

Introduction:

Human cancer is a diverse disease resulting from genetic alterations that include gene over-expression, amplification and mutation. These events have increasingly been the focus of drug development efforts resulting in effective therapies for specific cancer subtypes.

In order to better characterize patients by molecular alteration, large scale efforts are underway to evaluate cancers at the molecular level. This includes The Cancer Genome Atlas (TCGA) project, whose goal is to characterize all molecular alterations across 20 human cancers in large patient cohorts and make this data available to the research community in order to better understand and treat cancer.

We sought to develop an interactive web-based toolset to analyze and view multi-dimensional genomic alterations in order to identify aberrant patterns unique to individual cancer patients. Gene expression, DNA copy number and mutation data for glioblastomas and ovarian cancers were collected from the TCGA for analysis and summarization.

Methods:

Gene expression, DNA copy number and mutation data were collected from the TCGA portal (<http://cancergenome.nih.gov>). This data included mRNA expression data measured on the Affymetrix HT-U133A platform for 382 glioblastomas and 509 ovarian serous cystadenocarcinomas and DNA copy number data measured on the Agilent 244A platform for 459 glioblastomas and the Agilent 1x1M platform for 563 ovarian serous cystadenocarcinomas. Mutation data was collected for 147 glioblastomas and 326 ovarian serous cystadenocarcinomas that was generated from sequencing technologies including ABI's SOLiD and Illumina's Genome Analyzer. A total of 1,026 patients and 19,640 genes were included in the analysis.

Expression data was log₂ transformed and median centered per sample. DNA copy number data was log₂ transformed and processed by circular binary segmentation; Compendia then mapped the segmented values to genes based on genomic location. Mutation data was assembled with TCGA's collection of Level 3 MAF files. A supervised decomposition of select cancer networks—as defined by Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO)—produced 66 cancer pathways (averaging 10 genes/pathway).

Experimental data were summarized across tumors within each disease, establishing molecular aberration calls per patient per gene as follows:

- Amplifications at or above 4+ copy gain or homozygous deletions were called amplified or deleted genes, respectively.
- Expression higher than 8-fold above, or lower than 8-fold below the median value for that gene was called over-expressed or under-expressed, respectively.
- Mutation events were called positive or negative as reported by TCGA.
- Pathway Amp/Del/Mutation: DNA copy number or mutation events in any gene in the pathway were called as pathway aberrations.

Biomarker co-occurrence and exclusivity analysis was performed on a subset of The Cancer Gene Census genes. Fisher's exact test was performed and Odds Ratio and p-values for each pair of significant aberrations reported. Data integration and statistical computations are performed using Pipeline Pilot 7.5.2 (Accelrys, San Diego, CA).

Data visualization and analytics were developed using TIBCO Spotfire® v3.1 and designed to be deployed over the web using TIBCO Spotfire® Web Player and dynamic data-on-demand functionality (TIBCO Spotfire®, Somerville, MA).

The processing, summarization and deployment of these analyses are illustrated in Figure 1.

Conclusions:

Using this toolset we confirmed previous findings that molecular aberrations occur in multiple signaling pathways that include RTK/RAS/PI3K, P53 and RB, within many individual glioblastoma patients (Figure 2). Investigating several frequently altered members of these pathways confirmed their roles as tumor suppressors or activated oncogenes. We showed that TP53, RB1 and PTEN aberrations occur exclusively as mutations or deletions, suggesting loss of function in their role as tumor suppressors. CDKN2A is also shown as deleted in most cases and, to a lesser extent, mutated. Predominant EGFR aberrations included DNA copy number gain and co-occurring mutations, while MDM2 and CDK4 displayed both DNA copy number gain and over-expression (Figure 3). These results confirmed the findings for multiple aberrations in specific members of the RTK/RAS/PI3K, TP53 and RB signaling pathways in glioblastoma¹.

