TCGA Glioblastoma Pilot Project: Initial Report June 2008

THE CANCER GENOME ATLAS

NCI-NHGRI Partnership

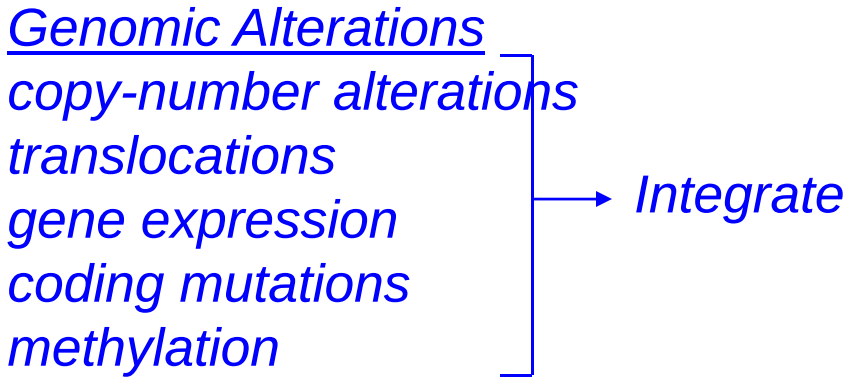
Cancer Genomics: Ultimate Goal

Create comprehensive public catalog

of all genomic alterations present at significant frequency

for all major cancer types

NCAB Report Feb 2005



NCI-NHGRI Partnership

TCGA Pilot Project

Launched, 4Q2006

Cancers:

- Glioblastoma multiforme
- Squamous cell lung cancer
- Ovarian cancer (serous cystadenomacarcinoma)

Goal:

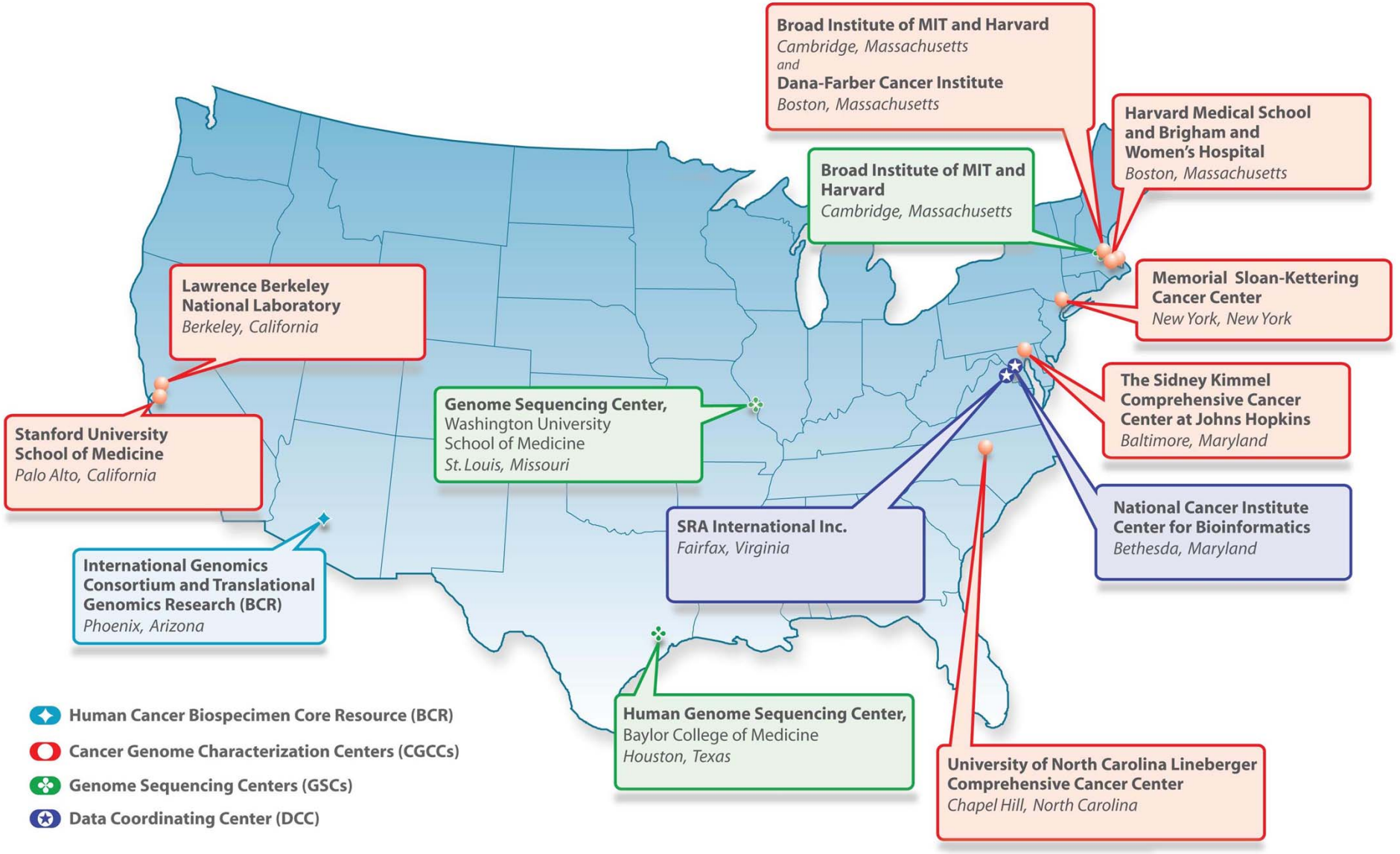
- Assemble high-quality samples of each type
- Characterize tumor genome by various approaches
- Rapidly share data with scientific community
- Compare and improve technologies
- Integrate and analyze data to illuminate genetic basis of cancers

