The Cancer Genome Atlas (TCGA)
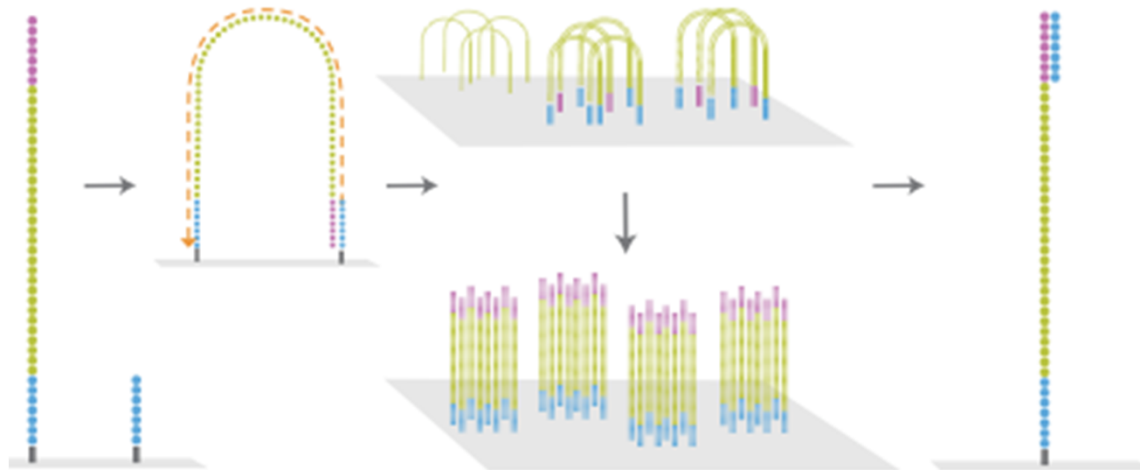
NHGRI Genome Sequencing Centers

Report to
National Cancer Advisory Board
Mark Guyer, Ph.D.

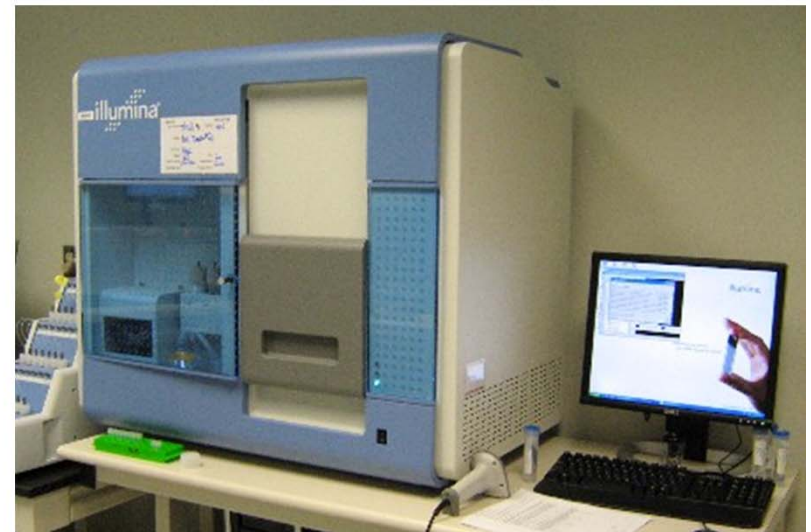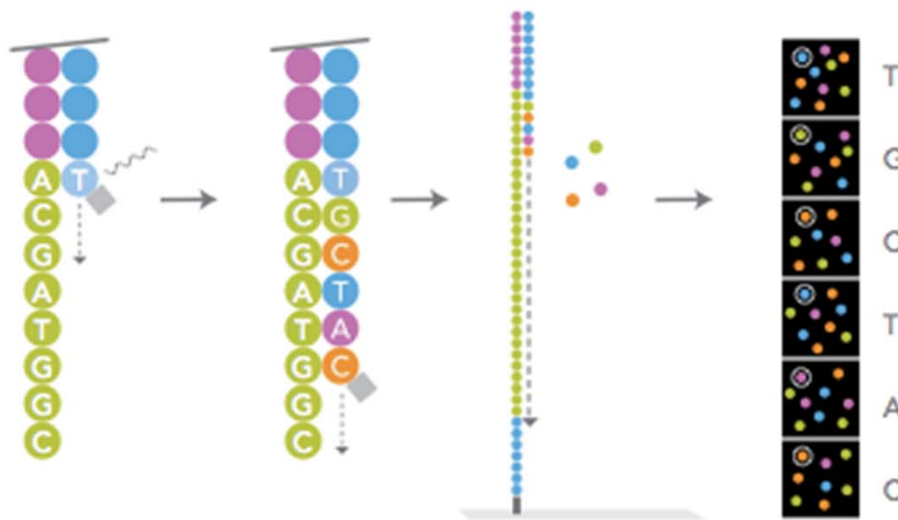**September 15, 2009**

