ABSTRACT

Title of Document: USING STATISTICAL METHOD TO REVEAL BIOLOGICAL ASPECT OF HUMAN DISEASE: STUDY OF GLIOBLASTOMA BY USING COMPARATIVE GENOMIC HYBRIDIZATION (CGH) METHOD

Yonghong Wang, MS, 2010

Directed By: Dr. Paul Smith, Mathematics Department

Glioblastoma is a WHO grade IV tumor with high mortality rate. In order to identify the underlying biological causation of this disease, a comparative genomic hybridization dataset generated from 170 patients' tumor samples was analyzed. Of many available segmentation algorithms, I focused mainly on two most acceptable methods: Homogeneous Hidden Markov Models (HHMM) and Circular Binary Segmentation (CBS). Simulations show that CBS tends to give better segmentation result with low false discovery rate. HHMM failed to identify many obvious breakpoints that CBS identified. On the other hand, HHMM succeeds in identifying many single probe aberrations.

Applying other statistical algorithms revealed distinct biological fingerprints of Glioblastoma disease, which includes many signature genes and biological pathways. Survival analysis also reveals that several segments actually correlate to the extended survival time of some patients.

In summary, this work shows the importance of statistical model or algorithms in the modern genomic research.

USING STATISTICAL METHOD TO REVEAL BIOLOGICAL ASPECT OF HUMAN DISEASES: STUDY OF GLIOBLASTOMA BY USING COMPARATIVE GENOMIC HYBRIDIZATION (CGH) METHOD

By

Yonghong Wang, Ph.D

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Science
2010

Advisory Committee:
Professor Paul Smith, Chair
Dr. Paul Meltzer
Professor Eric Slud

Acknowledgements

First of all, I would like to thank Dr. Paul Smith for leading me through this thesis writing process, especially for all his advises for the paper.

In the mean time, I would also give my thankfulness to my supervisor, Dr. Paul Meltzer. During the last several years, I have received all kind of supports as well as helps. Without his support, it is basically impossible for me to get to this stage of my study.

I would also like to thank all the faculty members at Mathematics Department of University of Maryland, including Dr. Slud and Dr. Smith. With their help, I have learnt some new knowledge that will be definitely helpful in my scientific career.

It has been more than 15 years since I got my Ph.D in Chemistry. It is always a big challenge for me to go back to the classroom. Without the encouragement and help from my family, it is impossible for me to become a student again. I will give my thanks to my wife and two lovely boys. Thanks them for always with me.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

Cancer is a class of diseases in which a group of cells display uncontrolled growth, invasion, and sometimes metastasis (http://en.wikipedia.org/wiki/Cancer). More than 200 different types of cancers have been identified so far and all these cancers account for 13% of all human deaths across the whole world.

Cancer could start from any tissue of human body, and many factors have been found to contribute to the development of the disease. These include environmental factors (pollution, smoking, food, ultraviolet radiation from the sun, et al.), genetic factors, as well as other factors such as infections from viruses. Genetically, the direct consequences of the effects of all these factors are the alteration of genetic information including the uncontrolled gene expression of the oncogenes or inhibition of the activity of tumor suppression genes.

The human genome contains 23 pairs of chromosomes with 22 pairs of autosomal chromosomes and 1 pair of sex chromosomes (XX for female and XY for male). There are about 3 billion base pairs of nucleotides for the entire human genome, which contains about 20-30 thousands protein-encoding genes. In fact, only 1.5% of the entire human genome sequence encodes protein-coding information, while the function of most of the remaining sequence remains unknown.

Cancer is the result of the accumulation of multiple genetic changes (Armitage and Doll 1954), which potentially become cytogenetically visible with the development of the cancer. Comparative Genomic Hybridization (CGH) is the study of chromosome aberrations and their correlations to human diseases. The common tumor chromosome aberrations are generally classified as numerical and structural.

Numerical aberration mainly involves the changes of chromosome numbers (Aneuploidy) and it is believed that some cancers are involved in the aneuploidy (Sen 2000). Structure aberration refers to copy number variations (copy number gains or losses) of some segments of chromosomes. Many tumors have shown to be related to chromosome aberration, for instance renal cell carcinoma (chromosome positions 3p13-21) (Mulshine, et al. 1988), lung adenocarcinoma (chromosome positions 3p13-23) (Mulshine, et al. 1988), etc.

Traditionally, due to the limitation of human genome sequence information, staining has been the main method to study the karyotypes of human genome, usually using dyes, such as Giemsa, which results in the G-banding method. For this method, chromosomes are first stained by Giemsa and then followed by digestions with trypsin. In this way, the chromosomes show a series of lightly and darkly stained bands (Figure 1.1). The dark regions tend to be heterochromatic, late-replicating and AT rich, and the light regions are euchromatic, early replicating and GC rich. Usually G-banding can produce 300-400 bands in a normal human genome, with the average resolution around 10M base pairs (bp). FISH or Fluorescence In-Situ Hybridization is a rapid reliable technique in molecular cytogenetics (Celep, Karaguzel, Ozgur and Yildiz 2003) and can be used to identify isolated abnormal cells among a large group of normal cells. It is also a method to locate a specific gene on the chromosome or to study translocations of a known gene. As with G-banding methods to detect genomic variation, the main problems of FISH are also slowness, low throughput and low resolution.

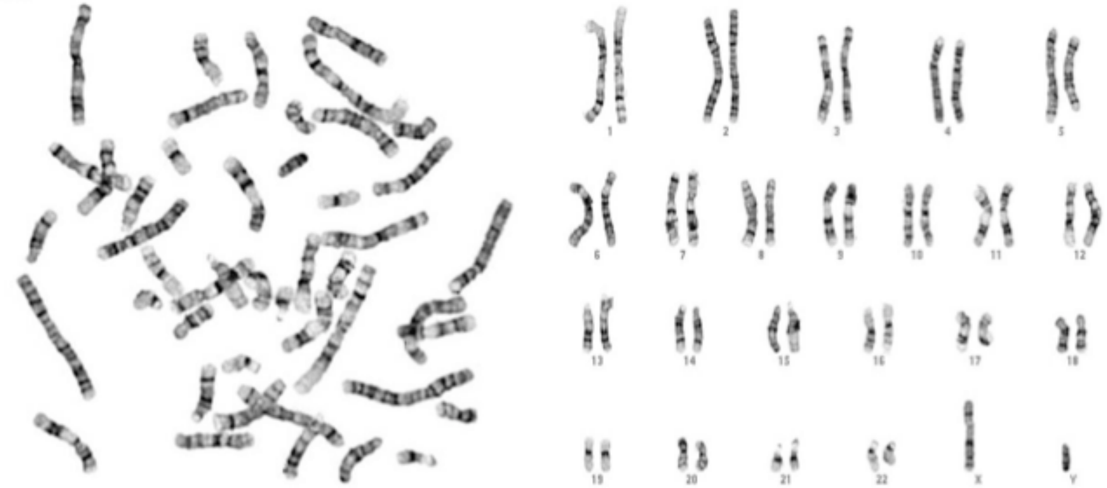


Figure 1.1 Chromosome spread and the karyotypes stained by the G-banding method. Chromosomes shown here are from a normal male.

Array-based comparative genomic hybridization (aCGH) is similar to the traditional karyotyping methods but with the advantages of high throughput and high resolution. For this method, hundreds of thousands of oligonucleotides based on human genome sequences are spotted on an area of roughly 1cm^2 on a 1"x 3" glass slide and tumor and reference samples labeled with different dyes are hybridized to the chip. Genomic gains and losses can be analyzed genome widely by comparing sample genomic DNA (in our case, tumor samples) to reference genomic DNA sample in one experiment; this will be discussed in detail in the experimental section.

The ultimate goal of this thesis is to study the underlying causes in the genomic level of human cancer Glioblastoma by using the array CGH method. Since we are dealing with millions of data points, in order to get more reliable biological results, we will apply various statistical methods to process the data set and compare the results generated by these methods.

Chapter 2: Experiment and collected data structure

2.1 Experiments

All the data used are downloaded from the Cancer Genome Atlas ("cancergenome.nih.gov"). The Agilent array platform – Human Genome CGH 244k chip – is used in this study. On each chip, 244,000 probes are spotted with 236,000 distinct biological features. There are also 1000 replicated biological features and 5045 internal quality control features. Each feature represents one 60mer oligonucleotide designed based on human genome sequence. The average distance on the chromosomes between two features is about 6.4kb. Brain Glioblastoma samples collected from 170 patients were analyzed by CGH at Broad Institute/Dana-Farber Cancer Institute. In addition, 24 normal samples were analyzed by CGH at Broad Institute and this data set is used as control in this analysis. Table 2.1.1 lists brief clinical information about the 170 patients. Detailed sample information can be obtained from the published paper (TCGA 2008) (also see supplementary file Table S2.1.1. For the array CGH experiment (hybridization), individual patient's genomic DNA and a common reference genomic DNA are labeled with two different fluorescence dyes (which emit red and green light after laser ignition) and hybridized to the Human Genome CGH 244k chips. Figure 2.1.1 shows the brief procedures to perform the CGH experiment.

total patients with tumor	170
gender	FEMALE: 67; MALE:103
pretreatment history	No: 148; Yes: 19
histological type	Treated primary GBM: 15; Untreated primary (De Nova) GBM:154
vital status1	DECEASED:159; LIVING: 10
vital status2	ALIVE: 15; DEAD: 155
days to birth	Min.: -31627; Q1: -24502; Median: -20931; Mean: -20375; Q3: -17080; Max.: -5303; NA's: 1
days to death	Min.: 8.0; Q1: 209.0; Median: 370.0; Mean: 516.7; Q3: 597.0; Max.: 3040.0; NA's: 11.0
days to last follow up	Min.: 3.0; Q1: 161.0; Median: 357.0; Mean: 519.6; Q3:578.0; Max.: 3040.0; NA's: 1.0
age at diagnosis	Min.: 14.00; Q1:46.00; Median: 57.00; Mean: 55.32; Q3: 67.00; Max.: 86.00; NA's: 1.00
day to progression	Min.: -1409.0; Q1: 88.0; Median: 183.0; Mean: 269.1; Q3: 334.0; Max.: 2339.0; NA's: 61.0
days to recurrence	Min.: 7.0; Q1: 113.0; Median: 283.0; Mean: 523.7; Q3: 524.0; Max.: 2296.0; NA's: 144.0
secondary or recurrent	No: 149; Rec: 16; Sec: 5
race	AMERICAN INDIAN OR ALASKAN NATIVE: 1; ASIAN: 1; BLACK OR AFRICAN AMERICAN: 2; UNK: 88; WHITE: 78
age at procedure	Min.: 15.00; Q1.: 47.00; Median: 57.00; Mean: 55.68; Q3: 66.50; Max.: 86.00; NA's: 3
age at death	Min.:15.00; Q1: 49.25; Median: 58.00; Mean: 56.82; Q3: 68.00; Max.: 87.00;

Table 2.1.1: Summarized information for all 170 patients used in this study. The WHO categories for all patients are grade IV and the median survival rate is approximately one year as states in the published paper (TCGA 2008). All the date information is relative to the zero date when patients seek treatment. The “AGE at DIAGNOSIS” indicates patient’s age at initial pathologic diagnosis. “Day to progression” stands for “days to tumor progression”, and “Day to recurrence” for “days to tumor recurrence”. There are two vital status columns available in the table. The discrepancy is due to the inconsistent information from published paper (TCGA 2008) and the TCGA website (http://tcga-data.nci.nih.gov/tcga/homepage.htm).

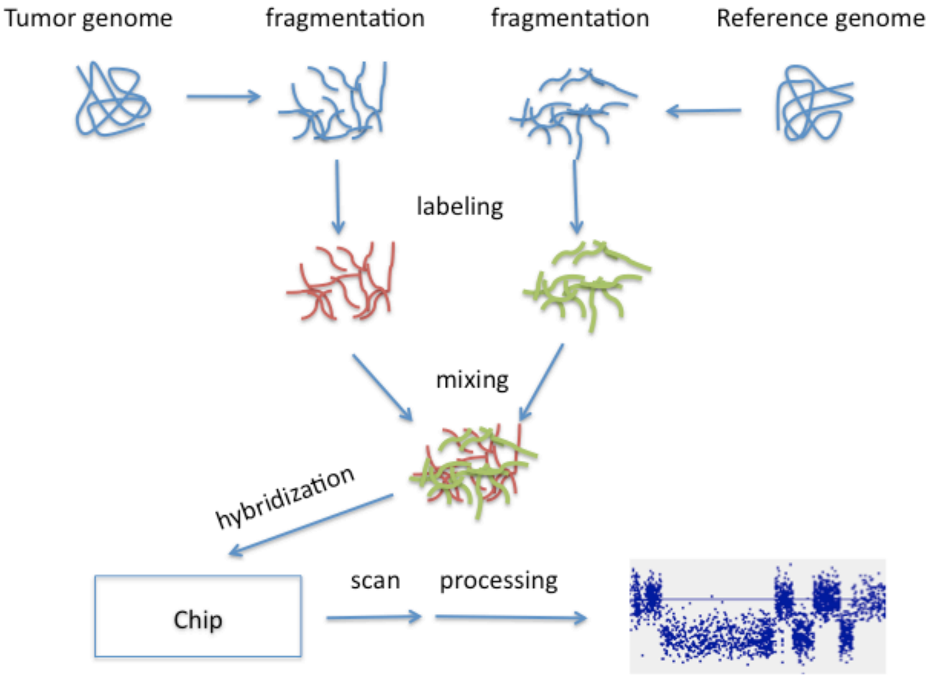


Figure 2.1.1 Schematic description of array CGH experiments. First the sample genomic DNA is isolated from patients' tissue samples (in this study, glioblastoma tumor) and then fragmented to yield the desirable around 200 bp short DNA population. The reference genomic DNA purchased from commercial sources is also processed in the similar way. After then, both the fragmented tumor DNA and the reference DNA are labeled with Cy5 and Cy3 dyes respectively and the mixture of these two samples is hybridized onto the human genome CGH 244k chip, which is followed by the scanning procedure. Cy5 and Cy3 dyes can generate red and green light, respectively, upon the excitation by laser light, and the intensities of the two dyes of all features are measured for further data analysis.

2.2 Data Structure

After hybridization, the chip is scanned and the intensities of red and green colors as well as the corresponding background intensities for each spot are recorded. The $\log_2$ transformed Ratio of the two background-subtracted fluorescent dyes of the same probe was used to detect the copy number variations for this probe or feature.

As mentioned above, each chip (from each patient out of a total 170 patients) could yield 236k distinct data sets (each set includes Cy3 and Cy5 intensities, the corresponding backgrounds, as well as other data information). In addition, there are also another 24 normal samples from 24 patients. But only four patients have both tumor samples and normal samples in our analyzed data set. Each sample (tumor or normal) is used to perform hybridization on the chip and the collected data sets are used for data processing and analysis.

2.3 Data processing and analyzing method

The data processing mainly relies on the R software "BioConductor" packages (www.bioconductor.org). Specifically, for the CGH part, the "snapCGH" package developed by Smith (Smith 2006) was used. For the survival analysis, "survival" package developed by Therneau (Terry Therneau 2009) and "KMsurve" package by Yan (Yan 2010) were used. The main codes for these analyses are listed as supplementary file. Many other R packages were also being called in the above individual package in order for the full functioning. These R packages include LIMMA (Gordon K. Smyth 2010), and aCGH (Willenbrock and Fridlyand 2005). Another powerful computation tool used in the data processing is Perl scripts and other R functions. Occasionally other commercially available software such as "NEXUS" is used, mainly for presentation purposes.

Chapter 3: Literature Review

3.1 Copy number variation

Array CGH is a high throughput technology to study the chromosome aberration of target samples. Specifically, suitable segmentation methods are used to identify the breakpoint of the DNA copy numbers and to assign gains, losses or no change to all segments based on the $\log_2$(Ratio) and the cutoff criteria applied. For diploid cells, there are $2n$ numbers of chromosomes. In human cells, especially somatic cells, this n is 23. Theoretically, the copy number variation is inferred by the $\log_2$(Ratio) of intensities of tumor samples vs. reference sample. For instance, one copy number loss would yield $\log_2(1/2) = -1$. Table 3.1.1 lists the relations of copy number variation and $\log_2$(Ratio). Practically, due to many random and systematic errors involved in each experiment, the observed $\log_2$(Ratio) values of copy number variation are far from their theoretical values. Nevertheless, we can still derive the chromosome aberration information with proper segmentation methods based on aggregation patterns.

Example	Copy number variation	Ratio	$\log_2$(Ratio)
00/AA	two copy losses	0	$-\infty$
A/AA	one copy loss	1/2	-1
AA/AA	no change	1	0
AAA/AA	one copy gain	3/2	0.58
AAAA/AA	two copy gains	2	1

Table 3.1.1 the relations of copy number variation with the observed $\log_2$(Ratio) in the idealized situation of the diploid genome. In this study, we assume that the reference sample has a perfect diploid genome that is represented by "AA" in the denominator position. The numerator positions represent the ploidy situations of individual sample (or segments within the chromosome regions). For instance, "00", stand for homozygous deletions and "A" indicates that only one copy of the diploid segment remains, which suggests heterozygous deletion. "AA" at the numerator shows no copy number changes, "AAA" indicates one copy gain and "AAAA" indicated 2 copy number gains. It is possible to have multiple copy number gains for some genes or segments.

3.2 Segmentation methods

During the past several years, many segmentation algorithms have been developed, for instance, Homogeneous Hidden Markov Models (HHMM) (Fridlyand et al. 2004) and Heterogeneous Hidden Markov Modes (BioHMM) (Marioni, Thorne and Tavare 2006), assign probes to underlying segments or states, which represent different copy numbers. Circular Binary Segmentation (CBS) (Olshen, et al., 2004) is a non-parametric change point method based on the original work of Sen and Srivastava (1975). Hupe proposed a Gaussian model-based algorithm (GLAD) (Hupe 2004) and Hsu suggested a wavelet approach (Hsu, et al. 2005). There are still some other approaches such as hierarchical cluster based methods (CLAC) (Wang, Kim, Pollack, Narasimhan and Tibshirani 2005), Shifting Level Model (SLM) (Magi, et al. 2009), and the Bayesian based model of Daruwala (Daruwala, et al. 2004), et al. In this section, I will briefly introduce some of the segmentation methods and some pros and cons involved in some of the major methods.

3.2.1 Homogeneous Hidden Markov Models (HHMM)

The following description of HHMM segmentation method is taken mainly from Fridlyand et al. (2004). Since each segment could be treated as a state with equal copy number and the total numbers of segments or states is unknown but fixed, the hidden Markov models could be used as a suitable statistical model for the identification of all the unknown states. In this model, each state is reachable from any other state and the state emission distributions are Gaussian with state specific mean and common variance. The initial state distribution and the transition state probability distributions comply with Markov Chain modeling requirement as shown below (cited mainly from Fridlyand et al., 2004):

a) The number of states, K, is unknown but fixed. The Markov chain is assumed irreducible with individual states space S = {S_1, S_2, …, S_K} and the state at location l is S_l, where $1 \leq l \leq L$.

b) The initial state distribution is $\pi = \{ \pi_k\}$ where $\pi_k = P\{s_1 = S_k\}$, $1 \leq k \leq K$.

c) The state transition probability matrix is $A=[a_{mp}]$ where $a_{mp}=P\{s_{l+1}=S_p|s_l=S_m\}$, $1 \leq m, p \leq K$.

d) The distribution of emission b_k is Gaussian with unknown mean and variance:

$b_k \sim N(\mu_k, \sigma_k^2)$ for $1 \leq k \leq K$

A forward-backward procedure is used to calculate the likelihood of parameters (Lik(λ|O), where λ is the parameter vector for the model and O represents the sequence of observed values on a given chromosome. The number of states K is estimated using the parameter correction penalty function $\Psi(K)$ similar to Akaike's Information Criterion (AIC) (Akaike 1969) or Bayesian Information Criterion (BIC) (Schwarz 1978) as shown below:

$$\Psi(\mathrm{K}) = -\log(\mathrm{Lik}(\lambda|\mathrm{O})) + \mathrm{q_K D(L)/L}.$$

Briefly, for k = 1 to $K_{\max}$, where $K_{\max}$ = 5 (based on the observation from real data sets), first fit the k-state HHMM model and calculate the penalized negative log-likelihood $\Psi(k)$. Choose the model corresponding to the number of states with the smallest $\Psi(k)$, and so that $K = \mathrm{argmin}_k\Psi(k)$. If K = 1, then stop the modeling. Otherwise, perform merge states test by calculating the absolute median differences of two states and compare to a cutoff threshold. If the difference is smaller than the cutoff threshold, merge the two states and also set $K = K$-1. Repeat the same testing until no merging is possible.

3.2.2 Heterogeneous Hidden Markov Models (BioHMM)

In homogenous hidden Markov models, the spatial information (the distances between the adjacent probes) is not taken into account. But in fact, due to specific properties of chromosomes, nucleotides are not evenly distributed and in many areas it is very hard to design suitable probes for the array CGH study. Therefore the distances between two adjacent probes vary significantly. This uneven tiling of probes on the chromosome reflects the biological properties embedded in each chromosome. BioHMM (Marioni, et al. 2006) is a segmentation method constructed based on the HHMM algorithm, but adding probe distance as an additional factor in the transition probabilities. Practically, the distance is defined as differences between two adjacent probes and this information is incorporated into the transition probabilities as shown below:

$$A_i = \begin{pmatrix} 1 - p_1(1 - \mathrm{e}^{-f_i}) & p_1(1 - \mathrm{e}^{-f_i}) \\ p_2(1 - \mathrm{e}^{-f_i}) & 1 - p_2(1 - \mathrm{e}^{-f_i}) \end{pmatrix}$$

where $f_i = x_\mathrm{i}^\mathrm{r}$, x_i is the vector containing the distance information and r belongs to R with initial value as 1. Detailed algorithms for this method can be found in the cited paper (Marioni et al., 2006).

3.2.3 Circular Binary Segmentation (CBS) method

CBS was proposed by Olshen (Olshen et al., 2004) based on the binary segmentation model proposed Sen and Srivastava (1975). The underlying framework for this model is that the gains and losses are discrete events and the aberrations occur in contiguous regions of the chromosomes. Segmentation is in fact a process to find the change points or breakpoints along the chromosome. For the binary segmentation method proposed by Sen and Srivastava, the third breakpoint located within the two known breakpoints will split the original segment into two segments. Both segments were assumed to follow normal distributions with equal variance but unknown means. Suppose $\mathrm{X_i}$ is the $\log_2$(Ratio) of n consecutive probes $\{X_1, X_2, \ldots, X_n\}$, Maximum likelihood ratio statistics for testing the null hypothesis is given by $Z_b = max_{1<i<n}|Z_i|$, where $Z_i = \{1/i+1/(n-i)\}^{-1/2}\{S_i/i-(S_n-S_i)/(n-i)\}$ and $S_i = X_1+X_2+X_i$, 1<i<n.

One problem with the binary segmentation method is that it can only detect one breakpoint each time. Olshen et al. (2004) proposed a modified test called the Circular Binary Segmentation (CBS) method. Instead of finding one breakpoint suggested by binary segmentation method, CBS assumes the original two breakpoints were linked together to form a cycle and then trying to assign two additional breakpoints on the cycle following the similar procedures as shown in binary segmentation method. In this case, the maximum likelihood ratio statistic for testing the null hypothesis is modified to $Z_c = max_{1\leq i<j\leq n}|Z_{ij}|$, in which

$$Z_{ij}=\{1/(j-i)+1/(n-j+i)\}-1/2\{(S_j-S_i)/(j-i)-(S_n-S_j+S_i)/(n-j+i)\}$$

where $1<i<j<n$, $S_n = X_1 + X_2+\ldots+X_n$ and X_i denotes $\log_2$(Ratio) of i^{th} probe. Similar to the binary segmentation method, the critical value is also derived by Monte Carlo simulation (Siegmund 1986). This procedure is applied recursively until all change points are identified.

3.2.4 Gain and Loss Analysis of DNA (GLAD) method

GLAD is a breakpoint detection method developed by Hupe et al. (Hupe 2004), mainly based on the adaptive weights smoothing (AWS) procedure proposed by Polzehl (Polzehl and Spokoiny 2000), which is “an iterative, data adaptive smoothing technique that was designed for smoothing in regression problems involving discontinuous regression function” (Hupe 2004). The statistical model for the GLAD method is based on the locally constant Gaussian regression model $Y_i = \theta(X_i) + \varepsilon_i$, where ε_i are i.i.d with $N(0, \sigma^2)$ distribution. (X_i, Y_i) are independent observations, X_i determines the chromosome locations of each individual probe or feature and the corresponding Y_i is the measured $\log_2$(Ratio). AWS is applied to find the maximal possible neighborhood around location X_i based on the local likelihood model in which $\theta(X_i)$ is constant. The regression function of $Y_i = \theta(X_i) + \varepsilon$ could be estimated by using the weighted maximum-likelihood estimate of local probes. The "contrasts" and the "edges" generated from AWS on the chromosomes provide potential segmentation (breakpoint) and its copy number information.

3.2.5 Cluster Along Chromosomes (CLAC) method

The basic idea of CLAC (Wang, et al. 2005) is to build hierarchical cluster trees along each chromosome arm and assign the gains or losses based on the information associated with each node, such as the height of the node in the tree, the size of the sub-tree and the mean value of the sub-tree. Unlike the standard agglomerative clustering method, for the CGH data set, the order of the probe sequences of the chromosome are fixed, and so is the order the cluster. In order to compare the different measurement of two adjacent nodes, the authors proposed of using "relative difference" to measure the similarity of two adjacent sub-tree. If there are n probe on one chromosome with $\{x_1, x_2, \ldots, x_n\}$ represent the $\log_2$(Ratio) of each probe. The relative difference (rd) for two contiguous probes is defined as:

$$\mathtt{rd}(x_i, x_{i+1}) = \frac{|x_i - x_{i+1}|}{|x_i| + |x_{i+1}| + |x_i + x_{i+1}|}.$$

By comparison of the relative difference to the specific cutoff value (defined based on the height of the cluster, span on the chromosome as well as number of the probes), we can assign gains or losses to each segment. A brief clustering procedure within each chromosome is listed as follows:

1). Start clustering using all probes with one probe in each cluster.

2).Merge the two adjacent clusters with the smallest value of relative difference.

3). Repeat 2) until one big cluster is formed.

3.2.6 Bayesian model

In this Bayesian model, some assumptions are made. First, all probes on the chromosomes are categorized into two groups. One group contains all "regular probes", whose copy numbers are not affected by the disease, and whose $\log_2$ transformed ratios follow a Gaussian distribution with mean μ_r and standard deviation σ_r. In the other group, which includes the deviated probes, the $\log_2$(Ratio) values also follow the Gaussian distribution but have unknown mean μ_b and unknown standard deviation σ_b. Another assumption is that the segments that have copy number variation (gains or losses) follow the Poisson distribution with parameter p_bN, where N is the total number of probes. The segmentation algorithm is to find all the potential breakpoints on the chromosome, or equivalently, to subdivide all the probes along the chromosome into k non-overlapping intervals. Probes within the same interval follow the same distribution with same mean (or same copy number variations) and standard deviation. Specifically, for the j-th interval, the copy number values follow the Gaussian distribution $N(\mu_j, \sigma_j^2)$. The parameter related to this interval is $I_j = (\mu_j, i_j, \sigma_j)$ *where* i_j is the position of the last probe in the interval. In this Bayesian model, we also assume the standard deviations of all intervals are similar.

In this model, the prior distribution is constructed based on both the Poisson distribution to model the number of intervals with Poisson parameter p_bN and the Bernoulli trial where p_r represents the probability that a given probe is a "regular" probe.

$$Pr(\mathsf{I}_N) = e^{-p_bN} \frac{(p_bN)^k}{k!} p_r^{\#regular}(1 - p_r)^{\#deviated}$$

where "#regular" denotes the number of unchanged probes and "#deviated" denotes the number of which assign gains or losses. The likelihood function of the first n probe is:

$$Pr(\mathbf{x}|\mathsf{I}_N) = \prod_{i=1}^{n} \phi(x_i,\ \mu_j,\ \sigma^2)$$

with posterior likelihood function:

$$L(\mathsf{I}_N|\mathbf{x}) = e^{-p_bN} \frac{(p_bN)^k}{k!} p_r^{\#regular} \cdot (1 - p_r)^{\#deviated} \cdot \prod_{i=1}^{n} \phi(x_i,\ \mu_j,\ \sigma^2) .$$

3.2.7 Comparisons of different segmentation methods

Even though many segmentation algorithms have been proposed during the past several years, only a few papers perform comparisons among some of the algorithms. In the recent paper of Willenbrock and Fridlyand (2005) three methods (HHMM, CBS and GLAD) are compared. Of these three methods, CBS yields the best operational sensitivity and lowest false discovery rate compared to the HHMM and GLAD segmentation methods. The drawbacks of this method are its slow computation and its inability to identify many single probe aberrations. On the contrary, the HHMM algorithm is fast and identifies many single probe aberrations. The problem is its low sensitivity to the breakpoint, since it fails to identify many big segments of aberrations. GLAD is said to be superior to HHMM in identifying wide aberration. Lai, et al. (2005) compared 11 different segmentation methods, even though many of these methods are gradually becoming obsolete in current data analysis. Some of their results are very consistent with the conclusions of Willenbrock and Fridlyand. In general, Lai et al. found that CBS performed better in general than most of the other algorithms. The detailed results can be found in the published paper (Lai et al., 2005). As mentioned above, different methods have their own pros and cons. After many years of development, HHMM and CBS have been recommended as more reliable in identifying potential breakpoints on the chromosome in general (Willenbrock and Fridlyand 2005). Based on the nature and goal of this thesis, we choose to use HHMM and CBS to perform the data processing and to compare the segmentation results by using both simulated and real tumor data sets.

3.3 Segment merging methods

Since most of the segmentation methods work on the individual chromosome level, which means each chromosome is segmented independently, this poses a potential danger that aberration segments located on different chromosomes are incomparable. In some cases, this can cause information losses when the whole chromosome is deleted or amplified, which is the case in many tumor tissues. Even in the same chromosome segmentation profile, numerous segments with many different segment means make the identification of chromosome aberration complicated. The merging algorithm is used to merge those comparable segment values with insignificant differences into a common value. In GLAD algorithm, the adaptive weights smoothing (AWS) algorithm is used to remove the excessive breakpoints (Hupe 2004). Another merging algorithm was proposed by Willenbrock and Fridlyand (2005). Specifically, in this algorithm, if the difference between the $\log_2$(Ratio) values of two segments is not statistically significantly different, then these two segments are assigned to the same predicted value. In detail, the distances calculated based on the predicted (or segmented) values of $\log_2$(Ratio) values from any segments are ordered. From the smallest distance, using Wilcoxon rank sum test based on the observed $\log_2$(Ratio) of all probes within each segment to test whether both segments values are statistically different by using $p < 0.0001$ as threshold or if the distance is less than a give threshold if the probe number within one or both segments contain less than 3. If the predicted values of two segments are not significantly different or the distance is less than the cutoff threshold, these two segments are merged and the above process is repeated until no more segments can be merged together. If only 3 or fewer probes exist on one segment, then only the distance cutoff threshold will apply.

In order to determine the cutoff threshold, the above steps are repeated with increased threshold and compare the current residual (current merged values minus observed $\log_2$(Ratio)) with the original residual (original merged value minus observed log2(Ratio)). The Ansari-Bradley two-sample test for equality of dispersion is used to test the distributions of these residues. The optimal threshold is chosen as the largest threshold such that the Ansari-Bradley 2-sample test has p value > 0.05. The above test is repeated until no more segments can be merged together.

3.4 Centering methods

DNA copy number gains or losses are defined relative to some common baseline agreed to represent the "normal" or "no changes" condition. Centering means calibrating the "normal" DNA copy number to a standard value, such as zero for the $\log_2$(Ratio) values. Different centering methods have been proposed, many of them based on different assumptions. Traditionally, the overall mean or median is used as the center value and all other $\log_2$(Ratio) values are shifted accordingly. But this method fails in highly aberrant genomes such as the tumor samples shown in Figure 3.4.1. Other methods, such as minimum aberration location, the segment intensity with the longest run length probe or using the highest mode of the intensities have also been suggested (Chen, et al. 2008). The first two methods require the pre-segmentation while the last one doesn't. Since the female X chromosome always has one copy number gain aberration in many studies with male genome references, sex chromosomes are always excluded from the calculation when define the centering values. In this analysis, after segmentation, we assume the mode of the predicted value of segments represents the "unchanged" information; so all the data are centralized accordingly.

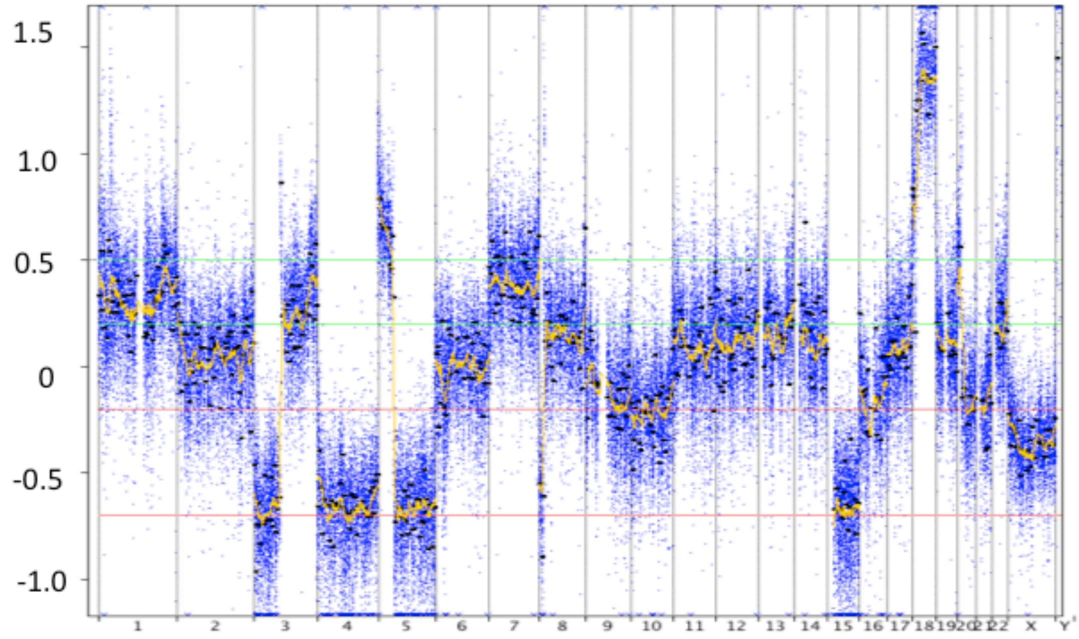


Figure 3.4.1 Genome aberration of Small Cell Lung Cancer samples shows the highly aberrant patterns across the whole genome. The y-axis represents the $\log_2$(Ratio) values of all probes and the black dotted line indicates the predicted segment values.

3.5 Chromosome Gains and Losses assignment

After finding all potential breakpoints, the next question is to decide which segments represent chromosome gains or losses. As discussed in Section 3.1, due to all kinds of systematic and random errors involved in each experiment, the predicted segmentation values are far different from the standard, ideal $\log_2$(Ratio) of copy number variation as shown in Table 3.1.1. After suitable data processing steps, finding the chromosome gains or losses becomes a question of setting the reasonable cutoff thresholds for the predicted segmentation values. In some papers (Veltman, et al. 2003, Nakao, et al. 2004), a series of normal vs. normal hybridizations is performed in addition to the tumor sample hybridization. Ideally, all the $\log_2$(Ratio) values from normal vs. normal comparisons are zero if no systematic or random errors are involved in the experiment. So the $\log_2$(Ratio) values in the normal vs. normal comparisons actually represent the noise information from the similar experiments. From the distribution of the $\log_2$(Ratio) of this controlled experiment, the cutoff thresholds could be decided by using 2 or 3 folds standard deviations of mean of $\log_2$(Ratio) values.

Derivative Log Ratio (DLR) is the difference between the $\log_2$(Ratio) values of consecutive probes along the chromosomes. Derivative Log Ratio spread (DLRs) can be used to estimate the probe-to-probe noise information. Based on the IQR of the DLRs, we can set a suitable cutoff threshold to estimate the noise level of the specific array CGH experiment.

3.6 Literature review of the molecular biology background for Glioblastoma disease

Glioma is a type of tumor that starts mostly in the brain. It is the most common primary brain tumor in adults. Histopathologically, Glioma has been subtyped into four categories, WHO grade I to IV, with grade I indicating a more benign tumor and grade IV being malignant featuring uncontrolled cellular proliferation, and resistance to apoptosis and high invasiveness. Glioblastoma is the most common grade IV brain tumor and is also a deadly disease with average survival time approximately one year. More than two decades of research work by many labs worldwide has revealed many molecular mechanisms about this disease. A detailed review of this research appears in the paper published on Gene & Development (Furnari, et al. 2007). Briefly, some genes or gene products have been shown to play very important roles in tumor development. These genes include, but are not limited to, phosphatase and tensin homolog (PTEN), retinoblastoma protein (RB1), TP53, epidermal growth factor receptor (EGFR). PTEN, Rb1, and TP53 have been classified as tumor suppressor genes, all of them involving cell cycle regulations. For instance, PTEN is involved in the regulation of the cell cycle, preventing cells from growing and dividing too rapidly (Chu and Tarnawski 2004). Rb1 is a protein that helps to prevent excessive cell growth by inhibiting cell cycle progression (Murphree and Benedict 1984). Like PTEN and Rb1, the protein product of TP53, p53, acts as a checkpoint regulator at the G1 to S phase (Furnari, et al. 2007). Loss of these functions through either mutations or deletions may cause the cancer cells to grow uncontrollably. Many molecular pathways have also been identified for the development of Glioblastoma. These pathways include PTEN/PI3K/AKT (Cantley and Neel 1999), TGF-Beta pathway (Xu and Kapoun 2009), and EGFR regulated pathway (Comincini, et al. 2009). Many other potential molecular pathways have also proposed and been detailed in Furnari's review paper (Furnari, et al. 2007).