Key questions posed at start of project

1. Can samples of adequate quality and quantity be assembled?
2. Can high-quality, high-throughput data be generated with current platforms?
3. How sensitive, specific and comparable are current platforms?
4. How can diverse data sets be integrated -- and what can be learned from integration?
5. Can recurrent events be distinguished from random background noise?
6. Can we identify new genes associated with cancer types?
7. Can we identify new subtypes of cancer?
8. Does new knowledge suggest therapeutic implications?
9. Can a network project drive technology progress in cancer?

TCGA Components

THE CANCER GENOME ATLAS



GBM Samples: Strict Criteria

(1) >80% tumor cell content

(2) <40% necrosis

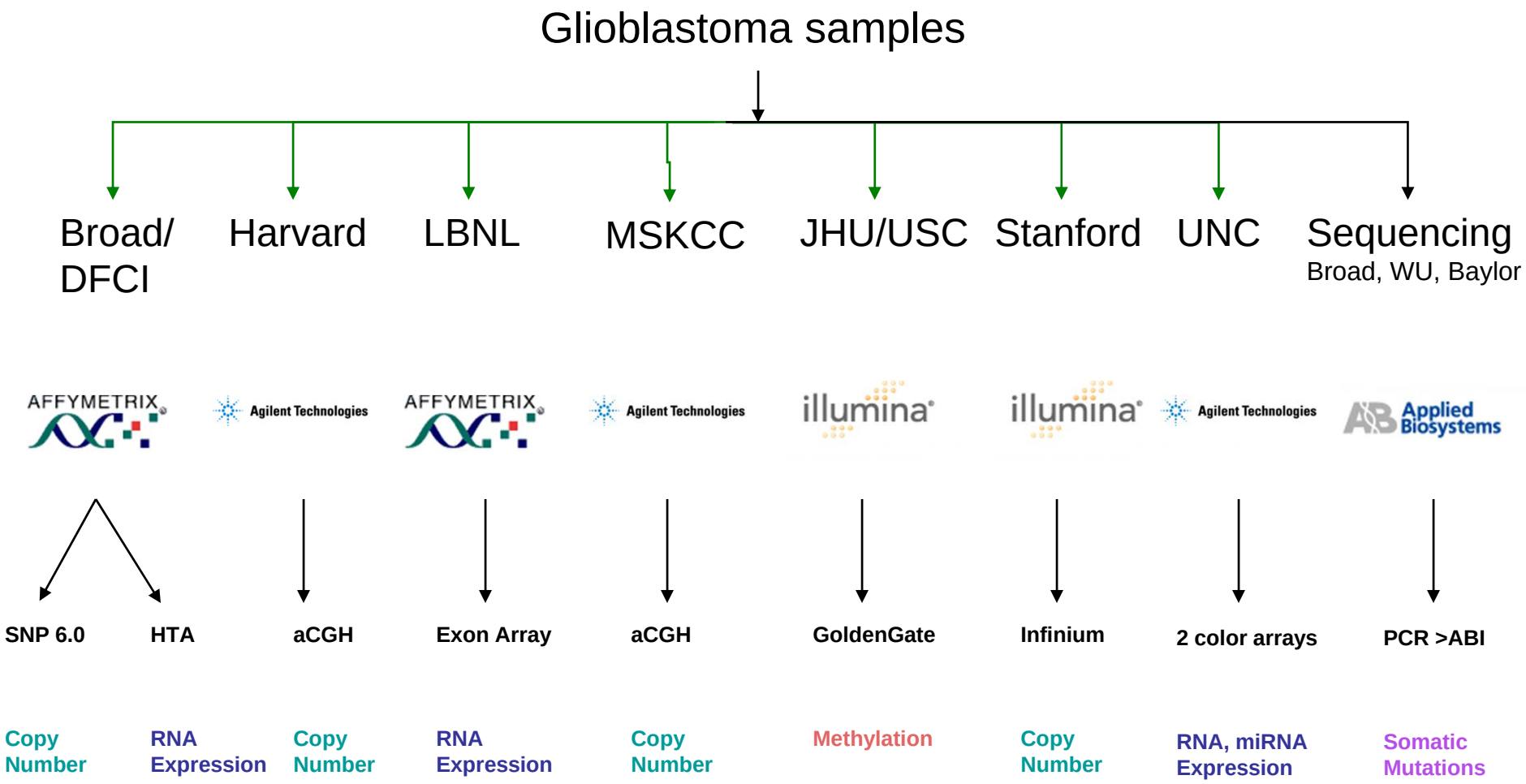
(3) Matched normal DNA sample

