

Comparative Analysis of Chromosomal Variation Altered New Region for Detecting Biomarker in Liver Cancer

Esraa M. Hashem¹, Mai S. Mabrouk^{1*}, Ayman M. Eldeib²

¹Biomedical Engineering Department, Misr University for Science and Technology (MUST University), Egypt
²Systems and Biomedical Engineering, Faculty of engineering, Cairo University, Giza, Egypt

Abstract Liver cancer is one of the most widely recognized malignancies in adults, hepatocellular carcinoma (HCC) is an out face malignancy. Nearly instances of HCC are identified with a preceding stage, and, in this manner, medication choices are restricted. Most Hepatic carcinoma emerge from the accumulation of genetic abnormalities and potentially induced by exterior etiological factors. Genome-wide gene expression profiling is a generally utilized system for discovering new potential biomarkers for prediction and diagnosis of disease severity. In this work, two statistical techniques (circular binary segmentation and discrete stationary wavelet transform) are utilized for detection of copy number alterations to test single-nucleotide polymorphism array of ten cell lines hepatic tumor patients on certain chromosomes, which are frequently detected in hepatic cancer have been proposed. Results detect a new mutated chromosomal amplification region (4q22.1) has been identified through our investigation utilizing both methods. This region sets up, SPP1, and AFF1 tumor suppressor genes and has not observed previously, that opening new frontiers that can be involved in liver carcinogenesis.

Keywords Hepatocellular carcinoma, Circular binary segmentation, Discrete stationary wavelet transform, Single-nucleotide polymorphism

1. Introduction

Cancer has been over and again distinguished as a controlled disease. Certainly, the procedure of tumor growth from primary to malignant structures, including metastasis to different cells includes the conflict between various cell lineages contend to adjust to the changing conditions both inside and around the tumor. This complex operation is almost associated with genetic alterations and the incidence of mutations [1]. Greater than 80% of hepatic cancers are hepatocellular carcinoma (HCCs), It is once in a while identified early [2]. The end stage of many different chronic liver diseases is liver cirrhosis, it is normally diagnosed at an end-stage when effective treatment is impossible. Patient with cirrhosis have an over risk of liver tumor. Mostly cirrhosis created because of: hepatitis C (HCV), alcoholic liver disease (ALD), hepatitis B (HBV), auto-immune hepatitis (AIH), and other risk factors [3], as shown in Figure 1-1. As stated by those current studies, the large part of liver cancer patients contracted the disease from the accumulation the abnormalities of genetics, possibly rise by exterior etiological factors especially HCV and HBV contaminations [4]. So far, no significant proof has found that enhances

survival utility with surveillance of high-risk patients. The diagnostic tools which utilized commonly to attempt to enhance patient survival are: biomarkers, liver biopsy, radiographic imaging and serum tumor marker alpha-fetoprotein (AFP) [5].

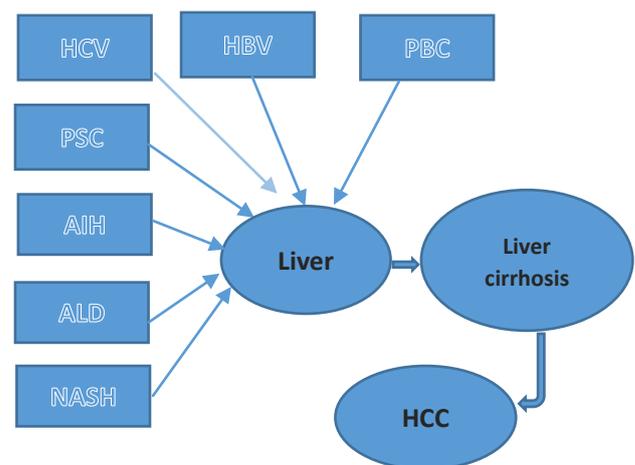


Figure 1.1. Causes of liver cirrhosis

Mainly human malignancy cancers are genetic diseases raise by genome aberrations. Chromosomal instability is one of the characteristics in cancer, its outcomes in the structural aberration of DNA copy number alterations (CNAs). That alteration changes the scale of gene expression, which alter

* Corresponding author:
msm_eng@yahoo.com (Mai S. Mabrouk)
Published online at <http://journal.sapub.org/computer>
Copyright © 2017 Scientific & Academic Publishing. All Rights Reserved

normal growth control and survival pathway [5]. Genetic variation between a similar tumor sort in various people may result in metabolic differences between one malignancy tissue and another that can influence disease progression or response to therapy.