Chapter 4: Data simulation

Simulated data sets have been widely used in evaluation of the performance of different CGH segmentation algorithms (Hupe 2004, Hsu, et al. 2005, Lai, et al. 2005, Willenbrock, et al. 2005, Huang, et al. 2007). The basic requirement is that the artificially generated data set should closely mimic the real data set, which includes probe tiling information on the chromosomes, breakpoint information, simulated intensity distribution, and even sample contamination information. In this analysis, in order to get a rough idea of segmentation properties of HHMM and CBS algorithms and also for the purpose of further comparison using the real data set, we perform a simulation study using the algorithm embedded in the R "snapCGH" package (Smith 2006) mainly based on the simulation method proposed by Willenbrock and Fridlyand (2005). First thing we need to consider for the simulation is the array numbers. Sample numbers always determine the statistical testing power. In order to reflect the condition in our data set, we should use comparable data set numbers for our simulation. Second, chromosome information is also considered. Human genome contain 22 pairs of euchromosomes, each pair has different length. Each chromosome also contain one short arm and one long arm separated by the centromere, this information has also been considered as suggested in the aCGH R package (Fridlyand 2008). Probe tiling is another issue we should consider. Even though sometimes it is difficult to design probes at specific regions due to some sequence structure reasons, we still try to design array probes evenly across all chromosomes in order to get a complete coverage for whole genome. In this simulation, we set the probabilities of assigning probes on both short and long arms of each chromosome to 50%. In addition to this, the simulation function of snapCGH also has the option to specify the minimum and maximum probe distances and the probabilities of each segment of chromosomes being tiled. In order to mimic the noise level of real data set, Gaussian noise with mean zero and varying variance were added based on the suggestion of Willenbrock and Fridlyand (2004). The standard deviation of each state is random sampled from uniform distribution of 0.1 to 0.2 as been observed from real data set (Fridlyand et al., 2004).

The ratio profiles for the simulated data set are generated based on the 145 real tumor samples in order to emulate the complexity of real tumor profile (Fridlyand et al., 2004). Briefly, the mean values of segments were binned into the interval between -0.2 and 0.2 and length for normal level were sampling from the level [-0.2, 0.2] bin. The altered segments values are sampled from other bins contain other segmentation values. This simulated dataset is considered as the known dataset with known breakpoints. The HHMM and CBS algorithms were applied to the simulated dataset estimated and known breakpoints were compared and the false discovery rate was returned.

For this study, all the simulated data sets are generated under the “default” condition (except otherwise specified) and then followed by HHMM and CBS segmentation processes. The default is the conditions that try to mimic the real data condition while still considering the limitation of simulation situation. For instance, the probability of probe tiles on both arms of each chromosome is 50% even though in reality the probes designed for short arms and long arms might be different. The default standard deviation of the simulated data in each of the states is randomly sampled uniformly between 0.1 and 0.2. The lower limit for the distance between adjacent clones in non-tiled region is 0.9Mb and upper limit is 1.1Mb. Sex chromosomes are always excluded from simulation due to their special situation.

For the remaining 22 pairs of chromosomes, about 8000 probes are generated in each simulated sample file. Since we have more than 170 hybridizations in this study, we first generated the simulated data set with 170 samples. Figure 4.1 shows the density plots of the generated $\log_2$(Ratio) from all 170 simulated samples. The left panel represents the "observed" $\log_2$(Ratio) and the right panel represents the "true" $\log_2$(Ratio). After segmentation with HHMM and CBS algorithms, we can compare the observed breakpoints and the true breakpoints and further derive the false discovery rate. First of all, we tested whether the sample size could affect the false discovery rate. Simultaneously, we also test whether the data quality has big effects to the false discovery rate (Figure 4.2). The R code is given in Appendix B.

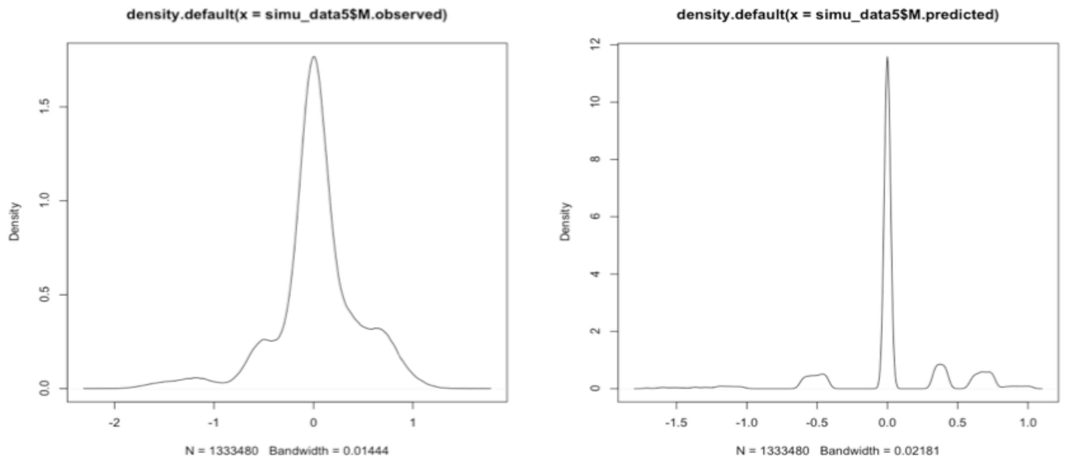


Figure 4.1 Density plot of the simulated $\log_2$(Ratio) from all 170 simulated sample set. The left panel represents the simulated "observed" $\log_2$(Ratio) and the right panel represents the "true" copy number variation.

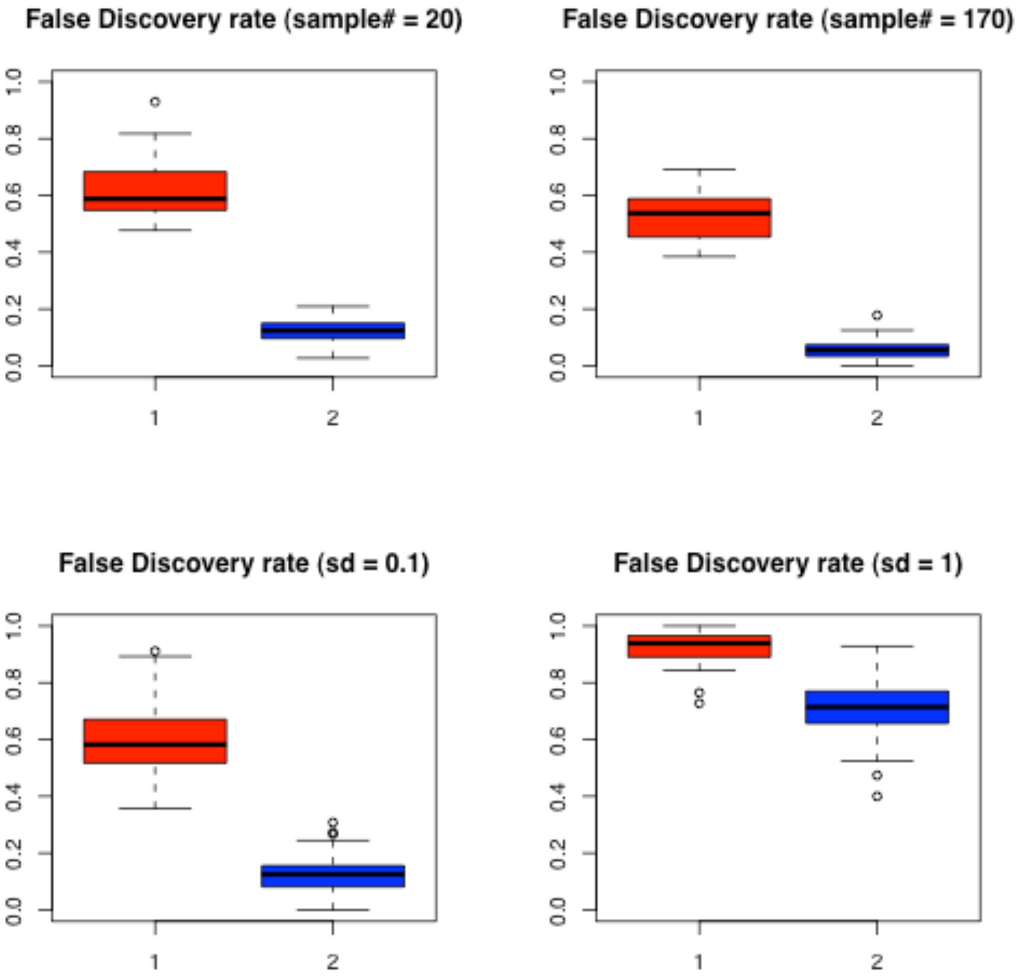


Figure 4.2 Comparison of HHMM (read color) and CBS (blue color) segmentation methods using simulated data set. The top two figures show the false discovery rate of similarly simulated data set but with different sample size (sample size number 20 vs 170). The two panels below indicate the false discovery rate of the two segmentation methods from the simulated data sets with the same sample sizes (sample numbers are 20) and different data quality (standard deviations 0.1 and 1 respectively).

First, within each panel of Figure 4.2, we can clearly see that the CBS segmentation method (blue color) generates more reliable results with relatively low false discovery rates. For the CBS segmentation method, the FDR for the CBS segmentation method is around 0.1, while for HHMM (red color) segmentation method is about 0.6. The top panel also indicates that for reasonable sample sizes (20 and 170 here), the FDRs for the same segmentation method are quite similar. For data sets with low quality, the ability to correctly find the breakpoints for both CBS and HHMM segmentation methods dropped significantly, reflected in the higher False Discovery Rate in the right figure of the lower panel. In summary, the results above clearly indicate that data segmented using the CBS method gives low false discovery rate in general even though it was suggested that HHMM surpass CBS in detecting the short segments (Willenbrock and Fridlyand, 2005). Of course, increasing the data variations could significantly increase the false discovery rate for both methods. This is also easy to understand, since the increasing of the variability of the data sets will increase the chance to commit the type I error.

Chapter 5: Data processing and analyzing

For the array-based experiment, there are several significant features compared to the traditional karyotyping experiments. The first feature is large data volume. Specifically in this study there are 236,381 probes on each chip that will generate the same volume of ratio values. Considering all the tumor data from 170 patients, the total data volume is tremendous. In order to derive reliable analytical results from this massive data set, we set up the following steps to process the data as shown in Figure 5.1.

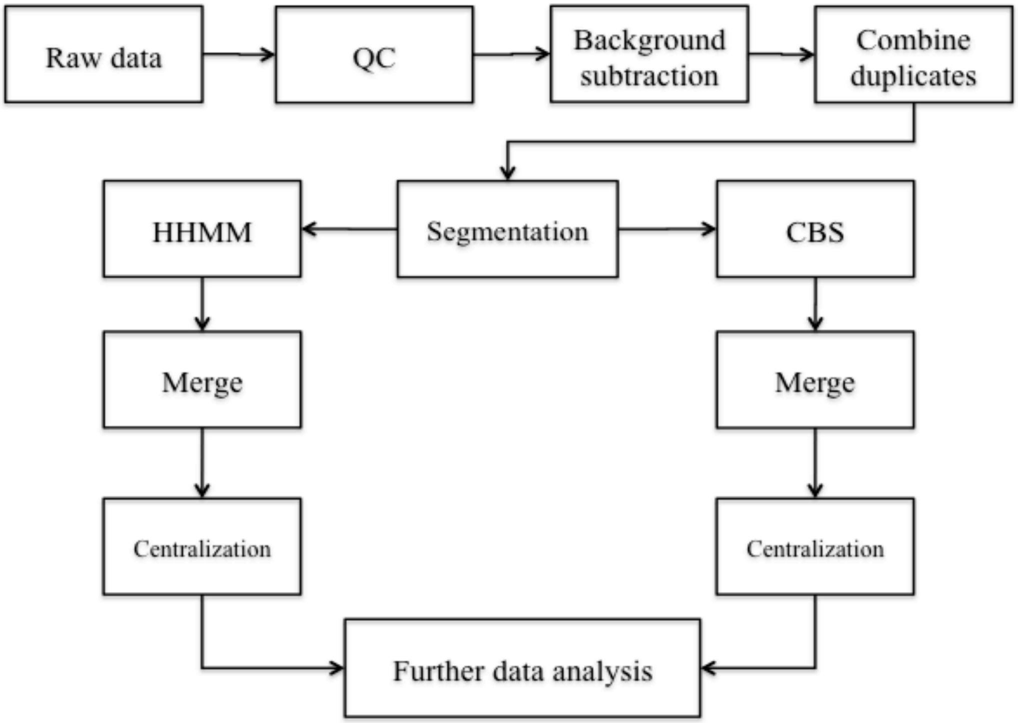


Figure 5.1 Array CGH data processing procedures. After segmentation, the data are processed using HHMM and CBS separately. QC represents "Quality Control" analysis.

5.1 Raw data

After hybridization, the chips are scanned and the intensities and backgrounds from both channels for all probes were recorded. This information will be used for further data processing.

5.2 Quality control (QC) analysis

As in many other experiments, data quality is vital to the reliability of the results generated. In microarrays, data quality of array CGH can be affected by many factors starting from the sample collection to the final chip scanning. Some of these factors include, but are not limited to, tissue quality, genomic DNA quality, and hybridization techniques as well as others. Good data quality will maximize the information we can get from each study. For the array CGH, one of the very important quality parameters to access the final data quality is Derivative Log Ratio spread (DLRs). Derivative Log Ratio (DLR) calculates the $\log_2$(Ratio) differences between consecutive probes on the chromosomes. Based on the inter-quartile range (IQR) of all DLR values on each chip data set, we can derive the DLRs with comparable robust standard deviation. DLRs value estimates probe-to-probe noise from the array itself calculated from the $\log_2$(Ratio). High DLRs values could result in low discrimination of different copy number variation. Figure 5.2.1 shows QC information of two hybridizations with different DLRs values. Clearly low DLRs values are required in order to get reliable result.

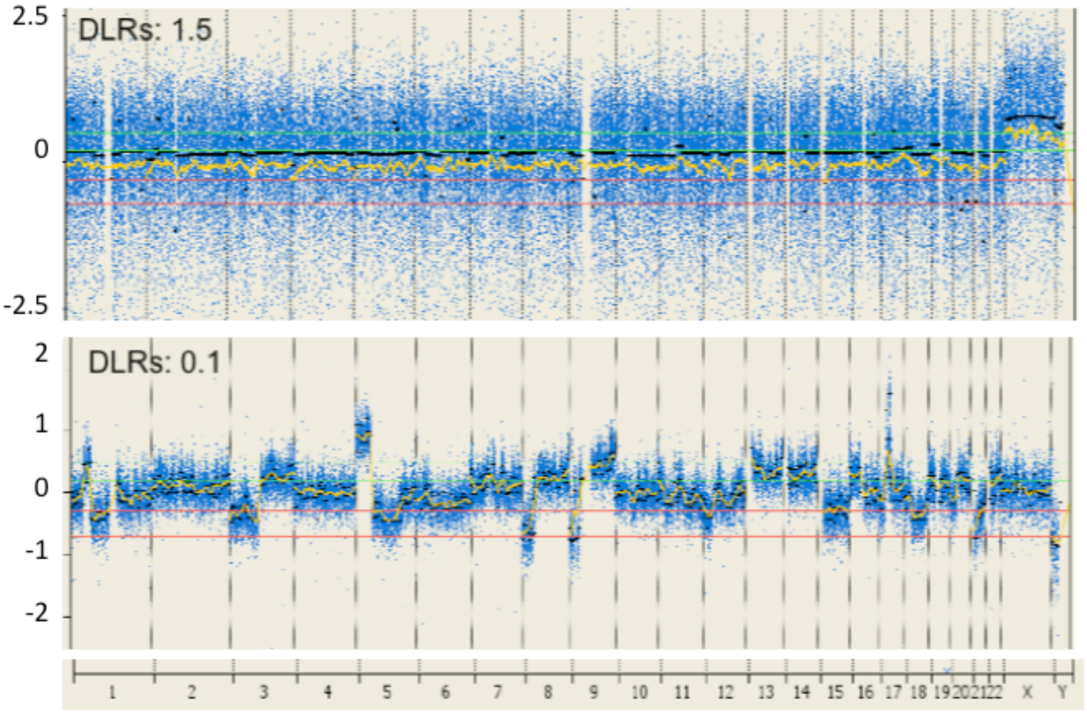


Figure 5.2.1 The relations between DLRs and the noise level in the array CGH experiments. The upper figure shows the probe expression levels for one array CGH experiment with very high DLRs value. This figure is generated based on real data set from different study. Clearly no chromosome aberration patterns can be observed if they exist. On the contrary, for the array CGH data with low DLRs illustrated in the lower figure, we can see clear chromosome gains and losses patterns across the whole genome. The yellow line represents the "center" or baseline. The black line represents the segments values for individual segments.

Table 5.2.1 list all the DLRs values for 170 hybridizations used in this part of the analysis. From this table, we can see that the DLRs values for most of the hybridizations are very low. Based on many experimental results and also the manufacturer's recommendation (Table 5.2.2), the highest DLRs (0.33) is also an acceptable value.

Patients	DLRs	Patients	DLRs	Patients	DLRs	Patients	DLRs
TCGA-08-0386	0.14	TCGA-02-0102	0.19	TCGA-02-0338	0.21	TCGA-02-0007	0.24
TCGA-08-0358	0.14	TCGA-06-0178	0.19	TCGA-02-0333	0.21	TCGA-06-0139	0.24
TCGA-06-0146	0.15	TCGA-02-0258	0.19	TCGA-02-0106	0.21	TCGA-08-0373	0.24
TCGA-06-0175	0.16	TCGA-08-0520	0.19	TCGA-12-0616	0.21	TCGA-02-0079	0.24
TCGA-08-0518	0.16	TCGA-02-0057	0.19	TCGA-08-0510	0.21	TCGA-06-0132	0.24
TCGA-06-0394	0.16	TCGA-02-0446	0.19	TCGA-08-0521	0.21	TCGA-02-0011	0.24
TCGA-06-0194	0.16	TCGA-08-0514	0.19	TCGA-02-0028	0.21	TCGA-02-0456	0.24
TCGA-08-0509	0.16	TCGA-02-0080	0.19	TCGA-12-0620	0.21	TCGA-06-0122	0.24
TCGA-08-0512	0.16	TCGA-02-0071	0.19	TCGA-08-0531	0.21	TCGA-06-0176	0.24
TCGA-08-0352	0.17	TCGA-02-0430	0.19	TCGA-02-0451	0.21	TCGA-02-0037	0.24
TCGA-06-0648	0.17	TCGA-06-0185	0.19	TCGA-06-0158	0.21	TCGA-02-0281	0.24
TCGA-02-0330	0.17	TCGA-06-0646	0.19	TCGA-06-0209	0.21	TCGA-02-0083	0.24
TCGA-02-0439	0.17	TCGA-12-0618	0.19	TCGA-02-0034	0.21	TCGA-02-0006	0.25
TCGA-08-0349	0.17	TCGA-02-0113	0.19	TCGA-02-0432	0.21	TCGA-06-0148	0.25
TCGA-06-0164	0.17	TCGA-02-0324	0.19	TCGA-06-0214	0.21	TCGA-02-0260	0.25
TCGA-02-0422	0.18	TCGA-06-0162	0.19	TCGA-06-0208	0.21	TCGA-02-0058	0.25
TCGA-06-0414	0.18	TCGA-02-0326	0.20	TCGA-02-0060	0.21	TCGA-06-0145	0.25
TCGA-06-0213	0.18	TCGA-02-0332	0.20	TCGA-06-0128	0.21	TCGA-02-0024	0.25
TCGA-06-0168	0.18	TCGA-02-0055	0.20	TCGA-02-0111	0.21	TCGA-02-0089	0.25
TCGA-06-0195	0.18	TCGA-02-0074	0.20	TCGA-06-0237	0.21	TCGA-06-0211	0.25
TCGA-06-0238	0.18	TCGA-02-0047	0.20	TCGA-02-0337	0.22	TCGA-06-0189	0.25
TCGA-06-0402	0.18	TCGA-02-0069	0.20	TCGA-06-0171	0.22	TCGA-06-0169	0.25
TCGA-02-0269	0.18	TCGA-06-0179	0.20	TCGA-06-0413	0.22	TCGA-06-0133	0.25
TCGA-02-0075	0.18	TCGA-08-0517	0.20	TCGA-06-0182	0.22	TCGA-02-0116	0.26
TCGA-06-0149	0.18	TCGA-08-0524	0.20	TCGA-06-0166	0.22	TCGA-02-0317	0.26
TCGA-02-0321	0.18	TCGA-06-0127	0.20	TCGA-06-0173	0.22	TCGA-02-0010	0.26
TCGA-02-0107	0.18	TCGA-02-0084	0.20	TCGA-08-0345	0.22	TCGA-02-0021	0.27
TCGA-06-0409	0.18	TCGA-02-0052	0.20	TCGA-02-0001	0.22	TCGA-02-0003	0.27
TCGA-06-0197	0.18	TCGA-06-0152	0.20	TCGA-02-0009	0.22	TCGA-06-0174	0.27
TCGA-06-0157	0.18	TCGA-06-0210	0.20	TCGA-06-0156	0.22	TCGA-06-0126	0.27
TCGA-08-0511	0.18	TCGA-02-0266	0.20	TCGA-08-0529	0.22	TCGA-06-0125	0.27
TCGA-08-0516	0.18	TCGA-02-0038	0.20	TCGA-06-0190	0.22	TCGA-06-0147	0.28
TCGA-02-0114	0.18	TCGA-02-0271	0.20	TCGA-02-0086	0.23	TCGA-06-0137	0.28
TCGA-06-0397	0.18	TCGA-02-0033	0.20	TCGA-02-0339	0.23	TCGA-06-0141	0.28
TCGA-02-0325	0.18	TCGA-06-0412	0.20	TCGA-02-0046	0.23	TCGA-06-0221	0.28
TCGA-02-0054	0.18	TCGA-06-0184	0.20	TCGA-06-0177	0.23	TCGA-06-0154	0.29
TCGA-02-0085	0.18	TCGA-02-0440	0.20	TCGA-06-0129	0.23	TCGA-02-0014	0.29
TCGA-06-0241	0.19	TCGA-02-0289	0.20	TCGA-06-0124	0.23	TCGA-06-0138	0.29
TCGA-06-0645	0.19	TCGA-02-0064	0.20	TCGA-02-0290	0.23	TCGA-02-0027	0.29
TCGA-12-0619	0.19	TCGA-02-0087	0.20	TCGA-06-0187	0.23	TCGA-06-0188	0.31
TCGA-02-0099	0.19	TCGA-06-0644	0.21	TCGA-02-0043	0.23	TCGA-02-0285	0.33
TCGA-02-0115	0.19	TCGA-06-0143	0.21	TCGA-08-0522	0.23		
TCGA-08-0525	0.19	TCGA-06-0410	0.21	TCGA-06-0130	0.24		

Table 5.2.1 DLRs values for all hybridizations from 170 patients tumor samples

Other data quality metrics include signal to noise ratio, signal intensities and background noise information for both red and green channels. Since there are also 1000 replicated biological features printed on each chip, for each channel, reproducibility is assessed by calculating the median percent coefficient of variation of background-subtracted signal for replicate non-control probes. Figure 5.2.2 summarizes the data distributions of all these data quality metrics including the DLRs presented as box plots. Exact values of these quantities are given in supplementary Table S5.2.1. Based on the manufacturer's guidelines (Table 5.2.2), all experiments have reasonably good quality and therefore all will be used for further data analysis.

Metric	Excellent	Good	Poor
BackgroundNoise	<5	5-10	>10
Signal Intensity	>150	50-150	<50
Signal to Noise	>100	30-100	<30
Reproducibility	<0.05	0.05-0.2	>0.2
DLRspread	<0.2	0.2-0.3	>0.3

Table 5.2.2 Manufacturer's recommended QC guidelines for the array CGH platform.

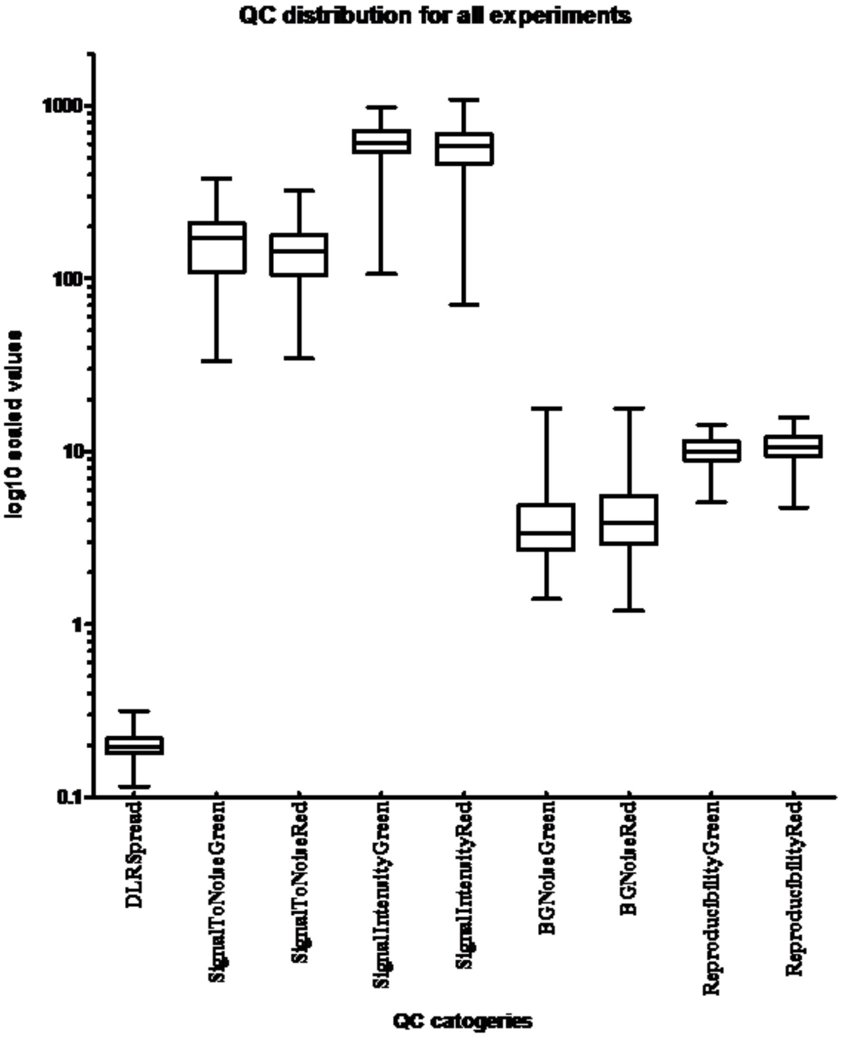


Figure 5.2.2 data quality metrics and their value information from all hybridizations used in this study. The y-axis values are $\log_{10}$ scaled. Since there are two color channels (Green and red for two samples, tumor and reference, separately), the qualities of each channel are accessed independently except for the DLRs. The reproducibility values shown here are the percentage values.

5.3 Background subtraction

Traditionally, background intensities are subtracted from the foreground intensities and generate the final intensities for each specific probe. The problem with this method is that the foreground signal intensities might be smaller than the corresponding background intensities in many cases. So the subtraction will result in “negative” intensities for some probes, which could cause further computational problems when performing log transformation. In this analysis, we use “minimum” background subtraction in which the background intensity is subtracted from foreground intensity of each probe unless the resulting intensity is zero or negative value. In this case, half the value of foreground intensity will be used. Using this “minimum” background subtraction can avoid generating either negative or zero value intensities and facilitate the computerization of the data processing steps.

5.4 Combination of intensities

As with many other Microarray platforms, duplicates or replicates of many probes are printed on the chip in order to access the data quality or the variability of each array experiment. The intensities from these replicates are normally merged into one value, mostly by mean or median, based on the unique ID. On the Agilent array platform – Human Genome CGH 244k chips, more than a thousand probes spotted on the chips are replicated. These replicates are averaged into the single final value for each probe.

5.5 Segmentation and Merging

As mentioned above, all the data are processed using both HHMM and CBS segmentation methods. Since HHMM and CBS are based on different statistical models, it is reasonably to expect that both the segmentation results should not be exactly same. Subtle differences in the mean $\log_2$(Ratio) values of two segments are not necessarily biologically relevant, but they do make the downstream analysis more complex. In this case, merging of these segments values with insignificant statistical differences becomes necessary especially when segments are located on different chromosomes. In this analysis, this process is performed by applying Willenbrock and Fridlyand's (2005) merging method. Briefly, the predicted levels of segments were merged into the same level if the differences between them were not significant according to the Wilcoxon rank test or less than the specific threshold generated by using the Ansari-Bradley 2-sample test (for details see Section 3.3).

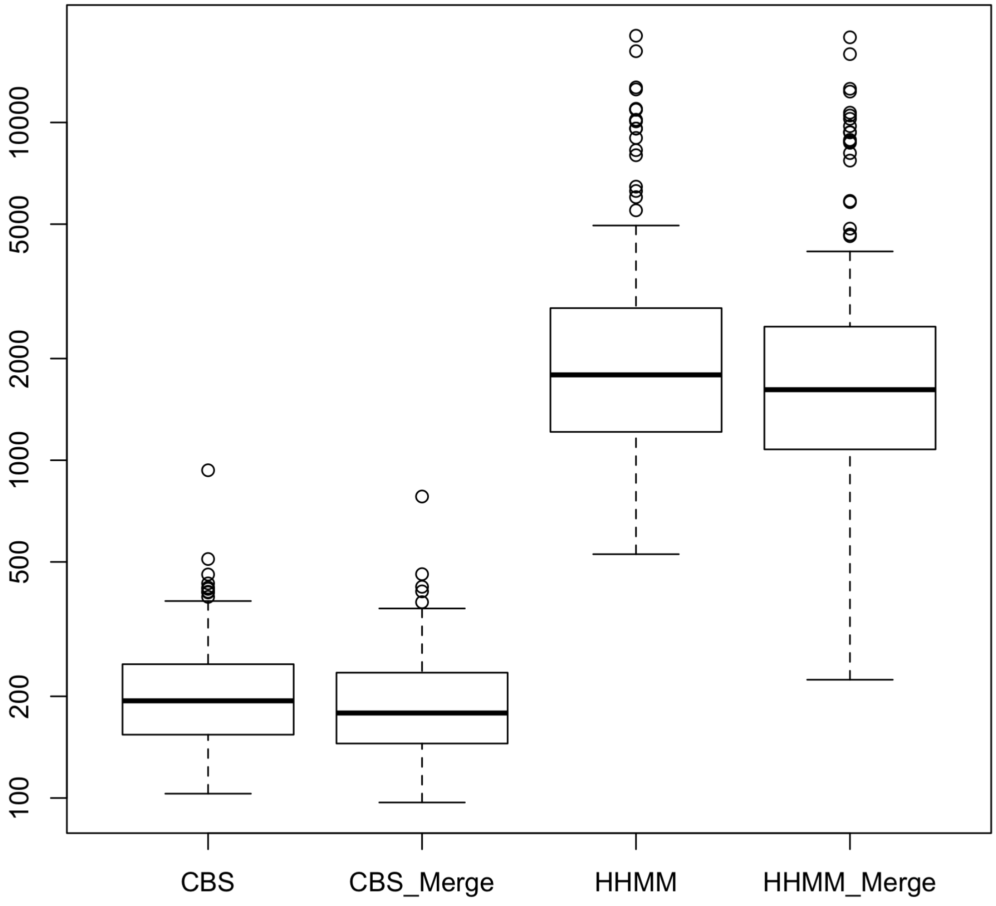
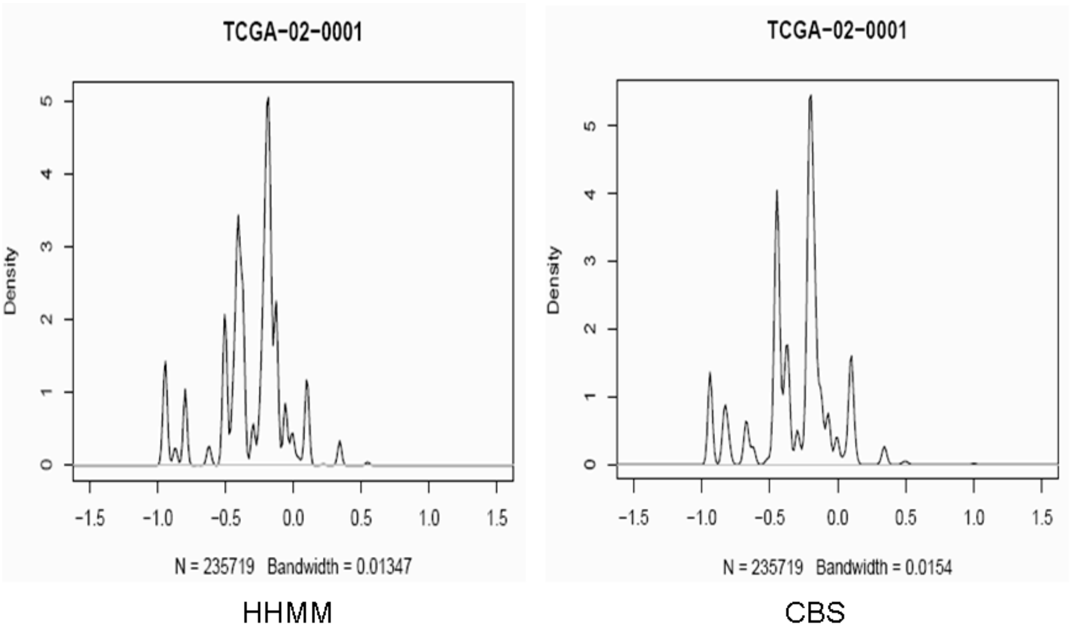


Figure 5.5.1 Boxplot of segmentation numbers result from HHMM and CBS methods before and after merging steps.

Figure 5.5.1 shows the segmentation and merging results from both the HHMM and CBS analyzing methods. The original segmentation numbers are listed in Supplementary Table S5.5.1. First, within each segmentation method, the merging process did reduce the segments numbers. Second, the two segmentation methods generate dramatically different segment numbers. For the CBS segmentation method, the mean segment numbers from 170 patients' samples after merging is 198, while as for HHMM segmentation method, this number increased to 2495. Density plots of the merged segmentation results also show subtle differences between the HHMM and CBS segmentation methods. Figure 5.5.2 is from two pairs of typical density plots from the same patient (patient TCGA-02-0001). From these two figures, we can see that the general patterns of the density plots generated from HHMM and CBS segmentation methods are quite similar, even though we can still see subtle differences between them.



HHMM

CBS

Figure 5.5.2 Density plot of segmented values after performing merging steps for sample TCGA-02-0001. The x-axis represents the $\log_2$(Ratio) values of all probes after performing the merging. From these two figures, we can see that the general patterns of the density plots generated from HHMM and CBS segmentation methods are quite similar, even though we can still see subtle differences between them.

5.6 Centering

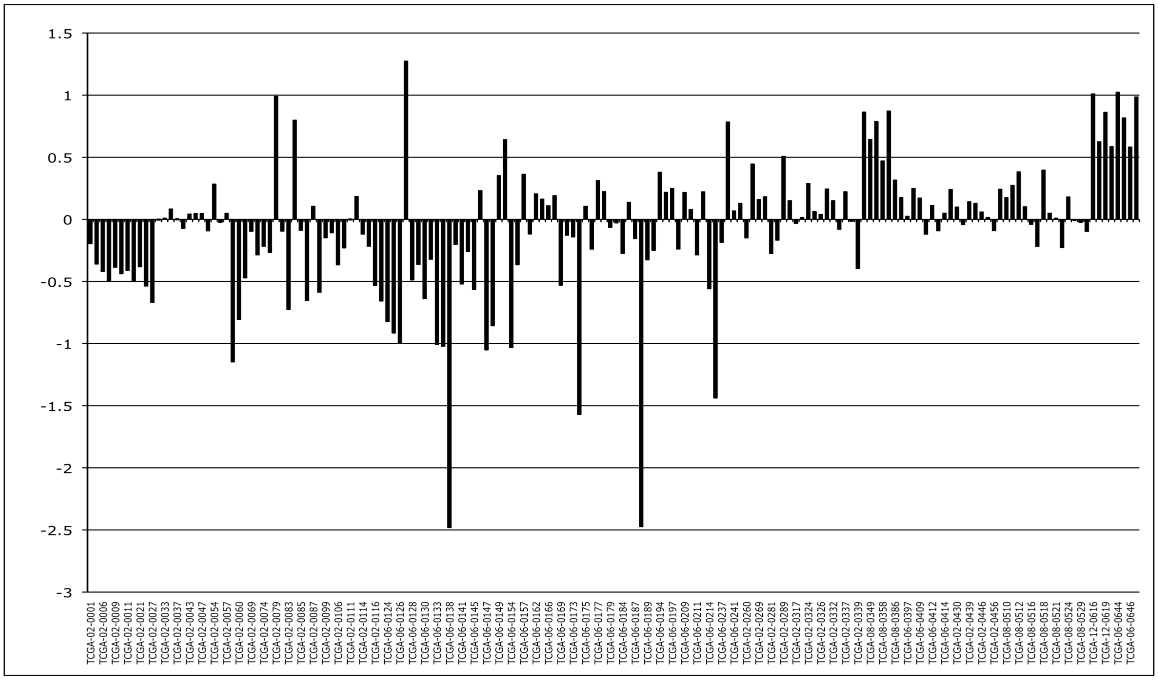


Figure 5.6.1 Log$_2$(Ratio) values of the modes of the density plots from all 170 patients samples (CBS segmentation methods). For most of the hybridizations, the mode of the density plot is shifted quite far from value and need to be adjusted accordingly.

As discussed previously, for the array CGH study, all the copy number changes (gains or losses) are relative to the unchanged copy number. If the segment shows no aberration, we expect to see the ratio of tumor sample and control sample equal to one (or log$_2$(Ratio) = 0). In this study, we assume the mode of densities for each patient sample represents the center of the log$_2$(Ratio) values and all the log$_2$(Ratio) values in each array CGH are shifted accordingly. The density plots in Figure 5.5.2 clearly show the shift of the mode of segments for one typical patient sample. Figure 5.6.1 shows the log$_2$(Ratio) of the modes from all 170 patients by using CBS segmentation method. Similar results are generated from HHMM segmentation and the plot is posted as supplementary figure (Figure S5.6.1). Since most of the modes from the density plots of all 170 hybridizations are quite far away from zero, all the log$_2$(Ratio) data should be adjusted. So, centering is necessarily in order to make all the experiments comparable. Figure 5.6.2 shows the density plots from the same patients as in Figure 5.5.2 after performing the centering (based on the CBS segmentation result).