(4) Clinical annotation

(5) Appropriate informed consent

Current collection > 200 high quality tumors

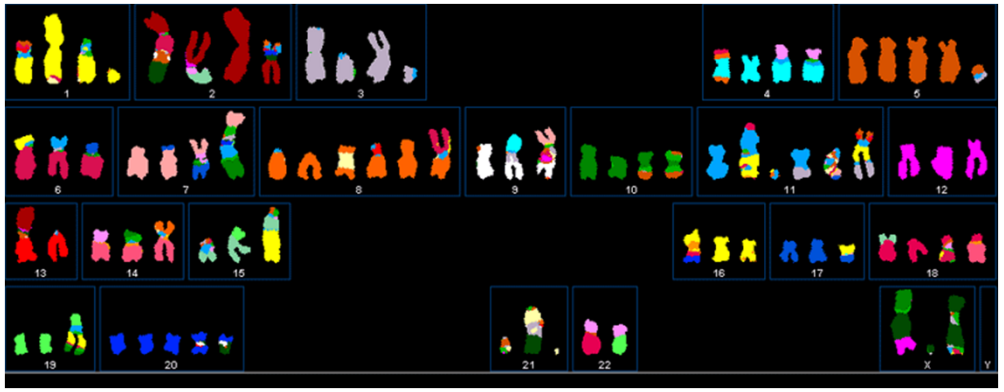
TCGA GBM: Center Overview

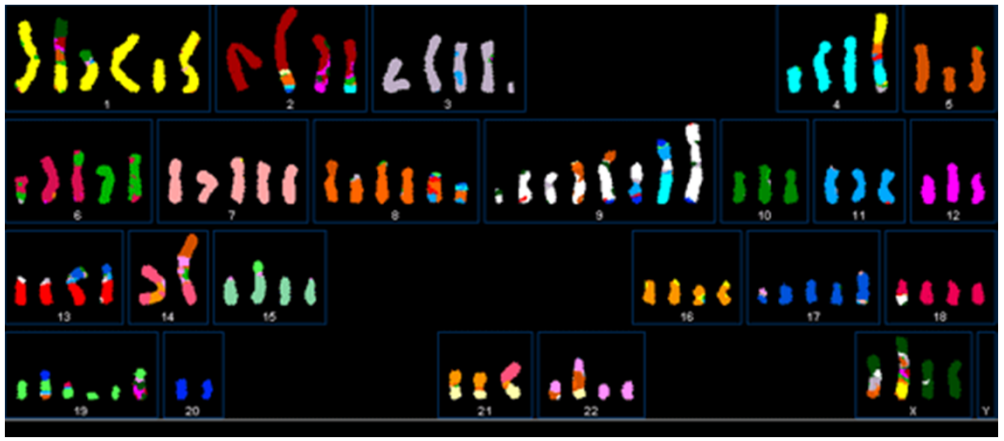


Cross-platform Data Integration, Comparison

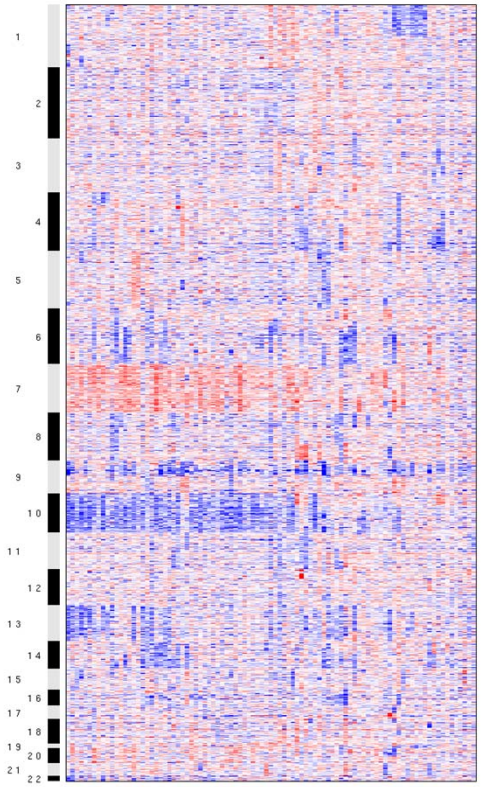
Cancer genomes are complex

Individual cancer genomes





Integrate across many samples



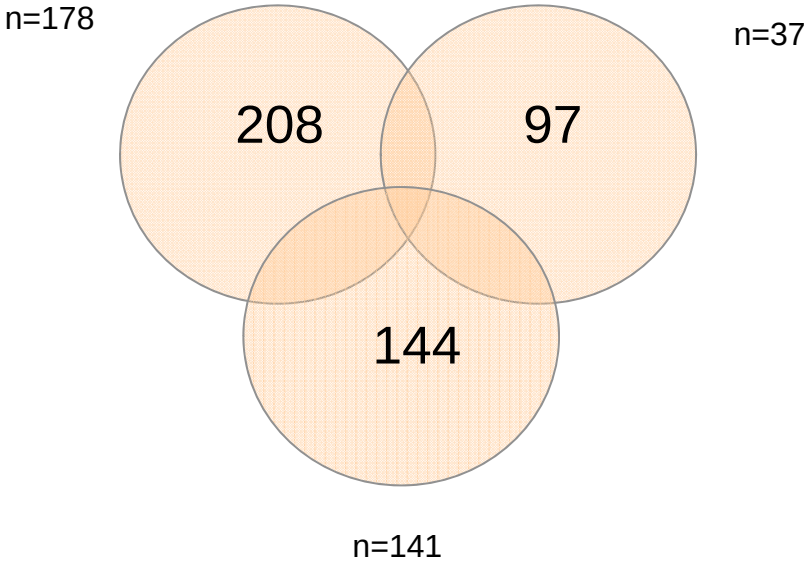
New analytical methods needed

Presentations

Cameron Brennan	GBM and Genome Characterization
Stephen Baylin	The GBM Epigenome
Rick Wilson	Identifying mutations in GBM and application of next gen sequencing technologies
Charles Perou	The Challenge of Integrative Analysis
Eric Lander	Summary and Discussion

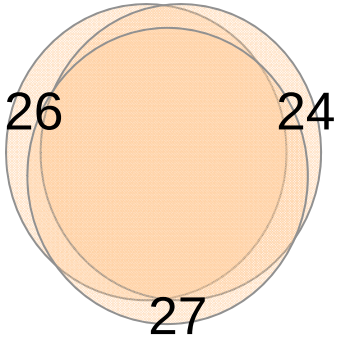
Copy-number alterations: Discordance in initial studies