We also found that for many pathway members, such as EGFR and CDKN2A, DNA copy number gain/loss correlated with over or under-expression, respectively. Our integrated analysis also identified an unexpected association of EGFR mutations in glioblastomas that also harbored significant DNA copy number gain and over-expression (Figure 4). Lastly, co-occurrence/exclusivity analysis of EGFR aberrations yielded a new finding that suggested mutual exclusivity of EGFR amplification and FGFR pathway aberrations. In this analysis no FGFR pathway aberration was present when EGFR DNA copy number gain was observed (Figure 5). Further investigation of the FGFR pathway members at the patient level showed that aberrations in FGFRs 1–4 never occurred together in the same patient (Figure 6). Using this toolset we confirmed the finding that glioblastomas harbor aberrations in multiple signaling pathways. Furthermore, we showed that multiple molecular alterations of members of the FGFR pathway in glioblastomas occur exclusively within individual patients suggesting reduced selective pressure for additional deregulation in that pathway.

In this study we describe a rich toolset for summarizing and identifying relationships between molecular alterations within a patient population. As larger patient populations across all cancers are characterized for all molecular alterations, these applications will reveal new genetic events and their co-occurrence or exclusivity in the population that can then be targeted by new therapies to improve the lives of cancer patients.

References

1. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. (2008) *Nature* 455:23, pp1061-1068.

Results:

Figure 1. Development and delivery of integrative analysis across multiple genomic aberrations.

TCGA experimental data covering gene expression, DNA copy number and mutation data from over 1,000 patients were normalized and annotated with consistent gene symbols and standard metadata. (The patient count per aberration type is illustrated in the Venn diagram.) Analytical pipelines specific to each data type summarized continuous experimental data as discrete calls based on constant thresholds. Additional pipelines extended gene-level biomarkers to include Compendia-curated pathways and established co-occurrence/exclusivity comparisons among biomarker pathways. Data were consolidated into a relational database and integrated in an interactive visualization, enabling dynamic exploration of all genes, pathways, aberration types, and diseases.