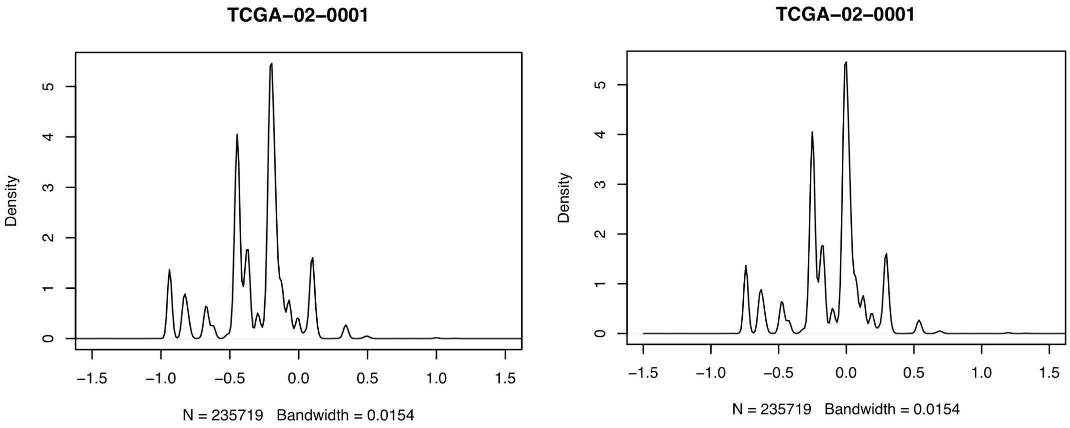


Figure 5.6.2 comparison of density plots before and after centering for one specific patient sample TCGA-02-0001. The left figure shows the density plot before the centering, and the right figure shows the density plot after the shift.

Chapter 6: Experimental results

6.1 Breakpoint identification of HHMM and CBS methods

Regardless of which segmentation algorithms we are using, the most important thing is that we can reliably identify those segments with copy number variation or identify the break points. Based on the simulated data set, we know CBS can identify breakpoints fairly reliably with low false discovery rate compared to the HHMM algorithm. For the real data set in this study, we also want to know whether CBS or HHMM segmentation methods can find the breakpoints reliably. By checking the chromosome aggregation plots (not included due to huge file sizes (10Gb)), we observed that the HHMM segmentation method failed to pick up many obvious change points as shown in one example in Figure 6.1.1, while CBS could. On the other hand, the HHMM segmentation method identified many single probe aberration change points, while the CBS algorithm failed to do so.

Detailed analysis of the length of all the segments reveals that HHMM identified many single probe aberrations while the CBS method failed to do so. The histogram of the lengths of all segments clearly shows this pattern (Figure 6.1.2). In this figure, the segments lengths, which are plotted along the x-axis, are log2 transformed. For the HHMM segmentation, in addition to a sharp peak at a short length position, there is another main peak around 15 to 20. But for the CBS segmentation, the only peak is around 20 to 25.

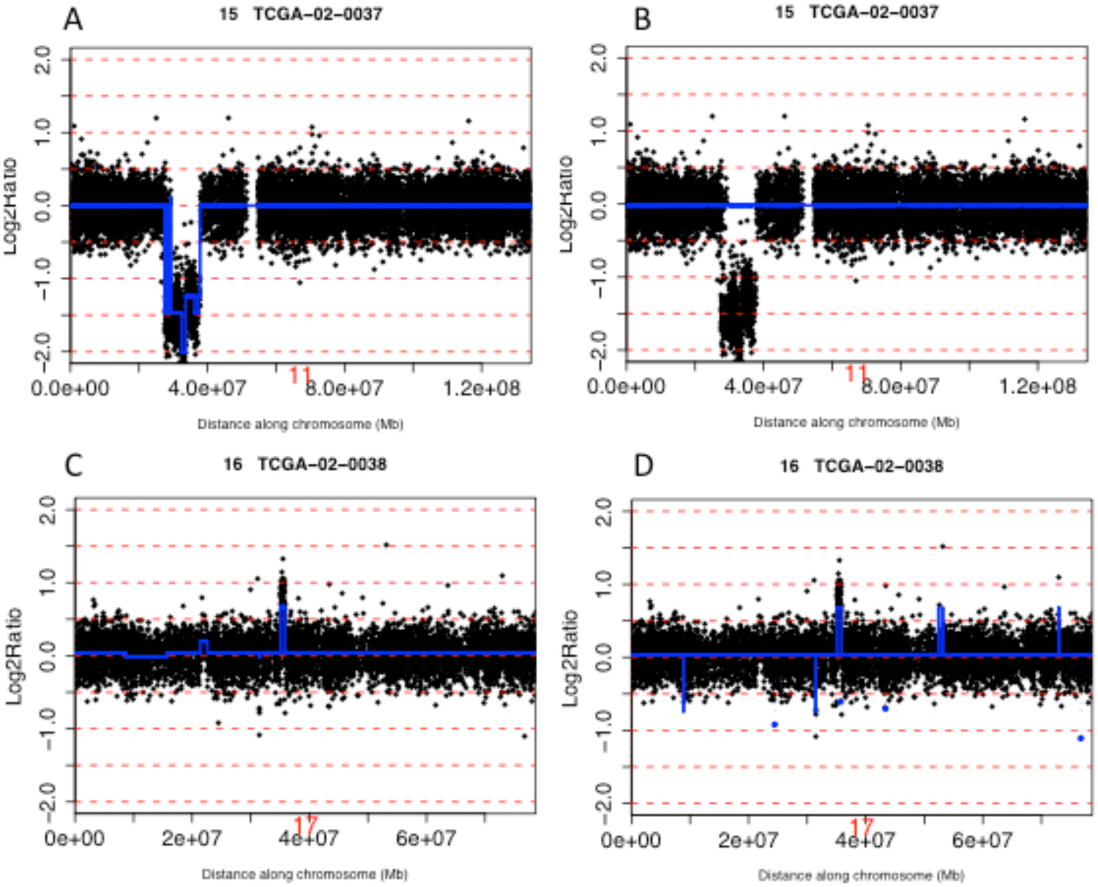


Figure 6.1.1 Breakpoint identification by using CBS and HHMM algorithms. Only two typical patient samples are displayed here, with only chromosome 11 for sample TCGA-02-0037 and chromosome 17 for sample TCGA-02-0038. Panels (A) and (C) are generated based on the CBS algorithm, and (B) and (D) are based on HHMM segmentation method. For the CBS algorithm, significant copy number loss at position around 40Mb is clearly identified in (A), while for the HHMM method, this segment is missing in (B). On the contrary, the HHMM algorithm identify many single probe aberration in (D), while CBS algorithm failed to do so as shown in (C).

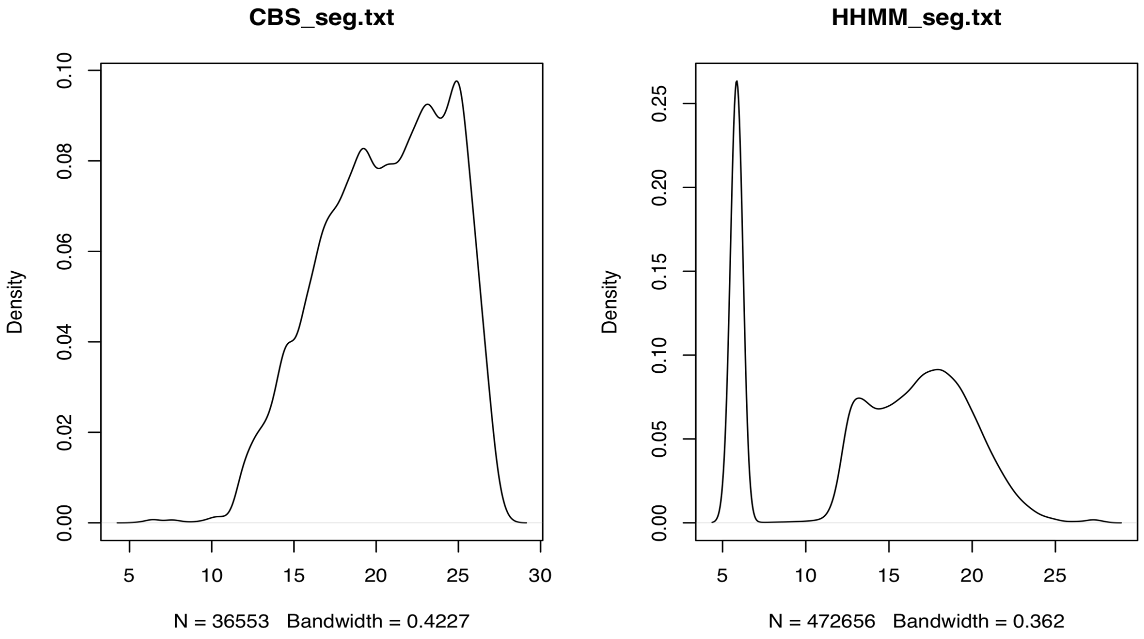


Figure 6.1.2 Density plots of all segment lengths (without merging) from both CBS and HHMM segmentation methods. The segments lengths are $\log_2$ transformed. The mode of the segments lengths densities of CBS segmentation methods are around 20 to 25, while as for the HHMM segmentation method, this mode is around 15 to 20.

Considering the absolute segmentation numbers from all 170 samples, for the CBS segmentation method, only 4 single probe aberrations are identified from the total 36,553 segments of all 170 samples. On average, only 215 segments are identified in each sample. For the HHMM segmentation method, the total number of segments from all samples was 472,656, or an average of 2,780 segments per sample. Of these 472,656 segments, 117,653 segments are single probe aberrations, or an average of 692 single probe aberrations identified per hybridization.

Since there are so many single probe aberrations identified by using the HHMM segmentation algorithm, the next question we would like to ask is how the $\log_2$(Ratio) distribution looks for these single probe aberration segments. In order to make all the data comparable, the predicted values from centralized (after merging) segments are used. Figure 6.1.3 shows the density plots of all the predicted $\log_2$(Ratio) values of all single probe aberration segments. Most of these 100k single probe aberration segments have relatively small chromosome variations. Nevertheless there are still 1000 single probe segments (not including segments on the sex chromosomes) having absolute $\log_2$(Ratio) values greater than 1 (For details see supplementary Table S6.1.1).

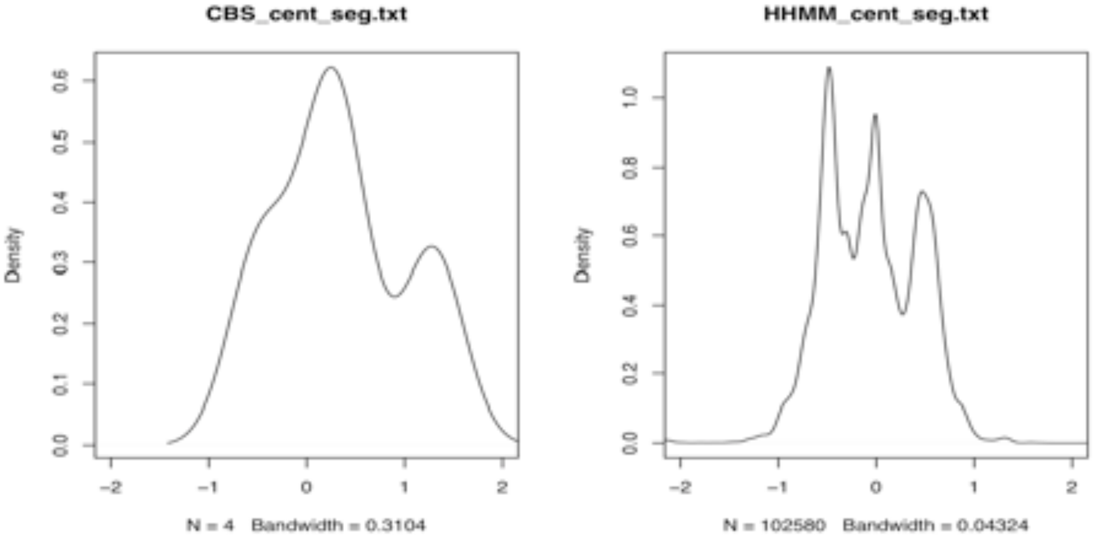


Figure 6.1.3 Density of $\log_2$(Ratio) of all single probe aberration segments. All the predicted $\log_2$(Ratio) values are from centralized segments. For the CBS algorithm, only four segments are single probe aberrations. But for the HHMM algorithm, 102,580 single probe segments are identified. Most of the segments have relatively small predicted $\log_2$(Ratio) values.

If we think DLRs reflects the noise information for each experiment, then the segments with $\log_2$(Ratio) outside the noise range may reflect the copy number variation. In this analysis, 1.5 times DLRs (1.5xDLRs) was used as cutoff threshold to identify the segments with copy number variations. The histogram of the segment lengths ($\log_2$ transformed) in Figure 6.1.4 shows that there are more “big” segments when using CBS while HHMM tends to identify short length segments. Again, in the segment profiles with 1.5xDLRs cutoff, HHMM identified a total of 47114 single probe segments out of 106,868 segments. On the contrary, CBS identified only four. For the segment length distribution, we can also see that CBS tends to identify big segments while HHMM fails to do so.

Even though we can’t say that all the single probe aberration identified by HHMM are biologically relevant, failure to identify these segments when using CBS segmentation methods clearly leads to the incomplete data analysis results. Similar conclusion could be drawn to the HHMM segmentation method that fails to identify many breakpoints of large sized segments

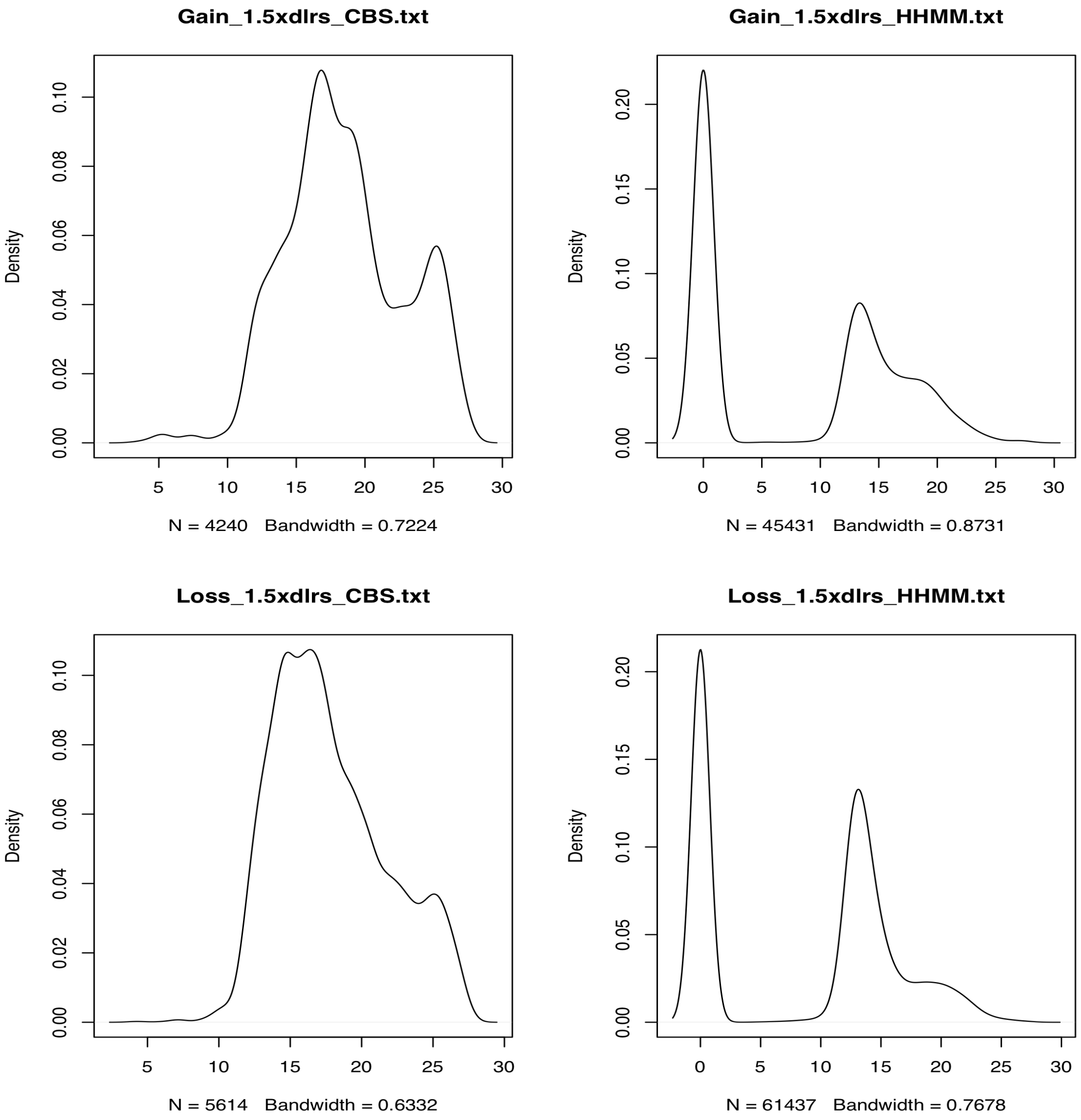


Figure 6.1.4 Density plots of all segments lengths ($\log_2$ transformed) from segments with copy number variation. The cutoff thresholds for the predicted segmentation values are 1.5xDLRs. Many segments with big length are identified in the CBS algorithm while HHMM segmentation method tends to find the small segments.

6.2 Segment analysis

Even though in many cases, copy number variation can lead to the development of certain cancers, due to the individual variation of each patient and also our poor understanding basically of all cancers, we should not expect all patients to have the same chromosome aberration patterns. In order to identify the underlying biological causes of Glioblastoma, we focus the analysis on the following two areas. The first area is to use frequency method to identify chromosome regions that show frequent copy number variations in many patients The second area is to focus on each individual patient and find those regions that clearly show high copy number gain or loss and to elucidate the relations of those genes within these segments to their biological phenotype (discussed in section 6.4).

6.2.1 Frequency analysis of copy number variation

Frequency analysis is attempts to identify those chromosome regions that show frequent aberrations (amplifications or deletions) in most of the disease patient samples. Regardless of which segmentation method is used, the frequency plots in Figure 6.2.1 clearly indicates that more than 60% of the patients have chromosome amplification of chromosome 7 and massive copy number losses of chromosomes 10 and 9p. About fourth of the patients have copy number losses on chromosome 13, 14 and 22 and/or copy number gain on chromosome 19, 20. Similar data processing and analyzing of 24 normal samples show no chromosome aberration at these regions (Figure 6.2.2). From the aberration plots, we can also see that for the HHMM segmentation algorithm, the data look more noise than the aberration plot generated based on CBS segmentation algorithm. This “noisy” pattern of HHMM frequency plot is probably due to the existence of many single probes.

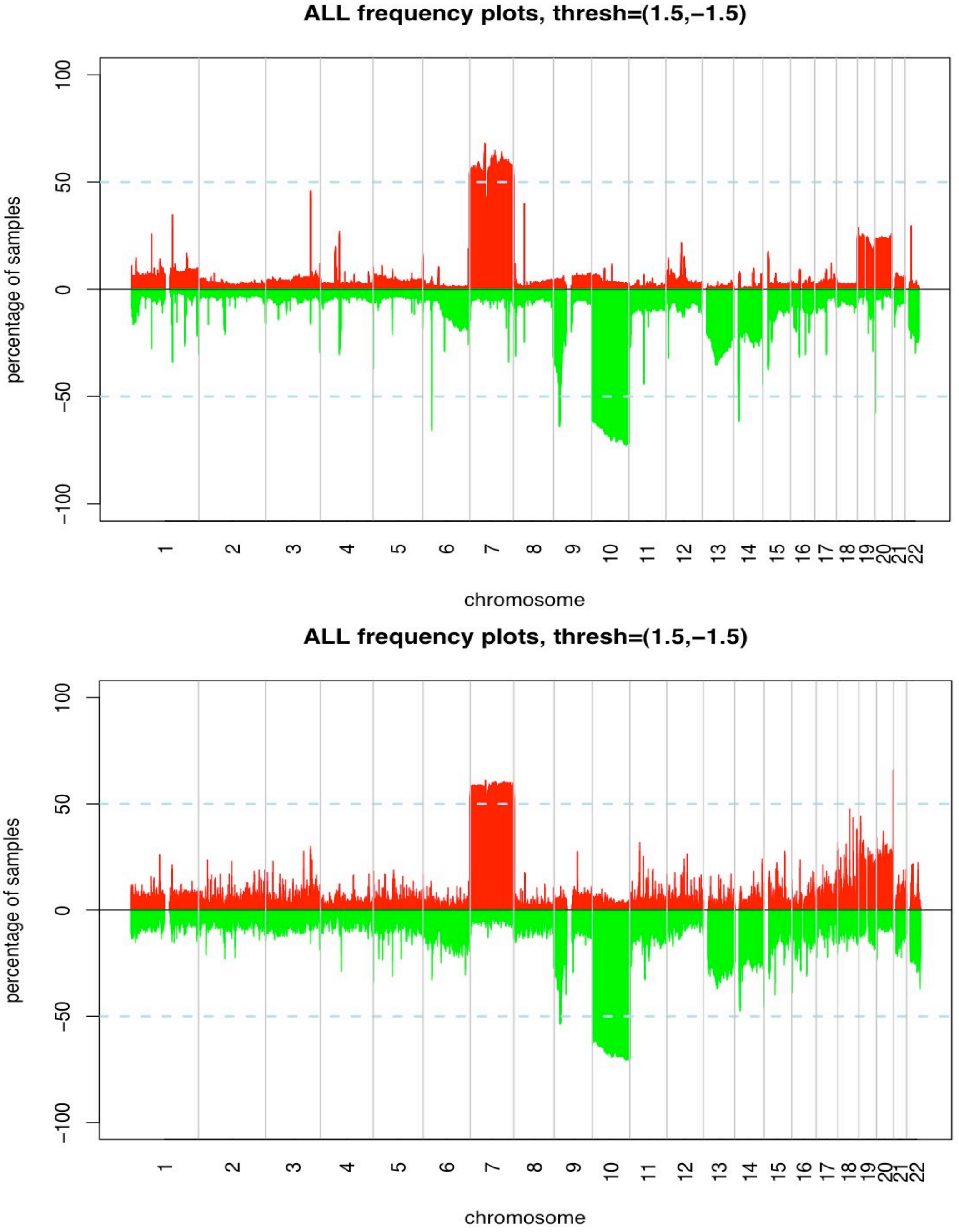


Figure 6.2.1 Frequency plots of all segments from 170 patients samples. Top figure generated based on CBS segmentation algorithm, and bottom figure based on HHMM segmentation methods.

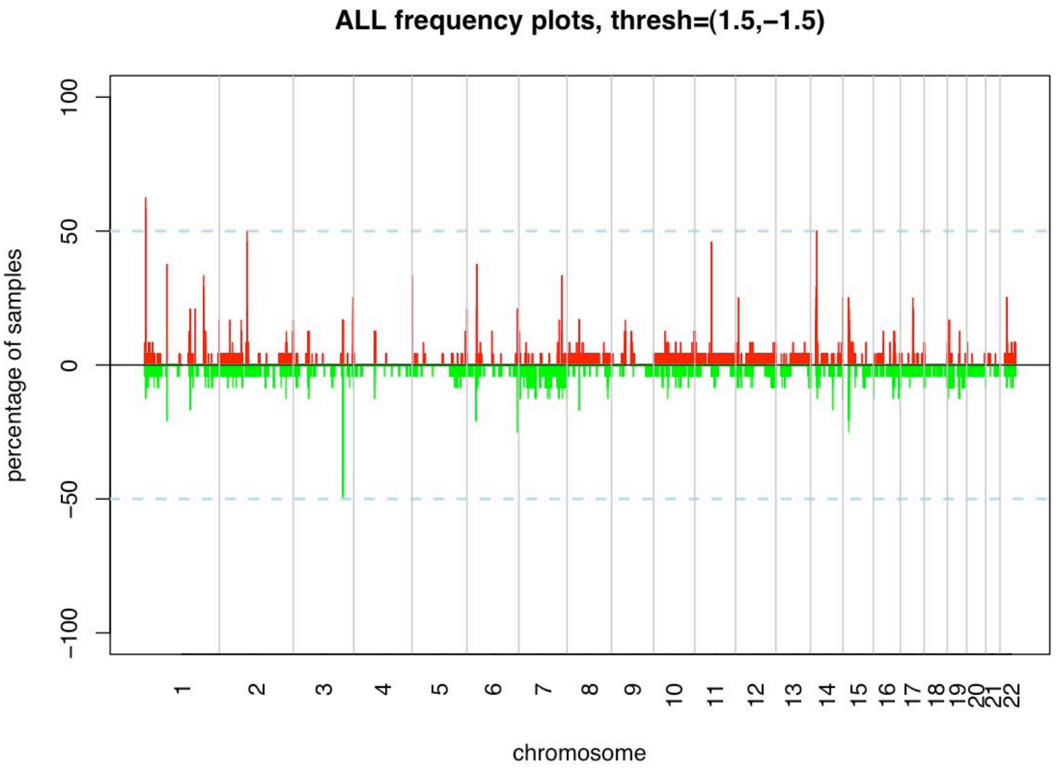


Figure 6.2.2 Frequency plot of all segments from 24 normal tissue samples. The segmentation is based on CBS algorithm. For the HHMM segmentation algorithm, a similar result was obtained

Table 6.2.1 shows the aberration regions with frequencies 25% or higher segmented by using the CBS algorithm method. For the HHMM segmentation algorithm, quite similar results are derived (see supplementary Table S6.2.1). We could notice from both table that some genes gain copy numbers in some patients (ADAM5P and ADAM3A at p11.23 of chromosome 8 in Table 6.2.1) while they lose copy numbers in other patients. The polymorphism properties of these genes (or other genes with the same property) are probably not of direct interest.

aberration pattern	chrom	start	end	cytoband	gene num
Gain	chr1	71641212	72520865	p31.1	1
Loss	chr1	71641212	72520865	p31.1	1
Loss	chr1	149370786	149385728	q21.2	1
Gain	chr1	149579739	149586393	q21.2	1
Loss	chr1	238719495	238842085	q43	1
Loss	chr1	245069024	245148302	q44	1
Loss	chr3	196933423	196950211	q29	1
Gain	chr4	69546931	69570979	q13.2	1
Loss	chr4	69546931	70396215	q13.2	2
Loss	chr5	848719	904101	p15.33	1
Loss	chr6	32593131	32719407	p21.32	4
Loss	chr6	165660767	165995574	q27	1
Gain	chr7	288051	158630410	p22.3-q36.3	964
Loss	chr7	141517438	141567562	q34	1
Loss	chr8	7408720	7791647	p23.1	14
Gain	chr8	39291338	39499627	p11.23	2
Loss	chr8	39291338	39499627	p11.23	2
Loss	chr9	111037	33029062	p24.3-p13.3	114
Loss	chr10	170642	135341873	p15.3-q26.3	782
Loss	chr11	5755465	5756440	p15.4	1
Loss	chr11	55127492	55190148	q11	4
Gain	chr12	56374152	56462591	q14.1	8
Loss	chr13	35241122	101852125	q13.3-q33.1	158
Loss	chr14	18447593	92225087	q11.1-q32.12	13
Loss	chr15	18997107	19915749	q11.2	8
Loss	chr16	54352011	54366325	q12.2	1
Loss	chr17	41463128	41605371	q21.31	1
Gain	chr19	232043	63722733	p13.3-q13.43	784
Loss	chr19	56837617	56841944	q13.33	1
Gain	chr20	16350	62378023	p13-q13.33	567
Loss	chr20	1493029	1548689	p13	1
Gain	chr22	22670594	22731899	q11.23	4
Loss	chr22	22703116	22714284	q11.23	2
Loss	chr22	37683472	37718729	q13.1	2

Table 6.2.1 Segmentation result generated based on CBS segmentation algorithm. All the segments have frequencies at least 25% of all 170 patients samples

6.3 Functional impact of candidate genes or related pathways in the development of Glioblastoma tumor

PTEN has been found to be a tumor suppressor gene and the mutation or deletion of this gene is associated with the development of many cancers. PTEN has also been deleted in 60% -- 70% of high-grade Glioblastoma patients' tumors (Johnston, et al. 2006). In this study, PTEN was deleted in around 100 patients out of all 170 patient samples (depending on the segmentation method; see Table 6.3.1 for detailed information). The pathway involving PI3K/PTEN/AKT has been suggested in previous studies (Fan, et al. 2007, Furnari, et al. 2007). Deletion of the tumor suppressor gene PTEN could result in uncontrolled PI3K signaling (Knobbe and Reifenberger 2003, Ohgaki, et al. 2004). Many regulatory subunits of PI3K have been identified, which include PIK3CA, PIK3CB and PIK3CD. Copy number gains of PIK3CA, PIK3CB and PIK3CD in some patient samples indicate the potential involvement of this pathway. In our study, we also observed that more than 90 patients' samples contain copy number gains of PIK3CG. PIK3CG has been shown to interact with PIK3CD (Vanhaesebroeck, et al. 1997). Amplification of PIK3CG probably also involves the PI3K/PTEN/AKT signal pathways. Other PI3K subunits including PIK3R2 (gain in about 40 samples), PIK3AP1 (losses in 95 patients samples) were also identified in our study.

Some other genes acting as "induction of programmed cell death" are deleted in most of the patient samples. These genes include FAS, TIAL1, SMNDC1, BNIP3 and CUL2. FAS belongs to TNF-receptor superfamily and has been shown to activate NF-kappa-B and MAPK3/ERK1 pathways. NF-kappa-B has been linked to inflammatory events associated with many diseases. In this study, more than half of the patients have NF-kappa-B2 deleted. Since this gene is also located on chromosome 10, whether the deletion of this gene is a causal factor for the disease or is simply an involuntary deletion is unclear. FAS contains a death domain and is shown to play a central role in programmed cell death. Human leukocyte antigen (HLA) variants have been shown associated to be with the onset and prognosis of adult Glioblastoma multiforme (Tang, et al. 2005), in this analysis with CBS segmentation algorithm, HLA-DRB5 has been deleted in 109 patients samples and HLA-DRB6 been deleted in 78 patients. There are also 38 patients see the deletion of HLA-DRB1 in the tumor samples.

Epidermal growth factor receptor (EGFR) is a cell surface receptor. It has been shown that amplification of EGFR occurs in about 40% of Glioblastoma patients (Furnari, et al. 2007). In our study, about 100 patients have amplified EGFR. Matrix metallopeptidase 9 (MMP9) is involved in the breakdown of extracellular matrix. Amplified MMP9 might facilitate tumor cell growth and migration. More than 30 patients also displayed the elevated expression level of this gene.

P53 is another tumor suppressor gene acting as a check point regulator at the G1 to S phase. Loss of p53 could be either through point mutations or chromosome deletions whereas point mutation is inaccessible by using the array CGH method. But in this study, there are also 11 patients who have copy number deletion for this gene. It has also been shown that the protein encoded by CDKN2A is an important accessory to p53 activation under conditions of oncogenic stress due to its neutralization of the p53 ubiquitin ligase, MDM2 (Kamijo, et al. 1998, Pomerantz, et al. 1998, Stott, et al. 1998, Furnari, et al. 2007). The overwhelming deletion of tumor suppressor gene CDKN2A on chromosome 9 in about 100 patient samples suggestions a potential pathway of development of Glioblastoma tumor involving p53 regulation. Concordantly, MDM2 has been detected amplified in 25 patient samples when segmented based on CBS algorithm. We also observed that only 8 patients samples show amplification when data were analyzed by the HHMM algorithm. This suggests that there is a potential discrepancy between the CBS and HHMM algorithms. It is the best idea to consider all possibilities in order to get a better picture of the result. MDM4 could inhibit p53 transcription and enhances the ubiquitin ligase activity of MDM2 (Furnari, et al. 2007). We also noticed that 14 patients samples have amplified MDM4.

Retinoblastoma protein (Rb1) is another tumor suppressor protein frequently found deleted in many types of cancer (Murphree, et al. 1984). It has been shown that Rb1 can block the proliferation of tumor cells by binding to the E2F family of transcription factors (Sherr and McCormick 2002). Amplification of CDK4 and CDK6 accounts for the inactivation of Rb1 in many patient samples (Serrano, Hannon and Beach 1993). In this study, 99 patient samples have amplified copy numbers of CDK6. Copy number gains of CDK6 probably at least partially contribute to the inactivation of Rb1. We observed 41 patient sample deletions using the CBS segmentation algorithm and 49 sample deletions using HHMM segmentation method. P16 (or CDKN2A) is another tumor suppressor gene, which functions as an inhibitor of CDK4. In 104 patient samples, copy number losses of p16 were observed based on the CBS segmentation algorithm and 90 patients showed similar patterns based on the HHMM methods. CDKN2B lies adjacent to the CDKN2A and encodes also a cyclin-dependent kinase inhibitor, which forms a complex with CDK4 or CDK6. We observed 103 and 89 patients with copy number losses of CDKN2B from the CBS and HHMM segmentation algorithms respectively.

DMTF1 functions as negative regulation of cell cycles. Gene expression level of DMTF1 was amplified in at least 93 patient samples. Other cell cycle regulating genes that have been amplified in at least half of the patient samples include RINT1, MAD1L1, HUS1, CDC2L5, INHBA and ASNS and all of them are located on chromosome 7.

DMBT1 is a "Deleted in Malignant Brain Tumors 1" protein and the deletion of this gene has been associated with the progression of human cancers. DMBT1 was originally isolated based on the deletion in a medulloblastoma cell line, but was also later found deleted in many cases of Glioblastoma tumor. In this study, from HHMM segmentation analysis with 1.5xDLRs cutoff threshold, 121 patients have DMBT1 deleted, and 105 patients have the same gene deleted based on the CBS segmentation method. Other genes that involve cell cycle regulating are BUB3, KIF20B, BCCIP, ZWINT, PDCD4, RASSF4, ZMYND11, CDC123, GTPBP4. All these genes were deleted in more than half of the patient sample.

Gene	Gain (CBS)	Loss (CBS)	Gain (HHMM)	Loss (HHMM)	Chromosomal Location	Gene Ontology Biological Process
ASNS	87	4	90	2	chr7q21.3	asparagine biosynthetic process; positive regulation of mitotic cell cycle
BCCIP	4	97	4	95	chr10q26.1	regulation of cyclin-dependent protein kinase activity; cell cycle
BNIP3	4	93	8	90	chr10q26.3	cell death; negative regulation of survival gene product expression
BUB3	4	98	4	94	chr10q26	mitotic sister chromatid segregation; mitotic cell cycle checkpoint

Gene	Gain (CBS)	Loss (CBS)	Gain (HHMM)	Loss (HHMM)	Chromosomal Location	Gene Ontology Biological Process
CDC123	10	84	9	83	chr10p13	cell cycle arrest; regulation of mitotic cell cycle; positive regulation of cell proliferation
CDC2L5	89	4	92	2	chr7p13	positive regulation of cell proliferation
CDK4	37	3	15	1	chr12q14	G1/S transition of mitotic cell cycle; positive regulation of cell proliferation
CDK6	99	4	93	2	chr7q21-q22	G1 phase of mitotic cell cycle; negative regulation of cell cycle
CDKN2A	6	104	7	90	chr9p21	cell cycle checkpoint; G1/S transition of mitotic cell cycle; induction of apoptosis
CDKN2B	6	103	7	89	chr9p21	negative regulation of cell proliferation ; G1/S transition checkpoint
CUL2	6	89	6	86	chr10p11.21	G1/S transition of mitotic cell cycle; negative regulation of cell proliferation
DMBT1	4	99	4	105	chr10q26.13	multicellular organismal development; epithelial cell differentiation
DMTF1	94	4	93	2	chr7q21	cell cycle
EGFR	112	7	96	3	chr7p12	activation of MAPKK activity; negative regulation of mitotic cell cycle
FAS	4	97	4	96	chr10q24.1	induction of apoptosis; positive regulation of necrotic cell death
GTPBP4	10	82	8	80	chr10p15-p14	regulation of cyclin-dependent protein kinase activity
HLA-DRB1	7	38	5	23	chr6p21.3	immune response
HLA-DRB5	10	109	10	47	chr6p21.3	immune response
HLA-DRB6	9	78	8	33	chr6p21.3	---
HUS1	89	4	91	3	chr7p13-p12	DNA damage checkpoint
INHBA	88	4	91	2	chr7p15-p13	G1/S transition of mitotic cell cycle; negative regulation of cell cycle
KIF20B	4	96	4	93	chr10q23.31	M phase of mitotic cell cycl; cell cycle arrest
MAD1L1	90	4	94	5	chr7p22	mitotic cell cycle checkpoint
MDM2	25	4	8	2	chr12q14.3-q15	negative regulation of transcription from RNA polymerase II promote
MDM4	14	0	9	0	chr1q32	negative regulation of transcription from RNA polymerase II promote; G0 to G1 transition
MMP9	37	0	38	0	chr20q11.2-q13.1	positive regulation of apoptosis
NFKB2	5	91	5	72	chr10q24	regulation of transcription, DNA-dependent
PDCD4	4	94	6	90	chr10q24	apoptosis; negative regulation of cell cycle

Gene	Gain (CBS)	Loss (CBS)	Gain (HHMM)	Loss (HHMM)	Chromosomal Location	Gene Ontology Biological Process
PIK3CA	8	2	11	4	chr3q26.3	negative regulation of apoptosis
PIK3CB	6	5	10	4	chr3q22.3	activation of MAPK activity
PIK3CD	8	17	8	5	chr1p36.2	B cell homeostasis
PIK3CG	90	4	92	2	chr7q22.3	negative regulation of apoptosis
PTEN	4	98	4	97	chr10q23.3	regulation of cyclin-dependent protein kinase activity; negative regulation of apoptosis
RASSF4	3	92	3	88	chr10q11.21	cell cycle
RB1	1	41	6	48	chr13q14.2	cell cycle checkpoint
RINT1	91	4	92	2	chr7q22.3	G2/M transition DNA damage checkpoint
SMNDC1	4	93	4	92	chr10q23	induction of apoptosis
TIAL1	4	98	4	97	chr10q	induction of apoptosis; positive regulation of cell proliferation
ZMYND11	10	82	8	83	chr10p14	negative regulation of transcription from RNA polymerase II promoter; cell cycle
ZWINT	5	91	5	88	chr10q21-q22	cell cycle; phosphoinositide-mediated signaling

Table 6.3.1 Gene information discussed in this section. In the column of Gene Oncology biological process, only partial GO terms are listed. All these GO terms are annotated based on Affymetrix gene chip annotation table downloaded from Affymetrix website (http://www.affymetrix.com).

