International Journal of Research
Available at https://edupediapublications.org/journals

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 14
November 2017

Forecasting and Marking of Lung Cancer in Patients via MRI/CT Images and Genomic Bio Markers using Neural Baye's Methodology

Penupotula Bhavika,
Dr.D.Haritha ph.D
Perceiving M.Tech in JNTU,Kakinada.

Abstract

This study aims to develop a new quantitative image feature analysis scheme and investigate its role along with 2 genomic biomarkers namely, protein expression of the excision repair cross-complementing 1 (ERCC1) genes and a regulatory subunit of ribonucleotide reductase (RRM1), in predicting cancer recurrence risk of Stage I non-small-cell lung cancer (NSCLC) patients after surgery. Methods: By using chest computed tomography images, we developed a computer-aided detection scheme to segment lung tumors and computed tumor-related image features. After feature selection, we trained a Naïve Bayesian network based classifier using 8 image features and a Multilayer Perceptron classifier using 2 genomic biomarkers to predict cancer recurrence risk, respectively. Two classifiers were trained and tested using a dataset with 79 Stage I NSCLC cases, a synthetic minority oversampling technique and a leave one case out validation method. A fusion method was also applied to combine prediction scores of two classifiers. Results: AUC (areas under ROC curves) values are 0.78±0.06 and 0.68±0.07 when using the image feature and genomic biomarker based classifiers, respectively. AUC value significantly increased to 0.84±0.05 ($p<0.05$) when fusion of two classifier-generated prediction scores using an equal weighting factor. Conclusion: A quantitative image feature based classifier yielded significantly higher discriminatory power than a genomic biomarker based classifier in predicting cancer recurrence risk. Fusion of prediction scores generated by the two classifiers further improved prediction performance. Significance: We demonstrated a new approach that has potential to assist clinicians in more effectively managingStage I NSCLC patients to reduce cancer recurrence risk.

Introduction:

Lung cancer is the leading cause of cancer-related deaths in the world. Epigenetic events are early and frequent in carcinogenesis, which makes DNA methylation an attractive biomarker for cancer. Epigenetic events could also provide a tractable link between the genome and the environment, with the epigenome serving as a biochemical record of relevant life events, e.g. cigarette smoking. Lung cancer is morphologically divided into non-small cell and small cell lung cancer (NSCLC and SCLC). NSCLC accounts for about 80% of the lung cancers and is a heterogeneous clinical entity with major histological subtypes such as squamous cell carcinoma (SCC), adenocarcinoma (AC) and large cell carcinoma. A common feature of the different subtypes of NSCLC is the somewhat slower growth and spread compared to SCLC, enabling surgical eradication in its early stages. Only a minor fraction of NSCLC cases are currently diagnosed in clinical stages I to IIb, where surgical removal is the therapy of choice. The biomarker-driven approach at preinvasive phases could aid in diagnosing or ruling out lung cancer. Current markers, including squamous cell carcinoma antigen, carcinoembryonic antigen, cytokeratin 19 fragment antigen 21-1 and neuron-specific enolase were shown to lack satisfactory diagnostic power. In a recent study, only 37.3% of early-stage lung cancers could be diagnosed using the combination assays of the above-mentioned tumor markers[1,2,3,4,5,6]. Our study was aimed at the genome-wide identification of DNA methylation-based biomarker candidates in early-stage NSCLC. DNA methylation occurs vastly in the context of cytosine-guanine dinucleotides (CpGs). CpG-rich short stretches (CpG islands) are usually located in the promoter region of genes and are normally kept in the demethylated state. In cancer, CpG islands located in the promoter area of tumor suppressor genes and "house-keeping" genes become hypermethylated, which can lead to decreased expression of these genes. At the same time, the genome is globally demethylated, which in turn can lead to the activation of oncogenes [7,8,9,10]. Methylation of CpG island shores – regions with lower CpG density within approximately 2 kb of CpG islands - is also closely associated with transcriptional inactivation. Recently, significant progress has been made in the genome-wide DNA methylation analysis. The methods include bisulfite conversion of DNA, immunoprecipitation or affinity purification of methylated DNA followed by microarray analysis or high-throughput sequencing. It has been shown that in terms of accuracy, bisulfite-based methods perform slightly better than enrichment-based methods and do not require a statistical correction for CpG bias [11,12,13].

We have performed a genome-wide DNA methylation study of stage I NSCLC to obtain an insight into early-stage epigenetic alterations in lung cancer and identify potential diagnostic or prognostic biomarkers. In our study we used the HumanMethylation27 BeadChips (Illumina, Inc) that enable cost-effective quantitative comparisons across many samples.

Analysis:

Methylation related to gene expression changes

Using Pearson's correlation analysis, we were able to determine the expected inverse correlation between the differential methylation levels and gene expression values for 378 (43.5%) of the 869 differentially methylated CpGs between NSCLC and control samples. In different histological groups we were able to find 337 of 780 CpGs (43.2%), the methylation levels of which were inversely correlated to the gene expression levels.