CNA is necessary for each of the major understanding of Hepatic carcinoma and its detection. Among such alterations, frequently detected of DNA copy number gains at chromosome 1, 8, 11, 17, and q [6-8] and losses at chromosome 4, 8, 9, 13, 16, and 17 have been specified in HCC [9-11] using traditional methodology.

Single-nucleotide polymorphism array (SNP) are single base alterations between genomes within a species. SNPs is helpful markers for diseases in haplotype-based association studies and utilized to recognize chromosomal alterations to identify mutations [12]. Since the most recent decade, the entire genome SNP array has been the basic element in a highly productive synergistic relationship between advances in biological understanding and computational approaches. The Illumina Bead Array has steadily expanded limit throughout the years from 100 000 SNPs to the current one million [13]. SNP array is a variation at one position in a DNA sequence among individuals. Recall that the DNA sequence is formed from a chain of nucleotide bases: A, C, G, and T. If more than 1% of a population does not carry the same nucleotide at a specific position in the DNA sequence, then this aberration can be classified as an SNP.

It is needed to identify new biomarkers for early diagnosis of HCC and it is important to find new methods that can evaluate the prognosis of HCC patients.

Numerous automated algorithms are used to define the characteristic of genomic profiles; Cooper et al. utilize SNP conditional mixture modeling (SCIMM) by using a mixture-likelihood clustering methodology within the R package to characterize deletions of CNA [14]. Franke et al. give a combined program focused on SNP interpretation [15].

A lot of automated algorithms to detect biomarkers in CNAs have a solid restriction it is the noise generated from signals [16]. To fill this obstacle, Circular Binary segmentation (CBS) and Discrete Stationary Wavelet transforms (DSWT) are utilized, that aids in the perception of the genomic basis of the disease progression and improve a method to find biomarkers for early identification of liver cancer.

2. Materials and Methods

The most profuse genetic variants are SNPs, which supply the ultimate comprehensive resource for verification of genetic diversity [17]. The progress in high-throughput sequencing technology has made copy number discovery much easier, the application of known CNA information means that we can target structural variation in a sample using less expensive techniques such as the SNP array without a large reduction in genome-wide coverage.

2.1. Experimental Setup

The data used in this study is SNP Genotyping Array, performed on Human Duov1 DNA investigation Bead Chip (HumanCNV370v1), includes the usual SNP content items on a HumanHap300-Duo Genotyping Bead Chip, with an extra 52,167 marker intended to specifically target close-by 14,000 CNA areas of the genome, for an aggregate of more than 370,000 markers. Ten HCC cell lines process according to the manufacturers' instructions. Raw data processed utilizing an in-house pipeline to take out a segment of copy number and gene-summarized copy number assessments. The data on the normalized log R ratios are obtainable in GEO [<http://www.ncbi.nlm.nih.gov/geo>] with accession number [GSE38207].

2.2. Statistical Methods

Circular Binary Segmentation (CBS):

Because of the complication of genetic variations in tumor cells, high density SNP array investigation is an extremely requesting method in the field of cancer research. The implementation of SNP arrays in combination with tissue array, PCR and functional analysis advances will finally prompt to the full revelation of the complex genetic alterations in human cancers.

CBS can give a characteristic approach to segment a chromosome into contiguous regions iteratively searching for parts of chromosomal regions and bypasses parametric modeling of the data with its utilize of a variation reference division. Once the chromosome partitioned, a copy number can appraise from the segments and discover the change points on SNP array data with the help of assistance information such as the ploidy of the chromosome [18]. The algorithm stops when no more statistically significant change-points are situated, according to a heuristically chosen significance threshold. This information will give the locations of copy number aberrations, assigned by Z_i as appeared in (1). Where $S_i = X_1 + \dots + X_i$ $1 \leq i \leq n$ are the partial sums and $X_1 \dots X_n$ is the log R ratios of the intensities. The probability proportion measurement for checking the null hypothesis that there is no change against the alternative that there is exactly one change at an unknown location i is given by (2) [19].

$$Z_i = \left[\frac{1}{i} + \frac{1}{n-i} \right] - \frac{1}{2} \left[\frac{S_i}{i} - \frac{S_{n-i}}{n-i} \right] \quad (1)$$

$$ZB = \max_{1 \leq i < n} \quad (2)$$

The CBS model makes losses or gains of CNVs discrete. These variations happen in adjacent regions of the chromosome that frequently hold different markers even entire chromosome regions. CBS precisely segment data by distinguishing change-points utilizing a maximal- t test [20]. The array information to be utilized for recognition change-point are the log R ratio of settled intensities listed by the marked positions. Note that there might be different change-points in a particular chromosome, everyone correlating to an adjustment in the copy number in the selected sample.