6.4 Genes that show significant variation in individual patients

Many genes show very big gene expression variation within individual samples. For instance, 951 genes have at least four-fold ratio differences between the tumor samples and the reference control within all 170 tumor samples. Among these genes, 245 of them have at least 16-fold aberration. As discussed before, EGFR plays a very important role in the Glioblastoma development. In 50 tumor samples out of 170 patients samples, EGFR is amplified at least 16 folds. Even though, for many genes, the biological function involving the development of Glioblastoma remains unknown, the potential function of some of them involving the pathways discussed before suggests the importance of these genes. Table 6.4.1 shows the partial gene list that has at least four-fold aberrations. The full list is given in supplementary Table S6.4.1.

genes	chrom	sample number	pattern	genes	chrom	sample number	pattern
K03193	7	62	Gain	AJ001612	7	15	Gain
EGFR	7	62	Gain	MBD6	12	15	Gain
SEC61G	7	42	Gain	DDIT3	12	14	Gain
LANCL2	7	32	Gain	GEFT	12	14	Gain
AK128355	7	25	Gain	DTX3	12	14	Gain
CDK4	12	25	Gain	SLC35E3	12	14	Gain
CENTG1	12	25	Gain	SLC26A10	12	14	Gain
TSPAN31	12	25	Gain	KIF5A	12	14	Gain
ECOP	7	24	Gain	FLJ44060	7	13	Gain
38784	12	24	Gain	DCTN2	12	13	Gain
CYP27B1	12	24	Gain	B4GALNT1	12	13	Gain
AK093897	12	24	Gain	PIP5K2C	12	13	Gain
DKFZP586D0919	12	24	Gain	MARS	12	13	Gain
METTL1	12	24	Gain	NUP107	12	13	Gain
CR613464	7	23	Gain	GSH2	4	12	Gain
BC045679	7	23	Gain	PDGFRA	4	12	Gain
MGC33530	7	22	Gain	AY229892	4	12	Gain
TSFM	12	22	Gain	LOC402176	4	12	Gain
AVIL	12	21	Gain	CHIC2	4	12	Gain
OS9	12	20	Gain	BC094796	7	12	Gain
CTDSP2	12	20	Gain	ZNF713	7	11	Gain
AF119871	12	19	Gain	ARHGAP9	12	11	Gain
MDM2	12	18	Gain	CR590495	7	10	Gain
CPM	12	18	Gain	MRPS17	7	9	Gain
BC032840	12	17	Gain	GBAS	7	9	Gain
XRCC6BP1	12	17	Gain	CPSF6	12	9	Gain
CR602022	12	16	Gain	GLI1	12	9	Gain
AK096400	12	16	Gain	DDX5	17	9	Gain
AF109294	9	37	Loss	KLHL9	9	19	Loss
CDKN2B	9	36	Loss	IFNA14	9	18	Loss
CDKN2A	9	36	Loss	IFNE1	9	18	Loss
HLA-DRB5	6	35	Loss	BTNL3	5	15	Loss
AJ007770	7	34	Loss	IFNW1	9	14	Loss
X89654	8	31	Loss	ELAVL2	9	14	Loss
LOC651362	8	30	Loss	CR627240	9	14	Loss
MTAP	9	26	Loss	BC042393	9	14	Loss
LOC652848	14	26	Loss	UGT2B17	4	13	Loss

genes	chrom	sample number	pattern	genes	chrom	sample number	pattern
LOC647353	7	25	Loss	U06641	4	13	Loss
BC003593	6	24	Loss	AK127991	22	12	Loss
AK092601	9	22	Loss	GSTT1	22	12	Loss
DMRTA1	9	22	Loss	BC043197	13	11	Loss
AK124391	9	21	Loss	BC019327	3	10	Loss
IFNA8	9	19	Loss	KIAA1797	9	10	Loss
IFNA2	9	19	Loss	OR4P4	11	10	Loss
V00541	9	19	Loss	CR593785	11	10	Loss

Table 6.4.1 Partial gene list that have either more than 4 folds gains or homozygous deletions with $\log_2$(Ratio) more than 2 in at least 9 or 10 patients' samples.

MDM2 is potentially involved in p53 tumor suppressor pathway. Similar to EGFR and CDK4, MDM2 helps to regulate the cell cycle process. The common GO term "regulation of Ras protein signal transduction" of GEFT, GENTG1 and also SLC26A10 indicate the potential involvement of the Ras signal transduction pathway in at least some patients. Since the maximum copy number losses are 2, the "high" copy number losses simply suggest the homozygous deletions of the gene.

6.5 Survival analysis

Even though Glioblastoma is a WHO grade IV tumor with very high mortality rate. From Table 2.1.1, we can still see that the survival times of individual patients are quite different, ranging from a few days to more than a hundred months after the diagnosis and treatment. In this analysis, in order to analyze the potential correlation of individual genes or segments of chromosome to the survival information, the log rank test (or Chi square test) is performed to compare the differences of two groups (the gene (or segment) amplified or deleted vs unchanged within all samples). Table 6.5.1 and Table 6.5.2 shows the analysis result based on the CBS segmentation method. In Table 6.5.1, all the “segments” have amplified copy number aberration, while Table 6.5.2 shows the survival analysis results of “segments” and genes with copy number deletion aberration.

Chrom	Start	End	Genes	pvalue
Chr7	5312948	5429703	TNRC18	0.000775179
Chr7	28305464	28832036	CREB5	0.000944061
Chr7	29200645	30763743	CHN2 PRR15 CRHR2 INMT	0.000611761
Chr7	30777557	31714593	FLJ22374 AQP1 GHRHR……	0.000983259
Chr7	31795771	32077516	PDE1C	0.00094093
Chr7	32491469	32736120	LSM5 AVL9 LOC441208	0.000983259
Chr7	32874302	32949307	KBTBD2 RP9P	0.000611761
Chr7	32963529	33612205	FKBP9 NT5C3 RP9 BBS9	0.000924488
Chr7	53070842	53072112	DKFZp564N2472	0.000828016
Chr7	54236410	54237608	HPVC1	8.52E-05
Chr7	54577512	54604442	VSTM2A	0.000614989
Chr7	54787433	55242525	SEC61G EGFR	0.000458793
Chr7	55400634	55468929	LANCL2	0.000566615
Chr7	55505799	55607694	ECOP	0.000829675
Chr7	55680805	55682137	LOC442308	0.000895219
Chr7	55828730	55897976	SEP14	0.000349342
Chr7	55947824	55975927	ZNF713	0.000685318
chr17	20970849	21035428	DHRS7B	0.00011957
chr17	24107122	24357207	C17orf63 ERAL1…	0.000676561
chr17	37718868	37794039	STAT3	0.000168583
chr17	37807993	38105536	PTRF ATP6V0A1 NAGLU ……	0.00045271
chr17	38306340	38318912	G6PC	1.22E-05
chr17	42283966	42873676	WNT9B GOSR2 RPRML……	0.000106719
chr17	43373887	45633999	PNPO ATAD4 CDK5RAP3 ……	0.000206057
chr17	45706831	46553225	TMEM92 XYLT2 MRPL27……	2.53E-06
chr17	46585918	51209747	NME1-NME2 NME1……	0.000206057
chr17	63144522	63799000	NOL11 BPTF C17orf58……	0.000966895
chr17	78494955	78646011	B3GNTL1 METRNL	0.000825796
chr10	7837372	14412872	KIN ATP5C1 TAF3 ……	0.000339128
chr10	14600564	14856902	FAM107B	0.000195038
chr10	14901256	15801776	ARMETL1 HSPA14 SUV39H2 ……	0.000339128
chr10	15860180	15942525	C10orf97	0.000562784
chr10	16518972	17283687	PTER C1QL3 RSU1 CUBN TRDMT1	0.000339128
chr22	21731592	21990224	RTDR1 GNAZ RAB36 BCR	0.000326071
chr22	41226539	41245752	SERHL RRP7A	0.000381544
Chr5	70366555	70399256	GTF2H2B GTF2H2C	0.000226356

Chrom	Start	End	Genes	pvalue
Chr5	70366706	70399253	GTF2H2	2.67E-05
Chr5	70366932	70399233	GTF2H2D	0.000226356
Chr5	70405174	70424653	OCLN LOC647859	2.67E-05
Chr5	129268421	135720972	CHSY3 TRPC7	0.00075346
chr16	11255774	11353118	SOCS1 TNP2 PRM3…..	4.73E-06

Table 6.5.1 Survival analysis of segments with amplified copy number aberration based on the CBS segmentation result. Only segments with p value less than 0.001 are shown here. Due to space limitation, for some segments, only a partial gene list is shown. The segment information generate here is different from the segments from CGH segmentation. For each individual gene a log rank test was performed, and then all the gene on the same chromosome, located near each other, and most importantly having the exact same gain and no change patterns across all 170 samples (so with the same p value) are combined together to generate this table.

Chrom	Start	End	Genes	pvalue
chr3	37259741	47029961	GOLGA4 C3orf35 ITGA9…	0.003403045
chr3	47032903	47180471	SETD2	0.002168912
chr3	47244520	51509041	KIF9 KLHL18 PTPN23 …	0.003403045
chr8	35212516	35771722	UNC5D	0.004481786
chr1	793319	869824	FAM41C LOC100130417 SAMD11	0.003662627
chr1	869445	1041599	NOC2L KLHL17 PLEKHN1 HES4 ISG15 AGRN C1orf159	0.00135508
chr1	1099148	1171965	TTLL10 TNFRSF18 TNFRSF4 SDF4 B3GALT6 FAM132A	0.002042499
chr1	1179154	1500125	UBE2J2 SCNN1D ACAP3 …	0.004211136
chr1	158352143	158379996	ATP1A2	0.003451951
chr1	239005439	239587101	RGS7	0.003109638
chr10	42598264	43224702	BMS1 RET CSGALNACT2 RASGEF1A FXYD4 HNRNPF	0.004732608
chr10	43252579	43464332	ZNF487 ZNF239 ZNF485 ZNF32	0.0040157
chr10	43602865	43605871	HNRNPA3P1	0.002556484
chr10	44185610	45410360	CXCL12 TMEM72 RASSF4…	0.001861497
chr10	45431044	45488257	ANUBL1	0.002811846
chr10	47128239	47171452	ANTXRL	0.003682633
chr10	52420950	53725280	PRKG1	0.004829294

Table 6.5.2 Survival analysis of segments with deleted copy number aberration based on the CBS segmentation result. Only segments with p value less than 0.005 are shown here. Due to space limitation, for some segments, only partial gene list is shown. Similar to Table 6.5.1, the segments information generate here is different to the segments from CGH segmentation. A log rank test is performed on each individual gene; then all the genes on the same chromosome, located near each other, and most importantly having the exact same deletion and no change patterns across all 170 samples (so with the same p value) are combined together to generate this table.

Even though Tables 6.5.1 and 6.5.2 clearly show significant differences based on the log rank test, we can't tell from the table whether the copy number aberration helps survival or works in an opposite way. Kaplan-Meier plots can reveal this kind of information visually. Figure 6.5.1 shows several typical Kaplan-Meier plots of the analysis from amplified segments based on the CBS segmentation. From these plots, basically there are two different patterns; one is that amplified copy numbers helps patients surviving longer time. Segments with this pattern in Figure 6.5.1 include some genes located at Chr10:14901256-15801776 (p value: 0.00034), Chr10:15860180-15942525 (p value: 0.00056), and Chr10:14600564-14856902(pvalue: 0.00020). Although for most patients' samples, chromosome 10 suffers massive copy number deletion, for instance, the gene PTER is deleted in 88 patients' samples, but still in 11 patients' samples, the copy number of this gene is actually amplified. Some other genes with this pattern and within above segments include FAM107B, ARMETL1, HSPA14, SUV39H2, C10orf97, PTER, C1QL3, RSU1, CUBN, TRDMT1, et al.

Another set of genes actually does more damage than good to the patients. Many of them located on Chromosome 17, for instance, Chr17: 45706831-46553225 (p value: 2.5e-06), Chr17: 78494955-78646011 (p value: 0.00083) and Chr17: 37807993-38105536 (p value: 0.00045). Some genes located at these regions include B3GNTL1, METRNL, TMEM92, XYLT2, MRPL27, PTRF, ATP6V0A1, NAGLU, et al. For the other genes or segments with significant log rank p value ($p<0.0001$), amplifications of these genes in these segments won't help or damage too much to the survival as shown on the rest of the Kaplan-Meier plots Figure 6.5.1, even though, statistically both groups (patients with amplified genes and patients with no gene changes) are different.

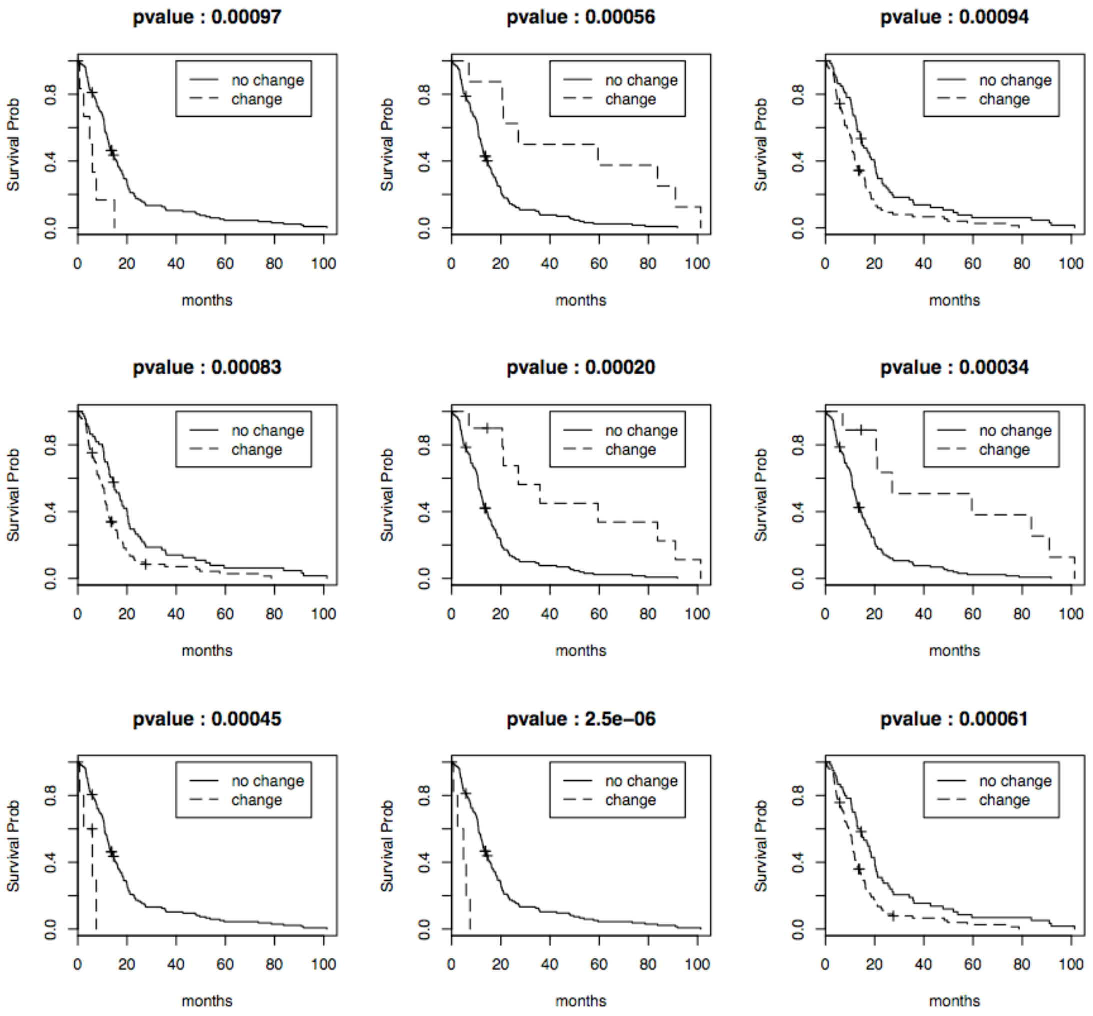


Figure 6.5.1 Typical Kaplan-Meier plots of survival analysis results based on the genes or segments with amplified copy number aberration based on the CBS segmentation analyzing result. In the figures, the y-axis represents the survival probability and the x-axis represents the survival time in months. In the legend, “change” represents that the specific gene or segment has amplified copy numbers and “no change” indicates no copy number aberration observed for this specific gene or segments in the patient sample. Here only show part of the Kaplan-Meier plots, but the rest plots are quite similar.

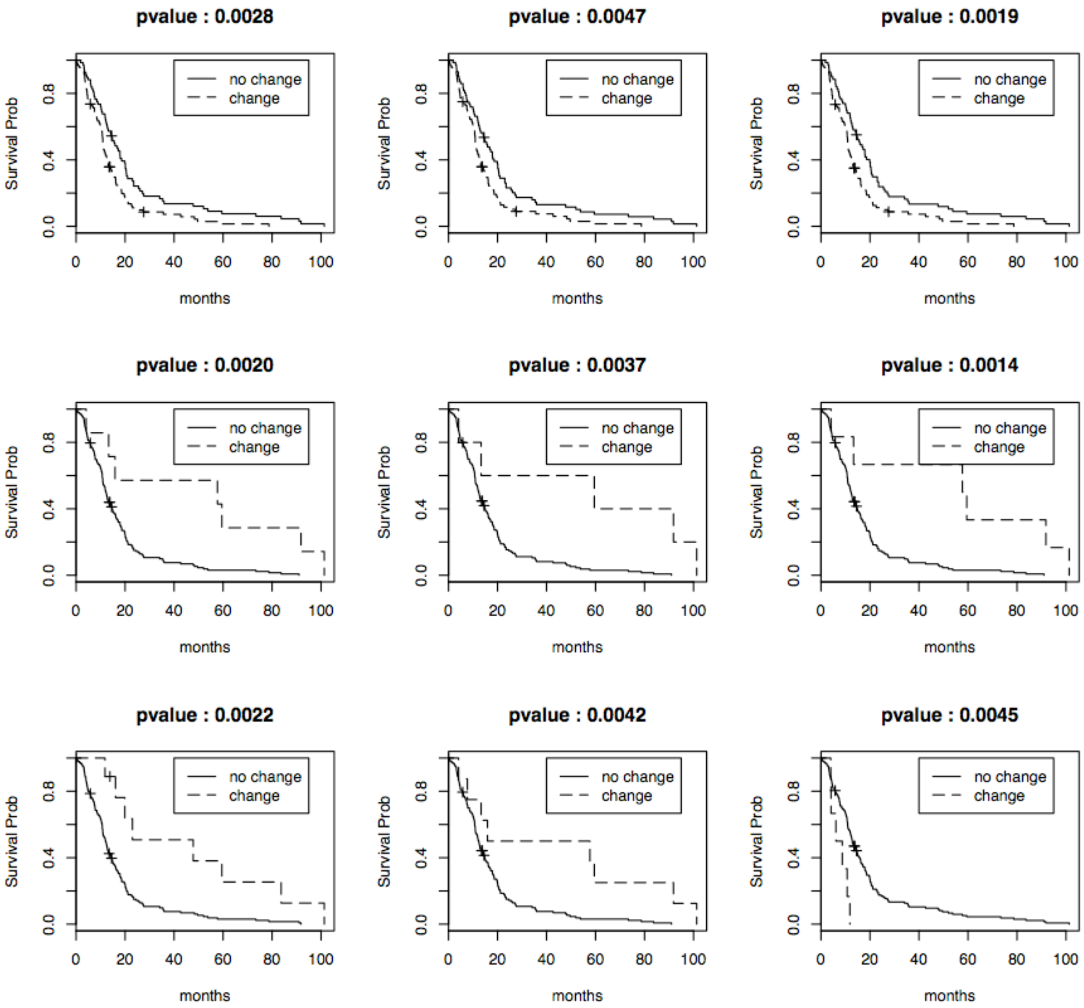
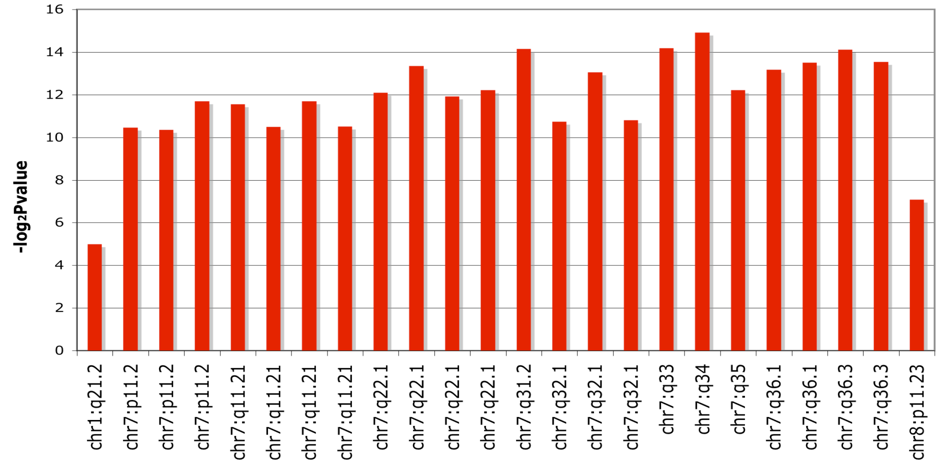


Figure 6.5.2 Typical Kaplan-Meier plots of survival analysis results based on the genes or segments with deleted copy number aberration based on the CBS segmentation analyzing result. In the figures, the y-axis represents the survival probability and the x-axis represents the survival time in months. In the legend, “change” represents that the specific gene or segment has amplified copy numbers and “no change” indicate no copy number aberration observed for this specific gene or segments in the patient sample. Here only show part of the Kaplan-Meier plots, but the rest plots are quite similar.

For those genes with deleted copy number aberration as shown in Table 6.5.2, Kaplan-Meier survival analysis (Figure 6.5.2) also shows patterns similar to Figure 6.5.1. Some segments at chromosome 1 and 3, for instance, Chr1: 793319-869824 (p value: 0.0037), Chr1: 869445-1041599 (p value: 0.0014), Chr1: 1099148-1171965 (p value: 0.0020), Chr1: 1179154-1500125 (p value: 0.0042), and Chr3: 3725941-51509047 (p value: 0.0021 and 0.0034), deletion of these genes actually correlate to extended survival of patients.

6.6 Fisher exact test

As discussed before, 1.5xDRLs is applied to identify those segments with copy number variation. Each gene within these segments could be assigned either gain or loss depending on the patterns of the copy number variation of each individual segments. For those segments with amplified copy number variation from all 170 tumor samples, each individual gene is assigned either gain or no gain. There are also 24 normal samples in which the same gene could also be assigned to gain or no gain. One-tail Fisher exact tests are performed for all genes (each gene has one 2x2 contingency table) and p values are derived for each gene. Similarly, for those segments with copy number deletion, each individual gene within these segments could also form a 2x2 contingency table and each table could apply Fisher exact test. After the testing, the adjusted p values for multiple comparisons are calculated based on the method of Benjamin and Hochberg (1995)and the result is shown in Figure 6.6.1. Many other genes, some of which were discussed in Section 6.3, also show significant p values. These genes include HUS1, CDC2L5, DMTF1, RINT1, EGFR, CDK6, MAD1L1, INHBA, PIK3CG, ASNS, and CDKN2B. All these genes have adjusted p values less than 0.05, and except for CDKN2B, all of them have gained copy number aberrations.



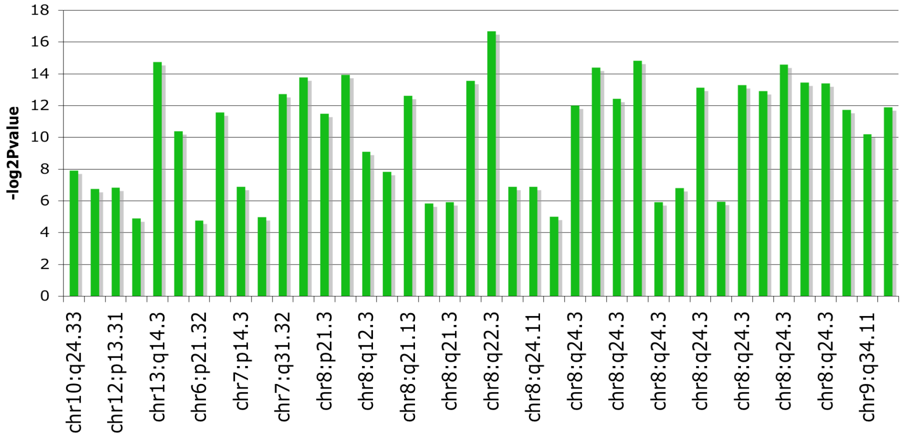


Figure 6.6.1 Fisher exact test based on the CBS segmentation result. Top figure is derived from the genes with copy number gains and the lower figure is from those genes with copy number deletion. Genes are grouped based on the same adjusted p value, and each group is located on the same segment. The log2 transformed adjusted p values (less than 0.05) are plotted against the cytobands shown as above.

Chapter 7 Discussion

This thesis includes two main parts; the first part discussed the raw data processing and the second part mainly focused on biological aspects from the analysis. Through out the entire analysis, extensive statistical methods are used, which include, but are not limited to, Hidden Markov Chain, log rank test, Fisher exact test, as well as many others. Briefly, in the data processing part, even though there are more than ten different data processing algorithms available, many of them are gradually becoming obsolete due to various reasons. HHMM and CBS are the two data processing algorithms that stand up among all methods. In this thesis, both HHMM and CBS algorithms are used and their performances are compared using the data set generated from real tumor samples. From the analysis we can see CBS tends to identify the relatively big segments, but HHMM tends to identify more single probe aberrations while missing many obvious big segments. Simply based on the current CGH experiments, it is hard to tell whether single probe aberrations represent true biological copy number variation. One potential solution to verify whether the single probe aberrations really represent the copy number variation is to increase the resolutions of the CGH array. Moreover, the biological significance of each copy number aberration needs to be verified based on the laboratory bench work.

Ideally, the copy number variation could easily be calculated mathematically as shown in Table 3.1.1. Due to many systematic and random errors involved in each experiment, we seldom can apply the ideal mathematic values in Table 3.1.1 to find the copy number changes. Instead, the noise level (DRLs) of each array experiment is calculated and an arbitrary 1.5x(DRLs) cutoff is applied to individual hybridization. All the segments with the absolute predicted $\log_2$(Ratio) values greater than this cutoff are treated as segments with copy number variation.

Frequency plots also show (Figure 6.2.1) that copy number gains on the whole or most part of chromosome 7 and losses on whole or part of chromosome 10 are the signatures of Glioblastoma. Other than these two whole chromosome copy number aberrations, other chromosomes also contain various copy number aberrations with various segment lengths as shown in Table 6.2.1. We also observed from the frequency plots that there is no single segment whose copy number variation is shared by all patients. Gene level analysis (Fisher test) also yields the same conclusion. This suggests that even though all patients suffered from the same or similar WHO grade IV Glioblastoma, the genotypes of these patients are similar but not identical. This also implies that the Glioblastoma in individual patients many not have the same causation at the genome level.

In this analysis, we also noticed that many tumor suppression genes are deleted and the copy numbers of many cell cycle related genes are changed. Some of the well-known tumor suppression genes identified in this study include, but are not limited to, PTEN, TP53, CDKN2A and RB1. Many of these genes are involved in the cell cycle checkpoints. Deletion of any of these tumor suppression genes in the tumor cells may directly cause the uncontrolled tumor growth in these patients.

From the segmentation results, we can also confirm some signal transduction pathways that are involved in the development of Glioblastoma. One of these pathways is PI3K/PTEN/AKT. EGFR has been found amplified in the tumors of more than half of the patients, which is consistent with published results. The NF-kappa-B involved signal transduction pathway has been linked to inflammatory events associated with many diseases. In this study, identification of some components involving this pathway suggests the importance of NF-kappa-B and MAPK3/ERK1signal transduction pathways in tumor development.

Due to the individual variation of each patient, the causations of the same disease are also very different. Even though we can generate a signature pattern for this disease using frequency plots as shown in Figure 6.2.1, the gene level fingerprints of each individual patients could also reveal the importance of many genes or segments to each individual patient, especially when the copy number variation is significantly changed (homozygous deletion or multi-fold copy number gains). This part of the analysis will clearly benefit us in the personalized gene therapy or personalized drug development study in the future.

Even though the mortality rate of the grade IV Glioblastoma disease is very high, survival analysis using the log rank test clearly shows the differences of some segments or genes to the survival rate. In this analysis, amplification of some genes at specific segments for instance Chr10: 14600564-15942525 or deletion of copy number of Chr1: 793319-1500152 and Chr3: 3725941-51509047, may help to extend the life expectation of some Glioblastoma patients. Even though many genes within these segments are known genes, their biological functions involving Glioblastoma are actually not quite clear and array CGH can’t answer this question. The only solution to this question is laboratory bench work for each individual gene. But at least this result could serve as guidance for future laboratory work.

Appendix A   Supplementary Tables and Figures

Table S2.1.1 Detailed patients' information downloaded from public available sources. Due to descripancy of two difference data sources, there are two vital status of each patient

Case ID	Gender	PRETREATMENT HISTORY	VITAL STATUS	Vital Status	DAYS TO BIRTH	DAYS TO DEATH	DAYS TO LAST FOLLOWUP	DAYS TO TUMOR PROGRESSION	DAYS TO TUMOR RECURRENCE	Secondary or Recurrent	Age at Procedure	Age at Death
TCGA-02-0001	FEMALE	Yes	DECEASED	DEAD	-16179	353	358	null	137	Rec	45	45
TCGA-02-0003	MALE	No	DECEASED	DEAD	-18340	144	144	null	40	No	50	50
TCGA-02-0006	FEMALE	No	DECEASED	DEAD	-20516	558	558	null	302	No	56	58
TCGA-02-0007	FEMALE	Yes	DECEASED	DEAD	-14805	705	705	518	null	Rec	42	43
TCGA-02-0009	FEMALE	No	DECEASED	DEAD	-22456	322	322	264	null	No	61	62
TCGA-02-0010	FEMALE	Yes	DECEASED	DEAD	-7451	1077	1077	351	null	Sec	22	23
TCGA-02-0011	FEMALE	No	DECEASED	DEAD	-6926	630	630	144	null	No	19	20
TCGA-02-0014	MALE	No	DECEASED	DEAD	-9369	2511	2511	-1409	null	Rec	25	32
TCGA-02-0021	FEMALE	Yes	DECEASED	DEAD	-16035	2361	2361	254	null	Rec	44	50
TCGA-02-0024	MALE	Yes	DECEASED	DEAD	-13116	1614	1614	null	1400	Rec	39	40
TCGA-02-0027	FEMALE	No	DECEASED	DEAD	-12369	370	315	257	null	No	34	35
TCGA-02-0028	MALE	Yes	DECEASED	DEAD	-14303	2755	2755	1921	null	Sec	46	46
TCGA-02-0033	MALE	No	DECEASED	DEAD	-20070	86	86	32	null	No	55	55
TCGA-02-0034	MALE	No	DECEASED	DEAD	-22166	430	430	386	null	No	61	62
TCGA-02-0037	FEMALE	No	DECEASED	DEAD	-27063	109	109	37	null	No	74	74
TCGA-02-0038	FEMALE	No	DECEASED	DEAD	-17749	326	326	238	null	No	49	50
TCGA-02-0043	FEMALE	Yes	DECEASED	DEAD	-19882	556	556	282	null	Rec	56	56
TCGA-02-0046	MALE	No	DECEASED	DEAD	-22417	208	208	194	null	No	61	62
TCGA-02-0047	MALE	No	DECEASED	DEAD	-28759	447	447	57	null	No	79	80
TCGA-02-0052	MALE	No	DECEASED	DEAD	-18060	383	383	204	null	No	50	51
TCGA-02-0054	FEMALE	No	DECEASED	DEAD	-16223	199	199	72	null	No	44	45
TCGA-02-0055	FEMALE	No	DECEASED	DEAD	-22798	76	76	6	null	No	62	62
TCGA-02-0057	FEMALE	Yes	DECEASED	DEAD	-24139	604	604	473	null	Rec	67	68
TCGA-02-0058	FEMALE	No	DECEASED	DEAD	-10517	254	254	171	null	Rec	29	29
TCGA-02-0060	FEMALE	No	DECEASED	DEAD	-24150	183	183	183	null	No	66	66
TCGA-02-0064	MALE	No	DECEASED	DEAD	-18280	600	600	null	496	No	49	51

Case ID	Gender	PRETREATMENT HISTORY	VITAL STATUS	Vital Status	DAYS TO BIRTH	DAYS TO DEATH	DAYS TO LAST FOLLOWUP	DAYS TO TUMOR PROGRESSION	DAYS TO TUMOR RECURRENCE	Secondary or Recurrent	Age at Procedure	Age at Death
TCGA-02-0069	FEMALE	No	LIVING	DEAD	-11668	null	873	873	null	No	31	34
TCGA-02-0071	MALE	No	DECEASED	DEAD	-19425	167	167	8	null	No	53	54
TCGA-02-0074	FEMALE	No	DECEASED	DEAD	-24906	310	310	154	null	No	68	69
TCGA-02-0075	MALE	No	DECEASED	DEAD	-23205	634	634	336	null	No	63	65
TCGA-02-0079	MALE	No	DECEASED	DEAD	-21173	828	828	797	null	No	56	58
TCGA-02-0080	MALE	Yes	DECEASED	DEAD	-10309	2729	2729	2339	null	Rec	34	35
TCGA-02-0083	FEMALE	Yes	DECEASED	DEAD	-21626	691	691	462	null	Rec	61	62
TCGA-02-0084	FEMALE	No	DECEASED	DEAD	-13263	384	7	null	null	No	34	37
TCGA-02-0085	FEMALE	No	DECEASED	DEAD	-23289	1560	1560	976	null	No	66	69
TCGA-02-0086	FEMALE	No	DECEASED	DEAD	-16763	268	268	97	null	No	46	47
TCGA-02-0087	FEMALE	No	LIVING	DEAD	-10185	null	1757	null	1757	No	28	30
TCGA-02-0089	MALE	Yes	DECEASED	DEAD	-19233	515	515	358	null	Rec	54	54
TCGA-02-0099	MALE	Yes	DECEASED	DEAD	-17080	106	106	null	null	Rec	46	47
TCGA-02-0102	MALE	Yes	DECEASED	DEAD	-15660	821	821	450	null	Sec	44	45
TCGA-02-0106	MALE	No	DECEASED	DEAD	-19883	355	355	null	195	No	55	55
TCGA-02-0107	MALE	Yes	DECEASED	DEAD	-20557	537	537	null	146	Rec	57	58
TCGA-02-0111	MALE	No	DECEASED	DEAD	-20813	704	704	74	null	No	57	59
TCGA-02-0113	FEMALE	Yes	LIVING	DEAD	-15836	null	2817	null	1559	Rec	49	51
TCGA-02-0114	FEMALE	Yes	DECEASED	DEAD	-13681	3040	3040	null	2296	Sec	45	46
TCGA-02-0115	MALE	No	DECEASED	DEAD	-19257	476	476	91	null	No	53	54
TCGA-02-0116	MALE	Yes	DECEASED	DEAD	-18656	1489	1489	1231	null	Rec	54	55
TCGA-02-0258	FEMALE	No	DECEASED	DEAD	-13268	503	503	null	503	No	36	37
TCGA-02-0260	MALE	No	DECEASED	DEAD	-19956	514	514	null	null	No	54	56
TCGA-02-0266	MALE	No	DECEASED	DEAD	-5303	538	538	293	null	No	15	16
TCGA-02-0269	MALE	No	DECEASED	DEAD	-25194	327	327	99	null	No	69	70
TCGA-02-0271	MALE	No	DECEASED	DEAD	-9578	440	440	0	null	No	26	27
TCGA-02-0281	FEMALE	No	DECEASED	DEAD	-28695	121	121	0	null	No	78	79
TCGA-02-0285	FEMALE	No	DECEASED	DEAD	-18353	422	422	0	null	No	50	52
TCGA-02-0289	MALE	No	DECEASED	DEAD	-21031	432	432	244	null	No	57	59
TCGA-02-0290	MALE	No	DECEASED	DEAD	-18066	485	485	374	null	No	49	50

