

## Integrating Cancer Molecular, Clinical, and Pathways Data with caBIG® Molecular Medicine Tri-Conference

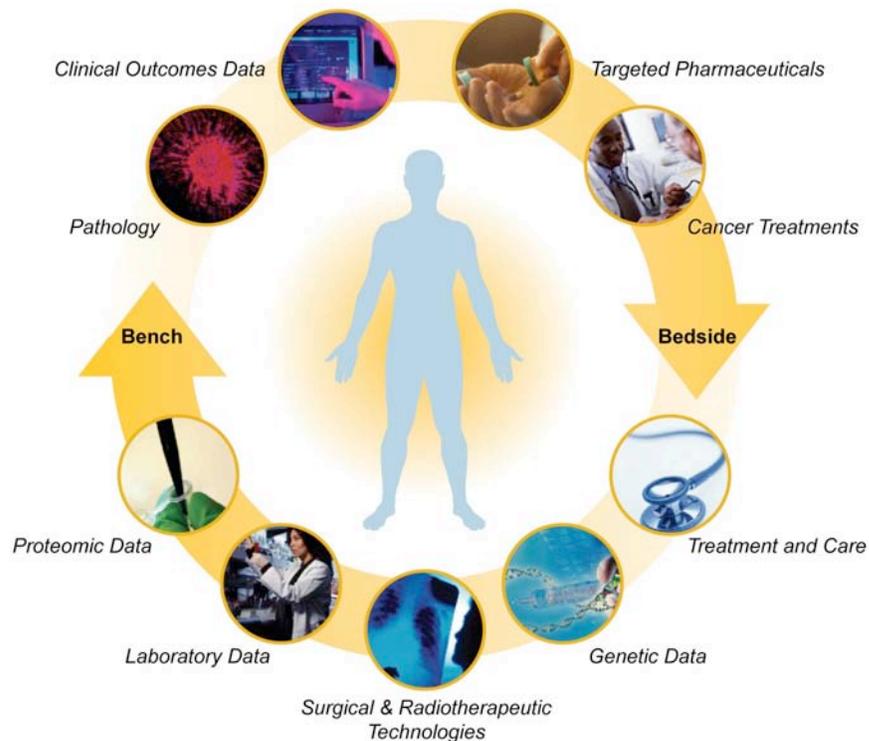
February 4, 2010

**Ken Buetow, Ph.D.**

*Director, Center for Biomedical Informatics and Information Technology  
National Cancer Institute*

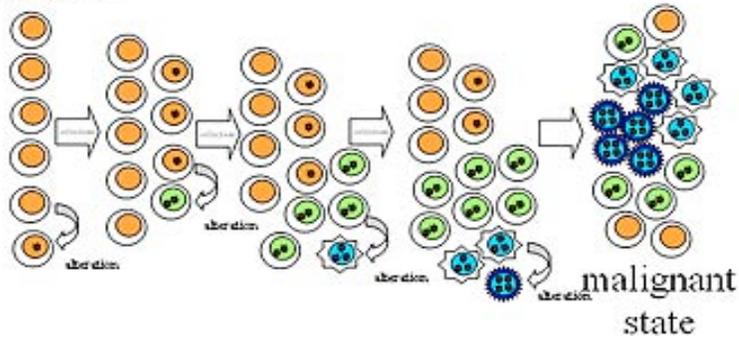
### 21st Century Biomedicine

- Personalized, Predictive, Preemptive, Participatory.....
- Unifies clinical research, clinical care, and discovery (bench-bedside-bed) into a seamless continuum
- Results in improved clinical outcomes
- Accelerates the time from discovery to patient benefit
- Enables a health care system, not a disparate “sector”
- Empowers consumers in managing their health over a lifetime



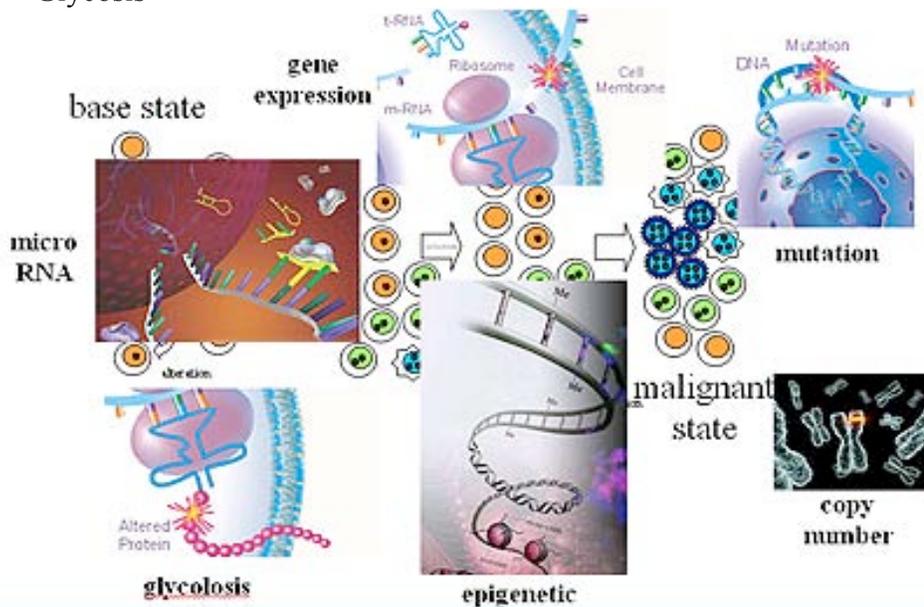
## Cancer is a Complex Adaptive System

base state



## Cancer is a Complex Adaptive System

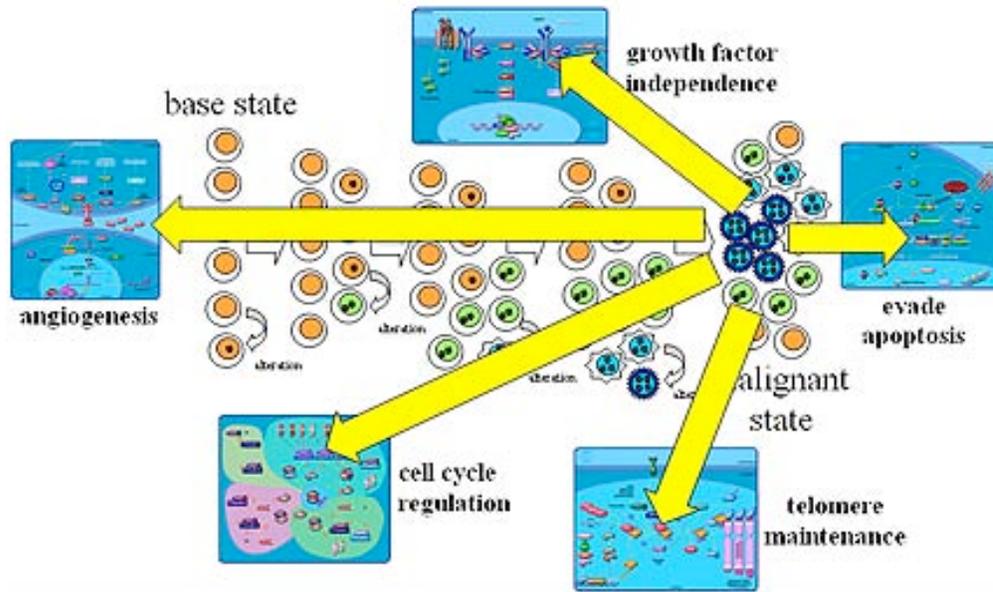
- Copy number
- Mutation
- Gene expression
- Epigenetic
- Micro RNA
- Glycolysis