Event counting



Little overlap in early studies
Include only some known genes

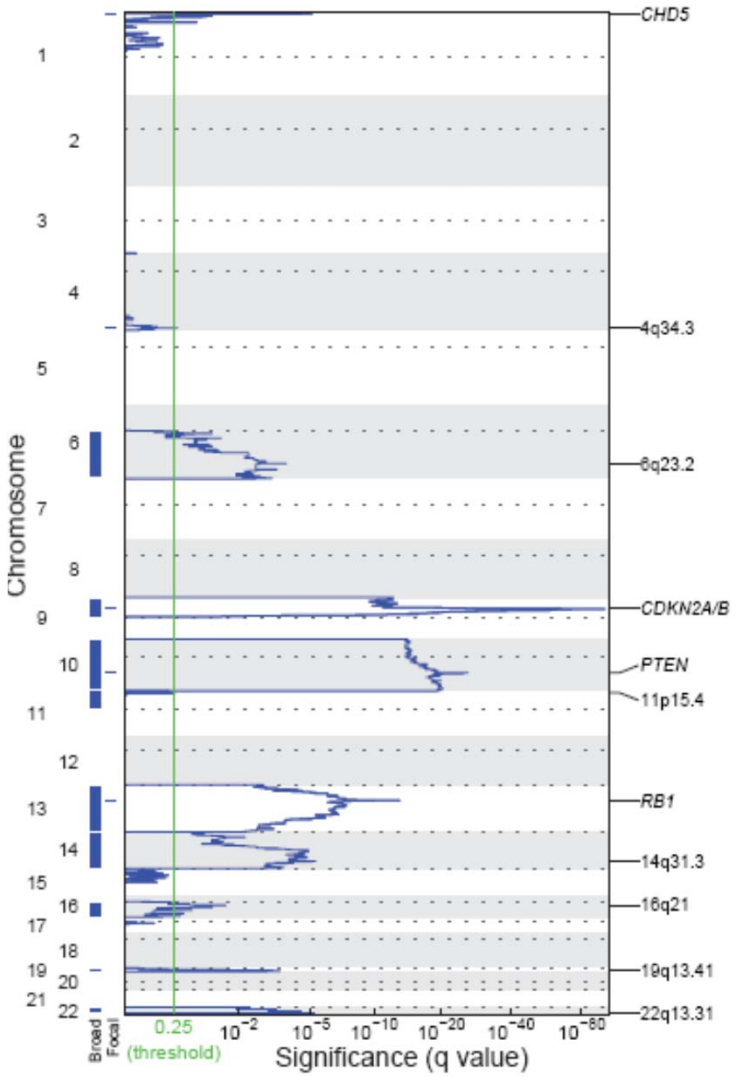
Statistical significance



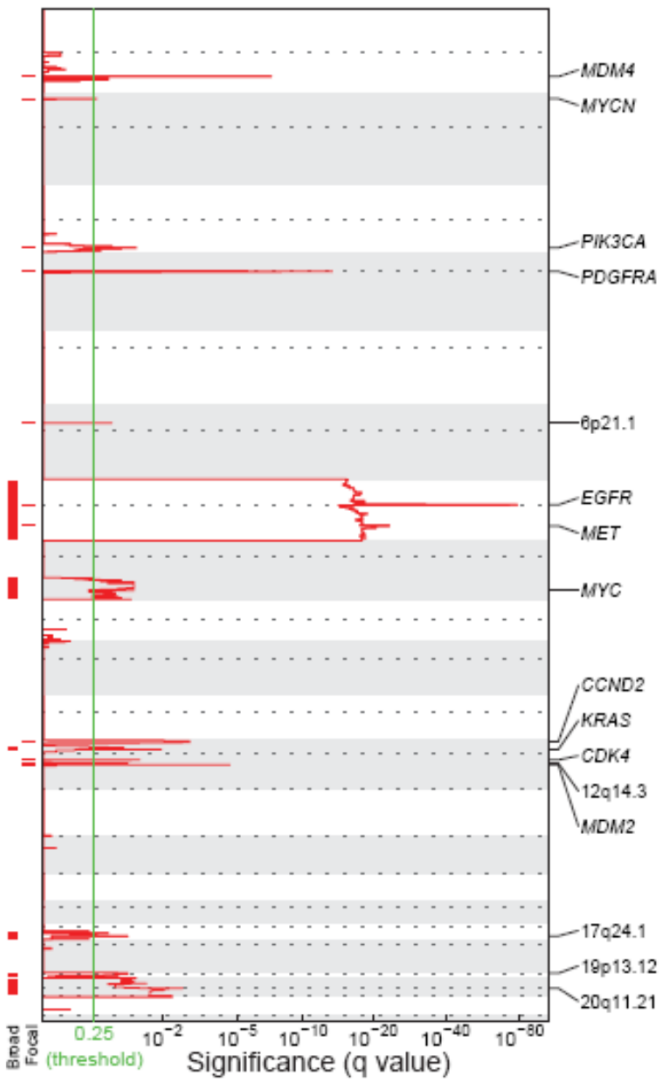
Many fewer events
Highly concordant
Includes all known genes