Discrete Stationary Wavelet transforms (DSWT):

A discrete wavelet is any wavelet transform, for which the wavelets are discretely tested. The other form of representing the signal is the transform of a signal. It does not modify the information, content currently in the signal [21]. The Wavelet change supplies a time-frequency illustration of the signal. The wavelet transforms enable high compression ratios with good quality of reconstruction.

Bases of wavelet are produced by the dilation and translation of a single wavelet role (x) called the Haar basis [22]. It is the simplest wavelet base, which produced by the Haar function in (3), it distinguishes statistically significant breakpoints in the data, utilizing the maxima of the Haar wavelet change, and segments accordingly [5].

$$\psi(x) = \begin{cases} -1/\sqrt{2} & 1 < x \leq 0 \\ 1/\sqrt{2} & 0 < x \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

After using Haar function there is need to eliminate noise from the signal through decomposition the input signal into an approximation sub-bands and a set of detail sub-bands at different resolution scales utilizing a set of high pass filter and low pass filters [23]. A limit level assigned to each dyadic resolution level by the guideline of minimizing the Stein Unbiased Estimate of Risk (SURE) for threshold estimates. De-blurring increases the sharpness signal features by boosting the high frequencies, whereas denoising tries to remove whatever noise is present regardless of the spectral content of a noisy signal. Restoration is a kind of denoising that tries to retrieve the original signal with optimal balancing of de-blurring and smoothing Wavelet transforms enable us to represent signals with a high degree of sparsity. This is the principle behind a non-linear wavelet based signal estimation technique known as **wavelet denoising**.

3. Results

HCC is a standout amongst the ultimate recurrent cancer entities worldwide, Hepatitis C and Hepatitis B are the main underlying basic causes of chronic liver disease which leads to cirrhosis. The tumor recurrence is the major restriction of long survival after liver transplantation. Most recent advancement of technologies for gene expression profiling and genome copy number investigation enable the comprehensive characterization of entire cancer genomes. This information will enhance our comprehension of the epigenetic and genetic aberrations leading to the evolution of HCC, as well as ultimately to provide information that helps in the treatment of patients [24].

By performing CBS algorithm in HCC cell lines. A region of 4q25-26 losses observed in 10 out of 10-HCC (100%), the allelic loss of chromosome 4q were significantly associated with HCC has elevated serum AFP [25]. A loss of 8p21.3 was observed in 10 out of 10-HCC (100%), loss of 9p was observed in 10 out of 10-HCC (100%) also loss of 16q11.2

and 17p13.1 observed in 10 out of 10-HCC (100%), all the results are summarized in table 3.1. While performing DSWT algorithm, it has been noticed that, the most commonly gained region of chromosome 1q22.1-23.1 in 4 out of 10-hepatocellular carcinoma (40%). Gains of 1q are the most frequent variation and occur early during tumor-genesis [26]. This region contains the tumor suppressors genes JTB, CCT3and COPA [27]. A loss of 4q was detected in 10 out of 10-HCC (100%). Loss for 9p additionally observed in 10 out of 10-HCC (100), the results summarized in table 3.1. This finding agrees with those of previous studies [28] leading to the assumption that these aberrations could have an essential part in the preparation of HCC, probably based on variations of genes lying on these chromosome segments.

Table 3.1. Prediction accuracy of each chromosome for 10-HCC sample

Chromosome CNAs type	Percentage of CBS sensitivity	Percentage of DSWT sensitivity
Gain1q21-23	3/10(30%)	4/10(40%)
Gain4q22.1	10/10(80%)	10/10(100%)
Gain8q24.12	4/10(40%)	2/10(20%)
Gain11q13.2-13.3	5/10(50%)	5/10(50%)
Gain17q12-21	2/10(20%)	5/10(50%)
Gain20q12	5/10(50%)	1/10(10%)
Loss4q25-26	10/10(100%)	10/10(100%)
Loss8p21.3	10/10(100%)	6/10(60%)
Loss9p21.3	10/10(100%)	10/10(100%)
Loss13q12.11	6/10(60%)	4/10(40%)
Loss16q11.2	10/10(100%)	4/10(40%)
Loss17p13.1	10/10(100%)	9/10(90%)

While performing CBS and DSWT the analysis shows that, a new altered region has been detected by amplification of chromosome 4 region **4q22.1** in 8 out of 10 HCC cell lines (80%) using CBS and 10 out of 10 HCC cell lines (100%) of SSP1 and AFF1 genes using DSWT as shown in figure (3-1,2). This observing has not previously reported being required in liver carcinogenesis before Esraa. M, et.al study [22].