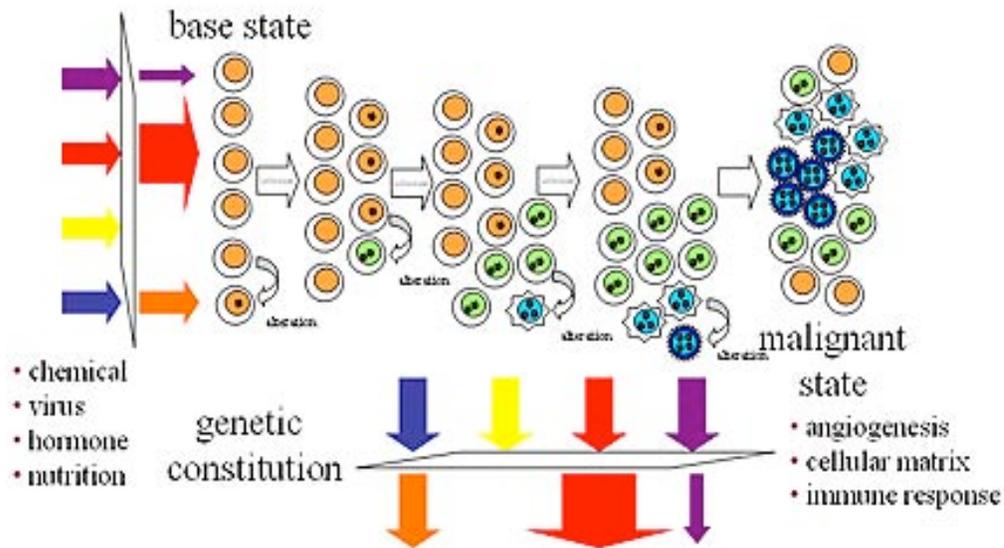
## Cancer is a Complex Adaptive System

- Angiogenesis
- Cell cycle regulation
- Telomere maintenance
- Evade Apoptosis

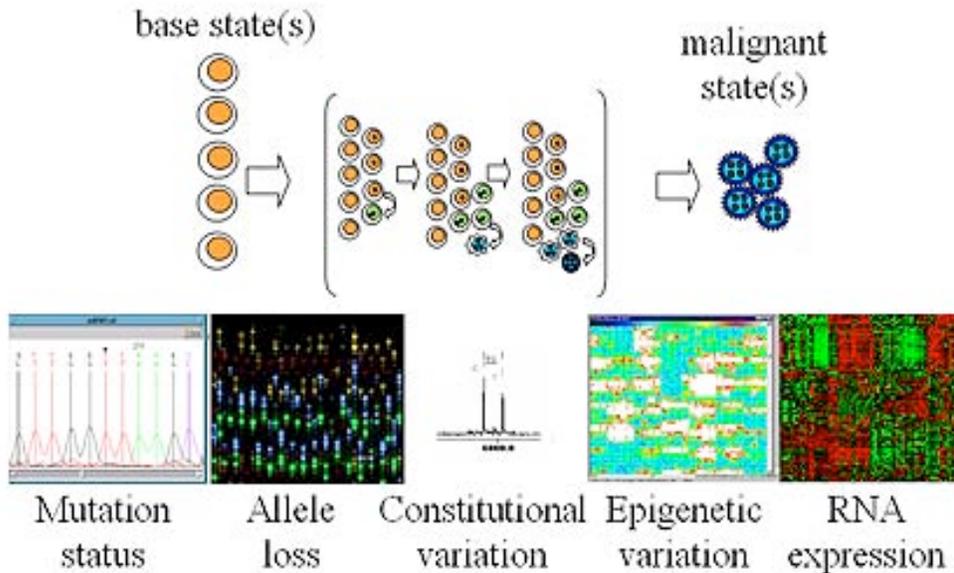
## Growth factor independence



## Cancer is a Complex Adaptive System



**Multiple systems technologies are needed to triangulate molecular state of disease**



**Molecular Medicine as a Complex Continuum**

- Clinical Research
- Imaging
- Molecular Biology
- Pathology

**caBIG®:**

**Biomedical Information Highway**

- The cancer Biomedical Informatics Grid® (caBIG®) is a virtual network of interconnected data, individuals, and organizations that redefines how research is conducted, care is provided, and patients/participants interact with the biomedical research enterprise.

**The caBIG® Initiative**

caBIG® empowers researchers to see the “BIG” picture by integrating increasingly complex layers of cancer biology, from gene to clinical phenotype, as a whole  
 Gene – Genome – Genomes – Pathways – Clinical Outcomes

*All from their computer*

## **caBIG® Capabilities Advance Discovery, Clinical Research, and Clinical Care**

### ***Clinical Research***

- Track clinical trial registrations
- Facilitate automatic capture of clinical laboratory data
- Manage reports describing adverse events during clinical trials

### ***Imaging***

- Use NBIA repository for medical images including CAT scans and MRIs
- Visualize images using DICOM-compliant tools
- Annotated Images with distributed tools

### ***Pathology***

- Access library of well characterized and clinically annotated biospecimens
- Use tools to keep an inventory of a user's own samples
- Track storage, distribution, and quality assurance of specimens

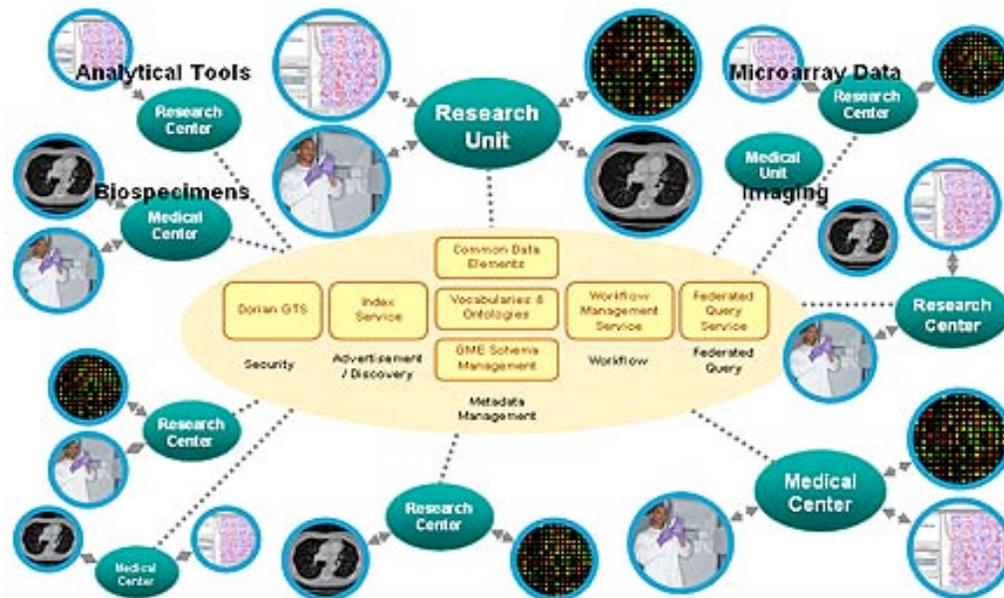
### ***Discovery Research***

- Combine proteomics, gene expression, and other basic research data
- Submit and annotate microarray data
- Integrate microarray data from multiple manufacturers and permit analysis and visualization of data