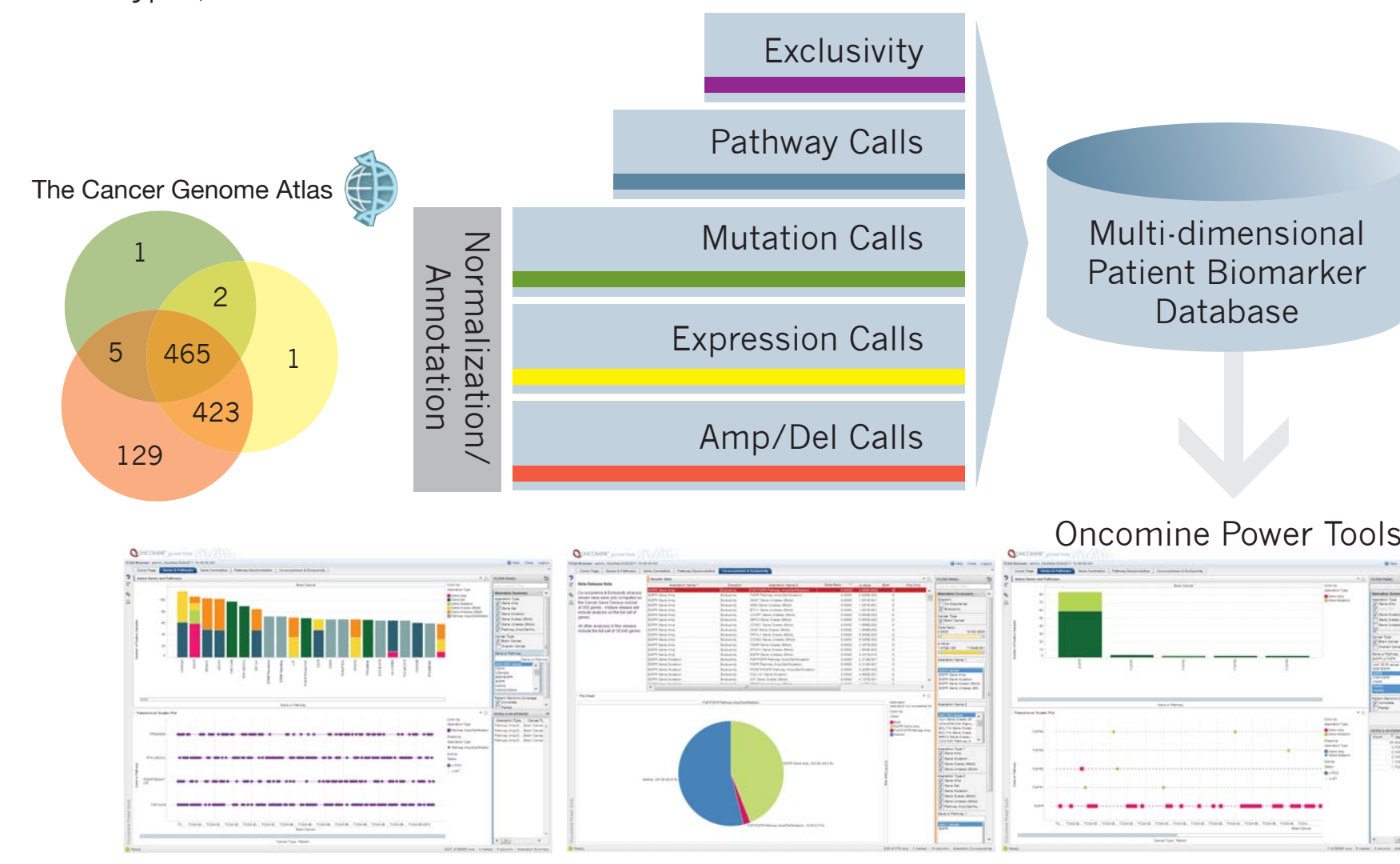


Figure 2. Pathway aberrations summarized in glioblastomas.

The bar chart provides a summary of all aberrations both at the gene and pathway level. Height of the bar and size of bar segments indicate the number of positive patients. Colors are used to denote the various aberrations that are summarized. A single aberration within a pathway is considered a positive result for the pathway. The dark green bars indicate the selection of four pathways, Cell Cycle, RTK (KEGG) PI3K/PTEN/mTOR and TP53/MDM, which retrieves patient level aberrations in the lower scatter plot. In this scatter plot most patients on the x-axis have multiple aberrations, represented as large markers, across selected pathways confirming earlier findings in glioblastoma.