4. Discussions

The liver has many identified functions related to metabolism, processing, and immunity, the liver disease indicates a different number of concerns for the delivery of medical consideration. Large-scale CN changes in the genomes of humans have been disregarded mostly because of technical challenges in scoring these changes genomes-wide with a favorable resolution.

Gene expression profiling can prompt to discover new markers that may help enhance diagnostic for detect HCC, represent novel targets for therapeutic agents, and permit us to identify regions in the HCC genome that have undergone recurrent high-level focal amplifications or deletions. It

should be considered for future investigation on detecting CNA regions.

5. Conclusions

Hepatic carcinoma is a tumor of the liver, which is generally emerging in the background of chronic liver diseases. The assessment for patients with liver cancer is poor because of, the low chance of therapeutic treatment.

The symptoms of HCC are usually absent. Further, the improvement of early detection markers is a vital missing component in methodologists to reduce HCC mortality.

Chromosome aberrations in cancer cells can be resolved with a rising number of large-scale genomic and molecular

genetics SNP array technology. However, the recent advancement of alteration of gene expression profiling and genome-wide copy number investigation have allowed the comprehensive selecting properties of whole cancer genomes. This finding enhances our comprehension of the epigenetic aberrations and genetic aberrations driving the development of liver cancer, instead of ultimately extend guidance to designed treatment of patients.

SNP array is the most widely recognized strategy to examine the copy number of DNA sequences. It provides accurate information of CNAs which is essential to distinguish deletion and amplification in the chromosome. A better perception of gene expression in hepatic cancer can aid to comprehend the pathophysiology of HCC and enhanced prognostication.