## **IT-enabled ecosystem**

Analytical Tools, biospecimens, array data, imaging, research units/centers and medical units/centers all connect to:

- **Security:** Dorian GTS
- **Advertisement/Discovery:** Index Service
- **Metadata Management:** Common Data Elements, Vocabularies & Ontologies, GME Schema Management
- **Workflow:** Workflow Management Service
- **Federated Query:** Federated Query Service

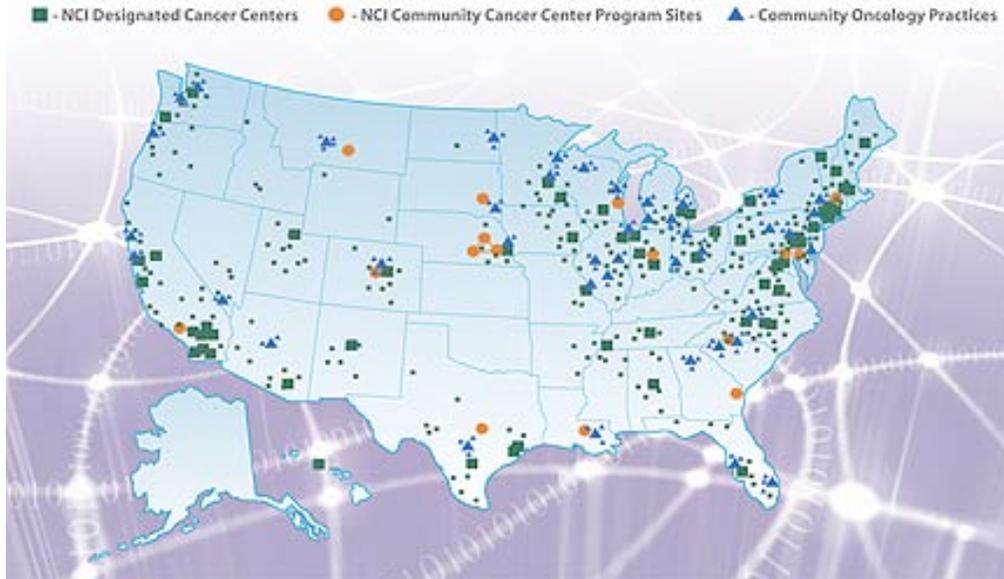


**IT-enabled ecosystem**

Together creates the:  
 Biomedical Knowledge Cloud  
 Grid Services Infrastructure



## caBIG® is Linking the Cancer Community



## caBIG® is Establishing Global Connections



United States, Mexico, Chile, Uruguay, Argentina, Brazil, UK, Netherlands, Germany, Czech Republic, Finland, Jordan, India, China, Australia, New Zealand

caBIG®, the world's largest biomedical research "highway", connecting a growing number of people and organizations across the globe

**Case Study:**  
**Hypothesis Generation Utilizing TCGA Resources**

## **Connecting multiple sources, experiments, and data types**

### ***Three forms of cancer***

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

### ***12 Organizations***

- Biospecimen Core Resource
- 7 Cancer Genomic Characterization Centers
- 3 Genome Sequencing Centers
- Data Coordinating Center

### ***Multiple data types***

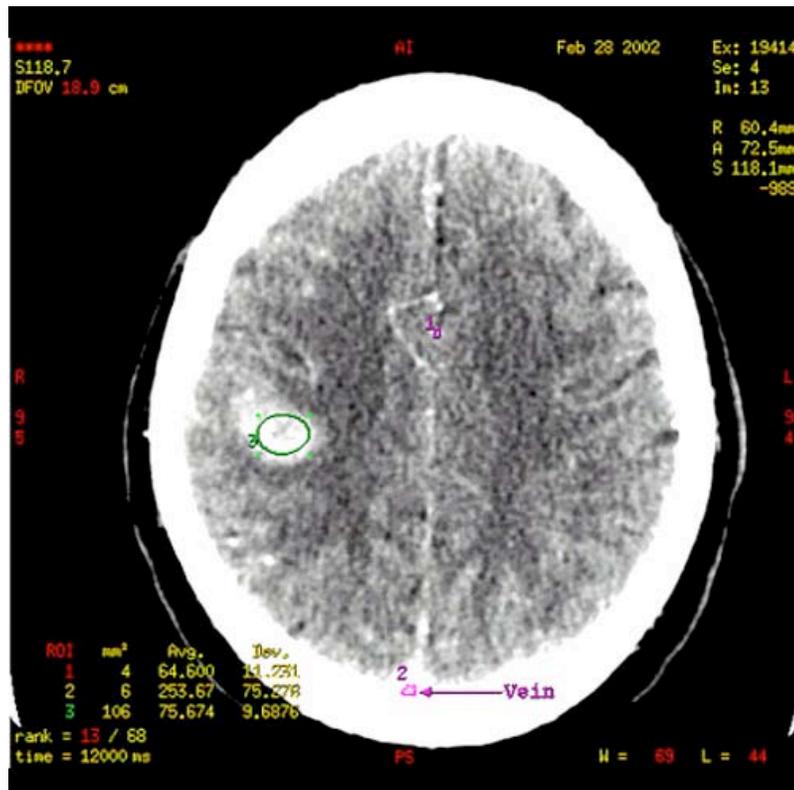
- Clinical diagnosis
- Treatment history
- Histologic diagnosis
- Pathologic status
- Tissue anatomic site
- Surgical history
- Gene expression
- Chromosomal copy number
- Loss of heterozygosity
- Methylation patterns
- miRNA expression
- DNA sequence

