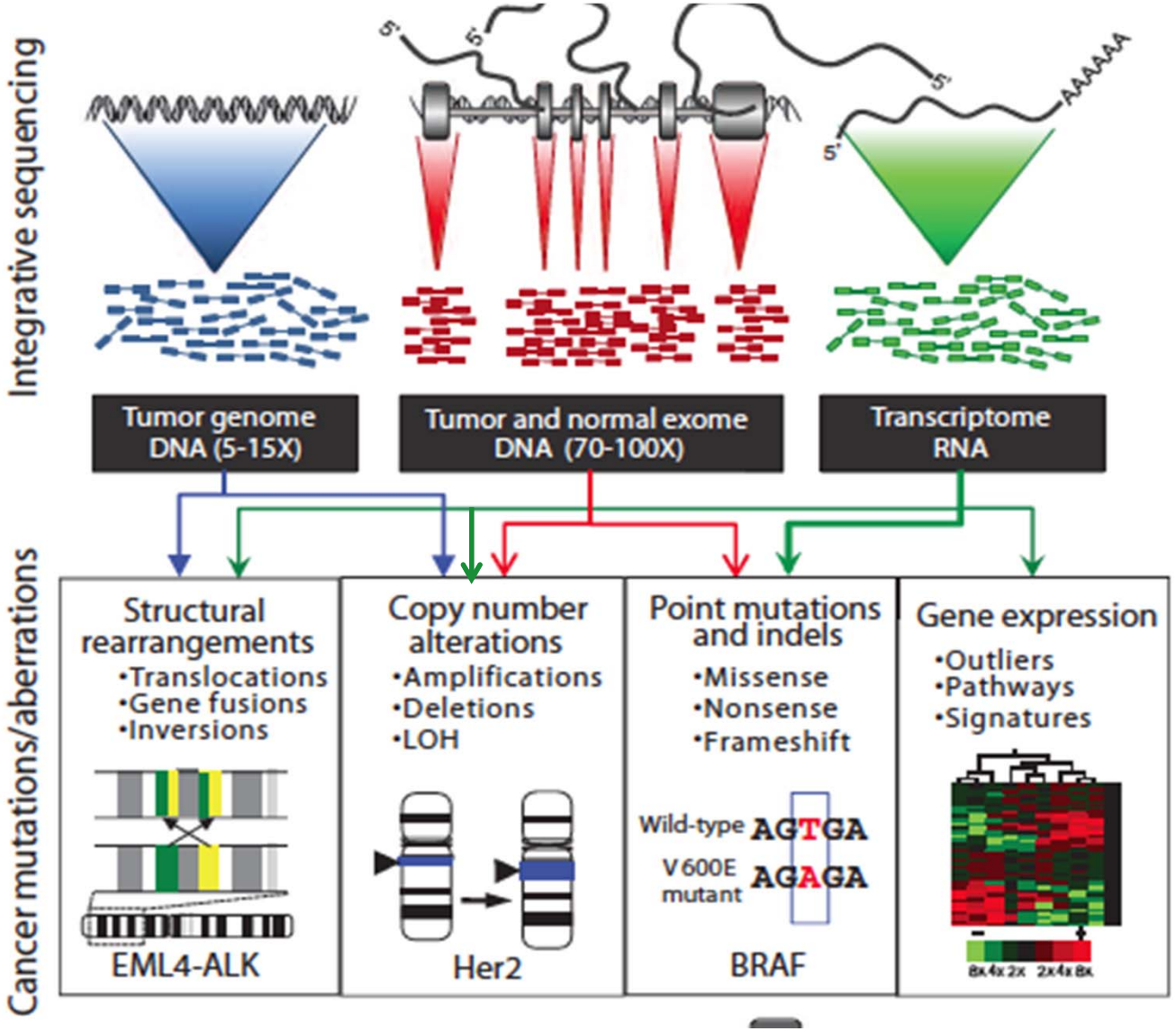
Arul Chinnaiyan and colleagues: Dec 1, 2011