Case ID	Gender	PRETREATMENT HISTORY	VITAL STATUS	Vital Status	DAYS TO BIRTH	DAYS TO DEATH	DAYS TO LAST FOLLOWUP	DAYS TO TUMOR PROGRESSION	DAYS TO TUMOR RECURRENCE	Secondary or Recurrent	Age at Procedure	Age at Death
TCGA-02-0317	MALE	No	DECEASED	DEAD	-14636	372	372	238	null	No	40	41
TCGA-02-0321	MALE	No	DECEASED	DEAD	-27125	300	300	73	null	No	15	15
TCGA-02-0324	FEMALE	No	DECEASED	DEAD	-25352	234	234	163	null	No	69	70
TCGA-02-0325	MALE	No	DECEASED	DEAD	-22598	323	323	284	null	No	62	63
TCGA-02-0326	FEMALE	No	DECEASED	DEAD	-30159	223	374	5	null	No	83	84
TCGA-02-0330	FEMALE	No	DECEASED	DEAD	-18654	484	484	120	null	No	51	52
TCGA-02-0332	FEMALE	No	DECEASED	DEAD	-17021	782	782	438	null	No	47	47
TCGA-02-0333	FEMALE	No	DECEASED	DEAD	-28449	133	133	4	null	No	78	78
TCGA-02-0337	MALE	No	DECEASED	DEAD	-17642	764	764	691	null	No	48	50
TCGA-02-0338	MALE	No	DECEASED	DEAD	-15231	322	322	167	null	No	42	43
TCGA-02-0339	MALE	No	DECEASED	DEAD	-24524	377	377	148	null	No	67	68
TCGA-02-0422	MALE	No	DECEASED	DEAD	-18367	441	441	125	null	No	50	51
TCGA-02-0430	FEMALE	No	DECEASED	DEAD	-24641	321	321	null	null	No	68	68
TCGA-02-0432	MALE	No	DECEASED	DEAD	-13260	1433	1433	1053	null	No	37	41
TCGA-02-0439	FEMALE	No	DECEASED	DEAD	-25623	20	20	13	null	No	70	70
TCGA-02-0440	MALE	No	DECEASED	DEAD	-22955	345	345	212	null	No	63	64
TCGA-02-0446	MALE	No	DECEASED	DEAD	-22541	281	281	null	15	No	61	62
TCGA-02-0451	FEMALE	No	DECEASED	DEAD	-22662	492	492	null	427	No	62	63
TCGA-02-0456	FEMALE	No	DECEASED	DEAD	-24796	102	102	null	12	No	UNK	68
TCGA-06-0122	FEMALE	No	DECEASED	DEAD	-30967	187	8	null	null	No	85	86
TCGA-06-0124	MALE	No	DECEASED	DEAD	-24591	619	123	null	null	No	67	69
TCGA-06-0125	FEMALE	No	DECEASED	DEAD	-23343	1448	1439	797	null	No	64	68
TCGA-06-0126	MALE	No	DECEASED	DEAD	-31627	210	3	null	null	No	86	87
TCGA-06-0127	MALE	No	DECEASED	DEAD	-24502	120	108	90	null	No	68	68
TCGA-06-0128	MALE	No	DECEASED	DEAD	-24217	691	691	189	null	No	66	68
TCGA-06-0129	MALE	No	DECEASED	DEAD	-11284	1024	988	147	null	No	31	33
TCGA-06-0130	MALE	No	DECEASED	DEAD	-19811	394	320	244	null	No	54	55
TCGA-06-0133	MALE	No	DECEASED	ALIVE	-23402	435	428	78	null	No	64	64
TCGA-06-0137	FEMALE	No	DECEASED	DEAD	-23273	812	701	487	null	No	63	66
TCGA-06-0138	MALE	No	DECEASED	DEAD	-15736	737	674	394	null	No	43	45

Case ID	Gender	PRETREATMENT HISTORY	VITAL STATUS	Vital Status	DAYS TO BIRTH	DAYS TO DEATH	DAYS TO LAST FOLLOWUP	DAYS TO TUMOR PROGRESSION	DAYS TO TUMOR RECURRENCE	Secondary or Recurrent	Age at Procedure	Age at Death
TCGA-06-0139	MALE	No	DECEASED	DEAD	-14728	362	327	152	null	No	41	42
TCGA-06-0141	MALE	No	DECEASED	DEAD	-22925	313	280	145	null	No	63	64
TCGA-06-0143	MALE	No	DECEASED	DEAD	-21386	357	357	null	264	No	58	59
TCGA-06-0145	FEMALE	No	DECEASED	DEAD	-19660	71	71	null	null	No	54	54
TCGA-06-0146	FEMALE	No	DECEASED	DEAD	-12252	611	611	530	null	No	33	35
TCGA-06-0147	FEMALE	No	DECEASED	DEAD	-18742	541	508	null	null	No	52	53
TCGA-06-0148	MALE	No	DECEASED	DEAD	-27842	307	298	188	null	No	76	77
TCGA-06-0149	FEMALE	No	DECEASED	DEAD	-27315	261	238	203	null	No	75	75
TCGA-06-0152	MALE	No	DECEASED	DEAD	-24844	375	359	299	null	No	68	69
TCGA-06-0154	MALE	No	DECEASED	DEAD	-20018	423	389	207	null	No	55	56
TCGA-06-0156	MALE	No	DECEASED	DEAD	-20931	178	164	null	null	No	57	57
TCGA-06-0157	FEMALE	No	DECEASED	DEAD	-23127	97	97	null	null	No	63	63
TCGA-06-0158	MALE	No	DECEASED	DEAD	-26855	329	166	90	null	No	74	74
TCGA-06-0162	FEMALE	No	DECEASED	DEAD	-17272	104	78	null	null	No	47	48
TCGA-06-0164	MALE	No	DECEASED	DEAD	-17510	1730	1729	1428	null	No	47	52
TCGA-06-0166	MALE	No	DECEASED	DEAD	-18902	178	161	66	null	No	52	52
TCGA-06-0168	FEMALE	No	DECEASED	DEAD	-21776	598	579	461	null	No	60	62
TCGA-06-0169	MALE	No	DECEASED	DEAD	-25127	100	95	92	null	No	69	69
TCGA-06-0171	MALE	No	DECEASED	DEAD	-24085	399	396	117	null	No	66	67
TCGA-06-0173	FEMALE	No	DECEASED	DEAD	-26548	136	7	null	null	No	73	74
TCGA-06-0174	MALE	No	DECEASED	DEAD	-19824	98	67	47	null	No	54	54
TCGA-06-0175	MALE	No	DECEASED	DEAD	-25558	123	83	39	null	No	70	71
TCGA-06-0176	MALE	No	LIVING	ALIVE	-12777	null	1561	41	null	No	35	37
TCGA-06-0177	MALE	No	DECEASED	DEAD	-23498	126	60	null	null	No	65	65
TCGA-06-0178	MALE	No	LIVING	ALIVE	-14235	null	1642	192	null	No	39	42
TCGA-06-0179	MALE	No	DECEASED	DEAD	-23449	616	578	250	null	No	64	66
TCGA-06-0182	MALE	null	DECEASED	DEAD	-27963	111	77	null	null	No	77	77
TCGA-06-0184	MALE	No	LIVING	ALIVE	-23317	null	1228	null	null	No	64	66
TCGA-06-0185	MALE	No	LIVING	ALIVE	-19922	null	1126	711	null	No	54	57
TCGA-06-0187	MALE	No	DECEASED	ALIVE	-25317	828	801	null	531	No	69	70

Case ID	Gender	PRETREATMENT HISTORY	VITAL STATUS	Vital Status	DAYS TO BIRTH	DAYS TO DEATH	DAYS TO LAST FOLLOWUP	DAYS TO TUMOR PROGRESSION	DAYS TO TUMOR RECURRENCE	Secondary or Recurrent	Age at Procedure	Age at Death
TCGA-06-0188	MALE	No	LIVING	ALIVE	-26079	null	866	null	null	No	72	73
TCGA-06-0189	MALE	No	DECEASED	DEAD	-20296	469	454	null	null	No	55	56
TCGA-06-0190	MALE	No	DECEASED	DEAD	-22835	317	313	88	null	No	62	63
TCGA-06-0194	FEMALE	No	DECEASED	DEAD	-13852	142	142	125	null	No	38	38
TCGA-06-0195	MALE	No	DECEASED	DEAD	-23131	225	214	147	null	No	64	64
TCGA-06-0197	FEMALE	No	DECEASED	DEAD	-24095	168	79	47	null	No	66	67
TCGA-06-0208	FEMALE	No	DECEASED	DEAD	-19108	255	166	148	null	No	53	53
TCGA-06-0209	MALE	No	DECEASED	DEAD	-27877	231	118	null	null	No	77	77
TCGA-06-0210	FEMALE	No	DECEASED	DEAD	-26600	225	151	67	null	No	73	74
TCGA-06-0211	MALE	No	DECEASED	DEAD	-17514	360	360	53	null	No	48	49
TCGA-06-0213	FEMALE	No	DECEASED	DEAD	-20134	16	6	null	null	No	55	55
TCGA-06-0214	MALE	No	DECEASED	DEAD	-24187	457	378	48	null	No	66	67
TCGA-06-0221	MALE	No	DECEASED	DEAD	-11332	603	548	260	null	No	31	32
TCGA-06-0237	FEMALE	No	DECEASED	ALIVE	-27735	415	314	null	null	No	75	77
TCGA-06-0238	MALE	No	DECEASED	ALIVE	-17037	404	359	311	null	No	47	48
TCGA-06-0241	FEMALE	No	LIVING	ALIVE	-24101	null	455	null	196	No	66	67
TCGA-06-0394	MALE	No	DECEASED	DEAD	-18913	329	313	87	null	No	51	52
TCGA-06-0397	FEMALE	No	DECEASED	DEAD	-20997	121	15	null	null	No	57	58
TCGA-06-0402	MALE	No	DECEASED	DEAD	-26059	8	8	null	null	No	71	71
TCGA-06-0409	MALE	No	DECEASED	DEAD	-16023	2201	2145	334	null	No	44	50
TCGA-06-0410	FEMALE	No	DECEASED	DEAD	-28084	142	142	null	7	No	77	77
TCGA-06-0412	FEMALE	No	DECEASED	DEAD	-20618	291	245	130	null	No	56	57
TCGA-06-0413	FEMALE	No	DECEASED	DEAD	-28433	96	5	null	null	No	77	78
TCGA-06-0414	MALE	No	DECEASED	DEAD	-23215	1068	1065	null	1013	No	64	67
TCGA-06-0644	MALE	No	LIVING	ALIVE	-26246	null	375	null	85	No	72	73
TCGA-06-0645	FEMALE	No	DECEASED	ALIVE	-20448	175	98	null	null	No	56	57
TCGA-06-0646	MALE	No	DECEASED	ALIVE	-22272	175	136	90	null	No	61	62
TCGA-06-0648	MALE	No	DECEASED	ALIVE	-28477	297	293	201	null	No	78	79
TCGA-08-0345	FEMALE	null	DECEASED	DEAD	-25960	53	53	null	null	No	71	72
TCGA-08-0349	MALE	No	DECEASED	DEAD	-16964	298	231	92	null	No	47	48

Case ID	Gender	PRETREATMENT HISTORY	VITAL STATUS	Vital Status	DAYS TO BIRTH	DAYS TO DEATH	DAYS TO LAST FOLLOWUP	DAYS TO TUMOR PROGRESSION	DAYS TO TUMOR RECURRENCE	Secondary or Recurrent	Age at Procedure	Age at Death
TCGA-08-0352	MALE	No	DECEASED	DEAD	-29106	39	39	null	null	No	80	80
TCGA-08-0358	MALE	No	DECEASED	DEAD	-18383	678	593	null	263	No	51	52
TCGA-08-0373	MALE	No	DECEASED	DEAD	-25273	134	134	null	null	No	69	69
TCGA-08-0386	MALE	No	DECEASED	DEAD	-27053	548	478	null	427	No	74	76
TCGA-08-0509	MALE	No	DECEASED	DEAD	-23273	382	17	null	null	No	63	65
TCGA-08-0510	MALE	No	DECEASED	DEAD	-27675	129	106	null	89	No	76	76
TCGA-08-0511	MALE	No	DECEASED	DEAD	-25299	235	235	null	null	No	69	70
TCGA-08-0512	MALE	No	DECEASED	DEAD	-17821	1282	358	231	null	No	49	53
TCGA-08-0514	FEMALE	No	DECEASED	DEAD	-25457	337	259	null	null	No	70	71
TCGA-08-0516	MALE	No	DECEASED	DEAD	-5308	596	561	40	null	No	15	17
TCGA-08-0517	FEMALE	Yes	DECEASED	DEAD	-12771	1785	1770	null	1024	Rec	UNK	40
TCGA-08-0518	FEMALE	No	DECEASED	DEAD	-21995	588	39	134	null	No	60	62
TCGA-08-0520	MALE	No	DECEASED	DEAD	-25766	326	326	null	105	No	70	71
TCGA-08-0521	MALE	No	DECEASED	DEAD	-6376	146	146	125	null	No	18	18
TCGA-08-0522	MALE	No	DECEASED	DEAD	-22413	635	555	266	null	No	61	63
TCGA-08-0524	FEMALE	No	DECEASED	DEAD	-6464	220	198	61	null	No	17	18
TCGA-08-0525	MALE	Yes	DECEASED	DEAD	-19012	486	486	343	null	Sec	UNK	54
TCGA-08-0529	FEMALE	No	DECEASED	DEAD	-20556	559	526	null	328	No	56	58
TCGA-08-0531	MALE	No	DECEASED	DEAD	-23481	230	168	null	null	No	65	65
TCGA-12-0616	FEMALE	No	DECEASED	DEAD	-13451	447	438	397	null	No	37	38
TCGA-12-0618	MALE	No	DECEASED	DEAD	-18071	394	49	null	null	No	50	51
TCGA-12-0619	MALE	No	DECEASED	DEAD	-21920	1062	316	203	null	No	60	63
TCGA-12-0620	MALE	No	DECEASED	DEAD	-21068	318	181	null	null	No	58	59
TCGA-06-0132	MALE			ALIVE						No	50	51

Table S5.2.1 QC parameters and their value information from all hybridizations used in this study. Since there are two color channels (Green and red for each sample, including both tumor and reference), the qualities of each channel are accessed independently except for the DLRspread. The reproducibility values shown here are the percentage values.

Array Name	DLRs	Signal To Noise Green	Signal To Noise Red	Signal Intensity Green	Signal Intensity Red	BG Noise Green	BG Noise Red	Reproducibility Green	Reproducibility Red
TCGA-02-0001	0.21	207.2	166.3	630.5	516.2	3	3.1	14	13.2
TCGA-02-0003	0.26	201.6	157.3	527.4	422	2.6	2.7	10.7	11.9
TCGA-02-0006	0.24	126.7	100.3	657	490.7	5.2	4.9	12.7	13.8
TCGA-02-0007	0.23	220.2	163.1	539	397.8	2.4	2.4	6.9	7.2
TCGA-02-0009	0.22	154.3	117.8	684.1	524.9	4.4	4.5	9.7	10.3
TCGA-02-0010	0.26	157.6	130.3	570.1	459.3	3.6	3.5	12.2	12.4
TCGA-02-0011	0.23	100.2	82.4	648.4	534	6.5	6.5	13.2	12.6
TCGA-02-0014	0.27	121.9	94.5	552	401.4	4.5	4.2	9.4	12.8
TCGA-02-0021	0.26	158.9	122.3	585.8	436.3	3.7	3.6	9.8	11.9
TCGA-02-0024	0.24	174	136.5	515.2	377.8	3	2.8	7.3	8.3
TCGA-02-0027	0.28	178.4	124.9	558.2	350.8	3.1	2.8	9.5	9.7
TCGA-02-0028	0.2	131	109.2	584.4	587	4.5	5.4	8.3	8.9
TCGA-02-0033	0.2	212.3	178.7	591.3	606.6	2.8	3.4	9.8	10.2
TCGA-02-0034	0.2	169.5	156.6	539.4	597.9	3.2	3.8	9.4	9.9
TCGA-02-0037	0.22	196	173.3	567.1	549.1	2.9	3.2	8.6	10.2
TCGA-02-0038	0.2	150.8	125.6	680.7	657	4.5	5.2	10.9	9.9
TCGA-02-0043	0.22	33.4	34.6	590.5	615.3	17.7	17.8	14.2	12.9
TCGA-02-0046	0.22	42.3	44.4	523.6	544.2	12.4	12.3	12	12.3
TCGA-02-0047	0.19	244.2	187.9	597.6	622.8	2.4	3.3	9.5	9.1
TCGA-02-0052	0.2	203.5	149.9	658.1	628.9	3.2	4.2	8.2	8.1
TCGA-02-0054	0.18	122.9	98.2	709.9	801.2	5.8	8.2	9.1	8.9
TCGA-02-0055	0.19	232.3	178.8	623.9	645.3	2.7	3.6	10.3	10.1
TCGA-02-0057	0.19	126.1	101.3	711	715.1	5.6	7.1	9.1	9.1
TCGA-02-0058	0.23	213.2	137.6	830.4	395.3	3.9	2.9	7.9	10.6
TCGA-02-0060	0.19	208.1	154.7	796.8	479.1	3.8	3.1	8.2	10.7
TCGA-02-0064	0.19	108	136.7	747.3	597.7	6.9	4.4	8.5	9.1
TCGA-02-0069	0.19	70.2	114.7	574.3	596	8.2	5.2	8.8	7.5
TCGA-02-0071	0.18	67.3	71.2	784.5	703.4	11.7	9.9	8.7	9
TCGA-02-0074	0.18	84.7	153.8	638	674.8	7.5	4.4	9	8.8
TCGA-02-0075	0.17	91.2	116.1	656.6	656.3	7.2	5.7	5.3	5
TCGA-02-0079	0.19	39.4	63.3	108.2	239.4	2.7	3.8	14	14.8
TCGA-02-0080	0.17	88.2	91.2	713.4	754.8	8.1	8.3	9.4	9.1
TCGA-02-0083	0.21	107.2	125.3	753.7	489.7	7	3.9	8.7	10.5
TCGA-02-0084	0.19	59.7	52.2	196.9	362.7	3.3	7	12.9	12.4
TCGA-02-0085	0.17	78.1	77.7	847.8	885.2	10.9	11.4	8.8	8.2
TCGA-02-0086	0.2	108.3	143.1	662	464	6.1	3.2	8.6	11.6
TCGA-02-0087	0.2	184.6	171.2	549.4	580	3	3.4	12.7	14.2
TCGA-02-0089	0.2	76.6	94.4	735	541.2	9.6	5.7	7.7	12.2
TCGA-02-0099	0.17	90.4	135.3	767.6	797.8	8.5	5.9	10	10.5

Array Name	DLRs	Signal To Noise Green	Signal To Noise Red	Signal Intensity Green	Signal Intensity Red	BG Noise Green	BG Noise Red	Reproducibility Green	Reproducibility Red
TCGA-02-0102	0.17	97.3	108.9	766.3	784.2	7.9	7.2	8	8
TCGA-02-0106	0.2	190.9	167.2	550	502	2.9	3	7.9	8.6
TCGA-02-0107	0.17	71.5	88.6	718.6	665	10.1	7.5	5.1	4.7
TCGA-02-0111	0.2	170.6	142.7	548.6	554.6	3.2	3.9	11.2	11.3
TCGA-02-0113	0.18	82.2	98.1	719.1	782	8.8	8	12.2	11.5
TCGA-02-0114	0.18	99.3	117	755.9	784.2	7.6	6.7	9.6	9.4
TCGA-02-0115	0.17	74.9	81.4	753.5	752.3	10.1	9.2	10.8	10.2
TCGA-02-0116	0.2	80	98.9	733.3	564.6	9.2	5.7	8.9	10.9
TCGA-02-0258	0.18	186.7	190.6	549	628.6	2.9	3.3	11.2	12.2
TCGA-02-0260	0.24	176.3	144.3	700.7	647.3	4	4.5	10.3	9.1
TCGA-02-0266	0.2	100	96.2	642.2	781.9	6.4	8.1	13.6	13.3
TCGA-02-0269	0.17	171.8	164.7	628.8	707.1	3.7	4.3	8.9	9.7
TCGA-02-0271	0.2	154.9	159.2	594.2	631.7	3.8	4	13.5	13.5
TCGA-02-0281	0.24	218.7	203.1	487.4	468.4	2.2	2.3	10.7	11.4
TCGA-02-0285	0.32	233.2	201.3	514.3	460.7	2.2	2.3	12.8	14.5
TCGA-02-0289	0.2	192.7	218	518.8	738.9	2.7	3.4	11.1	11.1
TCGA-02-0290	0.22	256	239	594.5	667.2	2.3	2.8	8.2	7.7
TCGA-02-0317	0.25	223.9	193.4	614.4	599.4	2.7	3.1	12.7	12.7
TCGA-02-0321	0.17	184.8	161.6	653.3	675.9	3.5	4.2	10.2	10.4
TCGA-02-0324	0.19	109.6	129.2	434.2	533.1	4	4.1	9.5	10.2
TCGA-02-0325	0.18	115.6	112	629.3	682.1	5.4	6.1	13.8	13.8
TCGA-02-0326	0.19	179.9	173.8	638.9	653.9	3.6	3.8	13.1	12.3
TCGA-02-0330	0.17	188.8	209.9	477.2	596.3	2.5	2.8	7.3	7.3
TCGA-02-0332	0.19	192	180.8	683.6	746.6	3.6	4.1	11.4	10.6
TCGA-02-0333	0.2	184.4	146	810.3	771.9	4.4	5.3	9.6	9.7
TCGA-02-0337	0.21	116.2	109.9	587.3	688.2	5.1	6.3	8.2	9
TCGA-02-0338	0.2	120.7	101.6	688.2	690.3	5.7	6.8	14	13.9
TCGA-02-0339	0.22	199.8	171.2	678.6	553.3	3.4	3.2	11.2	11.8
TCGA-02-0422	0.17	184.6	186.7	617.7	741.8	3.3	4	7.2	7.8
TCGA-02-0430	0.18	149	149.8	577.4	625.4	3.9	4.2	11	10.9
TCGA-02-0432	0.2	199.5	207.6	578.4	588.8	2.9	2.8	10.8	10.6
TCGA-02-0439	0.16	128.8	127.9	526.6	570.8	4.1	4.5	8.2	7.7
TCGA-02-0440	0.19	125	111.7	607.8	690.5	4.9	6.2	12.7	10.5
TCGA-02-0446	0.18	104.7	103.5	586.2	635.5	5.6	6.1	7.8	8.6
TCGA-02-0451	0.21	217	193.6	614.3	604.5	2.8	3.1	14.2	14.1
TCGA-02-0456	0.23	217.8	193.4	578.7	562.5	2.7	2.9	11.4	11.3
TCGA-06-0122	0.23	265.9	163.1	806.7	508.6	3	3.1	7	8.7
TCGA-06-0124	0.22	208.6	135.6	834.1	457.2	4	3.4	10	11.7
TCGA-06-0125	0.26	200	126.6	982.7	505.9	4.9	4	8.4	9.6
TCGA-06-0126	0.25	286.5	162.5	872.9	432.9	3	2.7	9.2	13.2
TCGA-06-0127	0.19	42	45	143.5	370	3.4	8.2	11.1	11.3
TCGA-06-0128	0.2	231.1	178.5	774	595.3	3.3	3.3	8.7	10.1
TCGA-06-0129	0.22	286.4	176.6	766.1	630.2	2.7	3.6	8.1	9.1
TCGA-06-0130	0.22	379.6	205.6	785.8	534.3	2.1	2.6	12.7	13.3
TCGA-06-0132	0.21	202.4	202.5	654.7	534.5	3.2	2.6	11.1	11
TCGA-06-0133	0.22	264.5	151.3	829.8	424	3.1	2.8	8.9	13.3
TCGA-06-0137	0.24	283.3	154.7	846.7	422.1	3	2.7	9.7	12.9

Array Name	DLRs	Signal To Noise Green	Signal To Noise Red	Signal Intensity Green	Signal Intensity Red	BG Noise Green	BG Noise Red	Reproducibility Green	Reproducibility Red
TCGA-06-0138	0.25	271.4	59	379.2	70.5	1.4	1.2	8.9	12.4
TCGA-06-0139	0.23	346.8	233.6	704.6	611.3	2	2.6	9.2	8.4
TCGA-06-0141	0.25	258.2	169.7	713.5	493.8	2.8	2.9	12.8	15.7
TCGA-06-0143	0.2	116.7	90.3	540.6	460.7	4.6	5.1	12.4	13
TCGA-06-0145	0.23	305.8	182.3	698.9	467.5	2.3	2.6	9.7	9.9
TCGA-06-0146	0.15	217.6	248	590.4	698	2.7	2.8	10.1	9.8
TCGA-06-0147	0.25	307.8	152.2	800.2	385.4	2.6	2.5	9.6	12.7
TCGA-06-0148	0.23	189.9	127.9	876.9	483	4.6	3.8	8.2	7.9
TCGA-06-0149	0.16	239	268.7	443.7	561.4	1.9	2.1	8.6	8.8
TCGA-06-0152	0.19	70.2	57.9	189.3	319.1	2.7	5.5	10.2	10.6
TCGA-06-0154	0.24	235	159.5	695.1	340.7	3	2.1	9.9	12.7
TCGA-06-0156	0.22	162.8	128.5	735.5	569.2	4.5	4.4	10.7	12.3
TCGA-06-0157	0.18	118.7	105.4	797.7	995.1	6.7	9.4	9.7	10.3
TCGA-06-0158	0.19	199.9	161.9	617.3	581.1	3.1	3.6	8.1	9.1
TCGA-06-0162	0.19	197.9	206.3	529.6	648.2	2.7	3.1	8.6	9.3
TCGA-06-0164	0.16	168.3	179.3	532.9	616.9	3.2	3.4	9	8.5
TCGA-06-0166	0.21	150.5	120.4	842.7	759.5	5.6	6.3	10.6	11.9
TCGA-06-0168	0.17	180.4	145.5	810.5	914.1	4.5	6.3	8.8	9.4
TCGA-06-0169	0.24	190.7	135.9	730	506.2	3.8	3.7	9.6	9.6
TCGA-06-0171	0.2	240.4	210.4	585.4	555.6	2.4	2.6	9.1	9.6
TCGA-06-0173	0.21	290.1	221.8	582.5	520.3	2	2.3	9.1	12
TCGA-06-0174	0.24	210.6	109.4	680.2	228.6	3.2	2.1	9.7	10.9
TCGA-06-0175	0.15	164.2	154.1	644.7	704.3	3.9	4.6	11.5	10.1
TCGA-06-0176	0.23	186.9	150.7	940.6	780.7	5	5.2	9	10.9
TCGA-06-0177	0.22	117.5	118.5	610	745.3	5.2	6.3	10.1	11.1
TCGA-06-0178	0.18	165.2	165.4	922.7	1079.5	5.6	6.5	10.9	11.9
TCGA-06-0179	0.2	173.8	143.9	459.2	432.7	2.6	3	8.2	8.8
TCGA-06-0182	0.21	234.1	207.7	474.3	459.2	2	2.2	12.8	12.1
TCGA-06-0184	0.2	119.9	100.7	935.3	808.3	7.8	8	9.9	10.1
TCGA-06-0185	0.19	237.2	225.8	570	632	2.4	2.8	10.8	10.4
TCGA-06-0187	0.22	166.4	177.3	566	519.2	3.4	2.9	8.7	11.4
TCGA-06-0188	0.26	322.4	87.2	635	107.7	2	1.2	9.4	15.1
TCGA-06-0189	0.23	171.3	121.6	421.9	334.4	2.5	2.7	12.7	12.9
TCGA-06-0190	0.21	250	193.2	740.3	636	3	3.3	10	11.4
TCGA-06-0194	0.16	205.2	236.5	587.1	752.6	2.9	3.2	10.2	11
TCGA-06-0195	0.17	176.4	149.9	905.9	1014.6	5.1	6.8	12.2	13
TCGA-06-0197	0.18	316.3	252.1	757.3	911.2	2.4	3.6	11.9	12.6
TCGA-06-0208	0.21	180.6	140.6	880.6	748.9	4.9	5.3	10.6	11.4
TCGA-06-0209	0.2	163.6	154.6	415.3	469.8	2.5	3	11.2	11.3
TCGA-06-0210	0.19	126.4	106.3	457.4	482.1	3.6	4.5	12.6	13.1
TCGA-06-0211	0.22	127.9	168	743.4	599.4	5.8	3.6	10.6	9
TCGA-06-0213	0.17	135.8	111.3	938.4	1080.3	6.9	9.7	9	9.5
TCGA-06-0214	0.2	133.1	113.1	873.3	611.1	6.6	5.4	11.7	12.5
TCGA-06-0221	0.23	272.4	149.2	716.3	272.7	2.6	1.8	7.7	12.2
TCGA-06-0237	0.2	168.8	134.2	351.2	315.9	2.1	2.4	13.2	13
TCGA-06-0238	0.17	63.8	72.7	194.5	359.1	3	4.9	9.8	9.7
TCGA-06-0241	0.18	192.7	149.5	868.9	922.9	4.5	6.2	10.1	9.5

Array Name	DLRs	Signal To Noise Green	Signal To Noise Red	Signal Intensity Green	Signal Intensity Red	BG Noise Green	BG Noise Red	Reproducibility Green	Reproducibility Red
TCGA-06-0394	0.15	289.7	285.6	510.1	581.5	1.8	2	11.5	11.4
TCGA-06-0397	0.17	161.8	143.4	692.7	675.7	4.3	4.7	11.9	11.2
TCGA-06-0402	0.17	218.4	212.8	538.5	645.2	2.5	3	11.2	10.9
TCGA-06-0409	0.17	112.4	118.8	628.6	692.4	5.6	5.8	10.1	8.8
TCGA-06-0410	0.2	171.3	165.4	532.3	591.8	3.1	3.6	11.2	10
TCGA-06-0412	0.2	208.9	180.7	583.1	624.1	2.8	3.5	11.9	11.9
TCGA-06-0413	0.21	197.9	172.4	554.4	490	2.8	2.8	11.3	12.6
TCGA-06-0414	0.17	240.9	250.6	550.8	657	2.3	2.6	9.5	9.8
TCGA-06-0644	0.2	56.6	58.4	164	356.5	2.9	6.1	12.7	10.8
TCGA-06-0645	0.17	64.1	71.8	163.6	297.5	2.6	4.1	11.9	11.2
TCGA-06-0646	0.19	48.4	102.5	168.5	281	3.5	2.7	11.8	11.9
TCGA-06-0648	0.17	54.7	67.2	180.6	385.8	3.3	5.7	10	8.8
TCGA-08-0345	0.21	49.5	40.9	177.2	334.5	3.6	8.2	7.8	7.7
TCGA-08-0349	0.16	73.7	74.4	201.8	337.2	2.7	4.5	10.2	10.3
TCGA-08-0352	0.15	70.9	79	179.9	324.2	2.5	4.1	9.9	11
TCGA-08-0358	0.12	80.9	87.4	205.5	307.4	2.5	3.5	9.9	11
TCGA-08-0373	0.22	40.3	72.1	106.2	232.7	2.6	3.2	12.6	14.5
TCGA-08-0386	0.13	57.1	57	247	338	4.3	5.9	11.4	11.7
TCGA-08-0509	0.15	200.1	203.9	546.9	654.7	2.7	3.2	9.2	8.9
TCGA-08-0510	0.19	173.3	161.2	680.9	712.7	3.9	4.4	11.4	11.2
TCGA-08-0511	0.17	140	128.8	564.1	696.3	4	5.4	10.3	9.2
TCGA-08-0512	0.16	162.9	170.3	558.7	761.7	3.4	4.5	11.4	10.9
TCGA-08-0514	0.19	75.8	69.5	674.1	722.2	8.9	10.4	11.5	11.5
TCGA-08-0516	0.18	348.1	320.9	554.4	555	1.6	1.7	13.6	13.8
TCGA-08-0517	0.19	141.6	119.9	697.9	632.2	4.9	5.3	12	11.8
TCGA-08-0518	0.15	177.8	191.4	535.3	683.9	3	3.6	9	8
TCGA-08-0520	0.18	303.4	299.4	625.4	695.3	2.1	2.3	10.6	9.4
TCGA-08-0521	0.2	267.6	246.1	623.5	636.2	2.3	2.6	10.3	9.7
TCGA-08-0522	0.23	244.8	186.8	724.4	649.4	3	3.5	8	7.7
TCGA-08-0524	0.19	125.4	128.9	613.2	680.3	4.9	5.3	9.7	9.3
TCGA-08-0525	0.18	167.7	165	674.6	710.8	4	4.3	13	12.4
TCGA-08-0529	0.22	206.9	184.7	404.6	380.1	2	2.1	10	10
TCGA-08-0531	0.19	266.2	243.4	611.5	552.3	2.3	2.3	12.9	10.6
TCGA-12-0616	0.2	53	48.1	251.1	511.3	4.7	10.6	10.5	9.4
TCGA-12-0618	0.18	67.8	62	183.4	298.7	2.7	4.8	10.6	11.1
TCGA-12-0619	0.18	64.8	69.3	170.6	323.3	2.6	4.7	11	10
TCGA-12-0620	0.2	77.6	57.8	205.8	330	2.7	5.7	9.2	9.8

Table S5.5.1 Segment numbers before and after merging processes for both CBS and HHMM segmentation methods. Both CBS merged and HHMM merged represent numbers after merging while as CBS and HHMM represent numbers before the merging.

probe	CBS	CBS Merged	HHMM	HHMM Merged
TCGA-02-0085	163	161	3387	3336
TCGA-02-0014	294	245	3060	2729
TCGA-06-0210	179	166	1290	1135
TCGA-02-0338	161	151	3983	3121
TCGA-06-0132	119	113	600	512
TCGA-02-0332	220	208	2291	1899
TCGA-06-0164	243	202	995	906
TCGA-06-0168	220	205	1053	1006
TCGA-06-0145	220	210	527	224
TCGA-02-0330	273	259	1634	1470
TCGA-06-0133	144	138	1925	1845
TCGA-02-0001	163	152	1124	1061
TCGA-08-0349	183	177	6268	5860
TCGA-02-0317	128	123	3202	3084
TCGA-06-0171	187	159	581	506
TCGA-06-0182	284	251	876	593
TCGA-06-0166	271	239	2270	1719
TCGA-02-0446	159	147	2279	1753
TCGA-06-0402	432	380	553	430
TCGA-06-0238	236	234	9595	9344
TCGA-02-0266	406	344	2751	2366
TCGA-06-0154	147	141	2997	2091
TCGA-02-0033	115	111	1181	1128
TCGA-02-0281	278	265	3413	3153
TCGA-02-0052	134	130	1535	1305
TCGA-06-0156	223	182	1107	683
TCGA-08-0529	245	237	973	843
TCGA-08-0522	141	136	1220	860
TCGA-02-0113	175	166	3932	2944
TCGA-02-0069	332	281	2619	2061
TCGA-02-0074	187	177	3476	3171
TCGA-08-0514	330	309	2821	2145
TCGA-02-0087	183	163	1628	1588
TCGA-06-0162	351	316	1952	1368
TCGA-06-0129	934	781	2889	2242
TCGA-06-0195	169	160	1526	1481
TCGA-02-0034	194	185	1304	1156
TCGA-02-0084	198	197	9592	8706
TCGA-06-0158	145	136	1364	1342
TCGA-08-0373	177	173	8999	8789
TCGA-06-0412	176	171	1361	1219
TCGA-02-0055	175	162	1084	1055
TCGA-02-0064	149	129	2645	2010
TCGA-02-0324	238	235	3141	2848
TCGA-02-0439	247	226	2232	1999
TCGA-02-0006	232	217	1512	1333
TCGA-02-0339	236	218	1450	1389
TCGA-06-0141	394	364	2294	1637
TCGA-08-0345	282	280	8280	8116
TCGA-06-0179	152	139	1026	952
TCGA-02-0071	150	144	1648	1536
TCGA-12-0620	416	409	10979	10254
TCGA-08-0531	185	167	758	668
TCGA-08-0510	237	195	1582	1347
TCGA-02-0432	165	130	1423	1133
TCGA-08-0516	206	194	722	686
TCGA-06-0127	207	207	18063	17867
TCGA-02-0037	145	134	1869	1627
TCGA-06-0126	119	113	3029	2044
TCGA-08-0386	291	272	3166	3063
TCGA-02-0028	167	166	1879	1867
TCGA-02-0083	208	176	4954	4151
TCGA-02-0116	111	108	3449	2971
TCGA-06-0175	300	278	1523	1155
TCGA-02-0011	159	140	3319	3218
TCGA-02-0038	209	205	1536	1317
TCGA-06-0648	283	283	10903	10486
TCGA-02-0054	167	155	1076	1044
TCGA-02-0440	166	164	1405	1282
TCGA-02-0456	145	137	2993	2755
TCGA-06-0148	196	177	1147	874
TCGA-06-0194	198	191	1265	1222
TCGA-02-0325	289	246	1993	1640
TCGA-06-0130	150	138	644	588
TCGA-06-0125	192	180	2151	1637
TCGA-02-0115	197	181	2784	2524
TCGA-02-0009	257	241	2178	1691
TCGA-06-0122	177	153	1300	1266
TCGA-02-0080	269	216	2095	1610
TCGA-02-0337	127	122	2587	2492

probe	CBS	CBS Merged	HHMM	HHMM Merged	probe	CBS	CBS Merged	HHMM	HHMM Merged
TCGA-08-0525	213	192	1156	978	TCGA-02-0422	235	208	1215	1166
TCGA-02-0269	408	334	959	710	TCGA-02-0285	119	117	2765	2350
TCGA-02-0099	236	212	2583	1687	TCGA-08-0352	284	278	5494	4850
TCGA-02-0089	138	127	2150	1924	TCGA-06-0176	147	143	2649	2208
TCGA-06-0187	160	148	1714	1056	TCGA-06-0147	125	117	2175	1634
TCGA-02-0058	148	120	1205	894	TCGA-02-0114	257	238	4835	4608
TCGA-02-0111	154	144	870	688	TCGA-06-0146	510	460	1494	1322
TCGA-06-0209	177	161	1270	1074	TCGA-02-0075	147	142	3491	2963
TCGA-06-0185	188	173	761	698	TCGA-06-0177	255	254	6457	5811
TCGA-06-0178	108	106	2151	2096	TCGA-08-0521	148	145	2533	2359
TCGA-06-0197	167	164	1390	1179	TCGA-06-0169	128	120	1792	1676
TCGA-08-0518	296	274	1706	1113	TCGA-02-0102	141	134	1183	931
TCGA-06-0644	184	176	12523	12341	TCGA-02-0260	232	213	3211	2822
TCGA-08-0517	305	267	853	730	TCGA-06-0137	191	152	1244	1085
TCGA-06-0128	194	170	2189	1986	TCGA-06-0397	195	181	1787	1696
TCGA-02-0010	275	224	1278	1202	TCGA-06-0237	262	234	1091	1044
TCGA-08-0511	235	219	1222	1166	TCGA-02-0046	162	152	2762	2486
TCGA-06-0645	206	203	10923	10692	TCGA-06-0409	159	154	879	817
TCGA-02-0024	215	171	898	865	TCGA-06-0124	150	139	1213	1085
TCGA-06-0241	287	274	2455	2096	TCGA-02-0326	169	159	1356	1130
TCGA-06-0410	245	236	2817	2726	TCGA-06-0174	119	104	1909	1835
TCGA-06-0213	225	199	2121	1662	TCGA-02-0271	135	126	2380	2092
TCGA-02-0003	125	117	929	842	TCGA-06-0184	205	188	832	795
TCGA-02-0107	112	110	1741	1468	TCGA-08-0358	459	422	3100	2805
TCGA-06-0190	165	157	1136	957	TCGA-02-0057	146	134	3056	2603
TCGA-12-0619	253	246	10186	9357	TCGA-06-0189	103	101	1930	1892
TCGA-06-0221	207	177	1295	772	TCGA-06-0149	223	215	1066	913
TCGA-06-0188	114	104	2690	2263	TCGA-06-0157	193	187	1743	1392
TCGA-06-0211	168	156	1140	689	TCGA-02-0289	164	164	4172	3450
TCGA-02-0007	175	159	1984	1626	TCGA-02-0079	283	269	7998	7711
TCGA-08-0512	149	145	1515	1291	TCGA-06-0208	227	212	1301	782
TCGA-02-0027	131	125	2957	2906	TCGA-02-0430	263	236	1678	1518
TCGA-02-0047	131	124	1149	1077	TCGA-02-0321	236	223	542	469
TCGA-02-0106	246	205	1133	932	TCGA-06-0414	154	149	1745	1608
TCGA-02-0021	193	186	2367	2287	TCGA-06-0214	259	236	1162	890
TCGA-06-0394	180	161	1191	1103	TCGA-08-0524	274	241	1398	1266
TCGA-12-0618	166	165	10075	8880	TCGA-02-0258	383	294	1186	1077
TCGA-12-0616	298	294	16256	15933	TCGA-02-0333	311	293	2191	1747
TCGA-08-0509	180	180	920	782	TCGA-06-0173	243	232	1539	1295
TCGA-06-0143	227	219	1828	1625	TCGA-02-0086	134	128	2255	2053
TCGA-06-0139	420	350	1508	1129	TCGA-02-0290	149	148	6011	4658
TCGA-06-0138	115	97	1128	1078	TCGA-06-0152	249	246	12700	12581
TCGA-06-0413	178	177	1762	1711	TCGA-08-0520	206	193	568	473
TCGA-02-0060	304	260	1620	1536	TCGA-06-0646	248	244	10156	9754
TCGA-02-0451	251	231	2299	1950	TCGA-02-0043	209	199	2184	2112

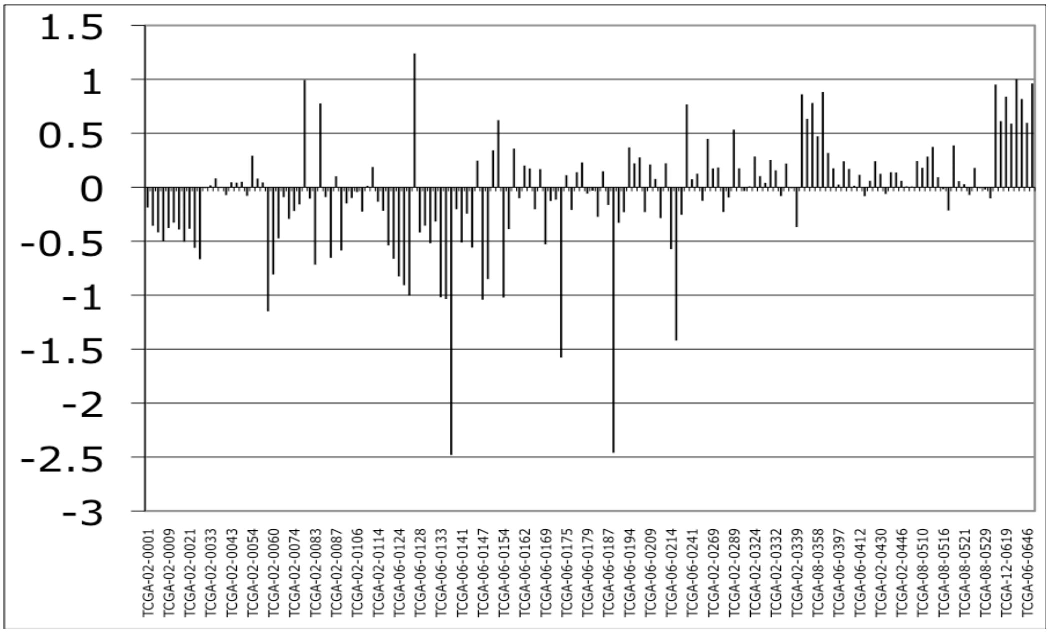


Figure S5.6.1 Log_2(Ratio) values of the modes of the density plots from all 170 patients samples (HHMM segmentation method). For most of the hybridizations, the mode of the density plot shifts quite away from zero value and need to be adjusted accordingly.