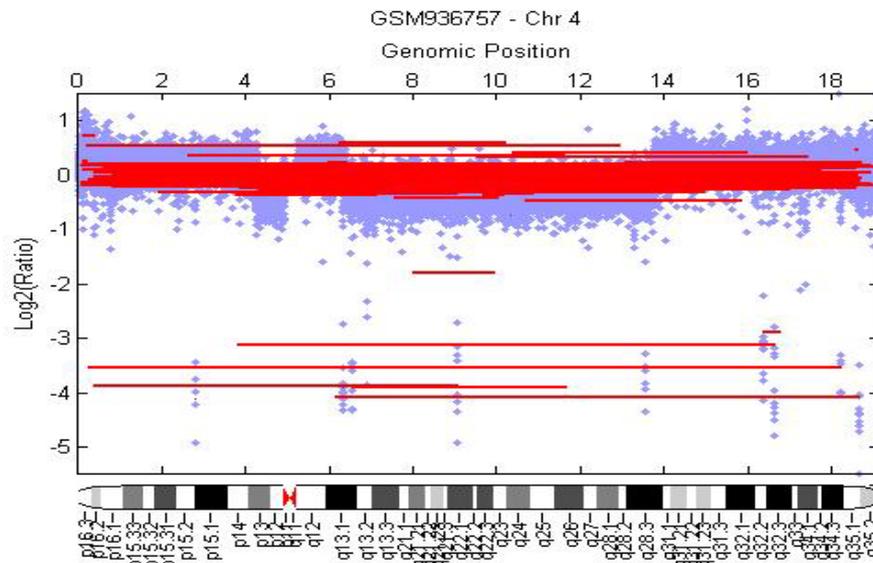


Figure 3.1. Gain of chromosome 4q22.1 using CBS algorithm

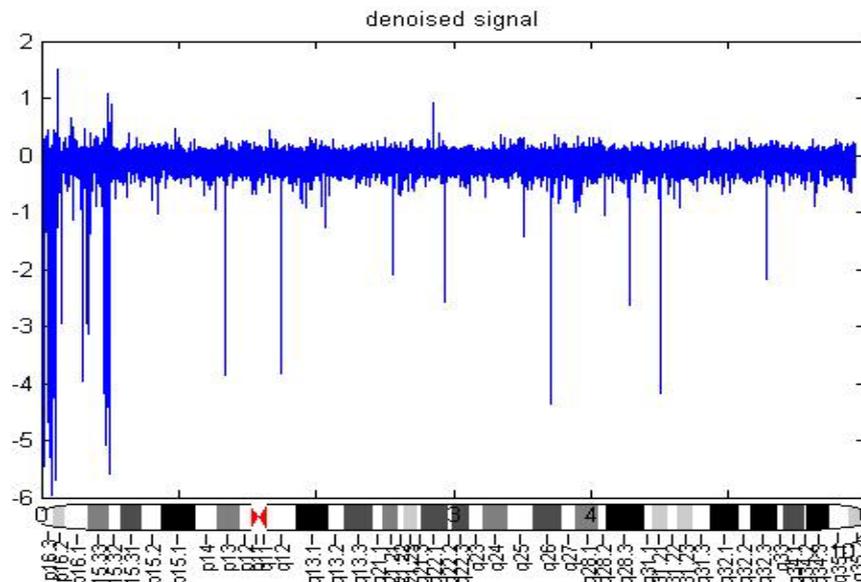


Figure 3.2. Gain of chromosome 4q22.1 using DSWT

This work demonstrates that genomic high-density SNP-array test can be utilized to detect and characterize genome-wide CNAs on more than one thousand clones emerging from several chromosomes. By testing CBS and one-dimensional DSWT techniques on 10 HCC cell lines it has been found that there is a certain number of chromosome aberration including gains of 1q,8q,11q,17q,20q chromosomes and losses of 4q,8p,13q,16q,17p. A new altered region has been detected by amplification of chromosome 4 region 4q22.1. This gene expression profiling SSP1(secreted phosphoprotein 1) and AFF1(AF4/FMR2 family member 1) can prompt to discover new biomarkers that may help enhance diagnostic for detecting HCC, represent novel objective for therapeutic agents, and allow us to identify regions in the HCC genome that have undergone recurrent high-level focal deletions or amplifications.