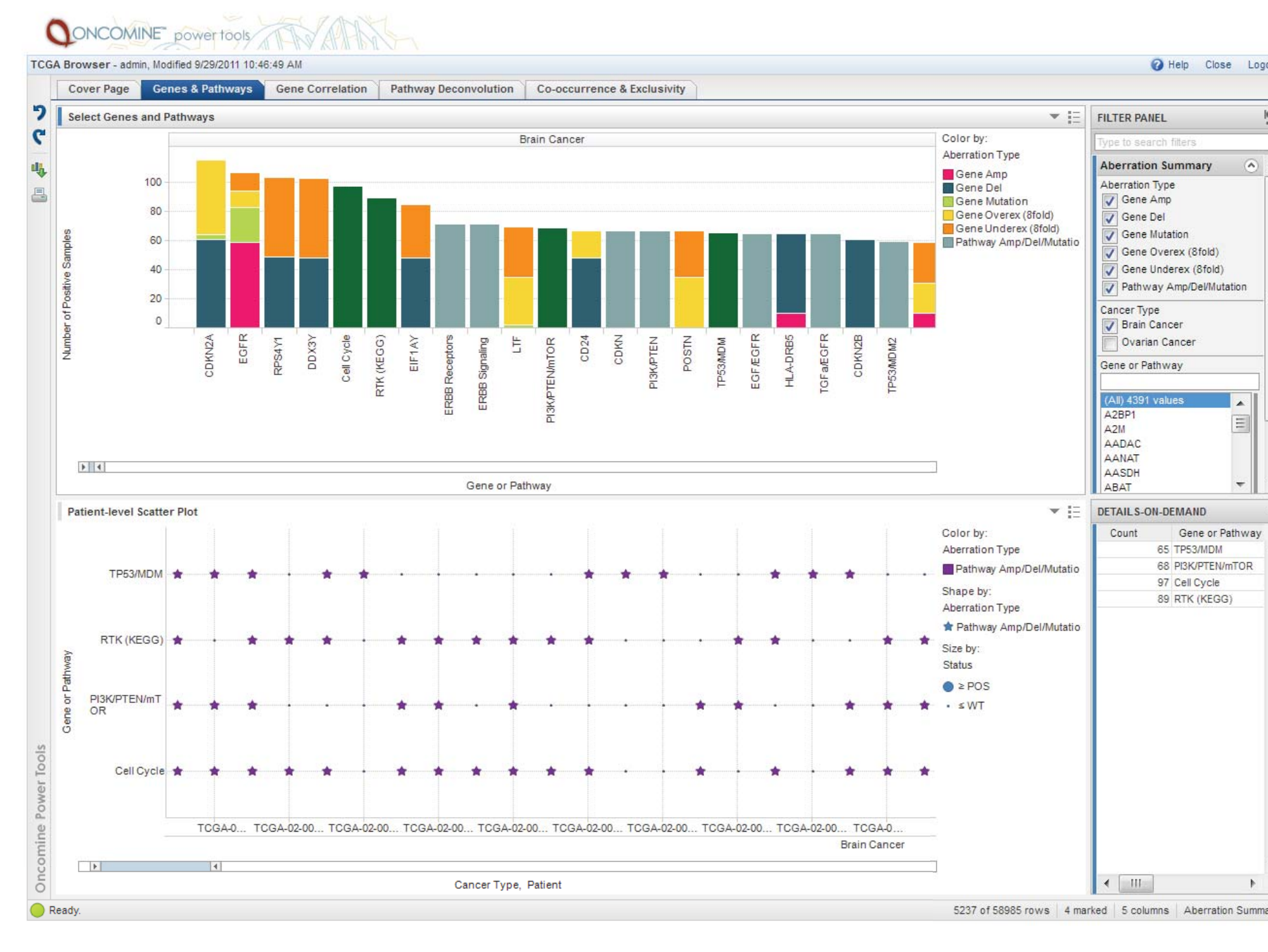


Figure 3. Dissecting pathway member aberrations.

In the bar chart, pathways and corresponding member genes are shown on the x-axis with height of bar and size of bar segment reflecting the number of positive patients for the corresponding aberration. Bars in dark blue reflect the selected individual members of several pathways that result in dynamic retrieval of patient-level data for these genes in the scatter plot view below. In the patient-level scatter plot we see each pathway member gene on the y-axis across all patients on the x-axis with aberration type differentiated by marker shape and color. This result shows that TP53, RB1 and PTEN aberrations are exclusively mutations or deletions, suggesting loss of function in their role as tumor suppressors. CDKN2A is also shown as deleted in most cases and, to a lesser extent, mutated. EGFR shows some variability in aberrations with mostly amplifications and co-occurring mutations, while MDM2 and CDK4 show DNA copy number gain and over-expression. In contrast, a low frequency FLT3 mutation was seen in a single patient. These results confirmed the finding of multiple aberrations in specific members of the RTK, TP53 and RB signaling pathways in glioblastoma.