Copy-number alterations: ~30 in GBM

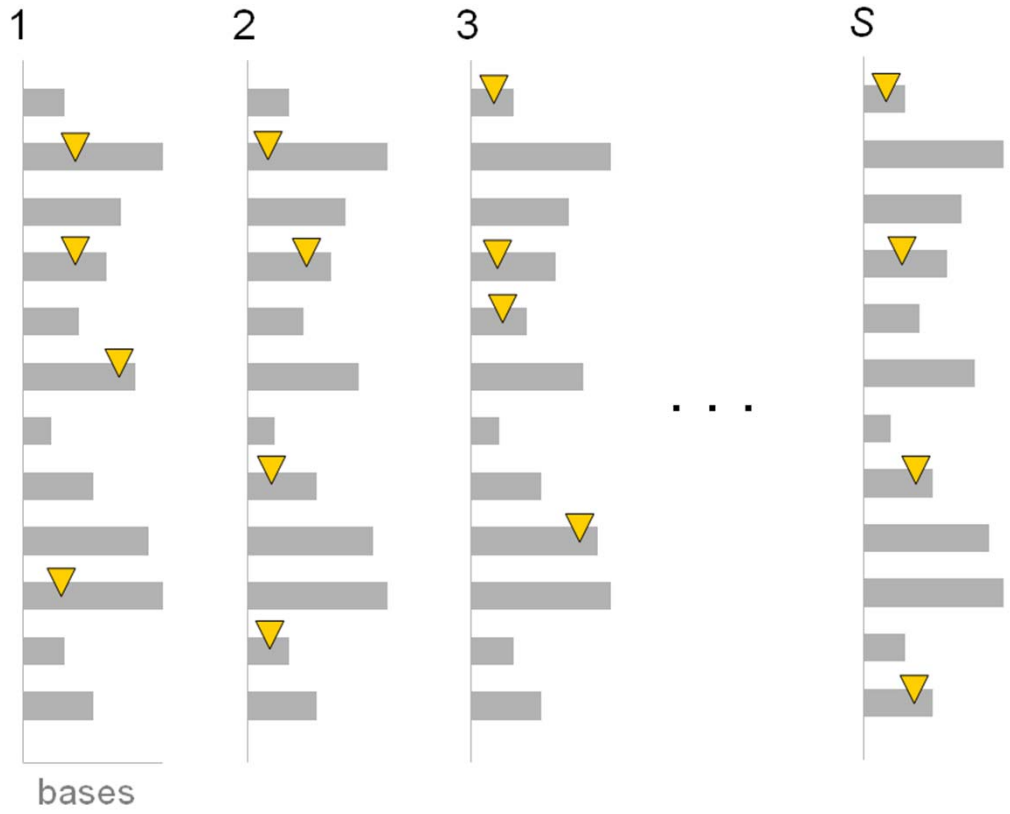
Deletions

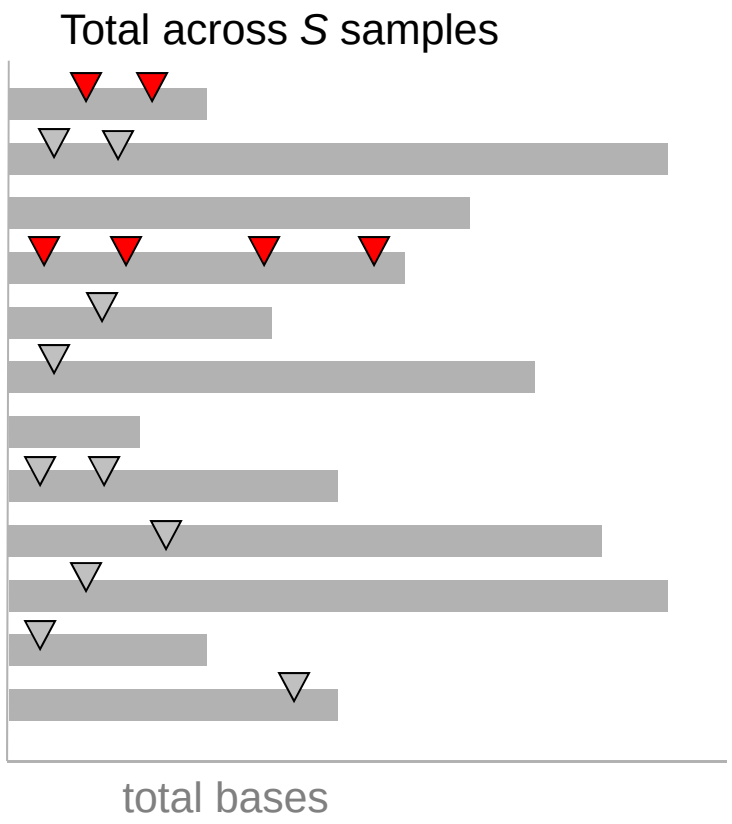


Amplifications