Roychowdhury et al, 2011 Dec 1 Science Transl. Med

3 different sequencing data-types help to identify relevant tumor genome lesions



Roychowdhury et al, 2011 Dec 1, Science Transl. Med

Tumor Board Review
Metastatic Colon Cancer: Liver biopsy genomics

Gene	Point mutation	Abs. copy number	Structure	Gene expression
NRAS	p.Q61L			
TP53	p.R175H	1.43 (loss)		
AURKA	p.W313R	2.89 (gain)		
FAS	p.T214N	allele of uncertain meaning		
MYH11	p.V1527A			
CDK8		7.02 (gain)		High
EGFR		3.35 (gain)		
PPP2R3B			Fusion with ASMTL-AS1	

Gene	Amino acid position	Status
KRAS	G12	WT
BRAF	V600	WT

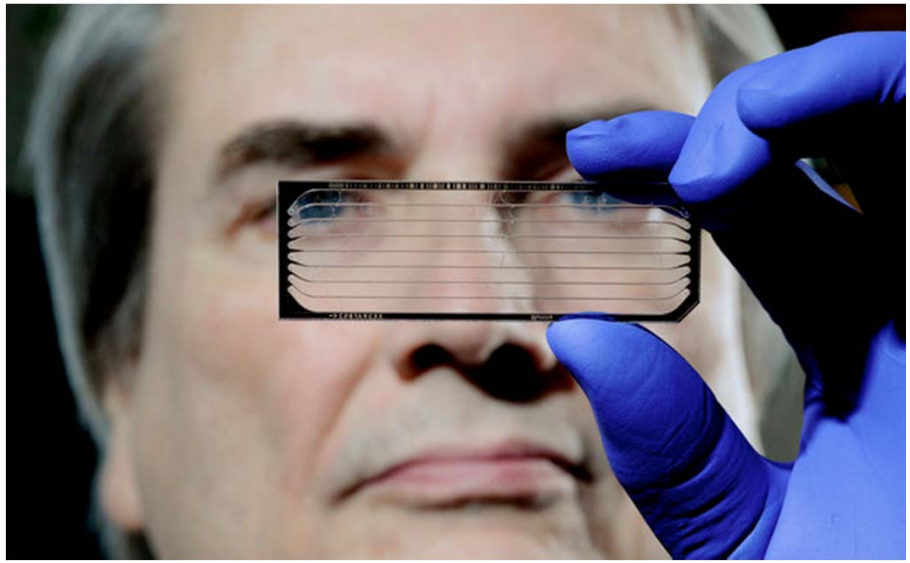
- NOT KRAS expected for CRC: Implications for trial eligibility MEK, PI3K inhibitors suggested
- Patient failed AURK inhibitor; might have been predicted & avoided
- CDK inhibitors suggested

Expected for CRC, but not in this patient

2012 and beyond

The New York Times
Expect the World®

DNA Sequencing Caught in Deluge of Data



Kathy Kmonicek
Published: November 30, 2011

"World capacity is now 13 quadrillion DNA bases a year, an amount that would fill a stack of DVDs two miles high"

Data access, storage, usability

A problem much bigger than TARGET and TCGA and even NCI

First consider storage and access and what we are doing about it

Cancer Genomics Hub: Due to accept and serve data Dec 15th 2011

- Stores BAM & VCF for TCGA, TARGET and CGAP/CGCI projects
- Designed for 25,000 cases with average of 200 gigabytes per case
- 5 petabytes (5×10^{15}) total, scalable to 20 petabytes
- co-location opportunities