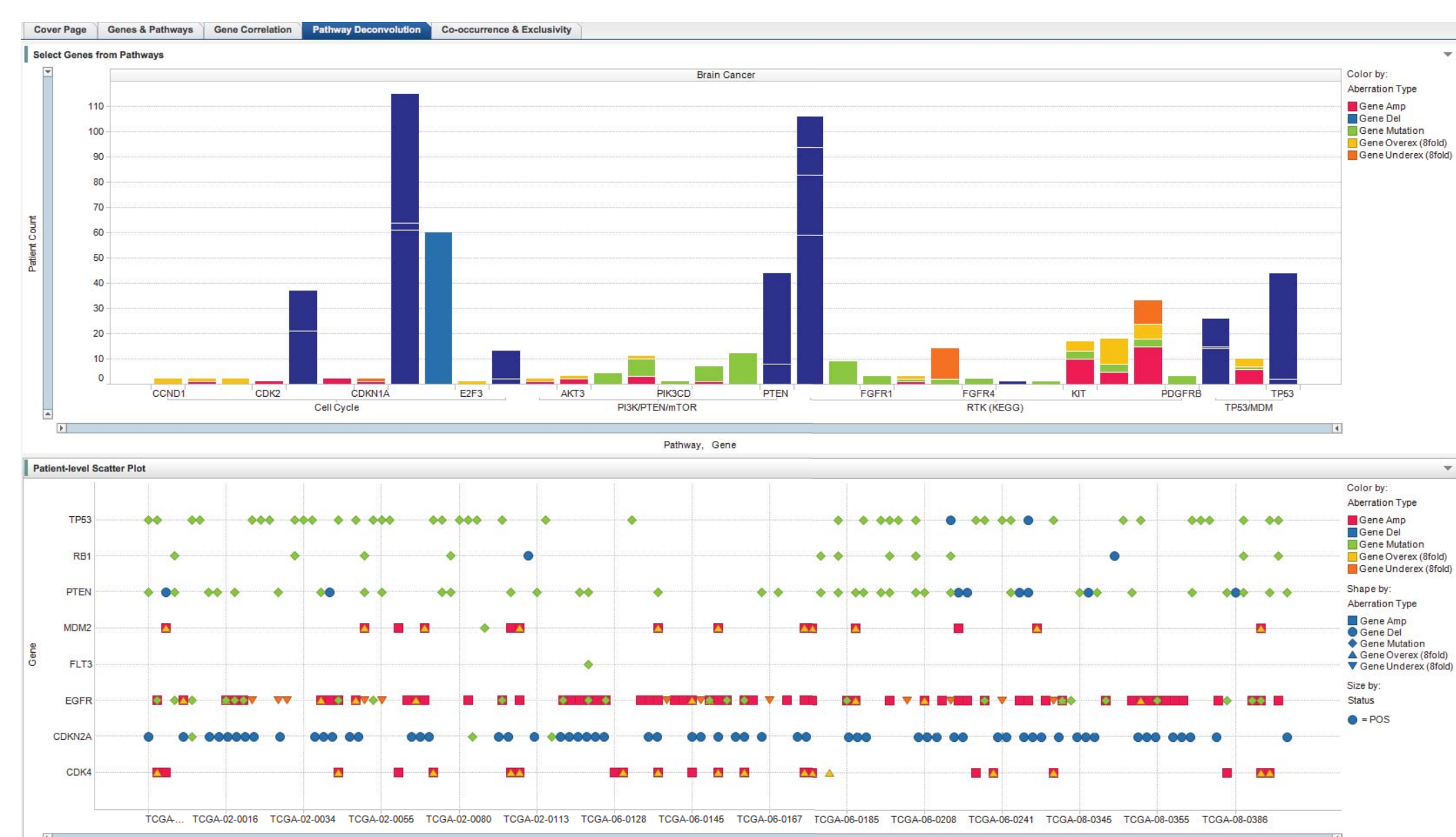


Figure 4. Gene correlation of expression, DNA copy number gain/loss showing mutations.