A Single Web-based Portal for all Analyses – <http://cma.nci.nih.gov>

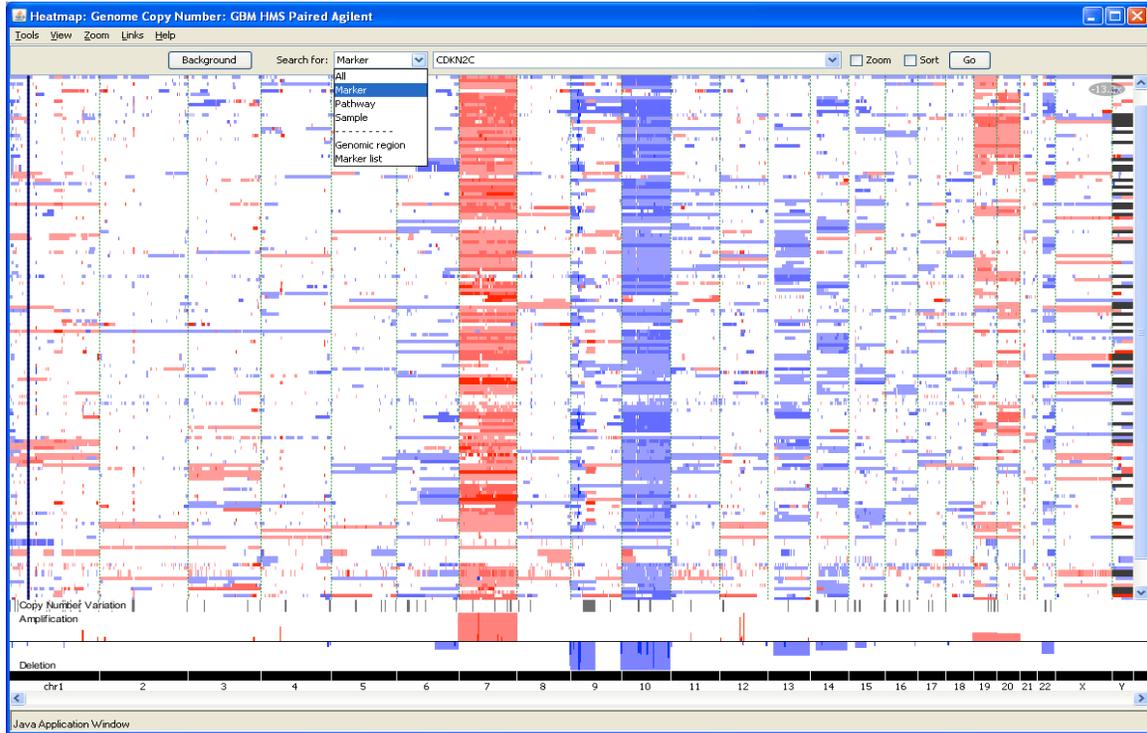
The screenshot shows the caBIG Cancer Molecular Analysis Portal. At the top left is the National Cancer Institute logo. The main header features the caBIG logo and the text "cancer Biomedical Informatics Grid™" and "Cancer Molecular Analysis Portal". A "Context: TCGA" dropdown menu is in the top right. A left sidebar contains navigation links: "Gene View", "Genome View", "Clinical View", and "Analysis Tools". The main content area is titled "Gene View" and includes a description: "Visualize gene expression, copy number, SNP, and pathway data on a gene by gene basis. Generate detailed study related reports for a given gene." Below this, it lists available resources: "Gene Expression Plots, KM Survival Plots, CGWB Integration, and Pathway Visualizations." Two visualizations are shown: a pathway diagram on the left and a Kaplan-Meier survival plot on the right. The survival plot is titled "Gene Expression-Based KM Plot (GEP 2.0)" and shows survival probability over time for different groups. On the right side of the page, there is a login section titled "Existing Users:" with fields for "user:" and "pass:" and a "login" button. Below that is an "Additional Information:" section with links for "Register" and "Provide your feedback". At the bottom, there are logos for the National Cancer Institute, the University of California, San Diego, and FIRSTGOV.

## Glioblastoma Multiforme (GBM)

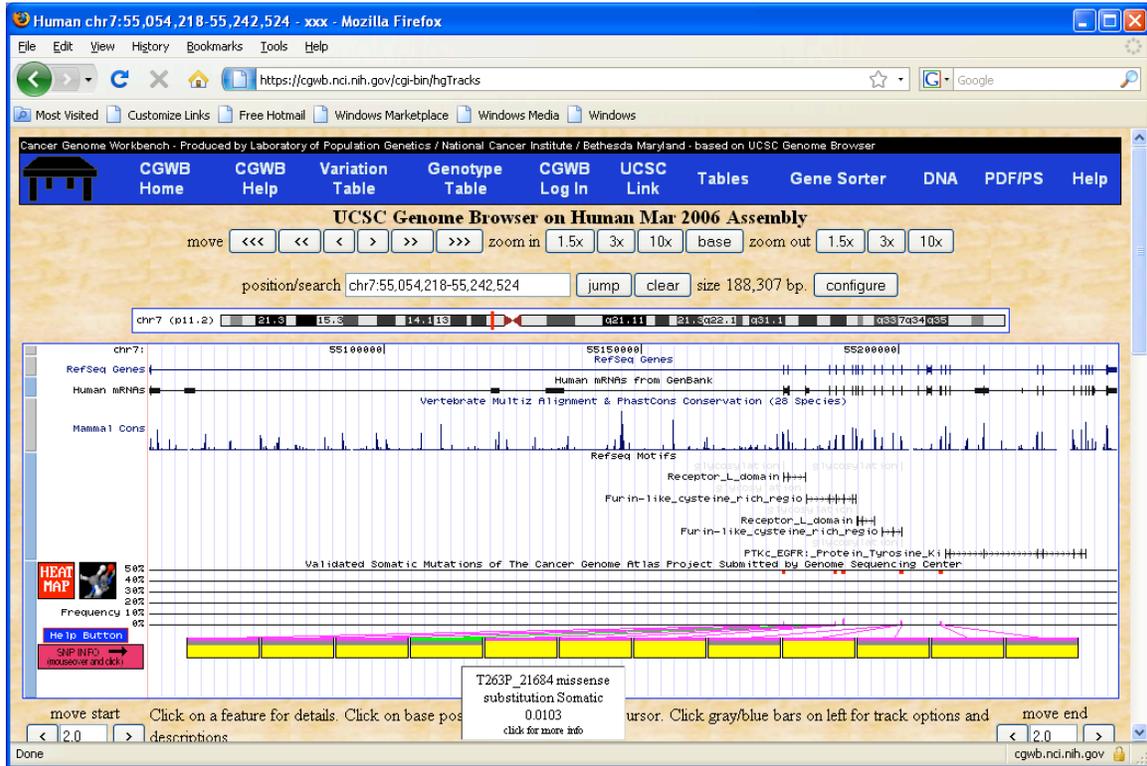
- GBM is the most common type of brain tumor. High grade gliomas are incurable and tumors expressing a mesenchymal phenotype are the most aggressive form