Coding mutations: Assessing statistical significance



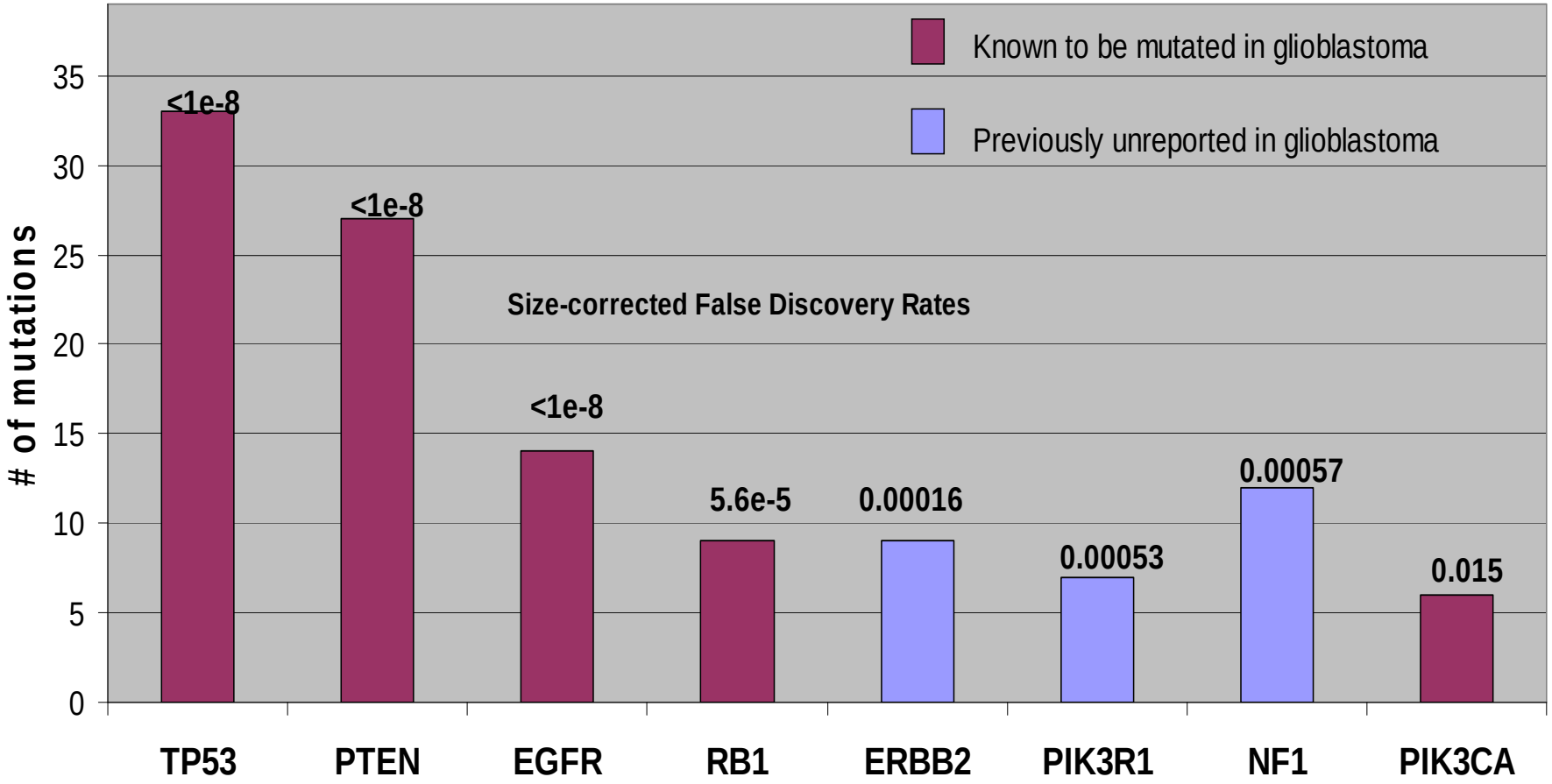


Random mutations: 2×10^{-6} per base
0.3% per typical coding region

