Supplementary File S1

Description of omic datasets and quality control

TCGA offers omics data at different processing levels (The Cancer Genome Atlas Research Network 2008). We used level 3 RNA gene expression and methylation β-values at CpG dinucleotides from tumor tissue, and level 2 SNP genotypes from normal tissue. Level 2 allele-specific copy number estimates were read with the Illumina 550 K Infinium HumanHap550 SNP Chip. The level 2 dataset corresponds to normalized intensity values for each allele of every SNP, with the log of total intensity summed over both alleles and considering B allele frequency values for each marker. SNP with more than 10% of missing genotypes, with pairwise genotypic correlation greater than 80%, and with a minor allele frequency (MAF) smaller than 0.05 were removed. Also, individuals with more than 10% of missing genotypes were not included. SNP data was edited using PLINK software (Purcell et al. 2007). Copy Number Variants (CNV) intensity was obtained from whole genome microarray by Affymetrix Genome-Wide Human SNP Array 6.0 platform at the Broad Institute Genome Characterization Center. Gene –level CNV intensity corresponded with the average copy number of a gene's genomic region. As for the other omics, samples with more than 10% missing information and with low variation were excluded. Level 3 DNA methylation data (from Illumina Infinium HumanMethylation27 platform), containing β-values (Sandoval et al. 2011) expressing the DNA methylation scaled for each probe. Probes with more than 10% missing information, or with low variation (standard deviation < 0.2), were discarded, as were individuals with more than 10% missing values. β-values were quantile-normalized and converted into M-values through a logistic transformation using the R package wateRmelon (Pidsley et al. 2013). The distribution of

missing information was completely at random after performing the Wald-Wolfowitz test of randomness for continuous data. The test was implemented using the randtests R package (Caeiro and Mateus, 2014). Therefore, missing values on remaining probes were imputed using the mean value of the sample. All probes were pre-corrected by batch effect with a mixed model (Lazar et al., 2013). Specifically, the gene expression for the j^{th} gene was considered as response in a model including the batch for that particular gene as the only random effect. Residuals obtained from the model for each gene were used in further analysis. RNA gene expression reads by locus corresponded to samples generated in 16 unique batches by the NCI, using the Affymetrix HT Human Genome U133A Array, log transformed and with the mean (or median) value subtracted from each expression level (The Cancer Genome Atlas Research Network 2008). Gene expression samples with more than 10% missing information or with low variation were discarded (Standard deviation < 0.2; Table S1 below). Finally, gene expression data was corrected by batch effect with a mixed model.

Table S1. Summary of omic information from normal and tumor tissue from GBM used in this study.

Omic profile and platform	Initial sample size	Filtering criteria	Sample size after edition
SNP genotypes from normal tissue (Illumina 550 K Infinium HumanHap550 SNP Chip	350 individuals 561,466 SNP.	Individuals and SNP with >10% missing genotypes, and SNP with minor allele frequency (MAF) <0.05.	350 individuals 509,197 SNP
DNA Methylation (Illumina Infinium HumanMethylation27)	287 individuals 27,578 CpG (27K)	β-values were normalized, transformed to M-values. Sites and individuals with >10% of missing values and sites with SD < 0.2 were discarded.	287 individuals and 23,972 CpG sites
Copy Number Variation (Affymetrix Genome-Wide Human SNP Array 6.0)	504 individuals, 22,937 genes	Samples with more than 10% missing information and with low variation.	443 individuals, 22,937 genes
mRNA gene expression (Affymetrix HT Human Genome U133A Array)	560 individuals 12,042 genes	Patients without clinical information, genes with low variation (SD < 0.2) and >10% missing information.	484 individuals 10,831 genes