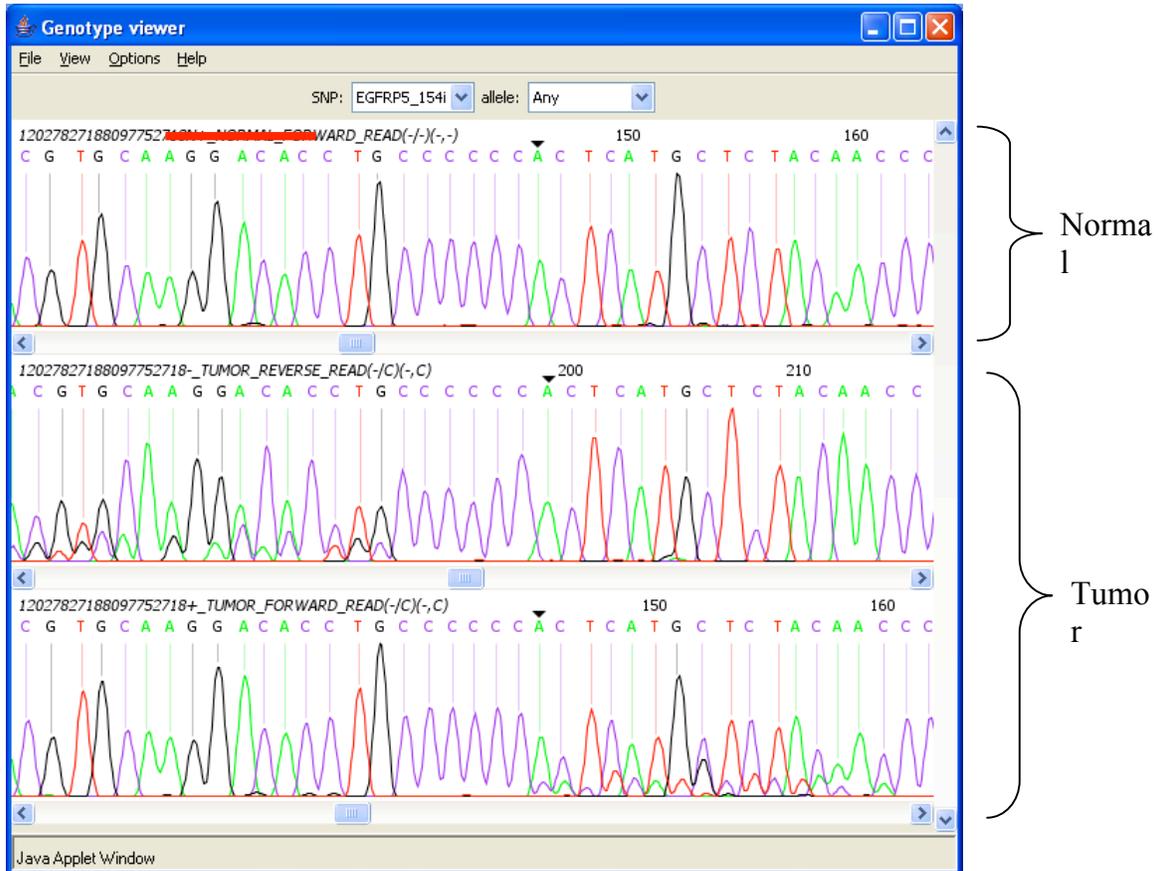
## Chromosome 7 and EGFR Seen as Frequent Targets of Alteration in GBM (*glioblastoma multiforme*)



## Summary View of EGFR Mutations Shows Clustering Around Extracellular Domain



**Putative Somatic Mutations can be Manually Reviewed**  
e.g.: Frameshift Mutation in EGFR in Paired Tumor/Normal



## Protein Structure View of EGFR Mutations

The screenshot displays the Protein Structure 3D viewer interface for the EGFR gene. The browser address bar shows the URL: <https://cgwb.nci.nih.gov/cgi-bin/3dViewer/Gene.cgi?proj=valid&sym=EGFR>. The page header includes the National Cancer Institute logo and the text "Protein Structure Viewer".

Key information displayed includes:

- Gene: EGFR GI: 29725609 1210 aa
- Also see GI: 41327736 41327732 41327734
- Protein Motifs: 4 (Solid - Pfam domain)
- Pub hits: 1 (cyan 3D Structure Viewing region(s): 25 to 638)
- SNP LogE & SIFT: Red: Predicted Deleterious, Blue: Predicted Tolerant, Black: Undecided, Gray: Not Analyzed
- Phosphorylate: 29 (Pubmed)

The main visualization is a linear map of the EGFR protein structure, showing various domains and mutations. A callout box labeled "Access proteins from alternative splicing" points to a specific region. A "Click to get 3D viewer" button is located below the map. The map shows mutations such as L858R, R1100C, R1150C, T790M, R1150V, R1150P, G709V, C797Y, V858H, and R1150G. A scale bar indicates positions up to 1210, with a 1000 marker.

The browser status bar at the bottom shows "Done" and the URL "cgwb.nci.nih.gov".

### 3D Structure Viewer

Big, highlighted atoms refer to the mutated amino acids. You can also click on the mutated amino acid to turn on or off a specific mutation

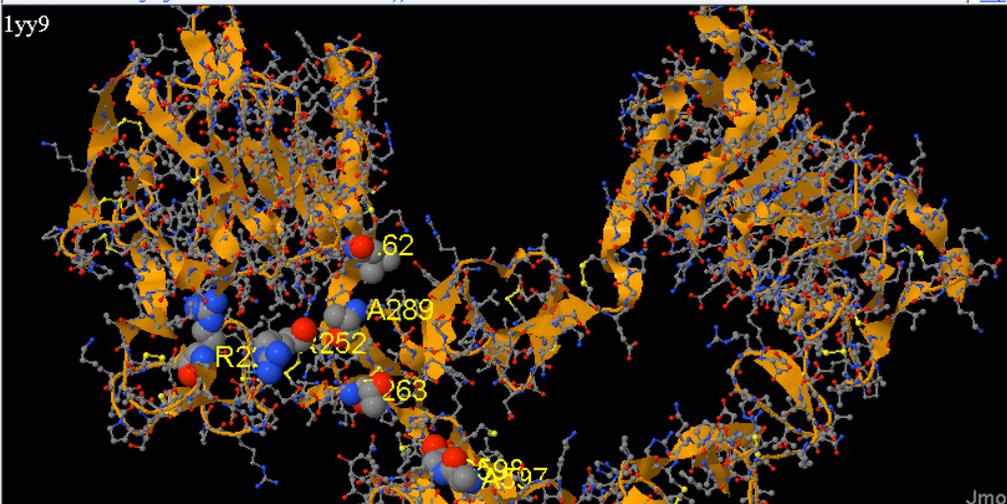
Protein Structure 3D Viewer - Mozilla Firefox

https://cgwb.nci.nih.gov/cgi-bin/3dviewer/ViewAA.cgi?proj=valid&gi=29725609&id=722&pdb=1yy9;A&sim=0.995&gstart=25&glen=613&phos=