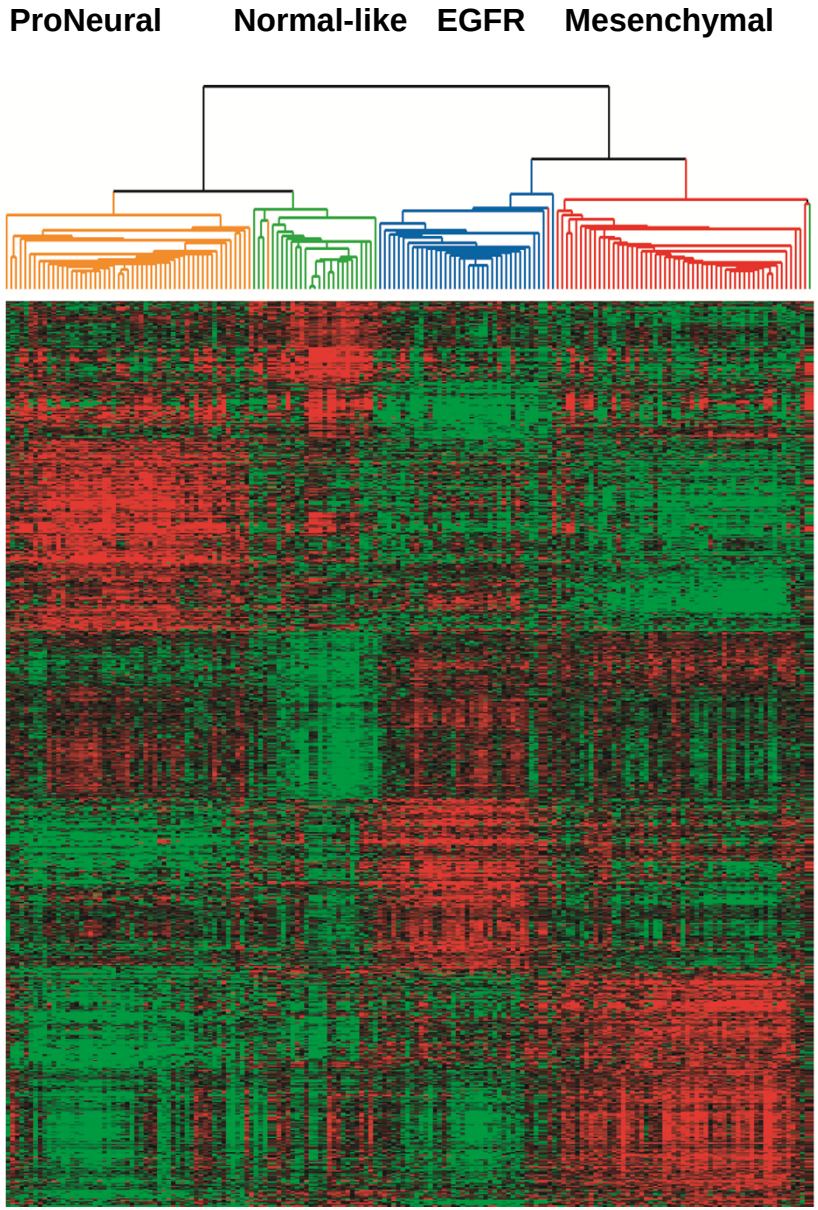
Cancer-related mutations: Aim to detect
3-5% per typical coding region

Significantly mutated genes in GBM

600 genes x 86 GBM (non-hypermutated)