Bold move by NCI

Such centers are few, expensive

Modified from D. Haussler

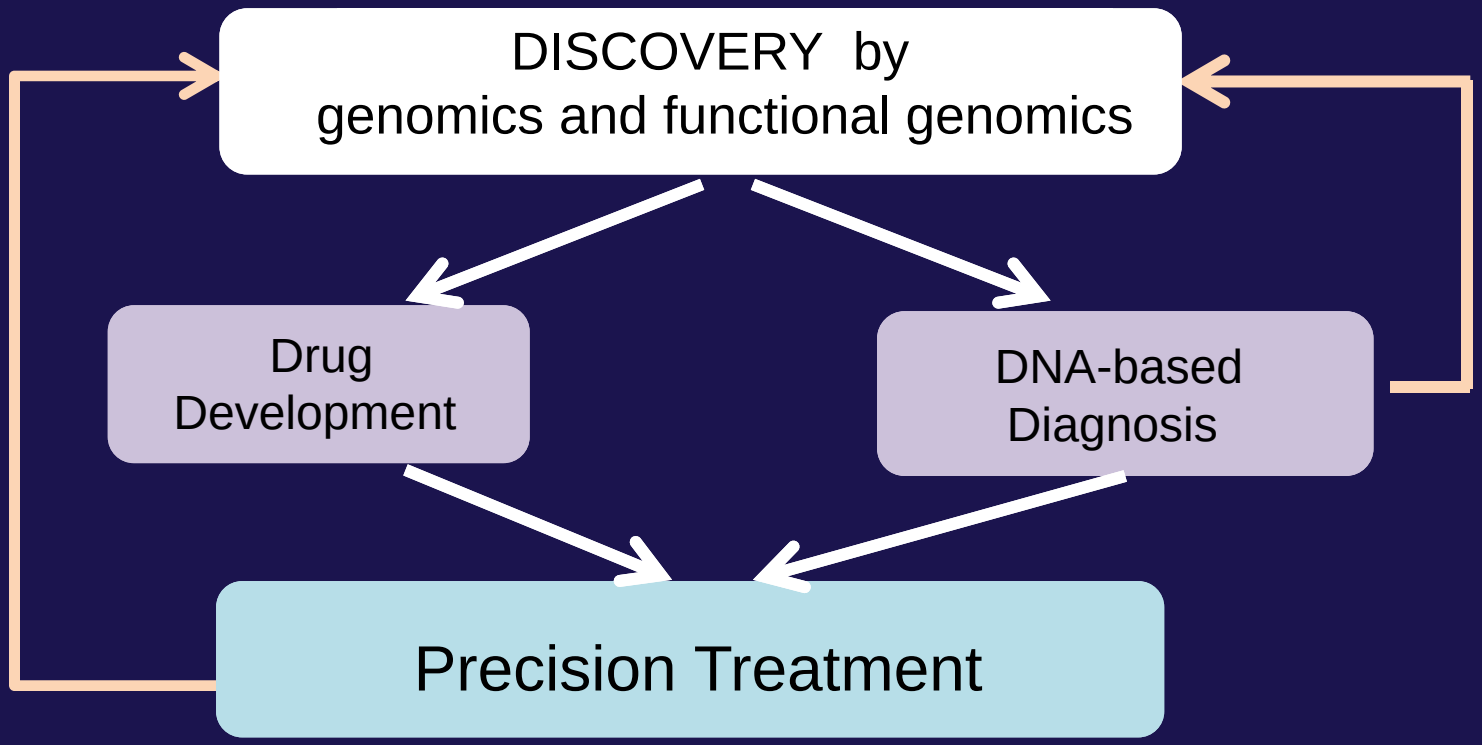
More New York Times

Essay
Computer Scientists May Have What It Takes to Help Cure Cancer

By DAVID PATTERSON
Published: December 5, 2011

“The war against cancer is increasingly moving into cyberspace.”

Clinic-to-Discovery
The opportunity we cannot afford to miss



Achieving Completeness: *Clinical can/should drive more and deeper discovery*

Global Cancer Alliance

Shared knowledge base where patients can choose to contribute their genomic data, clinical data

Many challenges, but essential

E. Lander, TCGA talk 2011

More Open and More Efficient Data Access is a Goal

Why we need it

Power of sample number.

Discover relationships never dreamed of.

What are the barriers?

Re-identification concerns in face of cybertech advances

Are GINA protections adequate?

Learn and appreciate the patient point of view

Open access to data goals for 2012

1. Assist advocacy community: make possible
   for patient choice for more open de-identified use

2. Document cancer patient community norm views

3. Investigate “data user” legal contract
   for de-identified data access

   Re-examine the risk /benefit equation in light
   of advances and protections now in place

   Investigate if unauthorized sequencing DNA of another
   person should/could be made illegal

Discovery-to-Clinic

Clinical trials organized around mutations and pathways

Fewer and better trials by NCI: set quality bar very high

Drug availability for genome-guided trials

Your advice ?

National Cancer Institute

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

Special Thanks to:

Dr. Kenna Shaw
Dr. Daniela Gerhard
and
our NHGRI collaborators