National Cancer Institute U.S. National Institutes of Health www.cancer.gov

GI: 29725609 Viewing region 25 - 638 99.5 % similar to Pdb: 1yy9 chain A 2 - 614

1yy9

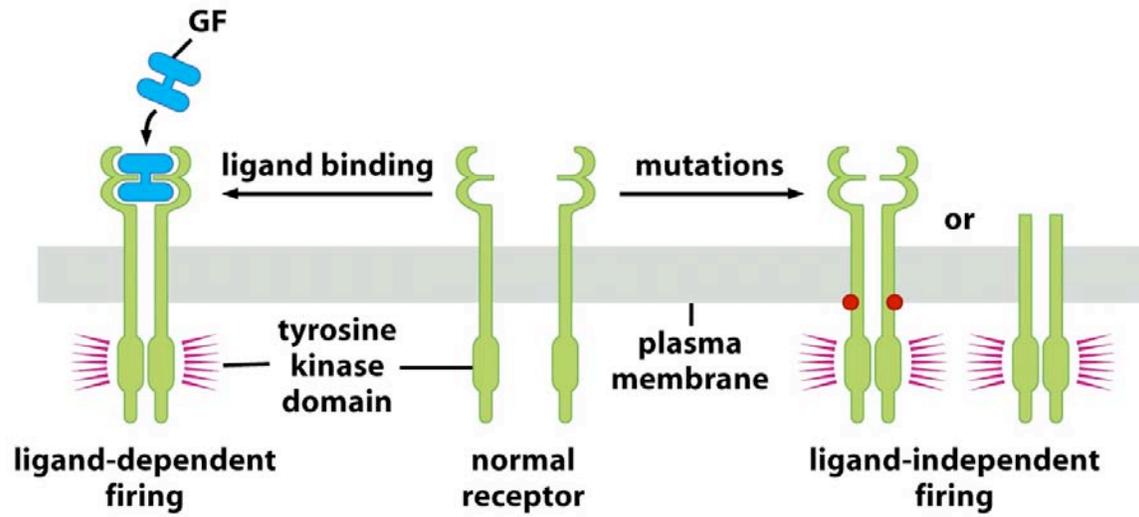


Jmol

start	Sequence: Click on a letter once will turn on the spacefill. Double-click on the same letter will turn the spacefill off.	end
2	EEKVKVCGQTSNKLTLQGFEDHFLSLQRMFNCEVVLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNT	71
72	VERIPLNLQIIRGNMYEENSALAVLSNYDANKTGLKELPMRNLQEIHLGAVRFSNNPALCNVESIQWRD	142
143	IVSSDFLSNMSMDFQNHLSGSCQKCDPSCPNWSCWGAGEENCQKLTKIICAQCCSGRCRGKSPSDCCHNQCA	213
214	AGCTGPRESDCLVCRKFRDEATCKDTCPPMLLYNPTTYQMDVNPPEGKYSFGATCVKCCPRNYVVDHGSCV	284
285	RACGADSYEMEEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFKNCTSSIGDLHILPVAFRGD	355
356	SFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAFENLEIIRGRTKQHGGQFSLAVVSLNITSLGL	426
427	RSLKEISDGDVLIISGNKNLCYANTINWKKLFGTSGQKTKIISNRGENKCKATGQVCHALCSPEGCWGPEPR	497
498	DCVSCRNVSRGRECVDKCKLLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCV	568
569	KTCPAGVMGENNTLVWVKYADAGHVCHLCHPNCTYGCTGPGLRGCPPT	614

Jmol script completed cgwb.nci.nih.gov

## Constitutive Activation of EGFR Leads to Abnormal Growth



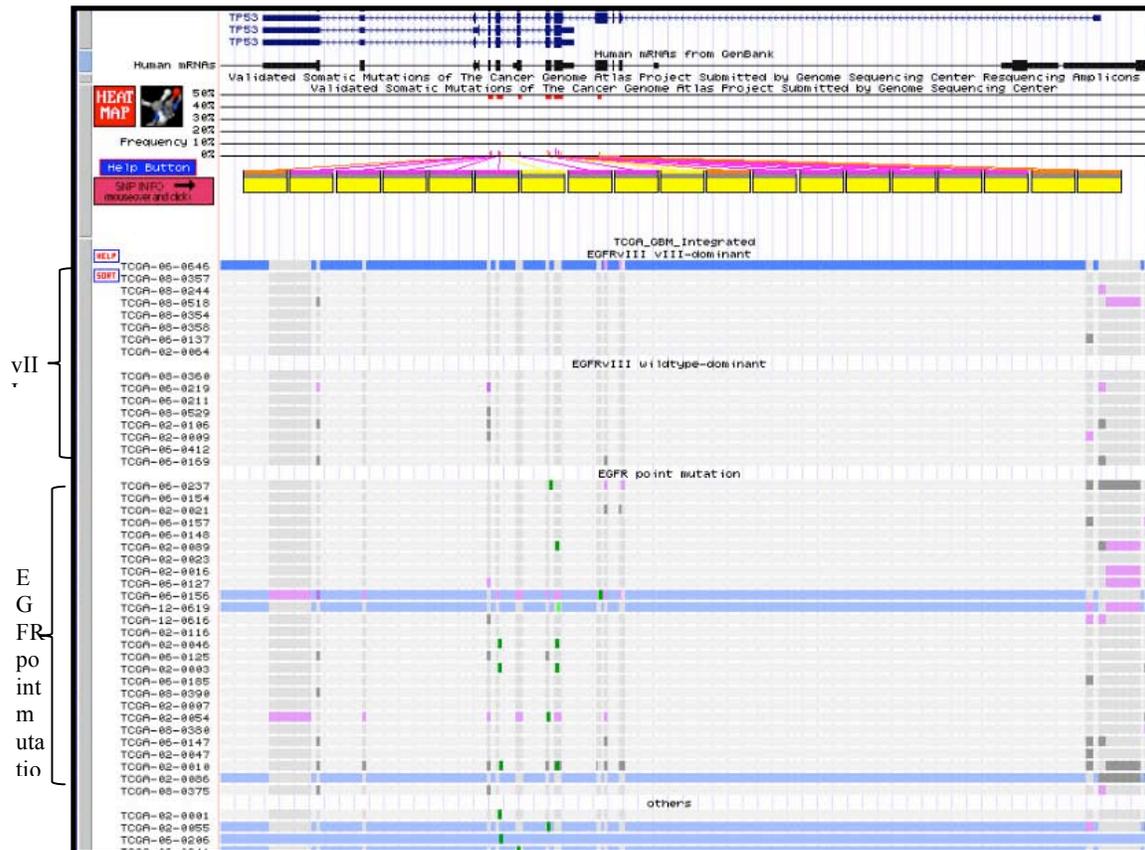
“EGFRvIII” mutation

*Modified from The Biology of Cancer* (© Garland Science 2007)

## View EGFR vIII Patient Subgroups to find vIII Mutations Occurring in Amplified Samples and Exclusive with EGFR Point Mutations