Table S6.1.1 Single probe segments with absolute predicted log$_2$(Ratio) great than one identified by using HHMM segmentation algorithm.

sample	chr	start	end	segval	genes
TCGA-06-0124	1	149390893	149390952	-2.47	
TCGA-02-0058	1	245054641	245054700	-2.27	
TCGA-12-0619	1	175222626	175222685	-2.14	C1orf49
TCGA-02-0028	1	5548432	5548490	-1.31	
TCGA-02-0028	1	14971358	14971417	-1.31	KIAA1026
TCGA-02-0028	1	37554771	37554830	-1.31	
TCGA-02-0028	1	50942883	50942942	-1.31	FAF1
TCGA-02-0028	1	104018651	104018710	-1.31	
TCGA-02-0028	1	120923696	120923755	-1.31	AK128714
TCGA-02-0028	1	177906433	177906492	-1.31	
TCGA-06-0124	1	149394958	149395017	-1.28	
TCGA-06-0124	1	149579637	149579696	-1.28	
TCGA-06-0124	1	197596962	197597021	-1.28	C1orf106
TCGA-06-0124	1	197982932	197982991	-1.28	
TCGA-12-0618	1	796956	797005	-1.18	
TCGA-12-0618	1	1116011	1116057	-1.18	BC028014
TCGA-12-0618	1	1161026	1161070	-1.18	TTLL10
TCGA-12-0618	1	1201344	1201388	-1.18	SDF4
TCGA-12-0618	1	1285365	1285409	-1.18	PUSL1
TCGA-12-0618	1	19410882	19410937	-1.18	CAPZB
TCGA-12-0618	1	22001371	22001430	-1.18	HSPG2
TCGA-12-0618	1	33988075	33988134	-1.18	CSMD2
TCGA-12-0618	1	59618935	59618994	-1.18	FLJ10986
TCGA-12-0618	1	112964625	112964684	-1.18	AF338193
TCGA-12-0618	1	147836482	147836541	-1.18	C1orf56
TCGA-12-0618	1	198567488	198567547	-1.18	IPO9
TCGA-12-0618	1	199154489	199154548	-1.18	PPP1R12B
TCGA-12-0618	1	219503635	219503694	-1.18	DISP1
TCGA-06-0152	1	796956	797005	-1.03	
TCGA-06-0152	1	1026414	1026458	-1.03	AGRN
TCGA-06-0152	1	1161026	1161070	-1.03	TTLL10
TCGA-06-0152	1	1201344	1201388	-1.03	SDF4
TCGA-06-0152	1	1285365	1285409	-1.03	PUSL1
TCGA-06-0152	1	2614922	2614970	-1.03	
TCGA-06-0152	1	19410882	19410937	-1.03	CAPZB
TCGA-06-0152	1	26288909	26288954	-1.03	CCDC21
TCGA-06-0152	1	30687222	30687281	-1.03	
TCGA-06-0152	1	33988075	33988134	-1.03	CSMD2
TCGA-06-0152	1	46410804	46410851	-1.03	C1orf190
TCGA-06-0152	1	59618935	59618994	-1.03	FLJ10986
TCGA-06-0152	1	72978376	72978435	-1.03	
TCGA-06-0152	1	112964625	112964684	-1.03	AF338193
TCGA-06-0152	1	147836482	147836541	-1.03	C1orf56
TCGA-02-0451	10	32885520	32885579	-1.28	CCDC7
TCGA-02-0451	10	66951636	66951695	-1.28	
TCGA-02-0451	10	122203932	122203991	-1.28	
TCGA-08-0509	10	7480070	7480129	-1.26	SFMBT2
TCGA-08-0509	10	55814784	55814843	-1.26	PCDH15
TCGA-08-0509	10	81766470	81766529	-1.26	
TCGA-08-0509	10	97984467	97984526	-1.26	BLNK
TCGA-06-0188	10	7739585	7739644	-1.24	ITIH5
TCGA-06-0188	10	14709241	14709300	-1.24	FAM107B
TCGA-06-0188	10	18548400	18548459	-1.24	CACNB2
TCGA-06-0188	10	70334318	70334377	-1.24	DDX50
TCGA-02-0089	10	13517548	13517607	-1.22	
TCGA-02-0089	10	90985518	90985577	-1.22	LIPA
TCGA-02-0089	10	130726959	130727018	-1.22	
TCGA-06-0648	10	1058642	1058692	-1.20	IDI2
TCGA-06-0648	10	3167933	3167992	-1.20	PFKP
TCGA-06-0648	10	12437331	12437390	-1.20	CAMK1D
TCGA-06-0648	10	26332271	26332330	-1.20	MYO3A
TCGA-06-0648	10	44193364	44193409	-1.20	CXCL12
TCGA-06-0648	10	45534651	45534696	-1.20	
TCGA-06-0648	10	49955253	49955306	-1.20	C10orf72
TCGA-06-0648	10	69627248	69627307	-1.20	FLJ14437
TCGA-06-0648	10	89199763	89199822	-1.20	
TCGA-06-0648	10	89615644	89615703	-1.20	PTEN
TCGA-06-0648	10	90291257	90291316	-1.20	C10orf59
TCGA-06-0648	10	90651370	90651429	-1.20	STAMBPL1
TCGA-06-0648	10	90685095	90685154	-1.20	ACTA2
TCGA-06-0648	10	99324268	99324325	-1.20	ANKRD2
TCGA-06-0648	10	101917068	101917127	-1.20	SPFH1
TCGA-06-0648	10	103517472	103517526	-1.20	
TCGA-06-0648	10	104147671	104147727	-1.20	NFKB2
TCGA-06-0648	10	105290373	105290432	-1.20	NEURL
TCGA-06-0648	10	132876794	132876853	-1.20	TCERG1L
TCGA-02-0006	10	66951636	66951695	-1.19	
TCGA-02-0006	10	70947915	70947974	-1.19	
TCGA-02-0006	10	89256027	89256086	-1.19	MINPP1
TCGA-02-0006	10	96842251	96842310	-1.19	
TCGA-06-0152	10	1386766	1386810	-1.16	ADARB2
TCGA-06-0152	10	3899933	3899992	-1.16	
TCGA-06-0152	10	7739585	7739644	-1.16	ITIH5
TCGA-06-0152	10	11674768	11674827	-1.16	D13644
TCGA-06-0152	10	11953319	11953378	-1.16	C10orf47
TCGA-06-0152	10	15249136	15249195	-1.16	NMT2

sample	chr	start	end	segval	genes
TCGA-06-0152	1	176128016	176128075	-1.03	C1orf125
TCGA-06-0152	1	219503635	219503694	-1.03	DISP1
TCGA-06-0152	1	233380409	233380468	-1.03	MTR
TCGA-02-0057	1	85468110	85468169	1.01	AK125723
TCGA-02-0057	1	119942294	119942353	1.01	
TCGA-02-0057	1	175227043	175227102	1.01	AK097518
TCGA-02-0057	1	230972236	230972295	1.01	BC040195
TCGA-02-0113	1	56126922	56126981	1.36	
TCGA-02-0113	1	72489681	72489740	1.36	
TCGA-02-0113	1	222038023	222038082	1.36	ENAH
TCGA-02-0071	1	72489681	72489740	1.50	
TCGA-02-0071	1	199904886	199904945	1.50	X75546
TCGA-08-0529	2	94950913	94950972	-1.16	
TCGA-08-0529	2	102195099	102195158	-1.16	
TCGA-08-0529	2	112717800	112717859	-1.16	ZC3H8
TCGA-08-0529	2	186437582	186437641	-1.16	FSIP2
TCGA-08-0529	2	200138001	200138060	-1.16	SATB2
TCGA-08-0529	2	231068830	231068889	-1.16	LOC93349
TCGA-08-0529	2	241125726	241125781	-1.16	GPC1
TCGA-06-0169	2	27432096	27432155	-1.06	TRIM54
TCGA-06-0169	2	99562517	99562576	-1.06	REV1L
TCGA-06-0169	2	112308318	112308377	-1.06	ANAPC1
TCGA-06-0169	2	131928689	131928748	-1.06	PLEKHB2
TCGA-06-0169	2	200138001	200138060	-1.06	SATB2
TCGA-02-0113	2	10886580	10886639	1.00	PDIA6
TCGA-02-0113	2	15211744	15211803	1.00	
TCGA-02-0113	2	26578752	26578811	1.00	MGC16372
TCGA-02-0113	2	30017514	30017573	1.00	ALK
TCGA-02-0113	2	32630298	32630357	1.00	BIRC6
TCGA-02-0113	2	43908309	43908368	1.00	AK096400
TCGA-02-0113	2	71396419	71396478	1.00	
TCGA-02-0113	2	89937411	89937470	1.00	AY942022
TCGA-02-0113	2	107379595	107379654	1.00	
TCGA-02-0113	2	113216035	113216094	1.00	CKAP2L
TCGA-02-0113	2	163200026	163200085	1.00	KCNH7
TCGA-02-0113	2	219879591	219879650	1.00	
TCGA-02-0113	2	240659146	240659205	1.00	NDUFA10
TCGA-02-0071	2	114719726	114719785	1.02	
TCGA-02-0071	2	169668054	169668113	1.02	ABCB11
TCGA-02-0071	2	205581237	205581296	1.02	ALS2CR19
TCGA-02-0071	2	210647216	210647275	1.02	AB058746
TCGA-02-0071	2	240659146	240659205	1.02	NDUFA10
TCGA-02-0071	2	241724916	241724971	1.02	AK074062
TCGA-02-0046	2	114719726	114719785	1.73	
TCGA-02-0046	2	182746745	182746804	1.73	CR600208
TCGA-02-0069	3	9896427	9896486	-1.50	AF230335
TCGA-06-0152	10	26030462	26030521	-1.16	AK096400
TCGA-06-0152	10	29704220	29704279	-1.16	
TCGA-06-0152	10	34438690	34438749	-1.16	AF196185
TCGA-06-0152	10	39018805	39018864	-1.16	
TCGA-06-0152	10	44193364	44193409	-1.16	CXCL12
TCGA-06-0152	10	49340053	49340111	-1.16	ARHGAP22
TCGA-06-0152	10	50178065	50178124	-1.16	C10orf71
TCGA-06-0152	10	52945287	52945346	-1.16	PRKG1
TCGA-06-0152	10	56071056	56071115	-1.16	PCDH15
TCGA-06-0152	10	59122592	59122651	-1.16	
TCGA-06-0152	10	63382145	63382204	-1.16	ARID5B
TCGA-06-0152	10	69627248	69627307	-1.16	FLJ14437
TCGA-06-0152	10	71541910	71541969	-1.16	H2AFY2
TCGA-06-0152	10	72029378	72029434	-1.16	PRF1
TCGA-06-0152	10	72324240	72324293	-1.16	
TCGA-06-0152	10	72351591	72351650	-1.16	
TCGA-06-0152	10	74539893	74539952	-1.16	
TCGA-06-0152	10	80668229	80668288	-1.16	RAI17
TCGA-06-0152	10	98350765	98350824	-1.16	PIK3AP1
TCGA-06-0152	10	99369370	99369421	-1.16	C10orf83
TCGA-06-0152	10	101284345	101284404	-1.16	NKX2-3
TCGA-06-0152	10	101591744	101591803	-1.16	ABCC2
TCGA-06-0152	10	102737790	102737837	-1.16	PEO1
TCGA-06-0152	10	103517472	103517526	-1.16	
TCGA-06-0152	10	103588026	103588085	-1.16	KCNIP2
TCGA-06-0152	10	104223603	104223647	-1.16	C10orf77
TCGA-06-0152	10	104425876	104425935	-1.16	ARL3
TCGA-06-0152	10	105040065	105040120	-1.16	AB209785
TCGA-06-0152	10	105290373	105290432	-1.16	NEURL
TCGA-06-0152	10	107369433	107369492	-1.16	
TCGA-06-0152	10	117346161	117346220	-1.16	ATRNL1
TCGA-06-0152	10	120447353	120447412	-1.16	C10orf46
TCGA-06-0152	10	122838691	122838750	-1.16	AB030073
TCGA-06-0152	10	125402689	125402748	-1.16	
TCGA-06-0152	10	125597203	125597262	-1.16	CPXM2
TCGA-06-0152	10	131317853	131317912	-1.16	MGMT
TCGA-06-0152	10	132876794	132876853	-1.16	TCERG1L
TCGA-06-0152	10	134972316	134972361	-1.16	ADAM8
TCGA-06-0645	10	1058642	1058692	-1.07	IDI2
TCGA-06-0645	10	1386766	1386810	-1.07	ADARB2
TCGA-06-0645	10	12437331	12437390	-1.07	CAMK1D
TCGA-06-0645	10	18949751	18949810	-1.07	NSUN6
TCGA-06-0645	10	26030462	26030521	-1.07	AK096400
TCGA-06-0645	10	32821797	32821856	-1.07	CCDC7
TCGA-06-0645	10	34430998	34431057	-1.07	
TCGA-06-0645	10	44987093	44987152	-1.07	AK098688

sample	chr	start	end	segval	genes
TCGA-02-0069	3	76799534	76799593	-1.50	AJ012503
TCGA-02-0038	3	166528671	166528730	-1.22	
TCGA-02-0038	3	167881635	167881694	-1.22	
TCGA-02-0055	3	191208239	191208298	-1.17	LEPREL1
TCGA-02-0271	3	6006012	6006071	-1.12	
TCGA-02-0271	3	36193272	36193331	-1.12	
TCGA-02-0271	3	64325557	64325616	-1.12	
TCGA-02-0271	3	111465930	111465989	-1.12	
TCGA-02-0271	3	169460868	169460927	-1.12	
TCGA-02-0271	3	175704921	175704980	-1.12	AJ607399
TCGA-02-0271	3	185573293	185573340	-1.12	THPO
TCGA-02-0271	3	190368566	190368625	-1.12	BX641108
TCGA-06-0122	3	88797733	88797792	-1.07	M19503
TCGA-02-0440	3	47209546	47209605	1.05	AK096853
TCGA-02-0440	3	99534264	99534323	1.05	BC104637
TCGA-02-0440	3	196258410	196258469	1.05	
TCGA-08-0345	3	33514508	33514567	1.19	CLASP2
TCGA-08-0345	3	48705552	48705611	1.19	IHPK2
TCGA-08-0345	3	49370234	49370284	1.19	GPX1
TCGA-08-0345	3	56669849	56669908	1.19	C3orf63
TCGA-08-0345	3	113831725	113831784	1.19	CCDC80
TCGA-08-0345	3	115523069	115523128	1.19	
TCGA-08-0345	3	128769344	128769403	1.19	
TCGA-08-0345	3	140574757	140574816	1.19	COPB2
TCGA-08-0345	3	186434695	186434754	1.19	EHHADH
TCGA-08-0345	3	187117470	187117529	1.19	SFRS10
TCGA-08-0345	3	188918925	188918984	1.19	BC045669
TCGA-08-0345	3	188936799	188936857	1.19	BCL6
TCGA-06-0646	3	50536585	50536644	1.24	
TCGA-06-0646	3	99893658	99893717	1.24	
TCGA-06-0238	3	151165586	151165645	1.25	PFN2
TCGA-02-0439	3	163987465	163987524	1.58	BC019327
TCGA-06-0148	4	41587341	41587400	-1.57	PHOX2B
TCGA-06-0148	4	83995737	83995796	-1.57	SCD5
TCGA-06-0148	4	173804227	173804286	-1.57	GALNT17
TCGA-06-0414	4	65081008	65081067	-1.42	SRD5A2L2
TCGA-06-0176	4	6519210	6519269	-1.06	PPP2R2C
TCGA-06-0176	4	142355308	142355367	-1.06	RNF150
TCGA-06-0176	4	154676070	154676129	-1.06	MND1
TCGA-08-0358	4	2969303	2969362	1.01	TETRAN
TCGA-08-0358	4	78326484	78326543	1.01	CCNI
TCGA-08-0358	4	88751976	88752035	1.01	SPARCL1
TCGA-08-0358	4	173814860	173814919	1.01	GALNT17
TCGA-02-0084	4	78326484	78326543	1.02	CCNI
TCGA-02-0084	4	83908026	83908085	1.02	SCD5
TCGA-02-0084	4	174684198	174684257	1.02	SCRG1
TCGA-06-0169	4	4314081	4314140	1.13	OTOP1

sample	chr	start	end	segval	genes
TCGA-06-0645	10	45468415	45468474	-1.07	ANUBL1
TCGA-06-0645	10	49955253	49955306	-1.07	C10orf72
TCGA-06-0645	10	50178065	50178124	-1.07	C10orf71
TCGA-06-0645	10	50485951	50486010	-1.07	
TCGA-06-0645	10	69627248	69627307	-1.07	FLJ14437
TCGA-06-0645	10	72029378	72029434	-1.07	PRF1
TCGA-06-0645	10	72324240	72324293	-1.07	
TCGA-06-0645	10	74692636	74692695	-1.07	TTC18
TCGA-06-0645	10	105290373	105290432	-1.07	NEURL
TCGA-12-0620	10	5530640	5530688	-1.05	CALML5
TCGA-12-0620	10	7372309	7372368	-1.05	SFMBT2
TCGA-12-0620	10	9864908	9864967	-1.05	
TCGA-12-0620	10	17931709	17931760	-1.05	MRC1
TCGA-12-0620	10	33163680	33163739	-1.05	C10orf68
TCGA-12-0620	10	35450870	35450929	-1.05	
TCGA-12-0620	10	39018805	39018864	-1.05	
TCGA-12-0620	10	72029378	72029434	-1.05	PRF1
TCGA-12-0620	10	72324240	72324293	-1.05	
TCGA-12-0620	10	74692636	74692695	-1.05	TTC18
TCGA-12-0620	10	79357832	79357891	-1.05	AF194537
TCGA-12-0620	10	88690277	88690334	-1.05	MMRN2
TCGA-12-0620	10	96987297	96987356	-1.05	PDLIM1
TCGA-12-0620	10	99324268	99324325	-1.05	ANKRD2
TCGA-12-0620	10	100789507	100789566	-1.05	HPSE2
TCGA-12-0620	10	102737790	102737837	-1.05	PEO1
TCGA-12-0620	10	102783080	102783139	-1.05	SFXN3
TCGA-12-0620	10	103528589	103528648	-1.05	
TCGA-12-0620	10	104147671	104147727	-1.05	NFKB2
TCGA-12-0620	10	105290373	105290432	-1.05	NEURL
TCGA-12-0620	10	116095056	116095104	-1.05	KIAA1914
TCGA-12-0620	10	119295569	119295616	-1.05	EMX2
TCGA-06-0410	10	37508402	37508461	-1.04	ANKRD30A
TCGA-02-0034	10	80943483	80943542	-1.01	BC070048
TCGA-02-0069	10	33177446	33177505	1.09	C10orf68
TCGA-02-0069	10	30701361	30701420	1.18	AL122121
TCGA-02-0010	11	40145100	40145159	-3.42	LRRC4C
TCGA-02-0010	11	43385450	43385509	-3.42	TTC17
TCGA-02-0010	11	43581380	43581439	-3.42	
TCGA-02-0271	11	116182194	116182253	-2.36	
TCGA-02-0010	11	5714298	5714357	-2.20	OR56B1
TCGA-02-0010	11	40138697	40138756	-2.20	LRRC4C
TCGA-02-0010	11	40636139	40636198	-2.20	DQ084202
TCGA-02-0010	11	40875860	40875919	-2.20	DQ084202
TCGA-02-0010	11	41967473	41967532	-2.20	
TCGA-02-0010	11	42320627	42320686	-2.20	
TCGA-02-0010	11	42441291	42441350	-2.20	
TCGA-02-0010	11	43350238	43350297	-2.20	TTC17

sample	chr	start	end	segval	genes
TCGA-06-0169	4	20771524	20771583	1.13	KCNIP4
TCGA-06-0169	4	173797255	173797314	1.13	GALNT17
TCGA-02-0069	4	119787238	119787297	1.25	
TCGA-08-0345	4	176930261	176930320	1.37	GPM6A
TCGA-08-0373	4	75786041	75786100	1.38	
TCGA-08-0373	4	111021272	111021331	1.38	CFI
TCGA-02-0446	4	56434374	56434433	2.10	
TCGA-06-0644	4	88751976	88752035	2.30	SPARCL1
TCGA-02-0011	5	65930516	65930575	-2.75	LOC375449
TCGA-02-0011	5	180351118	180351177	-2.75	BTNL3
TCGA-02-0439	5	763494	763553	-2.39	AF251188
TCGA-02-0060	5	41270804	41270863	-1.42	BC035723
TCGA-02-0060	5	55994367	55994426	-1.42	
TCGA-02-0069	5	15772846	15772905	-1.34	FBXL7
TCGA-02-0069	5	154503127	154503186	-1.34	
TCGA-08-0516	5	72852165	72852224	-1.20	
TCGA-02-0084	5	129005390	129005449	-1.16	ADAMTS19
TCGA-02-0084	5	175590732	175590791	-1.16	
TCGA-02-0084	5	175717047	175717106	-1.16	KIAA1191
TCGA-08-0529	5	20776045	20776104	-1.04	AK093362
TCGA-08-0529	5	69274433	69274492	-1.04	LOC153561
TCGA-08-0529	5	72852165	72852224	-1.04	
TCGA-08-0529	5	117403380	117403439	-1.04	
TCGA-06-0141	5	28853748	28853807	-1.02	
TCGA-02-0003	5	34407822	34407881	-1.01	
TCGA-02-0027	5	41541451	41541510	-1.01	PLCXD3
TCGA-02-0027	5	60277014	60277061	-1.01	mimitin
TCGA-02-0027	5	69274433	69274492	-1.01	LOC153561
TCGA-02-0027	5	100842157	100842216	-1.01	
TCGA-02-0027	5	125801997	125802056	-1.01	GRAMD3
TCGA-08-0521	5	34407822	34407881	-1.01	
TCGA-08-0521	5	117326075	117326134	-1.01	
TCGA-06-0414	5	34407822	34407881	-1.01	
TCGA-06-0414	5	55994367	55994426	-1.01	
TCGA-06-0414	5	118719760	118719808	-1.01	TNFAIP8
TCGA-06-0178	6	153446065	153446124	-2.50	RGS17
TCGA-02-0011	6	110760400	110760459	-1.81	AK127146
TCGA-06-0410	6	32629848	32629907	-1.72	BC003593
TCGA-08-0516	6	32654432	32654491	-1.65	BC003593
TCGA-02-0324	6	109812732	109812791	-1.39	
TCGA-02-0113	6	162554711	162554770	-1.27	PARK2
TCGA-02-0086	6	110760400	110760459	-1.23	AK127146
TCGA-06-0149	6	161002365	161002409	-1.10	LPA
TCGA-06-0645	6	35866245	35866302	-1.09	
TCGA-06-0645	6	126183513	126183572	-1.09	NCOA7
TCGA-06-0645	6	129614962	129615021	-1.09	LAMA2
TCGA-06-0645	6	167767660	167767704	-1.09	TCP10
TCGA-02-0010	11	43369996	43370055	-2.20	TTC17
TCGA-02-0010	11	43576127	43576185	-2.20	
TCGA-02-0010	11	66469412	66469471	-2.20	PC
TCGA-02-0086	11	12730077	12730136	-1.81	AL833289
TCGA-02-0116	11	4422151	4422209	1.13	
TCGA-02-0116	11	5039792	5039851	1.13	
TCGA-02-0116	11	50016006	50016065	1.13	
TCGA-02-0116	11	62081604	62081658	1.13	
TCGA-02-0116	11	125820910	125820969	1.13	KIRREL3
TCGA-08-0510	12	22246157	22246216	-1.86	ST8SIA1
TCGA-02-0075	12	6309670	6309724	-1.24	TNFRSF1A
TCGA-02-0075	12	83250693	83250752	-1.24	
TCGA-02-0432	12	20451634	20451693	-1.20	PDE3A
TCGA-02-0116	12	43930302	43930361	-1.17	TMEM16F
TCGA-02-0028	12	11077272	11077331	-1.05	BC071692
TCGA-02-0028	12	83250693	83250752	-1.05	
TCGA-02-0060	12	45828785	45828844	1.13	BC072670
TCGA-02-0060	12	47147923	47147970	1.13	
TCGA-02-0060	12	54363637	54363688	1.13	METTL7B
TCGA-02-0060	12	60982821	60982880	1.13	USP15
TCGA-02-0060	12	68301742	68301801	1.13	
TCGA-02-0060	12	68612610	68612669	1.13	C12orf28
TCGA-02-0060	12	70516795	70516854	1.13	
TCGA-02-0060	12	56445035	56445079	1.65	CYP27B1
TCGA-02-0060	12	68293939	68293998	1.65	
TCGA-02-0060	12	68627275	68627334	1.65	C12orf28
TCGA-02-0060	12	69119002	69119061	1.65	
TCGA-02-0060	12	73929394	73929453	1.65	AK093193
TCGA-02-0060	12	56449995	56450051	2.17	METTL1
TCGA-06-0414	13	59499981	59500040	-2.75	AY750055
TCGA-02-0115	13	98277054	98277113	-1.43	DOCK9
TCGA-06-0176	13	108149007	108149066	-1.36	MYR8
TCGA-02-0258	13	27564238	27564297	-1.22	FLT3
TCGA-02-0258	13	48770638	48770697	-1.22	
TCGA-02-0258	13	49224565	49224624	-1.22	KPNA3
TCGA-02-0258	13	50057943	50058002	-1.22	AJ412029
TCGA-02-0258	13	51077617	51077676	-1.22	WDFY2
TCGA-02-0258	13	58110118	58110177	-1.22	
TCGA-02-0258	13	66001904	66001963	-1.22	PCDH9
TCGA-02-0258	13	75553783	75553842	-1.22	
TCGA-02-0258	13	76189003	76189062	-1.22	
TCGA-08-0386	13	49085437	49085496	-1.13	
TCGA-08-0386	13	51491170	51491229	-1.13	ALG11
TCGA-08-0386	13	73171121	73171180	-1.13	KLF12
TCGA-08-0386	13	100296260	100296319	-1.13	AK025806
TCGA-08-0386	13	102046527	102046586	-1.13	
TCGA-06-0137	13	67882690	67882749	1.90	

sample	chr	start	end	segval	genes
TCGA-06-0645	6	169461396	169461455	-1.09	THBS2
TCGA-06-0127	6	21704616	21704667	-1.02	SOX4
TCGA-06-0127	6	22401348	22401407	-1.02	PRL
TCGA-06-0127	6	33649670	33649716	-1.02	BAK1
TCGA-06-0127	6	34509282	34509341	-1.02	
TCGA-06-0127	6	35327326	35327385	-1.02	BX537875
TCGA-06-0127	6	35479198	35479257	-1.02	PPARD
TCGA-06-0127	6	35853209	35853254	-1.02	UNQ3045
TCGA-06-0127	6	39397697	39397752	-1.02	KCNK16
TCGA-06-0127	6	45623047	45623106	-1.02	RUNX2
TCGA-06-0127	6	63574775	63574834	-1.02	
TCGA-06-0127	6	80042348	80042407	-1.02	
TCGA-06-0127	6	80846877	80846936	-1.02	
TCGA-06-0127	6	84706582	84706641	-1.02	CYB5R4
TCGA-06-0127	6	86406013	86406072	-1.02	SYNCRIP
TCGA-06-0127	6	104449055	104449114	-1.02	
TCGA-06-0127	6	105918008	105918066	-1.02	PREP
TCGA-06-0127	6	109466409	109466468	-1.02	SESN1
TCGA-06-0127	6	110408187	110408246	-1.02	GPR6
TCGA-06-0127	6	117931629	117931680	-1.02	DCBLD1
TCGA-06-0127	6	118697288	118697347	-1.02	SLC35F1
TCGA-06-0127	6	129614962	129615021	-1.02	LAMA2
TCGA-06-0127	6	143467919	143467978	-1.02	AIG1
TCGA-06-0127	6	167767660	167767704	-1.02	TCP10
TCGA-06-0127	6	169461396	169461455	-1.02	THBS2
TCGA-02-0266	6	31898633	31898692	1.02	BC011600
TCGA-02-0266	6	44347885	44347941	1.02	C6orf137
TCGA-02-0266	6	160175470	160175529	1.02	TCP1
TCGA-08-0373	6	35718556	35718615	1.13	FKBP5
TCGA-08-0373	6	76006782	76006841	1.13	COX7A2
TCGA-08-0373	6	76526464	76526523	1.13	MYO6
TCGA-08-0373	6	112142246	112142305	1.13	FYN
TCGA-08-0373	6	122807669	122807728	1.13	SERINC1
TCGA-08-0373	6	137561232	137561291	1.13	IFNGR1
TCGA-08-0373	6	144457695	144457754	1.13	SF3B5
TCGA-02-0084	6	18502918	18502977	1.16	IBRDC2
TCGA-02-0084	6	76025847	76025906	1.16	TMEM30A
TCGA-02-0084	6	122807669	122807728	1.16	SERINC1
TCGA-02-0086	7	141985466	141985512	-1.16	LOC647353
TCGA-02-0324	7	1263119	1263178	1.05	MICAL-L2
TCGA-02-0324	7	2333471	2333526	1.05	LFNG
TCGA-02-0324	7	27000106	27000165	1.05	BC025338
TCGA-02-0324	7	48477536	48477595	1.05	
TCGA-02-0324	7	75792244	75792291	1.05	UPK3B
TCGA-02-0324	7	83302247	83302306	1.05	SEMA3A
TCGA-02-0324	7	116363804	116363863	1.05	ST7
TCGA-02-0324	7	122349450	122349509	1.05	SLC13A1
TCGA-12-0616	14	67387689	67387733	-1.66	RAD51L
TCGA-12-0616	14	79742303	79742358	-1.66	DIO2
TCGA-12-0616	14	80737659	80737718	-1.66	GTF2A1
TCGA-12-0616	14	92173663	92173722	-1.66	RIN3
TCGA-12-0616	14	99330427	99330486	-1.66	EML1
TCGA-06-0646	14	96412034	96412093	-1.61	VRK1
TCGA-02-0011	14	19397709	19397768	-1.49	BC020999
TCGA-02-0011	14	105794161	105794220	-1.49	LOC652848
TCGA-08-0520	14	20418299	20418358	-1.26	
TCGA-02-0258	14	55960336	55960395	-1.13	
TCGA-06-0646	14	22765734	22765778	-1.11	
TCGA-06-0646	14	31509284	31509343	-1.11	
TCGA-06-0646	14	32544742	32544801	-1.11	NPAS3
TCGA-06-0646	14	32975012	32975071	-1.11	NPAS3
TCGA-06-0646	14	33049437	33049496	-1.11	NPAS3
TCGA-06-0646	14	34063670	34063718	-1.11	C14orf11
TCGA-06-0646	14	52112020	52112079	-1.11	AK123953
TCGA-06-0646	14	58891735	58891794	-1.11	DAAM1
TCGA-06-0646	14	64965154	64965213	-1.11	FUT8
TCGA-06-0646	14	67241027	67241086	-1.11	
TCGA-06-0646	14	67387689	67387733	-1.11	RAD51L1
TCGA-06-0646	14	71065782	71065838	-1.11	SIPA1L1
TCGA-06-0646	14	76641904	76641954	-1.11	KIAA1737
TCGA-06-0646	14	85586106	85586165	-1.11	
TCGA-06-0646	14	96404574	96404633	-1.11	VRK1
TCGA-06-0646	14	96417233	96417292	-1.11	VRK1
TCGA-06-0646	14	99375657	99375716	-1.11	EML1
TCGA-06-0646	14	102684040	102684099	-1.11	
TCGA-06-0646	14	105082760	105082811	-1.11	S55096
TCGA-06-0646	14	105594189	105594248	-1.11	LOC652848
TCGA-06-0646	14	105630089	105630148	-1.11	LOC652848
TCGA-06-0646	14	105881279	105881323	-1.11	AK125079
TCGA-06-0646	14	105946993	105947052	-1.11	AK125079
TCGA-06-0646	14	105985458	105985517	-1.11	AK125079
TCGA-06-0646	14	106125325	106125384	-1.11	AK125079
TCGA-06-0143	14	23523096	23523155	-1.09	AY616182
TCGA-06-0143	14	105857193	105857252	-1.09	LOC652848
TCGA-02-0099	14	21966870	21966929	-1.08	BC063432
TCGA-02-0034	14	37193624	37193683	-1.03	BC038110
TCGA-02-0034	14	52247460	52247519	-1.03	PSMC6
TCGA-06-0152	14	46778735	46778794	-1.01	MAMDC1
TCGA-06-0152	14	80737659	80737718	-1.01	GTF2A1
TCGA-06-0152	14	92173663	92173722	-1.01	RIN3
TCGA-06-0152	14	99317055	99317107	-1.01	
TCGA-06-0152	14	100073727	100073786	-1.01	KIAA1446
TCGA-06-0152	14	106125325	106125384	-1.01	AK125079
TCGA-02-0281	14	19505698	19505757	-1.01	BC020999