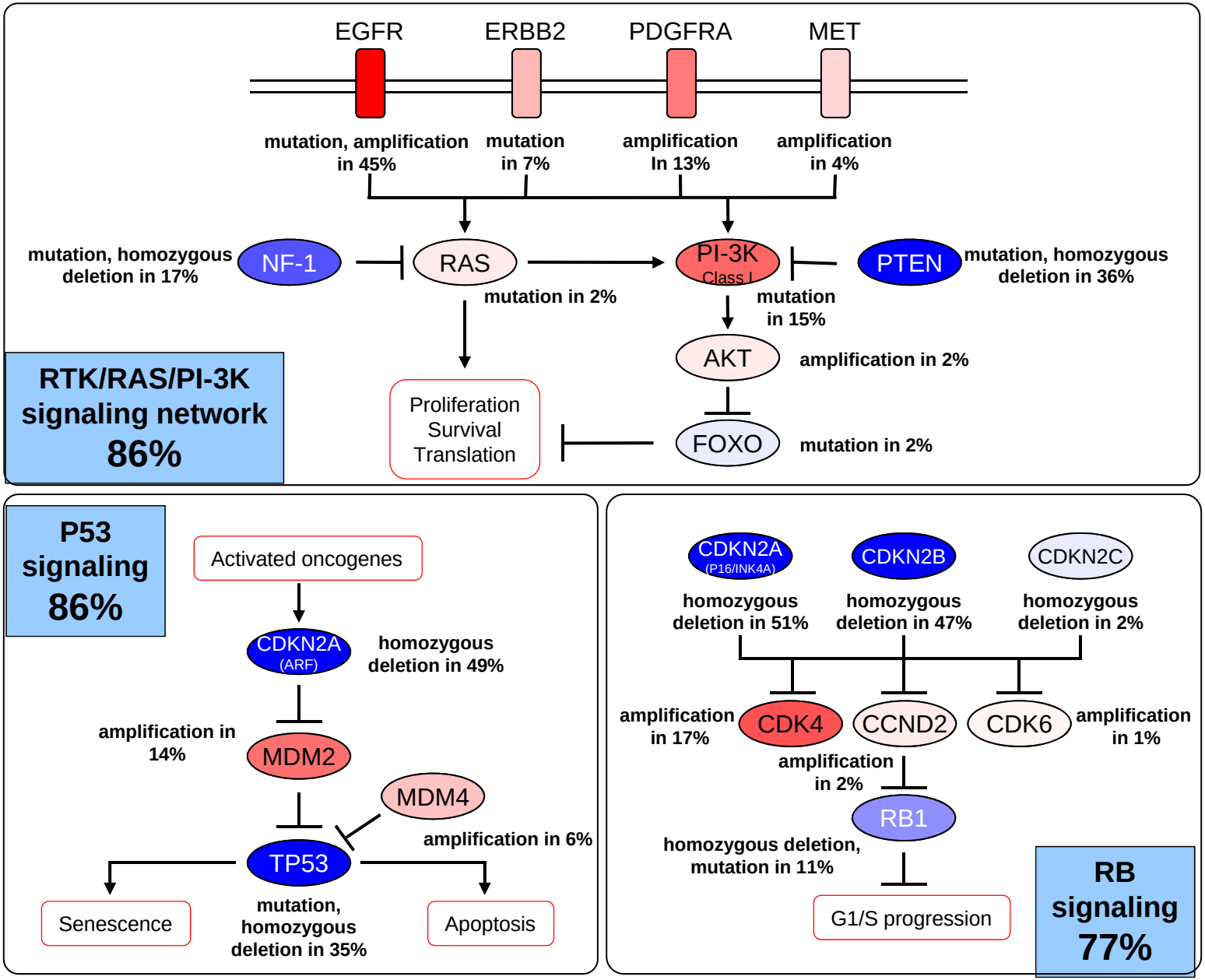


Integrated analysis defines four subtypes in GBM

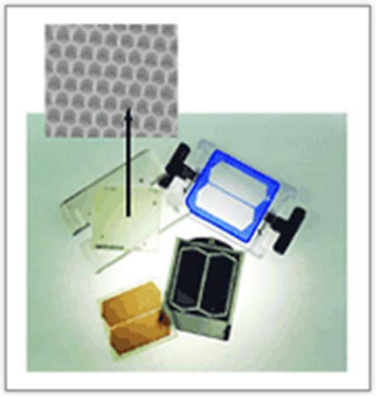
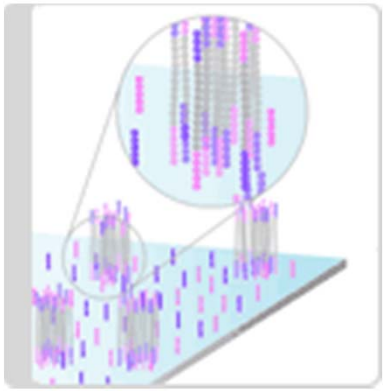


- Copy-number alteration
- RNA Expression
- DNA sequencing
- Methylation

Pathway Analysis in GBM



Next-generation sequencing technology

	ABI 3730XL	454	Solexa, ABI SOLiD
			
Read length	~700 bp	~200bp	30+bp reads
Read number	~100/run	400 K / run	~40 M / run
Total bases	70 K	400 Mb	8-12 Gb

+ others

Costs decreasing rapidly

Key questions posed at start of project

1. Can samples of adequate quality and quantity be assembled?
2. Can high-quality, high-throughput data be generated with current platforms?
3. How sensitive, specific and comparable are current platforms?
4. How can diverse data sets be integrated -- and what can be learned from integration?
5. Can recurrent events be distinguished from random background noise?
6. Can we identify new genes associated with cancer types?
7. Can we identify new subtypes of cancer?
8. Does new knowledge suggest therapeutic implications?
9. Can a network project drive technology progress in cancer?