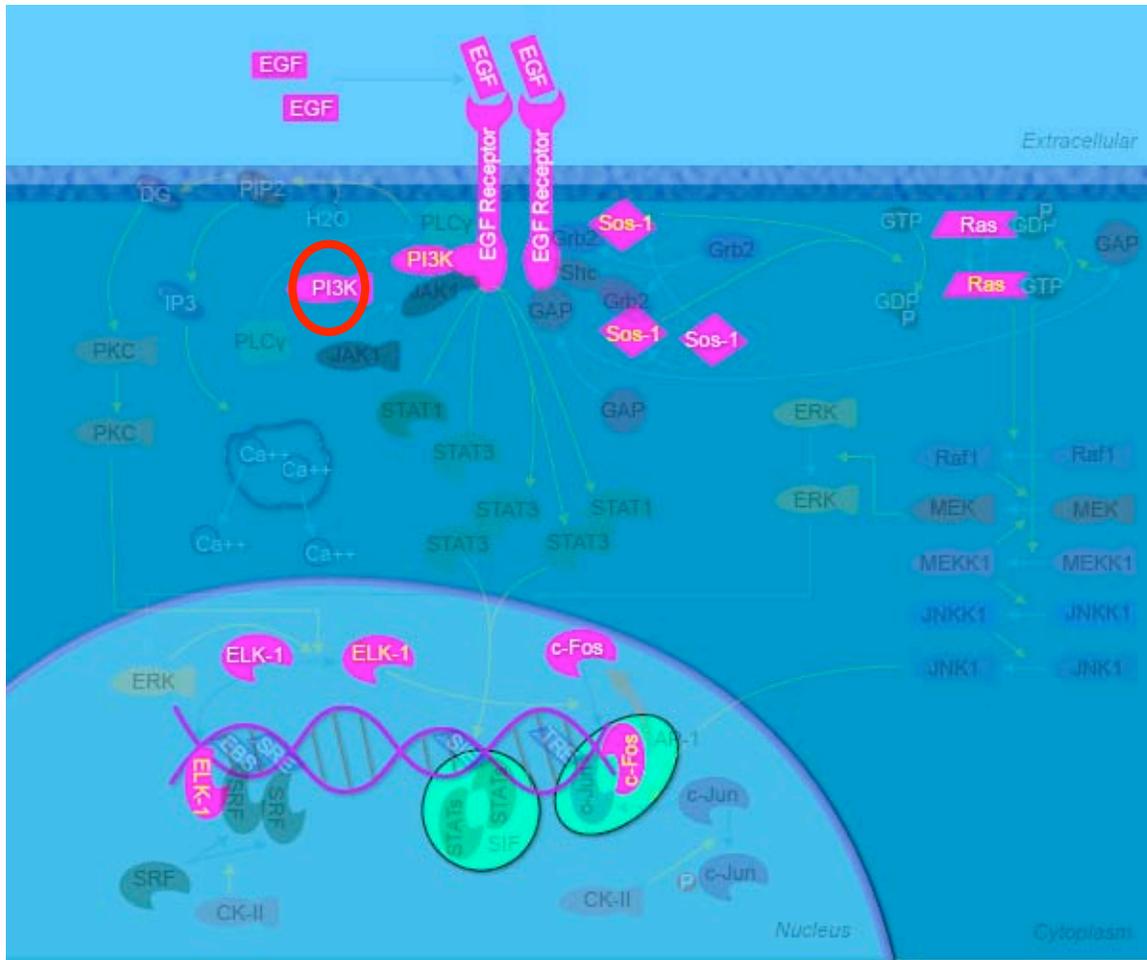
**EGFR Mutation Subgroups Viewed at the TP53 Locus**  
**No Mutation for EGFRvIII but 1/3 of EGFR Point Mutations have TP53 Mutations**  
**( $p=0.036$ )**



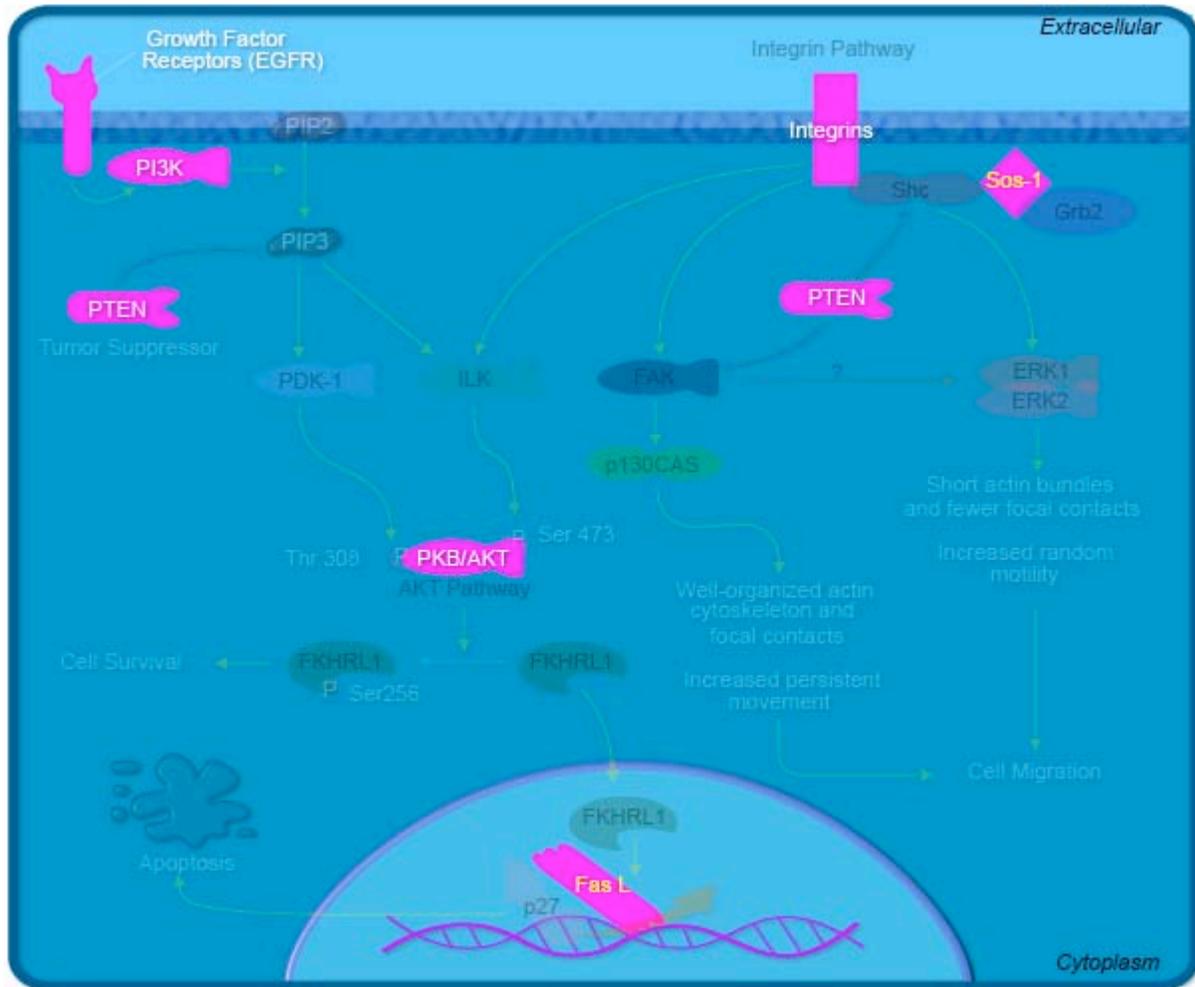
**Mutations in EGFR vIII and TP53 May be Anti-Correlated**

	EGFR amplification			No EGFR amplification		
	EGFR point mutation	EGFRvIII	No EGFR mutation	EGFR point mutation	EGFRvIII	No EGFR mutation
TP53	18	12	37	7	0	79
Fraction	5 28%	0 0%	4 11%	3 43%	0 N/A	35 44%