sample	chr	start	end	segval	genes
TCGA-02-0324	7	129532549	129532608	1.05	CPA4
TCGA-02-0324	7	129875106	129875151	1.05	KLF14
TCGA-06-0154	7	51730660	51730719	1.13	
TCGA-06-0154	7	52016452	52016511	1.13	
TCGA-06-0154	7	53947391	53947450	1.13	
TCGA-06-0154	7	55078607	55078666	1.13	K03193
TCGA-06-0154	7	55412238	55412297	1.13	ECOP
TCGA-06-0154	7	55739664	55739723	1.13	BC094796
TCGA-06-0154	7	65931904	65931963	1.13	RSAFD1
TCGA-06-0154	7	88764458	88764517	1.13	
TCGA-06-0154	7	92632452	92632511	1.13	FLJ20097
TCGA-06-0154	7	107423997	107424056	1.13	NRCAM
TCGA-06-0154	7	110498901	110498960	1.13	IMMP2L
TCGA-06-0154	7	113575827	113575886	1.13	AK131266
TCGA-06-0154	7	124008262	124008321	1.13	
TCGA-06-0154	7	127992874	127992933	1.13	CALU
TCGA-06-0154	7	129875106	129875151	1.13	KLF14
TCGA-06-0154	7	136947009	136947068	1.13	DGKI
TCGA-06-0154	7	146110679	146110738	1.13	CNTNAP2
TCGA-06-0154	7	150772061	150772120	1.13	PRKAG2
TCGA-06-0154	7	150894557	150894616	1.13	PRKAG2
TCGA-12-0620	7	810135	810191	1.18	MGC11257
TCGA-12-0620	7	13800963	13801022	1.18	ETV1
TCGA-12-0620	7	26003797	26003856	1.18	BX538099
TCGA-12-0620	7	101707816	101707862	1.18	POLR2J
TCGA-12-0620	7	107474529	107474588	1.18	NRCAM
TCGA-12-0620	7	111012308	111012367	1.18	DOCK4
TCGA-12-0620	7	121310523	121310582	1.18	AASS
TCGA-12-0620	7	130634163	130634222	1.18	MKLN1
TCGA-12-0620	7	156017585	156017644	1.18	LMBR1
TCGA-02-0001	7	7362407	7362466	1.38	
TCGA-02-0001	7	65931904	65931963	1.38	RSAFD1
TCGA-02-0001	7	82760953	82761012	1.38	SEMA3E
TCGA-02-0001	7	107423997	107424056	1.38	NRCAM
TCGA-02-0001	7	122705022	122705081	1.38	FLJ35834
TCGA-02-0001	7	129875106	129875151	1.38	KLF14
TCGA-02-0001	7	141238383	141238442	1.38	MGAM
TCGA-02-0001	7	141511878	141511937	1.38	AJ007770
TCGA-02-0001	7	146116115	146116174	1.38	CNTNAP2
TCGA-02-0086	7	20047659	20047718	1.57	
TCGA-02-0086	7	62806412	62806471	1.57	BC029561
TCGA-02-0086	7	95236836	95236895	1.57	DYNC1I1
TCGA-02-0086	7	122705022	122705081	1.57	FLJ35834
TCGA-02-0086	7	141511878	141511937	1.57	AJ007770
TCGA-02-0086	7	158617983	158618042	1.57	
TCGA-02-0271	7	139642681	139642740	1.69	
TCGA-02-0271	7	141974616	141974667	1.69	LOC647353

sample	chr	start	end	segval	genes
TCGA-02-0281	14	22385544	22385598	-1.01	MMP14
TCGA-02-0281	14	22439703	22439762	-1.01	BC106012
TCGA-02-0281	14	57266951	57267010	-1.01	
TCGA-02-0281	14	64811144	64811203	-1.01	
TCGA-06-0648	15	30423192	30423251	-2.30	
TCGA-06-0648	15	80451436	80451495	-2.30	CR936602
TCGA-02-0011	15	28719136	28719195	-2.05	BC071990
TCGA-12-0618	15	19819961	19820020	-1.31	AY941978
TCGA-12-0618	15	19910867	19910926	-1.31	AY941978
TCGA-12-0618	15	21610007	21610066	-1.31	
TCGA-12-0618	15	22860025	22860075	-1.31	AL832758
TCGA-12-0618	15	62292333	62292392	-1.31	CSNK1G1
TCGA-12-0618	15	63131441	63131496	-1.31	OSTbeta
TCGA-12-0618	15	80451436	80451495	-1.31	CR936602
TCGA-12-0618	15	88619485	88619529	-1.31	AK125370
TCGA-12-0618	15	98679517	98679576	-1.31	ADAMTS17
TCGA-06-0238	15	18722595	18722639	-1.18	
TCGA-06-0238	15	22364310	22364369	-1.18	AK058147
TCGA-06-0238	15	29640538	29640597	-1.18	OTUD7
TCGA-06-0238	15	30423192	30423251	-1.18	
TCGA-06-0238	15	46005914	46005973	-1.18	BC023624
TCGA-06-0238	15	47655530	47655589	-1.18	C15orf33
TCGA-06-0238	15	61329801	61329860	-1.18	RAB8B
TCGA-06-0238	15	73865562	73865606	-1.18	
TCGA-06-0402	15	40231199	40231256	-1.06	PLA2G4F
TCGA-06-0402	15	54579631	54579690	-1.06	
TCGA-06-0402	15	55616161	55616220	-1.06	CGNL1
TCGA-02-0439	15	19879661	19879720	-1.05	AY941978
TCGA-02-0439	15	86225234	86225293	-1.05	NTRK3
TCGA-06-0128	15	28441169	28441228	-1.00	
TCGA-06-0195	15	29272766	29272813	1.44	
TCGA-08-0520	16	3100485	3100544	-1.01	BC001809
TCGA-08-0520	16	79670474	79670533	-1.01	
TCGA-02-0326	16	76938723	76938782	1.24	WWOX
TCGA-06-0184	17	31474518	31474577	-1.05	
TCGA-12-0616	17	6532881	6532940	-1.03	SLC13A5
TCGA-12-0616	17	7318722	7318771	-1.03	ZBTB4
TCGA-12-0616	17	7694620	7694679	-1.03	AB002344
TCGA-12-0616	17	19522211	19522270	-1.03	FLJ31196
TCGA-12-0616	17	20234571	20234630	-1.03	
TCGA-12-0616	17	20625245	20625304	-1.03	
TCGA-12-0616	17	20684380	20684439	-1.03	
TCGA-12-0616	17	21442476	21442522	-1.03	
TCGA-12-0616	17	25911485	25911529	-1.03	DKFZP434O047
TCGA-12-0616	17	26501612	26501671	-1.03	NF1
TCGA-12-0616	17	31486540	31486599	-1.03	
TCGA-12-0616	17	39608802	39608861	-1.03	ASB16

sample	chr	start	end	segval	genes
TCGA-02-0116	7	54558881	54558940	2.71	
TCGA-02-0451	7	53857835	53857894	2.80	
TCGA-08-0514	7	64701922	64701968	3.98	
TCGA-02-0116	7	55862893	55862952	4.14	PSPH
TCGA-02-0106	8	12276178	12276237	-2.25	BC032892
TCGA-02-0084	8	7729311	7729370	-1.85	BC030211
TCGA-02-0107	8	1333233	1333292	-1.81	
TCGA-02-0107	8	37907045	37907104	-1.81	
TCGA-02-0107	8	101800174	101800233	-1.81	PABPC1
TCGA-02-0439	8	77057200	77057259	-1.48	
TCGA-02-0266	8	107602071	107602130	-1.38	AK124441
TCGA-02-0266	8	133094826	133094885	-1.38	KIAA0143
TCGA-06-0152	8	90816	90868	-1.22	BC071667
TCGA-06-0152	8	6929691	6929747	-1.22	
TCGA-06-0152	8	31607545	31607604	-1.22	
TCGA-06-0152	8	32141825	32141884	-1.22	NRG1
TCGA-06-0152	8	33568440	33568487	-1.22	DUSP26
TCGA-06-0152	8	37913540	37913593	-1.22	GOT1L1
TCGA-06-0152	8	39599448	39599507	-1.22	ADAM18
TCGA-06-0152	8	50776370	50776429	-1.22	
TCGA-06-0152	8	53424593	53424652	-1.22	ST18
TCGA-06-0152	8	55138187	55138246	-1.22	LYPLA1
TCGA-06-0152	8	75582884	75582943	-1.22	
TCGA-06-0152	8	92307137	92307196	-1.22	BC060784
TCGA-06-0152	8	118601645	118601704	-1.22	
TCGA-06-0152	8	123979018	123979077	-1.22	ZHX2
TCGA-06-0152	8	131103860	131103919	-1.22	
TCGA-06-0152	8	141106328	141106387	-1.22	NIBP
TCGA-06-0152	8	144166141	144166185	-1.22	BC007589
TCGA-06-0152	8	144203999	144204053	-1.22	C8orf31
TCGA-06-0152	8	144537484	144537534	-1.22	BC025767
TCGA-06-0152	8	145533392	145533436	-1.22	
TCGA-02-0115	8	137160246	137160305	-1.13	
TCGA-02-0046	8	140318249	140318308	-1.03	
TCGA-12-0619	8	90816	90868	-1.03	BC071667
TCGA-12-0619	8	974966	975025	-1.03	BC022082
TCGA-12-0619	8	6929691	6929747	-1.03	
TCGA-12-0619	8	33568440	33568487	-1.03	DUSP26
TCGA-12-0619	8	39599448	39599507	-1.03	ADAM18
TCGA-12-0619	8	50776370	50776429	-1.03	
TCGA-12-0619	8	53424593	53424652	-1.03	ST18
TCGA-12-0619	8	66640572	66640631	-1.03	
TCGA-12-0619	8	86064843	86064902	-1.03	
TCGA-12-0619	8	87405509	87405568	-1.03	
TCGA-12-0619	8	92307137	92307196	-1.03	BC060784
TCGA-12-0619	8	106441121	106441180	-1.03	ZFPM2
TCGA-12-0619	8	118601645	118601704	-1.03	
TCGA-12-0616	17	39752508	39752560	-1.03	CGI-69
TCGA-12-0616	17	41011530	41011589	-1.03	AK124512
TCGA-12-0616	17	46260128	46260181	-1.03	
TCGA-12-0616	17	55035194	55035253	-1.03	DHX40
TCGA-12-0616	17	57983727	57983786	-1.03	TLK2
TCGA-12-0616	17	59370192	59370251	-1.03	SCN4A
TCGA-12-0616	17	63739129	63739187	-1.03	
TCGA-12-0616	17	69818169	69818228	-1.03	DNAI2
TCGA-12-0616	17	76068937	76068984	-1.03	
TCGA-12-0616	17	76749430	76749479	-1.03	AK131529
TCGA-12-0616	17	77041927	77041971	-1.03	AB040880
TCGA-12-0616	17	77478901	77478945	-1.03	AF338198
TCGA-12-0616	17	78653545	78653589	-1.03	
TCGA-02-0079	17	1315406	1315465	-1.02	MYO1C
TCGA-02-0079	17	1877316	1877367	-1.02	
TCGA-02-0079	17	3727364	3727419	-1.02	CAMKK1
TCGA-02-0079	17	4424180	4424239	-1.02	
TCGA-02-0079	17	4789213	4789267	-1.02	RNF167
TCGA-02-0079	17	6871963	6872019	-1.02	BCL6B
TCGA-02-0079	17	7318722	7318771	-1.02	ZBTB4
TCGA-02-0079	17	7694620	7694679	-1.02	AB002344
TCGA-02-0079	17	8052025	8052084	-1.02	AURKB
TCGA-02-0079	17	16370645	16370704	-1.02	
TCGA-02-0079	17	20625245	20625304	-1.02	
TCGA-02-0079	17	23862418	23862477	-1.02	BC045622
TCGA-02-0079	17	26501612	26501671	-1.02	NF1
TCGA-02-0079	17	28177706	28177765	-1.02	MYO1D
TCGA-02-0079	17	41459223	41459276	-1.02	MAPT
TCGA-02-0079	17	44360021	44360080	-1.02	UBE2Z
TCGA-02-0079	17	45019015	45019071	-1.02	
TCGA-02-0079	17	46006096	46006155	-1.02	CACNA1G
TCGA-02-0079	17	69024557	69024616	-1.02	BC041474
TCGA-02-0079	17	69774078	69774137	-1.02	
TCGA-02-0079	17	69890659	69890716	-1.02	
TCGA-02-0079	17	70367463	70367509	-1.02	GRIN2C
TCGA-02-0079	17	70596987	70597031	-1.02	SLC16A5
TCGA-02-0079	17	77041927	77041971	-1.02	AB040880
TCGA-02-0258	17	3400142	3400201	-1.00	TRPV3
TCGA-02-0079	17	36351895	36351954	1.02	
TCGA-02-0079	17	37744308	37744367	1.02	STAT3
TCGA-02-0079	17	72239339	72239398	1.02	CR599912
TCGA-06-0208	17	69926294	69926352	1.04	
TCGA-06-0208	17	76449446	76449505	1.04	KIAA1303
TCGA-06-0208	17	78528645	78528704	1.04	B3GNTL1
TCGA-06-0414	17	36534226	36534285	1.05	
TCGA-06-0644	17	22315936	22315983	1.09	
TCGA-06-0644	17	36228012	36228071	1.09	KRT10

sample	chr	start	end	segval	genes
TCGA-12-0619	8	132063081	132063140	-1.03	ADCY8
TCGA-12-0619	8	141106328	141106387	-1.03	NIBP
TCGA-12-0619	8	144203999	144204053	-1.03	C8orf31
TCGA-12-0619	8	144537484	144537534	-1.03	BC025767
TCGA-12-0619	8	145695514	145695558	-1.03	PPP1R16A
TCGA-02-0266	8	144945861	144945906	1.31	SCRIB
TCGA-02-0043	9	23816193	23816243	-3.19	ELAVL2
TCGA-02-0043	9	127262922	127262981	-3.19	
TCGA-06-0148	9	346079	346138	-3.04	DOCK8
TCGA-06-0148	9	21179604	21179663	-3.04	
TCGA-06-0148	9	26913325	26913384	-3.04	PLAA
TCGA-06-0148	9	29245323	29245382	-3.04	
TCGA-02-0009	9	36611606	36611665	-2.94	MELK
TCGA-06-0125	9	110104594	110104653	-2.83	
TCGA-02-0021	9	23901804	23901863	-2.73	
TCGA-02-0021	9	23971151	23971210	-2.73	
TCGA-02-0021	9	25195505	25195564	-2.73	
TCGA-02-0021	9	43378350	43378400	-2.73	
TCGA-02-0037	9	43378350	43378400	-2.45	
TCGA-02-0430	9	43378350	43378400	-2.23	
TCGA-06-0148	9	330083	330142	-2.19	DOCK8
TCGA-06-0148	9	629010	629069	-2.19	ANKRD15
TCGA-06-0148	9	20208937	20208996	-2.19	
TCGA-06-0148	9	20684462	20684521	-2.19	KIAA1797
TCGA-06-0148	9	20739288	20739347	-2.19	KIAA1797
TCGA-06-0148	9	20871870	20871926	-2.19	KIAA1797
TCGA-06-0148	9	21282858	21282917	-2.19	
TCGA-06-0148	9	21466463	21466522	-2.19	AK124391
TCGA-06-0148	9	21510354	21510413	-2.19	AK124391
TCGA-06-0148	9	21958041	21958099	-2.19	CDKN2A
TCGA-06-0148	9	22723912	22723971	-2.19	AK092601
TCGA-06-0148	9	25678698	25678757	-2.19	
TCGA-06-0148	9	26104455	26104514	-2.19	FLJ16323
TCGA-06-0148	9	26123251	26123310	-2.19	
TCGA-06-0148	9	26613334	26613393	-2.19	
TCGA-06-0148	9	26821029	26821088	-2.19	
TCGA-06-0148	9	28196294	28196353	-2.19	LRRN6C
TCGA-06-0148	9	29289692	29289751	-2.19	
TCGA-06-0148	9	29358120	29358179	-2.19	
TCGA-06-0148	9	30237394	30237453	-2.19	
TCGA-06-0148	9	42913008	42913063	-2.19	BC031626
TCGA-02-0422	9	86952145	86952204	-2.13	
TCGA-02-0290	9	35971395	35971454	-2.13	
TCGA-02-0260	9	20431890	20431949	-2.11	MLLT3
TCGA-02-0260	9	20684462	20684521	-2.11	KIAA1797
TCGA-02-0260	9	21282858	21282917	-2.11	
TCGA-02-0260	9	21394289	21394347	-2.11	
TCGA-06-0644	17	72239339	72239398	1.09	CR599912
TCGA-06-0644	17	74381487	74381542	1.09	TIMP2
TCGA-02-0060	17	38398917	38398976	1.13	RUNDC1
TCGA-06-0210	17	76449446	76449505	1.17	KIAA1303
TCGA-06-0188	17	23682723	23682782	1.21	IFT20
TCGA-06-0188	17	23979594	23979653	1.21	KIAA0100
TCGA-06-0188	17	25520058	25520117	1.21	CCDC55
TCGA-06-0188	17	29640471	29640530	1.21	
TCGA-06-0188	17	33726698	33726757	1.21	MRPL45
TCGA-06-0188	17	36969988	36970044	1.21	
TCGA-06-0188	17	37513554	37513598	1.21	LGP2
TCGA-06-0188	17	38202283	38202338	1.21	WNK4
TCGA-06-0188	17	44212401	44212460	1.21	TTLL6
TCGA-06-0188	17	50986963	50987022	1.21	
TCGA-06-0188	17	53747474	53747533	1.21	BZRAP1
TCGA-06-0188	17	57148071	57148123	1.21	BRIP1
TCGA-06-0188	17	67531059	67531118	1.21	
TCGA-06-0188	17	68826587	68826646	1.21	
TCGA-02-0107	18	24048647	24048706	-1.29	
TCGA-02-0046	18	2382176	2382235	-1.07	
TCGA-02-0046	18	17945381	17945440	-1.07	
TCGA-02-0046	18	22436502	22436561	-1.07	KCTD1
TCGA-02-0046	18	33015923	33015982	-1.07	KIAA1328
TCGA-02-0046	18	56467989	56468048	-1.07	
TCGA-02-0439	18	64017240	64017299	-1.05	
TCGA-02-0024	18	14196224	14196283	-1.01	
TCGA-02-0010	18	22442191	22442250	1.19	KCTD1
TCGA-02-0271	18	42794580	42794639	1.22	KATNAL2
TCGA-08-0373	18	22690074	22690133	1.44	AQP4
TCGA-02-0271	19	737550	737609	-2.03	AK024373
TCGA-06-0176	19	13199527	13199586	-1.31	CACNA1A
TCGA-06-0176	19	57038202	57038261	-1.31	
TCGA-06-0213	19	40543992	40544045	-1.28	
TCGA-02-0086	19	61559132	61559191	-1.25	BC043232
TCGA-12-0616	19	1803379	1803438	-1.18	KLF16
TCGA-12-0616	19	3911050	3911095	-1.18	DAPK3
TCGA-12-0616	19	4760281	4760340	-1.18	
TCGA-12-0616	19	6155257	6155316	-1.18	
TCGA-12-0616	19	7015859	7015918	-1.18	
TCGA-12-0616	19	9861570	9861629	-1.18	OLFM2
TCGA-12-0616	19	20099547	20099606	-1.18	
TCGA-12-0616	19	20860538	20860597	-1.18	
TCGA-12-0616	19	23221921	23221980	-1.18	
TCGA-12-0616	19	40716218	40716265	-1.18	GAPDHS
TCGA-12-0616	19	40944669	40944728	-1.18	LOC148137
TCGA-12-0616	19	48221641	48221700	-1.18	PSG11
TCGA-12-0616	19	50609052	50609096	-1.18	ERCC1

sample	chr	start	end	segval	genes
TCGA-02-0260	9	21504931	21504990	-2.11	AK124391
TCGA-02-0260	9	21958041	21958099	-2.11	CDKN2A
TCGA-02-0260	9	22723912	22723971	-2.11	AK092601
TCGA-02-0260	9	22842920	22842979	-2.11	
TCGA-02-0260	9	23784955	23785014	-2.11	ELAVL2
TCGA-02-0116	9	34267524	34267583	-2.10	
TCGA-06-0125	9	21114816	21114875	-2.08	
TCGA-06-0125	9	21282858	21282917	-2.08	
TCGA-06-0125	9	21394289	21394347	-2.08	
TCGA-06-0125	9	22723912	22723971	-2.08	AK092601
TCGA-06-0125	9	23704617	23704676	-2.08	ELAVL2
TCGA-06-0125	9	23784955	23785014	-2.08	ELAVL2
TCGA-06-0125	9	25678698	25678757	-2.08	
TCGA-06-0125	9	25969531	25969590	-2.08	
TCGA-06-0125	9	27430770	27430827	-2.08	MOBKL2B
TCGA-06-0125	9	28196294	28196353	-2.08	LRRN6C
TCGA-06-0125	9	28294971	28295030	-2.08	LRRN6C
TCGA-06-0125	9	28335772	28335831	-2.08	LRRN6C
TCGA-06-0125	9	41470116	41470175	-2.08	
TCGA-06-0125	9	110100386	110100445	-2.08	
TCGA-06-0125	9	120634700	120634746	-2.08	FBXW2
TCGA-08-0521	9	109690961	109691020	-1.97	PALM2
TCGA-08-0521	9	138301699	138301758	-1.97	
TCGA-06-0412	9	109690961	109691020	-1.76	PALM2
TCGA-06-0154	9	113477324	113477383	-1.64	
TCGA-06-0210	9	21154498	21154557	-1.48	
TCGA-06-0210	9	86952145	86952204	-1.48	
TCGA-06-0143	9	41470116	41470175	-1.39	
TCGA-08-0525	9	26821029	26821088	-1.35	
TCGA-08-0525	9	26925960	26926019	-1.35	PLAA
TCGA-08-0525	9	27094550	27094609	-1.35	
TCGA-08-0525	9	30524794	30524853	-1.35	BC022036
TCGA-12-0620	9	32374264	32374323	-1.32	
TCGA-12-0620	9	35467102	35467161	-1.32	BC031276
TCGA-12-0620	9	94414166	94414225	-1.32	FBP2
TCGA-12-0620	9	108704113	108704159	-1.32	BC014610
TCGA-02-0089	9	26491363	26491422	-1.24	
TCGA-06-0646	9	23316739	23316798	-1.22	
TCGA-06-0646	9	34703703	34703762	-1.22	
TCGA-06-0646	9	36894709	36894768	-1.22	PAX5
TCGA-06-0646	9	79490663	79490722	-1.22	TLE4
TCGA-06-0646	9	83312126	83312177	-1.22	FRMD3
TCGA-06-0646	9	92816494	92816542	-1.22	FGD3
TCGA-06-0646	9	93619886	93619945	-1.22	
TCGA-06-0646	9	94414166	94414225	-1.22	FBP2
TCGA-06-0646	9	104697731	104697790	-1.22	ABCA1
TCGA-06-0646	9	111478822	111478881	-1.22	bA16L21.2.1
TCGA-12-0616	19	52722417	52722476	-1.18	ZNF541
TCGA-12-0616	19	55664690	55664734	-1.18	LOC112703
TCGA-12-0616	19	57791741	57791800	-1.18	ZNF137
TCGA-12-0616	19	57929244	57929303	-1.18	BC056265
TCGA-12-0616	19	58875960	58876019	-1.18	
TCGA-12-0616	19	59492650	59492707	-1.18	LILRA6
TCGA-12-0616	19	59617971	59618024	-1.18	AF315098
TCGA-12-0616	19	63190370	63190429	-1.18	ZNF606
TCGA-12-0616	19	63676549	63676604	-1.18	ZNF324
TCGA-06-0171	19	10432640	10432687	-1.11	PDE4A
TCGA-06-0171	19	59492650	59492707	-1.11	LILRA6
TCGA-08-0517	19	47574701	47574758	-1.08	SBP1
TCGA-08-0517	19	55146867	55146912	-1.08	SIGLEC11
TCGA-08-0517	19	56840346	56840401	-1.08	AY358369
TCGA-08-0517	19	57561995	57562054	-1.08	BC039903
TCGA-12-0618	19	3911050	3911095	-1.05	DAPK3
TCGA-12-0618	19	6155257	6155316	-1.05	
TCGA-12-0618	19	10074425	10074476	-1.05	AF230330
TCGA-12-0618	19	10420582	10420626	-1.05	AY593872
TCGA-12-0618	19	12889945	12889989	-1.05	CR620067
TCGA-12-0618	19	16858458	16858506	-1.05	
TCGA-12-0618	19	18400268	18400312	-1.05	SSBP4
TCGA-12-0618	19	20099547	20099606	-1.05	
TCGA-12-0618	19	20860538	20860597	-1.05	
TCGA-12-0618	19	40716218	40716265	-1.05	GAPDHS
TCGA-12-0618	19	50609052	50609096	-1.05	ERCC1
TCGA-12-0618	19	52615730	52615789	-1.05	
TCGA-12-0618	19	52722417	52722476	-1.05	ZNF541
TCGA-12-0618	19	53802678	53802732	-1.05	FAM83E
TCGA-12-0618	19	54232133	54232177	-1.05	J00117
TCGA-12-0618	19	55664690	55664734	-1.05	LOC112703
TCGA-12-0618	19	56018305	56018353	-1.05	KLK1
TCGA-12-0618	19	57791741	57791800	-1.05	ZNF137
TCGA-12-0618	19	57929244	57929303	-1.05	BC056265
TCGA-12-0618	19	59492650	59492707	-1.05	LILRA6
TCGA-12-0618	19	59617971	59618024	-1.05	AF315098
TCGA-12-0618	19	63190370	63190429	-1.05	ZNF606
TCGA-12-0618	19	63676549	63676604	-1.05	ZNF324
TCGA-06-0214	19	40543992	40544045	-1.02	
TCGA-06-0214	19	59146487	59146540	-1.02	AB209671
TCGA-06-0214	19	59492650	59492707	-1.02	LILRA6
TCGA-02-0106	19	57718358	57718417	-1.01	BC056265
TCGA-08-0345	19	611954	612011	-1.00	RNF126
TCGA-08-0345	19	626534	626591	-1.00	AK024373
TCGA-08-0345	19	807195	807247	-1.00	ELA2
TCGA-08-0345	19	2574069	2574116	-1.00	GNG7
TCGA-08-0345	19	3441792	3441848	-1.00	BC009863

sample	chr	start	end	segval	genes
TCGA-06-0646	9	114162328	114162387	-1.22	
TCGA-06-0646	9	115918913	115918972	-1.22	
TCGA-06-0646	9	125747778	125747837	-1.22	PBX3
TCGA-06-0646	9	132117335	132117394	-1.22	NTNG2
TCGA-06-0646	9	133011339	133011393	-1.22	RALGDS
TCGA-06-0646	9	135648788	135648839	-1.22	
TCGA-06-0646	9	136071499	136071558	-1.22	UBADC1
TCGA-06-0646	9	136839806	136839850	-1.22	EGFL7
TCGA-06-0646	9	137561971	137562015	-1.22	FLJ20433
TCGA-06-0646	9	138301699	138301758	-1.22	
TCGA-06-0190	9	21359384	21359443	-1.17	
TCGA-06-0190	9	21408944	21409003	-1.17	
TCGA-06-0190	9	24160666	24160725	-1.17	
TCGA-06-0210	9	20883453	20883512	-1.17	KIAA1797
TCGA-06-0210	9	21044957	21045016	-1.17	
TCGA-06-0210	9	21114816	21114875	-1.17	
TCGA-06-0164	9	36353448	36353507	-1.13	RNF38
TCGA-06-0164	9	41470116	41470175	-1.13	
TCGA-02-0269	9	114165838	114165888	-1.13	ORM1
TCGA-02-0064	9	11308830	11308889	-1.11	
TCGA-02-0064	9	43378350	43378400	-1.11	
TCGA-06-0194	9	114165838	114165888	-1.09	ORM1
TCGA-08-0345	9	30134402	30134461	-1.01	
TCGA-08-0345	9	39355550	39355594	-1.01	CNTNAP3
TCGA-08-0345	9	68962739	68962798	-1.01	
TCGA-08-0345	9	94414166	94414225	-1.01	FBP2
TCGA-08-0345	9	104690871	104690930	-1.01	ABCA1
TCGA-02-0034	9	39131835	39131894	-1.00	CNTNAP3
TCGA-02-0034	9	43017845	43017904	-1.00	
TCGA-02-0034	9	43948886	43948940	-1.00	
TCGA-06-0133	9	85298838	85298897	1.14	
TCGA-06-0413	9	42913008	42913063	1.73	BC031626
TCGA-06-0126	9	137529571	137529630	2.20	FLJ20433
TCGA-06-0137	9	43410478	43410537	2.98	AY098593
TCGA-02-0043	10	66951636	66951695	-3.03	
TCGA-02-0043	10	98125575	98125634	-3.03	TLL2
TCGA-02-0333	10	98125575	98125634	-2.79	TLL2
TCGA-06-0414	10	51264997	51265056	-2.74	TIMM23
TCGA-06-0138	10	32885520	32885579	-2.29	CCDC7
TCGA-06-0138	10	51581040	51581091	-2.29	
TCGA-06-0138	10	66951636	66951695	-2.29	
TCGA-06-0138	10	100678817	100678876	-2.29	HPSE2
TCGA-08-0520	10	118757869	118757928	-2.19	
TCGA-06-0648	10	45554717	45554776	-2.00	FAM21C
TCGA-06-0414	10	3438109	3438168	-1.98	BC037918
TCGA-06-0414	10	4104978	4105037	-1.98	AK055803
TCGA-06-0414	10	4217287	4217346	-1.98	
TCGA-08-0345	19	3935557	3935616	-1.00	EEF2
TCGA-08-0345	19	4245786	4245834	-1.00	MGC23244
TCGA-08-0345	19	4760281	4760340	-1.00	
TCGA-08-0345	19	5729696	5729750	-1.00	TMEM146
TCGA-08-0345	19	6155257	6155316	-1.00	
TCGA-08-0345	19	7015859	7015918	-1.00	
TCGA-08-0345	19	7498185	7498244	-1.00	MCOLN1
TCGA-08-0345	19	9800520	9800566	-1.00	UBL5
TCGA-08-0345	19	11298922	11298976	-1.00	RAB3D
TCGA-08-0345	19	16858458	16858506	-1.00	
TCGA-08-0345	19	18496572	18496624	-1.00	
TCGA-08-0345	19	19922085	19922142	-1.00	
TCGA-08-0345	19	20860538	20860597	-1.00	
TCGA-08-0345	19	22533332	22533391	-1.00	
TCGA-08-0345	19	38504014	38504073	-1.00	
TCGA-08-0345	19	40716218	40716265	-1.00	GAPDHS
TCGA-08-0345	19	42396997	42397056	-1.00	
TCGA-08-0345	19	43927527	43927583	-1.00	
TCGA-08-0345	19	49263677	49263736	-1.00	ZNF223
TCGA-08-0345	19	49950095	49950154	-1.00	BCL3
TCGA-08-0345	19	52615730	52615789	-1.00	
TCGA-08-0345	19	52722417	52722476	-1.00	ZNF541
TCGA-08-0345	19	55395280	55395334	-1.00	AK125082
TCGA-08-0345	19	55664690	55664734	-1.00	LOC112703
TCGA-08-0345	19	57791741	57791800	-1.00	ZNF137
TCGA-08-0345	19	57929244	57929303	-1.00	BC056265
TCGA-08-0345	19	58003930	58003989	-1.00	ZNF28
TCGA-08-0345	19	58088740	58088799	-1.00	X78928
TCGA-08-0345	19	58875960	58876019	-1.00	
TCGA-08-0345	19	59310832	59310881	-1.00	PRPF31
TCGA-08-0345	19	59617971	59618024	-1.00	AF315098
TCGA-08-0345	19	63190370	63190429	-1.00	ZNF606
TCGA-02-0107	19	52823576	52823635	1.19	BC032065
TCGA-06-0174	19	5934511	5934570	1.21	
TCGA-06-0174	19	13870803	13870862	1.21	MGC11271
TCGA-06-0174	19	16406261	16406318	1.21	EPS15L1
TCGA-06-0174	19	33356504	33356563	1.21	
TCGA-06-0174	19	39614667	39614726	1.21	UBA2
TCGA-06-0174	19	44473559	44473615	1.21	
TCGA-06-0174	19	49469272	49469331	1.21	ZNF233
TCGA-06-0174	19	52530292	52530351	1.21	
TCGA-06-0174	19	60941502	60941556	1.21	NALP9
TCGA-06-0412	20	1528899	1528958	-1.71	SIRPB1
TCGA-02-0027	20	51513440	51513499	-1.27	TSHZ2
TCGA-06-0156	20	51513440	51513499	-1.11	TSHZ2
TCGA-06-0129	20	25839919	25839978	-1.09	
TCGA-02-0085	20	25885935	25885994	1.08	BC052952

sample	chr	start	end	segval	genes
TCGA-06-0414	10	4742025	4742084	-1.98	
TCGA-06-0414	10	5154246	5154305	-1.98	
TCGA-06-0414	10	51306200	51306259	-1.98	AF005043
TCGA-06-0414	10	89241227	89241286	-1.98	
TCGA-06-0414	10	89295627	89295686	-1.98	MINPP1
TCGA-06-0414	10	89679976	89680035	-1.98	PTEN
TCGA-06-0414	10	89879205	89879264	-1.98	
TCGA-06-0414	10	90051433	90051492	-1.98	C10orf59
TCGA-06-0414	10	90174342	90174401	-1.98	C10orf59
TCGA-06-0414	10	97984467	97984526	-1.98	BLNK
TCGA-06-0414	10	99548159	99548218	-1.98	
TCGA-02-0083	10	60513350	60513409	-1.89	
TCGA-02-0324	10	37508402	37508461	-1.78	ANKRD30A
TCGA-02-0324	10	90976366	90976425	-1.78	LIPA
TCGA-02-0324	10	98125575	98125634	-1.78	TLL2
TCGA-12-0619	10	37508402	37508461	-1.64	ANKRD30A
TCGA-12-0619	10	89440846	89440905	-1.64	PAPSS2
TCGA-12-0619	10	90678178	90678237	-1.64	AB037794
TCGA-06-0646	10	9864908	9864967	-1.56	
TCGA-06-0646	10	25672743	25672802	-1.56	GPR158
TCGA-06-0646	10	33163680	33163739	-1.56	C10orf68
TCGA-06-0646	10	37508402	37508461	-1.56	ANKRD30A
TCGA-06-0646	10	105290373	105290432	-1.56	NEURL
TCGA-08-0520	10	5530640	5530688	-1.47	CALML5
TCGA-08-0520	10	37523207	37523266	-1.47	ANKRD30A
TCGA-08-0520	10	53695083	53695142	-1.47	PRKG1
TCGA-08-0520	10	55138890	55138949	-1.47	AF083130
TCGA-08-0520	10	55384225	55384284	-1.47	PCDH15
TCGA-08-0520	10	69688265	69688324	-1.47	
TCGA-08-0520	10	90976366	90976425	-1.47	LIPA
TCGA-08-0520	10	98125575	98125634	-1.47	TLL2
TCGA-08-0520	10	118752176	118752235	-1.47	KIAA1598
TCGA-02-0046	10	80943483	80943542	-1.45	BC070048
TCGA-02-0046	10	99247798	99247854	-1.45	MMS19L
TCGA-02-0046	10	126834997	126835056	-1.45	CTBP2
TCGA-06-0214	10	66951636	66951695	-1.45	
TCGA-06-0210	10	100678817	100678876	-1.42	HPSE2
TCGA-02-0071	10	42266351	42266410	-1.33	BC039000
TCGA-02-0071	10	69257482	69257541	-1.33	DNAJC12
TCGA-02-0071	10	80943483	80943542	-1.33	BC070048
TCGA-02-0071	10	88203067	88203126	-1.33	WAPAL
TCGA-12-0618	10	11953319	11953378	-1.30	C10orf47
TCGA-12-0618	10	23014261	23014320	-1.30	PIP5K2A
TCGA-12-0618	10	25129184	25129243	-1.30	
TCGA-12-0618	10	29704220	29704279	-1.30	
TCGA-12-0618	10	33163680	33163739	-1.30	C10orf68
TCGA-12-0618	10	33647766	33647825	-1.30	NRP1
TCGA-06-0412	20	25885935	25885994	1.09	BC052952
TCGA-06-0402	20	4144452	4144506	1.14	
TCGA-06-0397	21	28368917	28368976	-1.80	AK093119
TCGA-02-0102	21	28368917	28368976	-1.51	AK093119
TCGA-02-0106	21	28368917	28368976	-1.28	AK093119
TCGA-02-0432	21	28368917	28368976	-1.24	AK093119
TCGA-08-0531	21	28368917	28368976	-1.20	AK093119
TCGA-06-0209	21	28368917	28368976	-1.14	AK093119
TCGA-02-0043	21	39030818	39030877	-1.03	
TCGA-02-0043	21	40192300	40192359	-1.03	PCP4
TCGA-02-0043	21	42480546	42480605	-1.03	
TCGA-02-0043	21	44856311	44856357	-1.03	C21orf29
TCGA-02-0079	21	32896034	32896092	1.13	C21orf59
TCGA-12-0619	21	26175055	26175112	1.43	APP
TCGA-06-0174	22	22672513	22672572	-4.08	
TCGA-06-0188	22	22672513	22672572	-3.19	
TCGA-06-0188	22	24050904	24050963	-3.19	
TCGA-08-0349	22	22672513	22672572	-2.17	
TCGA-06-0402	22	14504218	14504277	-2.15	BC016035
TCGA-06-0402	22	45755429	45755488	-2.15	TBC1D22A
TCGA-02-0084	22	22672513	22672572	-1.89	
TCGA-02-0084	22	49356005	49356052	-1.89	ARSA
TCGA-06-0238	22	49356005	49356052	-1.78	ARSA
TCGA-08-0345	22	49356005	49356052	-1.74	ARSA
TCGA-06-0175	22	31690525	31690584	-1.71	SYN3
TCGA-06-0175	22	47880174	47880233	-1.71	
TCGA-02-0325	22	36070350	36070409	-1.67	AK055475
TCGA-02-0057	22	27667173	27667232	-1.62	AB051436
TCGA-02-0060	22	49349292	49349344	-1.50	
TCGA-06-0188	22	22665928	22665976	-1.41	HS322B1A
TCGA-06-0188	22	22676549	22676593	-1.41	
TCGA-06-0188	22	22690716	22690769	-1.41	
TCGA-06-0188	22	22719859	22719907	-1.41	AK127991
TCGA-06-0188	22	23978623	23978682	-1.41	
TCGA-06-0188	22	24228038	24228097	-1.41	BC004918
TCGA-06-0188	22	31690525	31690584	-1.41	SYN3
TCGA-06-0152	22	17317314	17317358	-1.32	
TCGA-06-0152	22	17494609	17494663	-1.32	DGCR13
TCGA-06-0152	22	18127109	18127159	-1.32	TBX1
TCGA-06-0152	22	22672513	22672572	-1.32	
TCGA-06-0152	22	25749756	25749815	-1.32	
TCGA-06-0152	22	26468793	26468852	-1.32	MN1
TCGA-06-0152	22	28007348	28007405	-1.32	EWSR1
TCGA-06-0152	22	29594981	29595040	-1.32	OSBP2
TCGA-06-0152	22	30207655	30207714	-1.32	EIF4ENIF1
TCGA-06-0152	22	37730228	37730287	-1.32	
TCGA-06-0152	22	41221190	41221234	-1.32	SERHL

sample	chr	start	end	segval	genes
TCGA-12-0618	10	49340053	49340111	-1.30	ARHGAP22
TCGA-12-0618	10	50178065	50178124	-1.30	C10orf71
TCGA-12-0618	10	50485951	50486010	-1.30	
TCGA-12-0618	10	54775509	54775568	-1.30	
TCGA-12-0618	10	60786104	60786163	-1.30	FAM13C1
TCGA-12-0618	10	69612724	69612783	-1.30	FLJ14437
TCGA-12-0618	10	72324240	72324293	-1.30	
TCGA-12-0618	10	88690277	88690334	-1.30	MMRN2
TCGA-12-0618	10	99324268	99324325	-1.30	ANKRD2
TCGA-12-0618	10	100789507	100789566	-1.30	HPSE2
TCGA-12-0618	10	101591744	101591803	-1.30	ABCC2
TCGA-12-0618	10	101933488	101933542	-1.30	SPFH1
TCGA-12-0618	10	103517472	103517526	-1.30	
TCGA-12-0618	10	104152486	104152530	-1.30	PSD
TCGA-12-0618	10	105290373	105290432	-1.30	NEURL
TCGA-12-0618	10	107411772	107411831	-1.30	
TCGA-12-0618	10	119298404	119298463	-1.30	EMX2
TCGA-06-0152	22	43210959	43211009	-1.32	LDOC1L
TCGA-06-0152	22	45818352	45818411	-1.32	TBC1D22A
TCGA-06-0152	22	48956760	48956818	-1.32	SELO
TCGA-06-0152	22	49356005	49356052	-1.32	ARSA
TCGA-02-0058	22	21174548	21174607	-1.23	SUHW2
TCGA-02-0058	22	48749925	48749984	-1.23	FLJ41993
TCGA-02-0111	22	17525955	17526014	-1.15	
TCGA-06-0237	22	14504218	14504277	-1.12	BC016035
TCGA-06-0237	22	22719859	22719907	-1.12	AK127991
TCGA-06-0237	22	47122623	47122682	-1.12	
TCGA-06-0410	22	22719859	22719907	-1.08	AK127991
TCGA-06-0410	22	27754812	27754871	-1.08	AB051436
TCGA-06-0410	22	27884866	27884925	-1.08	AK056425
TCGA-06-0410	22	32067638	32067697	-1.08	LARGE
TCGA-06-0410	22	38851943	38852002	-1.08	TNRC6B
TCGA-02-0102	22	18984101	18984160	1.33	