Further investigation of several members of signaling pathways involved in glioblastoma reveals the correlation of over- and under-expression (y-axis) and DNA copy number gain/loss (x-axis). DNA copy number loss and under-expression is particularly striking for CDKN2A in most patients. In contrast, EGFR has strong correlation of over-expression and DNA copy number gain, with a significant number of patients exhibiting multiple EGFR aberrations, including DNA copy number gain, over-expression and mutation.

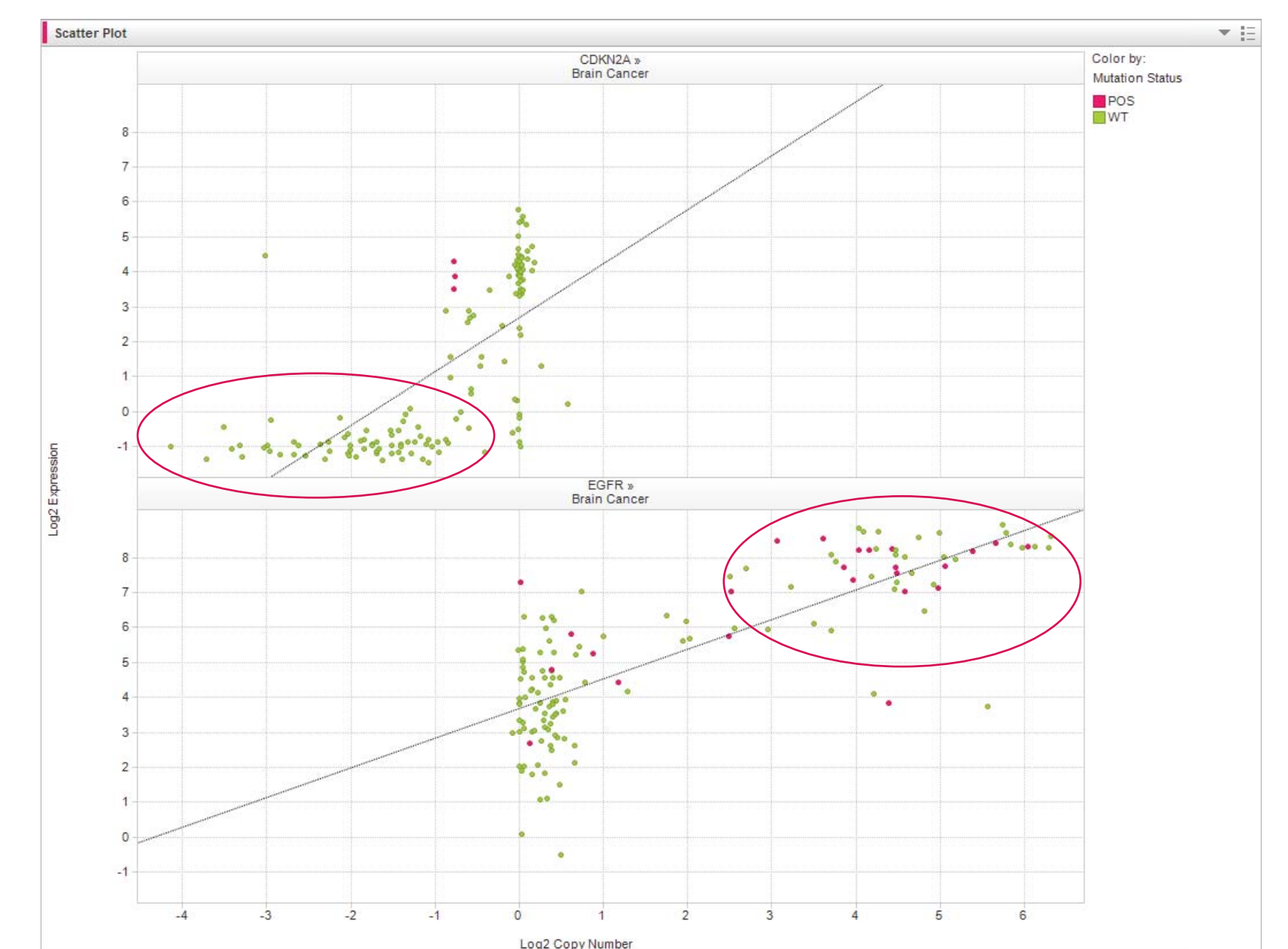


Figure 5. Co-occurrence and exclusivity analysis.