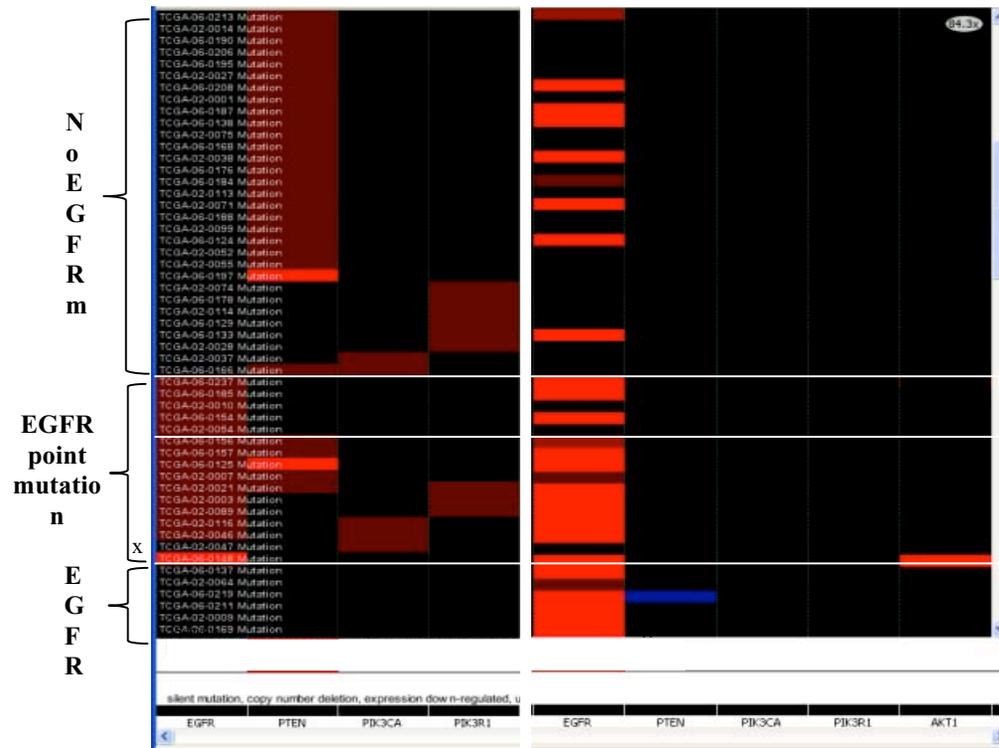
## EGFR Pathway Mutation Profile Through CMA



## Alterations in PI3K Pathway Through CMA



**Somatic Mutations (L) and Copy Number (R) Shows Frequent Co-occurrence of EGFR Point Mutations with Other Genes in PI-3K Pathway but not the EGFR vIII Mutations**

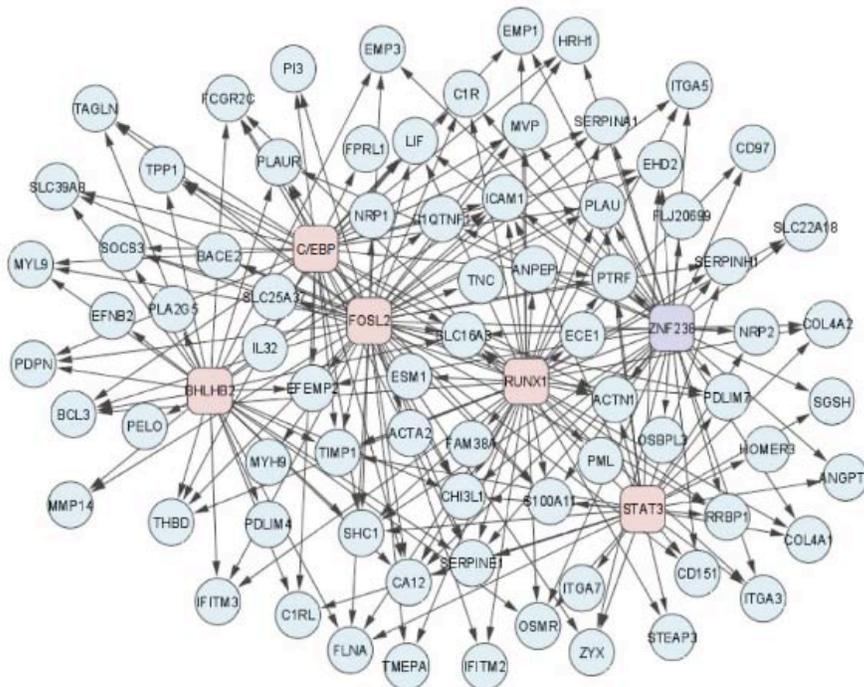


***In Silico* Hypotheses**

- 1) No P53 mutations were found in amplified samples with EGFRvIII while significant levels of P53 mutation were found in amplified samples with EGFR point mutations. Suggests alternative molecular etiologies.
- 1) EGFR point mutations co-exist with additional mutations in other genes involved in PI-3K pathway while EGFRvIII rarely have additional mutations in PI-3K pathway. This suggests the possibility of oncogene addiction in EGFRvIII tumors but not in tumors with EGFR point mutations even though both types of mutations target EGFR extracellular domains.

## Case Study: Identification of Transcriptional Networks in GBM

### Identification of Transcriptional Networks in GBM

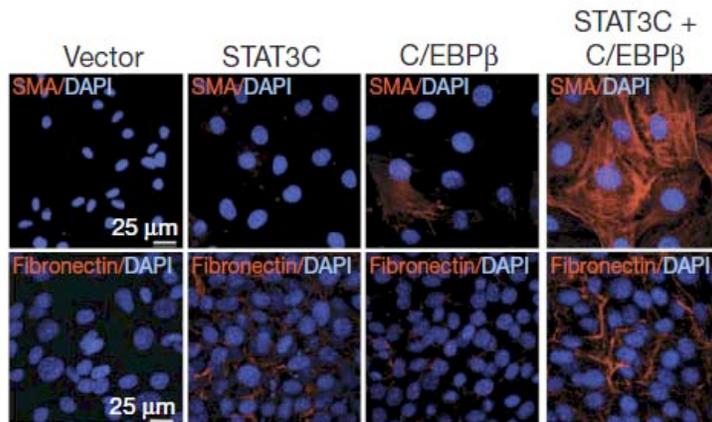
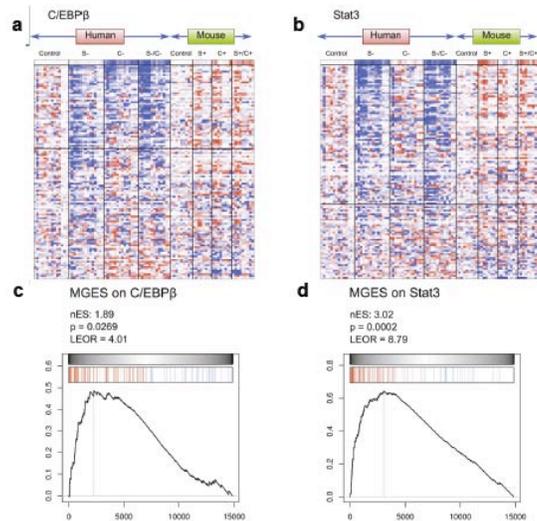


Carro, MS, et. al. The Transcriptional Network for Mesenchymal Transformation of Brain Tumors. Nature 463:21, Jan 2010, 318-327.

- Gene expression profiles, comparative genomic hybridization, and copy number data collected by the TCGA program were subjected to integrated analysis
- More than 92K predicted glioma-specific transcriptional interactions were identified, 53 of which were specific to mesenchymal pathway genes

## Identification of Transcriptional Networks in GBM

- Two genes, C/ERP $\beta$  and STAT3C, were specifically associated with activating genes in the mesenchymal pathway, with associated protein expression and phenotypic changes
- The group is part of the caBIG® *In Silico Research Centers* program from



## Summary

- Effective translational research requires new ways to manage and integrate biomedical data and new ways to conduct collaborative research
- Interoperable IT frameworks enable the next wave of translational research
- Beginning in cancer, caBIG® is providing these interoperable IT frameworks that will lead to the development of a Learning Healthcare System

For more information, please visit:

<http://caBIG.cancer.gov>