Table S6.4.1 Genes that have at least four folds aberrations in at least one patient tumor sample. Count for the last column indicates the sample numbers that specific genes being identified having big aberrations

Patient ID	Chromosomal Location	gene	count
TCGA-06-017	chr1q43-q44	SDCCAG8	1
TCGA-06-015	chr1q32	SOX13	1
TCGA-06-0412;TCGA-06-015	chr1q32.1	PPP1R15B	2
TCGA-06-017	chr1q44	C1orf101	1
TCGA-06-017	chr1q44-qter	ZNF238	1
TCGA-06-017	chr1q44	CEP170	1
TCGA-06-015	chr1q32.1	ETNK2	1
TCGA-06-0412;TCGA-06-015	chr1q32	PIK3C2B	2
TCGA-06-015	chr1q32.1	GOLT1A	1
TCGA-06-017	chr1q44	C1orf100	1
TCGA-06-015	chr1q32.1	NFASC	1
TCGA-06-015	chr1q32.1	PLEKHA6	1
TCGA-06-017	chr1cen-q12	ADSS	1
TCGA-06-0412;TCGA-06-015	chr1q32	MDM4	2
TCGA-06-017	chr1q43-q44	AKT3	1
TCGA-06-015	chr1q32	KISS1	1
TCGA-06-015	chr1q32	REN	1
TCGA-06-020	chr3q28	IL1RAP	1
TCGA-06-020	chr3q28-q29	CLDN1	1
TCGA-06-020	chr3q28	CLDN16	1
TCGA-06-0241;TCGA-08-0524;TCGA-06-017	chr4q11-q12	KIT	3
TCGA-06-017	chr4q12	SGCB	1
TCGA-06-0241;TCGA-06-017	chr4q11-q12	KDR	2
TCGA-02-0440;TCGA-06-017	chr4q12	SCFD2	2
TCGA-06-017	chr4q12	LNX1	1
TCGA-02-0440;TCGA-06-0241;TCGA-08-0524;TCGA-06-017	chr4q11	CHIC2	4
TCGA-06-017	chr4q12	SPATA18	1
TCGA-08-052	chr4p13	RPL9	1
TCGA-02-0440;TCGA-06-0241;TCGA-08-0524;TCGA-06-017	chr4q11-q13	PDGFRA	4
TCGA-06-017	chr4q12	DCUN1D4	1
TCGA-06-0139;TCGA-06-013	chr4q13	UGT2B17	2
TCGA-08-052	chr4p14	LIAS	1
TCGA-06-017	chr4q12	USP46	1
TCGA-06-017	chr4q12	RASL11B	1
TCGA-02-0440;TCGA-06-017	chr4q12	FIP1L1	2
TCGA-06-0210;TCGA-02-0285;TCGA-02-0321;TCGA-02-0007;TCGA-06-0168;TCGA-02-0024;TCGA-02-043	chr5q35.3	BTNL3	7
TCGA-06-0122;TCGA-02-0021;TCGA-06-018	chr7p12	GBAS	3
TCGA-02-008	chr7q36.1	ACTR3B	1
TCGA-06-018	chr7p11.2	CCT6A	1
TCGA-02-033	chr7q31.2	TFEC	1

Patient ID	Chromosomal Location	gene	count
TCGA-02-004	chr7q31.3	PTPRZ1	1
TCGA-02-0333;TCGA-02-0285;TCGA-02-0317;TCGA-06-0157;TCGA-02-0071;TCGA-06-0169;TCGA-12-0619;TCGA-08-0352;TCGA-06-0209;TCGA-02-0021;TCGA-02-0260;TCGA-02-0289;TCGA-06-0126;TCGA-02-0064;TCGA-02-0422;TCGA-02-0089;TCGA-06-0133;TCGA-06-0211;TCGA-08-0514;TCGA-08-0531;TCGA-08-0529;TCGA-02-0043;TCGA-08-0358;TCGA-06-0137;TCGA-06-0127;TCGA-06-0145;TCGA-02-0046;TCGA-06-0122;TCGA-08-0525;TCGA-02-0038;TCGA-06-0125;TCGA-02-0269;TCGA-02-0009;TCGA-02-0106;TCGA-02-0116;TCGA-06-0143;TCGA-02-0290;TCGA-02-0003;TCGA-06-0148;TCGA-02-0083;TCGA-06-0152;TCGA-06-0158;TCGA-06-0187;TCGA-06-0185;TCGA-06-0173;TCGA-06-0182;TCGA-06-0645;TCGA-06-0409;TCGA-02-0430;TCGA-02-010	chr7p12	EGFR	50
TCGA-02-0333;TCGA-08-0352;TCGA-06-0209;TCGA-02-0021;TCGA-02-0289;TCGA-02-0064;TCGA-06-0133;TCGA-02-0089;TCGA-06-0211;TCGA-08-0514;TCGA-02-0043;TCGA-06-0137;TCGA-06-0127;TCGA-06-0145;TCGA-02-0046;TCGA-06-0122;TCGA-02-0038;TCGA-02-0116;TCGA-02-0290;TCGA-02-0083;TCGA-06-0187;TCGA-06-0185;TCGA-06-0645;TCGA-06-0173;TCGA-02-0430;TCGA-02-010	chr7q31.1-q31.33	LANCL2	26
TCGA-06-018	chr7p15.2-p15.1	PSPH	1
TCGA-12-061	chr7p14.1	C7orf10	1
TCGA-06-0209;TCGA-02-0021;TCGA-06-0211;TCGA-06-0122;TCGA-06-018	chr7p11.2	ZNF713	5
TCGA-02-0057;TCGA-02-0010;TCGA-08-0525;TCGA-02-002	chr7q34	MGAM	4
TCGA-02-0317;TCGA-06-0157;TCGA-02-0071;TCGA-06-0169;TCGA-12-0619;TCGA-06-0209;TCGA-02-0021;TCGA-02-0260;TCGA-02-0289;TCGA-06-0126;TCGA-06-0133;TCGA-06-0211;TCGA-08-0514;TCGA-08-0531;TCGA-08-0358;TCGA-06-0127;TCGA-06-0137;TCGA-06-0145;TCGA-02-0046;TCGA-06-0122;TCGA-08-0525;TCGA-06-0125;TCGA-02-0009;TCGA-02-0116;TCGA-02-0003;TCGA-02-0290;TCGA-06-0158;TCGA-06-0152;TCGA-06-0187;TCGA-06-0185;TCGA-06-0409;TCGA-06-0645;TCGA-06-0182;TCGA-06-0173;TCGA-02-043	chr7p11.2	SEC61G	35
TCGA-08-052	chr9q34	GTF3C5	1
TCGA-08-052	chr9q34.13	C9orf98	1
TCGA-08-052	chr9q34	TSC1	1
TCGA-08-052	chr9q34.13-q34.3	GBGT1	1
TCGA-08-052	chr9q34.3	CEL	1
TCGA-08-052	chr9q34.3	RALGDS	1
TCGA-08-052	chr9q34.13	GFI1B	1
TCGA-08-052	chr9q34	C9orf9	1
TCGA-06-017	chr12q12-q13	KRT6B	1
TCGA-06-017	chr12q12-q13	KRT3	1
TCGA-06-0129;TCGA-06-0209;TCGA-12-0616;TCGA-06-0177;TCGA-06-0152;TCGA-06-018	chr12q14	CDK4	6
TCGA-06-0209;TCGA-02-0052;TCGA-06-0152;TCGA-06-0187;TCGA-06-0173;TCGA-06-0182;TCGA-02-010	chr12q15	SLC35E3	7
TCGA-06-017	chr12q13	KRT8	1
TCGA-06-017	chr12q14.3	WIF1	1
TCGA-06-017	chr12q13.13	GRASP	1
TCGA-06-0648;TCGA-06-018	chr12q14	CAND1	2
TCGA-06-012	chr12q24.33	GLT1D1	1
TCGA-06-017	chr12q13.3	HOXC11	1
TCGA-06-017	chr12q12-q13	KRT6A	1

Patient ID	Chromosomal Location	gene	count
TCGA-12-061	chr12q21.32	CEP290	1
TCGA-06-012	chr12q13.3	NDUFA4L2	1
TCGA-06-012	chr12p13.3	CACNA1C	1
TCGA-06-0129;TCGA-06-0209;TCGA-12-0616;TCGA-06-0177;TCGA-06-0152;TCGA-06-018	chr12q13.3	TSPAN31	6
TCGA-06-017	chr12q14.1	XRCC6BP1	1
TCGA-06-017	chr12q13-q21	TAC3	1
TCGA-06-020	chr12p12.3	AEBP2	1
TCGA-06-017	chr12q14.2	DPY19L2	1
TCGA-06-0209;TCGA-06-0152;TCGA-06-0187;TCGA-06-0182;TCGA-02-010	chr12q15	CPSF6	5
TCGA-06-0182;TCGA-02-010	chr12q15	PTPRR	2
TCGA-06-018	chr12q13.2-q13.3	GLI1	1
TCGA-06-017	chr12q13	ACVR1B	1
TCGA-06-017	chr12q13	KRT18	1
TCGA-06-017	chr12q13.3	RDH16	1
TCGA-06-017	chr12q13.3	HOXC10	1
TCGA-06-0152;TCGA-06-0187;TCGA-02-010	chr12q15	CCT2	3
TCGA-06-0177;TCGA-06-018	chr12q14.1	LRIG3	2
TCGA-06-0152;TCGA-06-0187;TCGA-06-018	chr12q15	CNOT2	3
TCGA-06-0129;TCGA-06-0209;TCGA-06-0177;TCGA-06-0152;TCGA-06-018	chr12q13.1-q13.3	CYP27B1	5
TCGA-06-018	chr12q13.2-q13.3	DCTN2	1
TCGA-06-017	chr12q12-q13	KRT4	1
TCGA-12-061	chr12q21.32	TMTC3	1
TCGA-06-0187;TCGA-06-018	chr12q	KCNMB4	2
TCGA-06-018	---	MBD6	1
TCGA-06-0129;TCGA-06-0209;TCGA-06-018	chr12q13	OS9	3
TCGA-06-0209;TCGA-12-0616;TCGA-02-0337;TCGA-06-0177;TCGA-02-0052;TCGA-06-0648;TCGA-06-0152;TCGA-06-0187;TCGA-06-0173;TCGA-06-0182;TCGA-02-010	chr12q14.3-q15	MDM2	11
TCGA-06-0209;TCGA-12-0616;TCGA-02-0337;TCGA-06-0177;TCGA-02-0052;TCGA-06-0648;TCGA-06-0152;TCGA-06-0187;TCGA-06-0173;TCGA-06-0182;TCGA-02-010	chr12q14.3	CPM	11
TCGA-06-018	chr12q14.3	IRAK3	1
TCGA-06-018	chr12q15	C12orf28	1
TCGA-06-0129;TCGA-06-017	chr12q13.3	R3HDM2	2
TCGA-06-0129;TCGA-06-0209;TCGA-06-017	chr12q13-q14	TSFM	3
TCGA-06-018	chr12q13.3	INHBE	1
TCGA-06-0177;TCGA-06-018	chr12q14.1	FAM19A2	2
TCGA-06-0177;TCGA-06-018	chr12q14.3	RAB3IP	2
TCGA-06-018	chr12q13.2	MARS	1
TCGA-12-061	chr12q21.32	C12orf29	1
TCGA-06-017	chr12q13-q15	CTDSP2	1
TCGA-06-012	chr12q21.33	C12orf12	1
TCGA-06-0209;TCGA-02-0052;TCGA-06-0152;TCGA-06-0173;TCGA-06-0182;TCGA-02-010	chr12q15	NUP107	6
TCGA-06-017	chr12q13.11	SLC38A1	1
TCGA-06-017	chr12q12-q13	KRT7	1
TCGA-06-0129;TCGA-06-017	chr12q14.1	AVIL	2

Patient ID	Chromosomal Location	gene	count
TCGA-06-017	chr12q13.2	WIBG	1
TCGA-06-0209;TCGA-06-0152;TCGA-06-018	chr12q13-q15	YEATS4	3
TCGA-06-0177;TCGA-06-0152;TCGA-06-0182;TCGA-02-010	chr12q14.2	SRGAP1	4
TCGA-06-017	chr12q13.3	HOXC8	1
TCGA-06-017	chr12q13.3	HOXC9	1
TCGA-06-0129;TCGA-06-0209;TCGA-06-0177;TCGA-06-0152;TCGA-06-018	chr12q13	METTL1	5
TCGA-06-017	chr12q13.3	ZBTB39	1
TCGA-06-0209;TCGA-06-0152;TCGA-06-0187;TCGA-02-010	chr12q15	FRS2	4
TCGA-06-018	chr12q14	ARHGAP9	1
TCGA-06-0209;TCGA-06-018	chr12q15	LYZ	2
TCGA-06-018	chr12q15-q21	PTPRB	1
TCGA-06-012	chr12q13.3	STAC3	1
TCGA-06-015	chr12q14	LEMD3	1
TCGA-06-017	chr12q13.3	HOXC4	1
TCGA-06-014	chr14q32.33	PLD4	1
TCGA-06-014	chr14q32.32|14q32.32	AKT1	1
TCGA-06-014	chr14q32.33	KIAA0284	1
TCGA-06-014	chr14q32.33	ADSSL1	1
TCGA-06-018	chr22q11.23	GSTT1	1
TCGA-02-031	chr22q13.1-q13.2	APOBEC3A	1

Appendix B R code for data processing

```r
library(snapCGH);
library(limma);
targets <- readTargets("targets");
files <- dir(pattern = ".txt")
cols =
list(Rf="rMedianSignal",Gf="gMedianSignal",Rb="rBGMedianSignal
",Gb="gBGMedianSignal")
data = read.maimages(targets$FileName, columns=cols, source =
'agilent')
save(data, file = "data")
save.image(file = "importeddata.RData")
RG2 <- backgroundCorrect(data, method="minimum")
RG3 <- RG2[grep("chr", data$genes$SystematicName),]
RG3$genes$Chr <- gsub(":.+", "", RG3$genes$SystematicName)
RG4 <- RG3[-grep("_random", RG3$genes$Chr),]
RG4$genes$Chr <- gsub("chr", "", RG4$genes$Chr)
RG4$genes$Chr <- gsub("X", "23", RG4$genes$Chr)
RG4$genes$Chr <- gsub("Y", "24", RG4$genes$Chr)
RG4$genes$Chr <- as.numeric(as.character(RG4$genes$Chr))

RG4$genes$Position <- gsub(".+:", "",
RG4$genes$SystematicName) RG4$genes$Position <- gsub("-.+",
"", RG4$genes$Position)
RG4$genes$Position <-
as.numeric(as.character(RG4$genes$Position))

RG4$genes$end <- gsub(".+-", "", RG4$genes$SystematicName)
RG4$genes$end <- as.numeric(as.character(RG4$genes$end))
RG5 <- RG4[order(RG4$genes$Chr, RG4$genes$Position),]
RG5$weights <- array(1, dim(RG5))
RG5$weights[RG5$genes$chrom > 22,] <- 0
colnames(RG5$weights) <- colnames(RG5$G)
save(RG5, file = "RG5");
rm(list = ls());
gc();
load("RG5");

MA <- normalizeWithinArrays(RG5, bc.method="none",
method="none")
MA$targets <- readTargets("targets");
MA$design <- rep(1,dim(MA)[2])
save(MA, file = "MA");
MA2 <- processCGH(MA, maxChromThreshold=24,
method.of.averaging=mean, ID="ProbeName")
save.image(file = "MA.RData");
gc();
```

```
Seg.CBS <- runDNAcopy(MA2)
Seg.HHMM <- runHomHMM(MA2)
save.image(file = "seg.RData");        gc();
Seg.CBS.merged <- mergeStates(Seg.CBS, MergeType=1)
Seg.HHMM.merged <- mergeStates(Seg.HHMM, MergeType = 1)
save(Seg.CBS.merged, Seg.CBS, file="SegCBS")
save(Seg.HHMM.merged, Seg.HHMM, file = "SegHHMM")
save.image(file = "mergesStates.RData")
```

R code for data simulation

```
library(snapCGH);
library(limma);
simu_data1 <- simulateData(nArrays = 4, seed = 1, output =
TRUE)
simu_data2 <- simulateData(nArrays = 20, seed = 1, output =
TRUE)
simu_data3 <- simulateData(nArrays = 50, seed = 1, output =
TRUE)
simu_data4 <- simulateData(nArrays = 100, seed = 1, output =
TRUE)
simu_data5 <- simulateData(nArrays = 170, seed = 1, output =
TRUE)

sim.HHMM1 <- runHomHMM(simu_data1)
sim.DNAcopy1 <- runDNAcopy(simu_data1)
sim.HHMM2 <- runHomHMM(simu_data2)
sim.DNAcopy2 <- runDNAcopy(simu_data2)
sim.HHMM3 <- runHomHMM(simu_data3)
sim.DNAcopy3 <- runDNAcopy(simu_data3)
sim.HHMM4 <- runHomHMM(simu_data4)
sim.DNAcopy4 <- runDNAcopy(simu_data4)
sim.HHMM5 <- runHomHMM(simu_data5)
sim.DNAcopy5 <- runDNAcopy(simu_data5)

comp1 <- compareSegmentations(simu_data1, offset = 0,
sim.HHMM1, sim.DNAcopy1)
comp2 <- compareSegmentations(simu_data2, offset = 0,
sim.HHMM2, sim.DNAcopy2)
comp3 <- compareSegmentations(simu_data3, offset = 0,
sim.HHMM3, sim.DNAcopy3)
comp4 <- compareSegmentations(simu_data4, offset = 0,
sim.HHMM4, sim.DNAcopy4)
comp5 <- compareSegmentations(simu_data5, offset = 0,
sim.HHMM5, sim.DNAcopy5)


par(mfrow = c(4,3))
```

```r
boxplot(comp1$TPR ~ row(comp1$TPR), col = c("red", "blue"),
main = "True Positive rate")
boxplot(comp1$FDR ~ row(comp1$FDR), col = c("red", "blue"),
main = "False Positive rate")
boxplot(comp2$TPR ~ row(comp2$TPR), col = c("red", "blue"),
main = "True Positive rate")
boxplot(comp2$FDR ~ row(comp2$FDR), col = c("red", "blue"),
main = "False Positive rate")
boxplot(comp3$TPR ~ row(comp3$TPR), col = c("red", "blue"),
main = "True Positive rate")
boxplot(comp3$FDR ~ row(comp3$FDR), col = c("red", "blue"),
main = "False Positive rate")
boxplot(comp4$TPR ~ row(comp4$TPR), col = c("red", "blue"),
main = "True Positive rate")
boxplot(comp4$FDR ~ row(comp4$FDR), col = c("red", "blue"),
main = "False Positive rate")
boxplot(comp5$TPR ~ row(comp5$TPR), col = c("red", "blue"),
main = "True Positive rate")
boxplot(comp5$FDR ~ row(comp5$FDR), col = c("red", "blue"),
main = "False Positive rate");

## R code for Survival analysis

library(survival)
library(KMsurv)

files <- list.files(pattern=".txt$");
pdf("survival_plot.pdf", width=8, height=10)
par(mfrow=c(4,3))
for(i in files) {
d <- read.table(i, header=TRUE, sep="\t");
fit <- survfit(Surv(d$suv, d$dead) ~d[,5], data = d)
fit2 <- survdiff(Surv(d$suv, d$dead) ~d[,5], data = d)
pval <- 1-pchisq(fit2$chisq, 1)
f <- length(d[d[,5]==1,4]);

p <- format(pval, digit = 2)
title = paste("pvalue", ":", p)
if((pval <0.005)&(f>3)){
write.table(pval, paste(i, c(".out"), sep=""), quote=FALSE,
sep="\t", col.names = "pvalue");
plot(fit, main = title, lty = c(1,3), xlab = "months", ylab =
"Survival Prob")
legend(40, 1, c("no change", "change"), lty = c(1,3))};
}
dev.off()
```

Glossary and abbreviation

Aneuploidy: The occurrence of one or more extra or missing chromosomes leading to an unbalanced chromosome complement, or, any chromosome number that is not an exact multiple of the haploid number

BioConductor: BioConductor represents the web site of www.bioconductor.org where hoses many R packages used for the microarray data processing and analyzing.

CBS: Circular Binary Segmentation

Cell cycle: The cell cycle, or cell-division cycle, is the series of events that takes place in a cell leading to its division and duplication (replication). In cells without a nucleus (prokaryotic), the cell cycle occurs via a process termed binary fission. In cells with a nucleus (eukaryotes), the cell cycle can be divided in two brief periods: interphase—during which the cell grows, accumulating nutrients needed for mitosis and duplicating its DNA—and the mitosis (M) phase, during which the cell splits itself into two distinct cells, often called "daughter cells".

Centromere: The specialized region of a chromosome to which spindle fibers attach during cell division

CGH: Comparative genomic hybridization

Comparative genomic hybridization (CGH) is a technique that allows the detection of losses and gains in DNA copy number across the entire genome without prior knowledge of specific chromosomal abnormalities. Comparative genomic hybridization utilizes the hybridization of differentially labeled tumor and reference DNA to generate a map of DNA copy number changes in tumor genomes. Comparative genomic hybridization is an ideal tool for analyzing chromosomal imbalances in archived tumor material and for examining possible correlations between these findings and tumor phenotypes.

Chromosome: In prokaryotes, the intact DNA molecule containing the genome. In eukaryotes, a DNA molecule complexed with RNA and protein into a threadlike structure containing a linear array of genes.

Chromosomal aberration: Any type of change in the chromosome structure or number (deficiencies, duplications, translocations, inversions, etc.). Although it can be a mechanism for enhancing genetic diversity, such alterations are usually fatal or ill-adaptive, especially in animals.

CLAC: Cluster Along Chromosomes

Diploid: The condition of having two of each chromosome. Somatic cells of higher plants and animals are normally diploid.

DLRs: Derivative Log Ratio spread

DNA: (deoxyribonucleic acid) The macromolecule that contains genetic information and comprises the genes. DNA consists of a chain of deoxyribonucleotides joined by phosphodiester linkages. Each deoxyribonucleotide consists of a nitrogenous base attached to the sugar deoxyribose, which in turn has a phosphate group attached at its 5' position.

FISH: Fluorescent in situ hybridization

Fluorescence in situ hybridization (FISH) is a laboratory technique for detecting and locating a specific DNA sequence on a chromosome. The technique relies on exposing chromosomes to a small DNA sequence called a probe that has a fluorescent molecule attached to it. The probe sequence binds to its corresponding sequence on the chromosome.

Gene: The fundamental unit of heredity; a segment of DNA found at a fixed location on a chromosome that codes for a single polypeptide.

Gene expression: The process of RNA and protein production by which genes exert their phenotypic effects on an organism.

Genome: A genome is an organism’s complete set of DNA, including all of its genes. Each genome contains all of the information needed to build and maintain that organism. In humans, a copy of the entire genome—more than 3 billion DNA base pairs—is contained in all cells that have a nucleus.

GLAD: Gain and Loss Analysis of DNA

GO: Genome Ontology

The Gene Ontology (GO) project is a collaborative effort to address the need for consistent descriptions of gene products in different databases. The GO collaborators are developing three structured, controlled vocabularies (ontologies) that describe gene products in terms of their associated biological processes, cellular components and molecular functions in a species-independent manner.

HHMM: Homogenous Hidden Markov Model

Karyotypes: A photographic representation of the chromosomes of a single cell, cut and arranged in pairs based on their banding pattern and size according to a standard classification

Microarray: Sets of miniaturized chemical reaction areas that may also be used to test DNA fragments, antibodies, or proteins, by using a chip having immobilised target and hybridising them with probed sample. the color we get from the chip after hybridisation is then scanned and the data is analysed by a soft ware to find the expression level.

Ploidy: A term referring to the basic set of chromosomes or multiples of that set.

QC: Quality Control

Sex chromosome: A chromosome involved in sex determination. An example of this are the X and Y chromosomes of humans.

TCGA: The Cancer Genome Atlas

The Cancer Genome Atlas (TCGA) is a comprehensive and coordinated effort to accelerate our understanding of the genetics of cancer using innovative genome analysis technologies.

WHO: World Health Organization

Bibliography

Akaike, H. (1969), "Fitting Autoregressive Models for Prediction," *Annals of the Institute of Statistical Mathematics*, 243-247.

Armitage, P., and Doll, R. (1954), "The Age Distribution of Cancer and a Multi-Stage Theory of Carcinogenesis," *Br J Cancer*, **8**, 1-12.

Benjamini, Y, Hochberg, Y. (1995), "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *J Roy Statist Soc Ser B (Methodological)*, **57**, 289-300.

Cantley, L. C., and Neel, B. G. (1999), "New Insights into Tumor Suppression: Pten Suppresses Tumor Formation by Restraining the Phosphoinositide 3-Kinase/Akt Pathway," *Proc Natl Acad Sci U S A*, **96**, 4240-4245.

Celep, F., Karaguzel, A., Ozgur, G. K., and Yildiz, K. (2003), "Detection of Chromosomal Aberrations in Prostate Cancer by Fluorescence in Situ Hybridization (Fish)," *Eur Urol*, **44**, 666-671.

Chen, H. I., Hsu, F. H., Jiang, Y., Tsai, M. H., Yang, P. C., Meltzer, P. S., Chuang, E. Y., and Chen, Y. (2008), "A Probe-Density-Based Analysis Method for Array Cgh Data: Simulation, Normalization and Centralization," *Bioinformatics*, **24**, 1749-1756.

Chu, E. C., and Tarnawski, A. S. (2004), "Pten Regulatory Functions in Tumor Suppression and Cell Biology," *Med Sci Monit*, **10**, RA235-241.

Comincini, S., Paolillo, M., Barbieri, G., Palumbo, S., Sbalchiero, E., Azzalin, A., Russo, M. A., and Schinelli, (2009), "Gene Expression Analysis of an Egfr Indirectly Related Pathway Identified Pten and Mmp9 as Reliable Diagnostic Markers for Human Glial Tumor Specimens," *J Biomed Biotechnol*, **2009**, 924565.

Daruwala, R. S., Rudra, A., Ostrer, H., Lucito, R., Wigler, M., and Mishra, B. (2004), "A Versatile Statistical Analysis Algorithm to Detect Genome Copy Number Variation," *Proc Natl Acad Sci U S A*, **101**, 16292-16297.

Fan, Q. W., Cheng, C. K., Nicolaides, T. P., Hackett, C. S., Knight, Z. A., Shokat, K. M., and Weiss, W. A. (2007), "A Dual Phosphoinositide-3-Kinase Alpha/Mtor Inhibitor Cooperates with Blockade of Epidermal Growth Factor Receptor in Pten-Mutant Glioma," *Cancer Res*, **67**, 7960-7965.

Fridlyand, J., and Dimitrov, P. (2008), "Bioconductor's aCGH Package," *BioConductor*.

Fridlyand, j., Snijders, A. M., Pinkel, D., Albertson, D. G., and Jain, A. N. (2004), "Hidden Markov Medels Approach to the Analysis of Array Cgh Data," *Journal of Multivariate Analysis*, 132-151.

Furnari, F. B., Fenton, T., Bachoo, R. M., Mukasa, A., Stommel, J. M., Stegh, A., Hahn, W. C., Ligon, K. L., Louis, D. N., Brennan, C., Chin, L., DePinho, R. A., and Cavenee, W. K. (2007), "Malignant Astrocytic Glioma: Genetics, Biology, and Paths to Treatment," *Genes Dev*, **21**, 2683-2710.

Gordon K. Smyth, M. R., Natalie Thorne, James Wettenhall and Wei Shi. (2010), "Limma: Linear Models for Microarray Data User's Guide," *BioConductor*.

Hsu, L., Self, S. G., Grove, D., Randolph, T., Wang, K., Delrow, J. J., Loo, L., and Porter, P. (2005), "Denoising Array-Based Comparative Genomic Hybridization Data Using Wavelets," *Biostatistics*, **6**, 211-226.

Huang, J., Gusnanto, A., O'Sullivan, K., Staaf, J., Borg, A., and Pawitan, Y. (2007), "Robust Smooth Segmentation Approach for Array Cgh Data Analysis," *Bioinformatics*, **23**, 2463-2469.

Hupe, P., Stransky N, Thiery JP, Radvanyi F, Barillot E. (2004), "Analysis of Array Cgh Data: From Signal Ratio to Gain and Loss of DNA Regions," *Bioinformatics*, **20**, 3412-3420.

Johnston, J. B., Navaratnam, S., Pitz, M. W., Maniate, J. M., Wiechec, E., Baust, H., Gingerich, J., Skliris, G. P., Murphy, L. C., and Los, M. (2006), "Targeting the Egfr Pathway for Cancer Therapy," *Curr Med Chem*, **13**, 3483-3492.

Kamijo, T., Weber, J. D., Zambetti, G., Zindy, F., Roussel, M. F., and Sherr, C. J. (1998), "Functional and Physical Interactions of the Arf Tumor Suppressor with P53 and Mdm2," *Proc Natl Acad Sci U S A*, **95**, 8292-8297.

Knobbe, C. B., and Reifenberger, G. (2003), "Genetic Alterations and Aberrant Expression of Genes Related to the Phosphatidyl-Inositol-3'-Kinase/Protein Kinase B (Akt) Signal Transduction Pathway in Glioblastomas," *Brain Pathol*, **13**, 507-518.

Lai, W. R., Johnson, M. D., Kucherlapati, R., and Park, P. J. (2005), "Comparative Analysis of Algorithms for Identifying Amplifications and Deletions in Array Cgh Data," *Bioinformatics*, **21**, 3763-3770.

Magi, A., Benelli, M., Marseglia, G., Nannetti, G., Scordo, M. R., and Torricelli, F. (2010), "A Shifting Level Model Algorithm That Identifies Aberrations in Array-Cgh Data," *Biostatistics*, **11**, 265-280

Marioni, J. C., Thorne, N. P., and Tavare, S. (2006), "Biohmm: A Heterogeneous Hidden Markov Model for Segmenting Array Cgh Data," *Bioinformatics*, **22**, 1144-1146.

Mulshine, J. L., Avis, I., Treston, A. M., Mobley, C., Kaspryzyk, P., Carrasquillo, J. A., Larson, S. M., Nakanishi, Y., Merchant, B., Minna, J. D., et al., (1988), "Clinical Use of a Monoclonal Antibody to Bombesin-Like Peptide in Patients with Lung Cancer," *Ann N Y Acad Sci*, **547**, 360-372.

Murphree, A. L., and Benedict, W. F. (1984), "Retinoblastoma: Clues to Human Oncogenesis," *Science*, **223**, 1028-1033.

Nakao, K., Mehta, K. R., Fridlyand, J., Moore, D. H., Jain, A. N., Lafuente, A., Wiencke, J. W., Terdiman, J. P., and Waldman, F. M. (2004), "High-Resolution Analysis of DNA Copy Number Alterations in Colorectal Cancer by Array-Based Comparative Genomic Hybridization," *Carcinogenesis*, **25**, 1345-1357.

Ohgaki, H., Dessen, P., Jourde, B., Horstmann, S., Nishikawa, T., Di Patre, P. L., Burkhard, C., Schuler, D., Probst-Hensch, N. M., Maiorka, P. C., Baeza, N., Pisani, P., Yonekawa, Y., Yasargil, M. G., Lutolf, U. M., and Kleihues, P. (2004), "Genetic Pathways to Glioblastoma: A Population-Based Study," *Cancer Res*, **64**, 6892-6899.

Olshen, A. B., Venkatraman, E. S., Lucito, R., and Wigler, M. (2004), "Circular Binary Segmentation for the Analysis of Array-Based DNA Copy Number Data," *Biostatistics*, **5**, 557-572.

Polzehl, J., and Spokoiny, V. G. (2000), "Adaptive Weights Smoothing with Applications to Image Restoration," *Journal of the Royal Statistical Society Series B-Statistical Methodology*, **62**, 335-354.

Pomerantz, J., Schreiber-Agus, N., Liegeois, N. J., Silverman, A., Alland, L., Chin, L., Potes, J., Chen, K., Orlow, I., Lee, H. W., Cordon-Cardo, C., and DePinho, R. A. (1998), "The Ink4a Tumor Suppressor Gene Product, P19arf, Interacts with Mdm2 and Neutralizes Mdm2's Inhibition of P53," *Cell*, **92**, 713-723.

Schwarz, G. (1978), "Estimating the Dimension Os a Model," *Ann. Statist.*, 461-464.

Sen A., Srivastava, M. S. (1975), "On Tests for Detecting Change in Mean," *The annals of Statistics*, **3**, 98-108.

Sen, S. (2000), "Aneuploidy and Cancer," *Curr Opin Oncol*, **12**, 82-88.

Serrano, M., Hannon, G. J., and Beach, D. (1993), "A New Regulatory Motif in Cell-Cycle Control Causing Specific Inhibition of Cyclin D/Cdk4," *Nature*, **366**, 704-707.

Sherr, C. J., and McCormick, F. (2002), "The Rb and P53 Pathways in Cancer," *Cancer Cell*, **2**, 103-112.

Siegmund, D. (1986), "Boundary Crossing Probabilities and Statistical Applications," *Annals of Statistics*, **14**, 361-404.

Smith, M. L., Marioni, J.C., Hardcastle, T.J., Thorne, N.P. (2006), "Snapcgh: Segmentation, Normalization and Processing of Acgh Data Users' Guide," *Bioconductor*.

Stott, F. J., Bates, S., James, M. C., McConnell, B. B., Starborg, M., Brookes, S., Palmero, I., Ryan, K., Hara, E., Vousden, K. H., and Peters, G. (1998), "The Alternative Product from the Human Cdkn2a Locus, P14(Arf), Participates in a Regulatory Feedback Loop with P53 and Mdm2," *EMBO J*, **17**, 5001-5014.

Tang, J., Shao, W., Dorak, M. T., Li, Y., Miike, R., Lobashevsky, E., Wiencke, J. K., Wrensch, M., Kaslow, R. A., and Cobbs, C. S. (2005), "Positive and Negative Associations of Human Leukocyte Antigen Variants with the Onset and Prognosis of Adult Glioblastoma Multiforme," *Cancer Epidemiol Biomarkers Prev*, **14**, 2040-2044.

TCGA. (2008), "Comprehensive Genomic Characterization Defines Human Glioblastoma Genes and Core Pathways," *Nature*, **455**, 1061-1068.

Therneau, T. Lumley, T. (2009), "Survival: Survival Analysis, Including Penalised Likelihood," *Bioconductor*.

Vanhaesebroeck, B., Welham, M. J., Kotani, K., Stein, R., Warne, P. H., Zvelebil, M. J., Higashi, K., Volinia, S., Downward, J., and Waterfield, M. D. (1997), "P110delta, a Novel Phosphoinositide 3-Kinase in Leukocytes," *Proc Natl Acad Sci U S A*, **94**, 4330-4335.

Veltman, J. A., Fridlyand, J., Pejavar, S., Olshen, A. B., Korkola, J. E., DeVries, S., Carroll, P., Kuo, W. L., Pinkel, D., Albertson, D., Cordon-Cardo, C., Jain, A. N., and Waldman, F. M. (2003), "Array-Based Comparative Genomic Hybridization for Genome-Wide Screening of DNA Copy Number in Bladder Tumors," *Cancer Res*, **63**, 2872-2880.

Wang, P., Kim, Y., Pollack, J., Narasimhan, B., and Tibshirani, R. (2005), "A Method for Calling Gains and Losses in Array Cgh Data," *Biostatistics*, **6**, 45-58.

Willenbrock, H., and Fridlyand, J. (2005), "A Comparison Study: Applying Segmentation to Array Cgh Data for Downstream Analyses," *Bioinformatics*, **21**, 4084-4091.

Xu, X. L., and Kapoun, A. M. (2009), "Heterogeneous Activation of the Tgfbeta Pathway in Glioblastomas Identified by Gene Expression-Based Classification Using Tgfbeta-Responsive Genes," *J Transl Med*, **7**, 12.

Yan, J. (2010), "Data Sets from Klein and Moeschberger (1997), Survival Analysis," *Bioconductor*.