REFERENCES

- [1] Alfonso. V and Manuel. H, Getting personalized cancer genome analysis into the clinic: the challenges in bioinformatics, *Valencia and Hidalgo Genome Medicine*, vol. 4, pp.1-17, 2012.
- [2] El-Serag. HB, Hepatocellular carcinoma, *New England Journal of Medicine*, vol.365, pp.1118–1127, 2011.
- [3] Guha. N. I., Iredale. J.P, Clinical and diagnostic aspects of cirrhosis. In: Textbook of hepatology from basic science to clinical practice. 3rd edition. Edited by: Rodés J, Benhamou J-P, Blei A, Reichen J, Rizzetto M. *Oxford, Blackwell Publishing*, pp.604-622, 2007.
- [4] Esraa. M. H, Mai. S. M, A study of Support Vector Machine Algorithm for Liver Disease Diagnosis, *American Journal of Intelligent Systems*, vol. 4, pp.9-14, 2014.
- [5] Mai. S. M, Esraa. S. H, Amr. S, Discrete Stationary Wavelet Transform of Array CGH Data for Biomarkers Identification of Hepatocellular Carcinoma, *Journal of Bioinformatics and Intelligent Control*, vol. 1 (2), pp.148-154,2012.
- [6] Beroukhim .R, Mermel. CH, Porter .D, Wei. G, Raychaudhuri. S, Donovan. J, et al. The landscape of somatic copy-number alteration across human cancers, *Nature*, vol.463, pp.899-905, 2010.
- [7] Lee. S.A, Ho. C, Roy. R, Kosinski. C, Patil. M.A, Tward. A.D, et al. Integration of genomic analysis and in vivo transfection to identify sprouty 2 as a candidate tumor suppressor in liver cancer, *HEPATOLOGY*, vol.47, pp. 1200-1210, 2008.
- [8] Tanaka .Y, Kanai. F, Tada. M, Tateishi. R. M. Sanada. M, Nannya. Y, Gain of GRHL2 is associated with early recurrence of hepatocellular carcinoma, *J Hepatol.*, vol. 49, pp.746-57, 2008.
- [9] Midorikawa. Y, Tang. W, Sugiyama. Y, High-resolution mapping of copy number aberrations and identification of target genes in hepatocellular carcinoma, *Biosci Trends*, vol. 1, pp. 26-32., 2007.
- [10] Patil. M. A, Gutgemann. I, Zhang. J, Ho. C, Cheung. S.T, Ginzinger. D, et al. Array-based comparative genomic hybridization reveals recurrent chromosomal aberrations and Jab1 as a potential target for 8q gain in hepatocellular carcinoma., *Carcinogenesis*, vol.26, pp.2050-2057, 2005.
- [11] Chen T.C, Hsieh. L.L., Kuo. T. T., Ng. K.F,W.C. YH, et.al, p16INK4 gene mutation and allelic loss of chromosome 9p21-22 in Taiwanese hepatocellular carcinoma, *Anticancer Res.*, vol.20, pp.1621-6, 2000.
- [12] DE. REICH, M. CARGILL, S. BOLK, et al.: Linkage disequilibrium in the human genome. *Nature*, vol. 411, pp. 199-204, 2001.
- [13] Framboise .T.L, Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances, *Nucleic Acids Research*, Vol. 37, pp. 4181–4193, 2009.
- [14] Kennedy. J, Efficient Algorithms for SNP Genotype Data Analysis using Hidden Markov Models of Haplotype Diversity, *M.Sc., Rensselaer, Hartford, CT, USA*, 2002.
- [15] Cooper. G.M, Zerr. T, Kidd. J.M, et al. Systematic assessment of copy number variant detection via genome-wide SNP genotyping, *Nat Genet*, vol.40 (10), pp.1199–203, 2008.
- [16] Franke. L, Kovel. C.G, Aulchenko. Y.S, et al. Detection, imputation, and association analysis of small deletions and null alleles on oligonucleotide arrays, *AmJ Hum Genet*, vol. 82, pp.1316–33, 2008.
- [17] Wiltshire. T, Pletcher. M. T, Batalov. S, Barnes. S. W, Tarantino. L. M, Genome-wide single-nucleotide polymorphism analysis defines haplotype patterns in mouse, *PNAS*, vol.100, pp 3380–3385, 2013.
- [18] Esraa. M.H, Mai. S. M, Amr. Sh, Circular Binary Segmentation Modeling of Array CGH Data on Hepatocellular carcinoma., *Radio Science Conference (NRSC), 29th National*, pp. 667 – 674, 2012.
- [19] Sen. A and Srivastava. M. S, On tests for detecting a change in mean, *Annals of Statistics*, vol.3, pp. 98–108, 1975.
- [20] Olshen. A.B and Venkatraman. E.S. Circular binary segmentation for the analysis of array based DNA copy number data, *Biostatistics*, vol.5, pp. 557–572, 2004.
- [21] Darshana. M, Asim. B, DISCRETE WAVELET TRANSFORM USING MATLAB, *IGCET*, Vol.4, pp. 252-259, 2013.
- [22] Esraa. M. H, Mai. S. M, Ayman. M. El, Novel Altered Region for Biomarker Discovery in Hepatocellular Carcinoma (HCC) Using Whole Genome SNP Array,(*IJACSA International Journal of Advanced Computer Science and Applications*, Vol. 7, No. 4,2016.
- [23] Esraa. S. H, Mai. S. M, Impact of parallel computing on identifying biomarkers of hepatocellular carcinoma, *J. Med. Imaging Health Inf.*, vol.4, pp. 1–5, 2014.
- [24] Esraa. M. H, Mai. S. M, Ayman. M. Eldeab, Clinical and Genomic strategies for Detecting Hepatocellular Carcinoma in Early Stages: A systematic review, *American Journal of Biomedical Engineering*, vol. 5, pp.101-115, 2015.
- [25] Singhal. A, Jayaraman. M, Dhanasekaran .D.N, Kohli. V,

Molecular and serum markers in hepatocellular carcinoma: Predictive tools for prognosis and recurrence, *Critical Reviews in Oncology/Hematology*, vol. 82, pp.116–140, 2012.

- [26] Subramanian. A, Tamayo. P, Mootha. V. K, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles, *Proc Natl Acad Sci USA*, vol.102, pp.15545–15550, 2005.
- [27] Wang. Y, Wu .M.C., Sham. J.S, Zhang. W, Wu. W. Q, and Guan. X.Y, Prognostic significance of c-myc and AIB1 amplification in hepatocellular carcinoma. A broad survey using high-throughput tissue microarray, *Cancer*, vol.95, pp.2346–52, 2002.
- [28] Wong. N, Lai. P. Lee. S-W, et al. Assessment of genetic changes in hepatocellular carcinoma by comparative genomic hybridization analysis,. *Am J Pathol*, vol. 154, pp.37-43, 1999.