"Next Gen" sequencing technology

"Solexa" (2006)

~1 Gb/wk

..

Illumina GA IIx (2009)

25 Gb/wk

3 Large-scale sequencing centers: The Broad Institute (Eric Lander)
Washington University (Richard Wilson)
Baylor College of Medicine (Richard Gibbs)

	ABI 3730	454	Illumina	ABI SOLiD	Helicos
Instruments	43	21	99	13	1
2008 Total	50Gb	350Gb	2,959Gb	454Gb	-
2009 To Date	10Gb	709Gb	13,126Gb	2,453Gb	19Gb
Phase	Production	Production	Production	Production	Prototype
Applications	Clone Seq Directed Seq Finishing	Viral Bacterial Fungal Metagenomics	Large Genomes SNP Discovery CNV Hybrid Selection ChIP	Large Genomes SNP Discovery CNV Hybrid Selection	ChIP Expression Barcode Counts SNP Discovery

* All projects, Gb = "good" bases by platform-specific definition

TCGA Sequencing production status

THE CANCER GENOME ATLAS

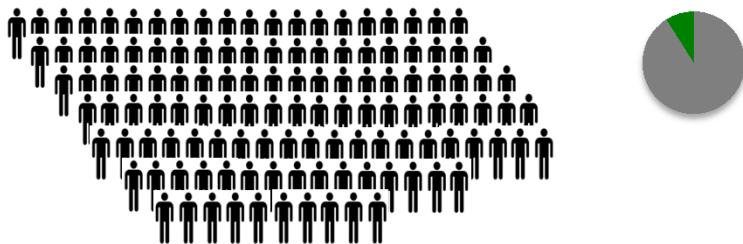## Glioblastoma multiforme

Whole Genome Sequencing

10 complete   2 in progress
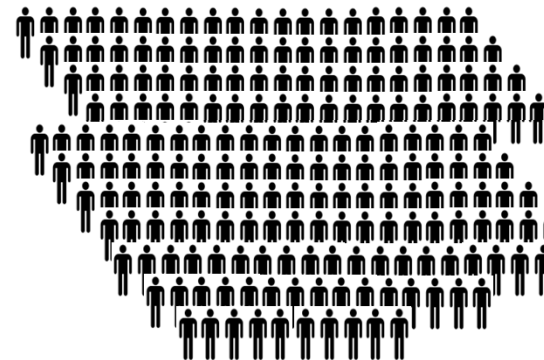
Targeted Sequencing

~144 cases   ~1300 genes

Ovarian serous

Whole Genome Sequencing

12 complete

Targeted Sequencing

238 cases

2000 Genes
26 cases

6000 Genes
229 cases

Whole Exome
9 cases

August '09

Whole Genome Shotgun	6000 Gene Capture

#	Center	Case	coverage T	coverage N
1	Broad	TCGA-13-0751	35.0	34.9
2	Broad	TCGA-13-0725	35.2	37.0
3	Broad	TCGA-04-1371	44.2	42.4
4	Broad	TCGA-24-0982	39.3	42.7
5	Broad	TCGA-25-1319	43.1	43.4
6	WUGSC	TCGA-13-0890	38.6	28.8
7	WUGSC	TCGA-13-0723	39.9	29.7
8	WUGSC	TCGA-24-0980	34.0	43.3
9	WUGSC	TCGA-24-1103	30.1	35.4
10	WUGSC	TCGA-13-1411	32.6	14.3
11	BCM	TCGA-13-0720	38.3	38.0
12	BCM	TCGA-10-0927	36.6	36.3