A table of significant co-occurrence and exclusivity can be sorted and filtered to identify new relationships between genes and pathway aberrations. Here we have focused on EGFR DNA copy number gain and note a new finding of mutual exclusivity with FGF/FGFR pathway aberrations. Selecting a result from the summary table displays the pie chart showing that in 247 patients neither EGFR amplification nor FGFR pathway aberrations occurred. FGF/FGFR pathway aberrations occurred in 10 cases and EGFR amplifications in 202 cases. No cases harbored both aberrations, thus they are considered mutually exclusive.

Aberration Name 1	Direction	Aberration Name 2	Odds Ratio	p-value	Both	First Only	Second Only	Neither
EGFR Gene Amp	Exclusivity	FGF/FGFR Pathway Amp/Del/Mutation	0.0000	2.00E-003	0	222	10	247
EGFR Gene Amp	Exclusivity	FGFR Pathway Amp/Del/Mutation	0.0000	2.00E-003	0	202	7	247
EGFR Gene Amp	Exclusivity	CDKN2A Gene Underex (IFRA)	0.0000	8.00E-003	0	163	9	266
EGFR Gene Amp	Exclusivity	CDKN2A Gene Overex (IFRA)	0.0000	8.00E-003	0	163	5	232
EGFR Gene Amp	Exclusivity	EGFR Gene Underex (IFRA)	0.0000	4.47E-010	0	163	36	181
EGFR Gene Amp	Exclusivity	EGFR Gene Overex (IFRA)	0.0000	1.01E-001	0	163	4	213
EGFR Gene Amp	Exclusivity	PTEN Gene Overex (IFRA)	0.0000	8.00E-003	0	163	9	208
EGFR Gene Amp	Exclusivity	CDK4 Gene Overex (IFRA)	0.0000	1.01E-001	0	163	4	213
EGFR Gene Amp	Exclusivity	MDM2 Gene Underex (IFRA)	0.0000	1.01E-001	0	163	8	208
EGFR Gene Amp	Exclusivity	MDM2 Gene Overex (IFRA)	0.0000	1.01E-001	0	163	11	206
EGFR Gene Amp	Exclusivity	CDK2 Gene Underex (IFRA)	0.0000	8.00E-003	0	163	10	207
EGFR Gene Amp	Exclusivity	CDK2 Gene Overex (IFRA)	0.0000	8.00E-003	0	163	29	180
EGFR Gene Amp	Exclusivity	PTEN Gene Underex (IFRA)	0.0000	8.00E-003	0	163	22	195
EGFR Gene Amp	Exclusivity	PTEN Gene Overex (IFRA)	0.1101	1.74E-004	2	161	22	195
EGFR Gene Amp	Exclusivity	RTK Gene Overex (IFRA)	0.1427	3.97E-002	1	162	9	208