Center	Cases Assigned	Samples Assigned	Samples Through First Pass Sequencing
Broad	95	190	190
WUGSC	94	188	185
BCM	49	98	98
Total	**238**	**476**	**473**

•10 Cases complete to full 30x T & N
•2 Normal samples in progress of Top-off

• Nearly all cases completed first pass (236/238)

➢ >8,000,000,000,000 nucleotides (**8 Terabases**) sequenced in 4 months
➢ **Unprecedented application of genomic sequencing to clinical specimens**
➢ Data analysis challenge: magnitude and complexity

OVARIAN

Coverage(T/N) Callable Purity Ploidy
31x / 30x **81%** **90%** **2.8**

Point Mutations

Rate/Mb **0.75**
Total   **1786**
Coding   **9**

MIS   5
STOP   1
INDEL ---

HIGHLIGHTS

| GENE	MUTATION	FUNCTION
TP53	Insertion	Tumor suppressor
EXOC6B	Missense	protein transport, exocytosis
ANKRD6	Missense	ankyrin
AHNAK	Missense	CNS development
C11orf52	Nonsense	?
GABRB3	Missense	GABA receptor

Lost BRCA1 germline indel

Chr. Aberrations

CNA Breaks **---**
TX-Inter   **9**
TX-Intra   **15**

HIGHLIGHTS

NF1-EFCAB5 fusion gene probably inactivating validated by RNA-seq
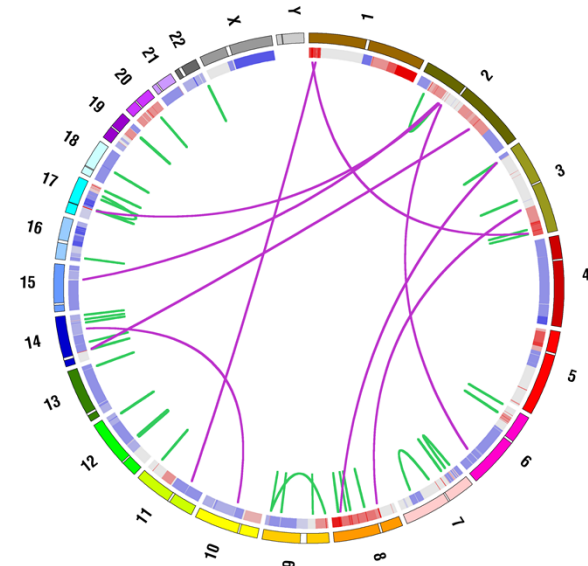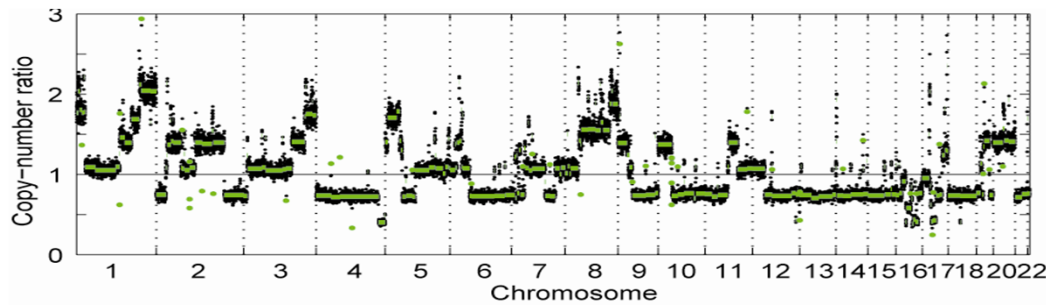
Courtesy of Gad Getz – Unpublished

Technical

- Unprecedented data production
- Platforms still improving, becoming more economical
- More attention to analysis, data sharing, data management
- Sample range – e.g., paraffin

Strategic

- Whole genomes vs. whole exomes
- Cancer types: Depth vs. breadth
- **Ready for bold goals for TCGA**