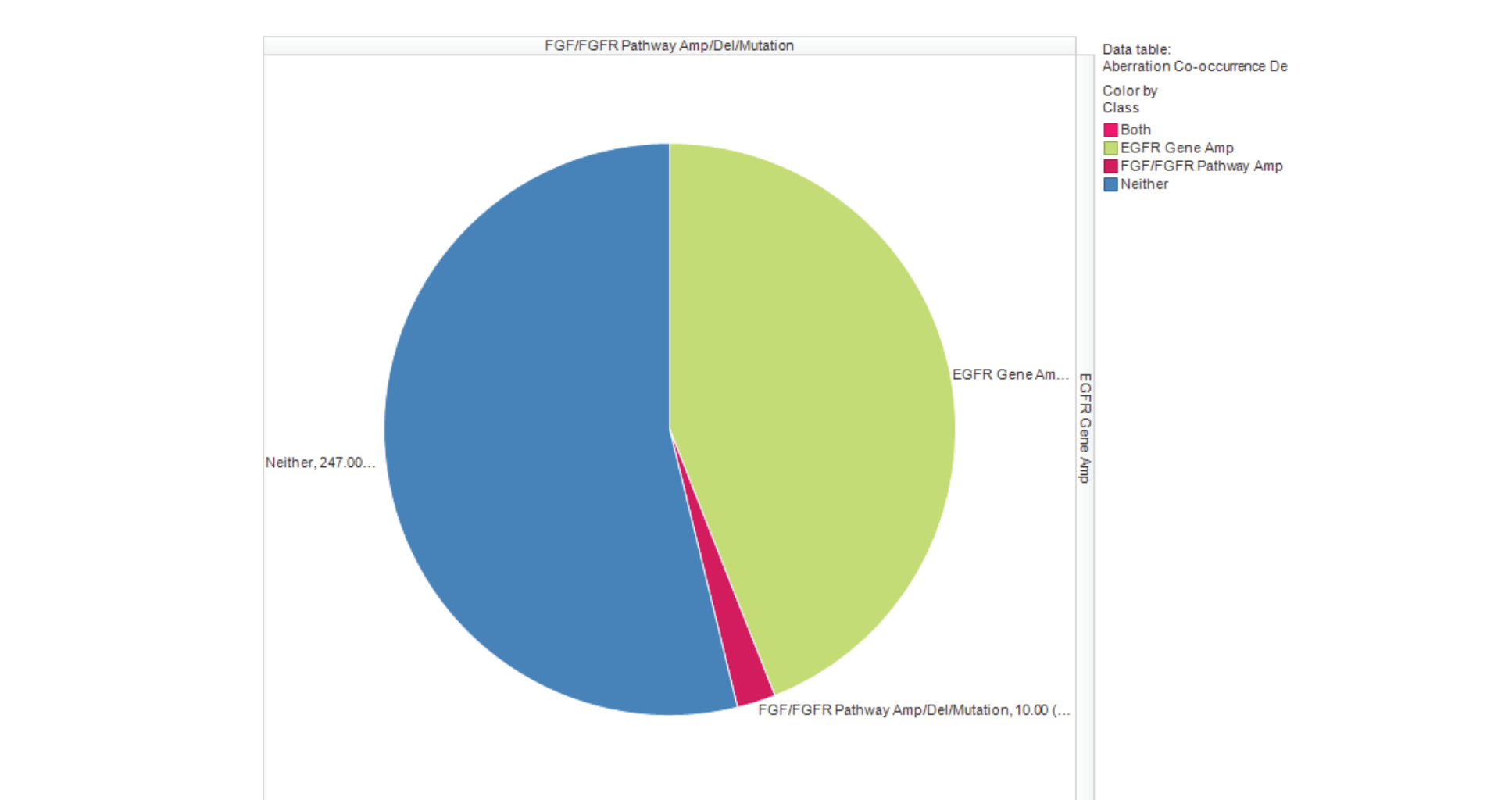


Figure 6. FGFR pathway member aberrations and EGFR DNA copy number gain in glioblastomas.

In this view EGFR and FGFR pathway member genes are selected in the bar chart, resulting in dynamic retrieval of patient-level data in the scatter plot below. The patient-level scatter plot shows the distinct occurrence of FGFRs 1–4 and EGFR DNA copy number gain. None of the four FGFRs has co-occurring aberrations in any one glioblastoma suggesting FGFR aberrations are unique events within a subset of glioblastomas independent of EGFR DNA copy number gain.