We performed a qPCR analysis of the different *TP73* isoforms to test whether differential methylation next to the TSSs of different isoforms affects their expression level. The qPCR analysis did not reveal a statistically significant difference (p-value 0.36, paired t-test) between the expression levels of the two isoforms in our tumor samples.

Ingenuity Pathway Analysis

We performed *in silico* functional and interaction analyses of the differentially methylated genes using Ingenuity Pathway Analysis (IPA) software (Ingenuity Systems, Redwood City, CA), and found 78 network eligible genes and 451 Functions/Pathways eligible genes. By including the known direct and indirect interactions, the most prominently represented gene network was related to tumor necrosis factor (TNF, Figure S5). Most of the genes (n=22) in the network were hypomethylated, but some genes (n=7) were also hypermethylated.

DNA methylation related to smoking behavior

Based on tobacco smoking pack-years data, we asked whether smoking affects the DNA methylation patterns in a tumor. One patient who lacked smoking data was excluded. Linear regression analysis was performed using the Bioconductor Limma package. Analysis within tumor samples did not show any differentially methylated genes related to the extent of tobacco smoking. Comparing the limited number of non-smokers (n=3, 6.4%) with smokers (n=44, 93.6%) we found four differentially methylated CpG-sites in three genes (p<0.05, FDR adjusted), which are all hypomethylated in smokers group: *CXorf38*, *MTHFD2* and *TLL2*.

Altered methylation and long-term survival

We performed two types of survival analyses to find potential prognostic methylation markers. The patients with only up to one month survival after surgery (n=2) were excluded to avoid any potential confounding influence of postoperative complications.

Firstly, we performed a Kaplan-Meier survival analysis on each of the CpG sites by dividing the Beta values into low, medium and high methylation groups. We only report results where all groups were larger than five patients. As a result, we found 10 CpGs in 10 genes, with methylation level differences in different survival groups (Figure 4). Patients with a medium methylation level of *UGT1A7*, *GPR171*, *P2RY12*, *FLJ35784* and *C20orf185* had better survival than patients with high-level methylation. Patients with a medium methylation level of *CLEC11A* and *GRIK3* had better survival compared to low-level methylation. Patients with a low methylation level of *CYP1A1* and *INGX* had better survival than those with medium-level methylation. Patients with a high methylation level for *PIK3R5* had better survival than those with medium-level methylation.

Secondly, we performed the differential methylation analysis by combining Cox proportional hazard analysis and Wilcoxon rank-sum test. We found 18 CpGs in 15 genes in patients with 1 to 24 months survival (n=12) vs. patients with 60 months and longer survival (n=15), p<0.05, and the methylation difference cut-off applied. From the differentially methylated genes, *DXS9879E* (*LAGE3*), *RTEL1* and *MTM1* were hypermethylated, and *SCUBE3*, *SYT2*, *KCNC3*, *KCNC4*, *GRIK3*, *CRB1*, *SOCS2*, *ACTA1*, *ZNF660*, *MDFI*, *ALDH1A3* and *SRD5A2* were hypomethylated in the group with poor survival (Figure S6).

Proposed System:

We have performed a genome-wide DNA methylation study in 48 stage I NSCLC patients and 18 macroscopically cancer-free control samples by using cluster analysis to search for genes that distinguish between the cancerous and normal lung tissue and compared these genes' methylation levels with their expression levels. In addition, we performed *in silico* functional and interaction analysis of differently methylated genes using IPA software. Linear regression was used to find genes related to smoking, and Kaplan-Meier survival analysis was performed to identify the differential methylation of genes related to patient survival.

As a result, we detected 496 CpGs in 379 hypermethylated genes and 373 CpGs in 336 genes that were hypomethylated in NSCLC. A perfect separation of the control lung tissue samples from NSCLC samples was not obtained, as one normal lung sample clustered together with the cancer samples and six cancer samples (one in replicate) showed methylation patterns somewhat resembling the tumor-free lung tissue. Since we used non-dissected tumor samples, this finding may be at least partially caused by the confounding effect of non-neoplastic tissue present in these samples. Pathological examination of the NSCLC samples with DNA methylation profiles similar to the normal lung samples revealed either low tumor content (10 to 30%) or a very heterogeneous composition of tumor cells in the samples. These co-clustering cancer samples and cancer-free lung samples were therefore excluded from further analyses.

We analyzed Union for International Cancer Control (UICC) stage I NSCLC samples [35] from 48 patients and macroscopically cancer-free “normal” lung control samples from 18 patients. All the specimens had been isolated during lung surgery at Tartu University Hospital, Estonia. The patients with adenocarcinoma (n=6, 12.5%) and its subtype bronchioloalveolar carcinoma (n=10, 20.8%) were analyzed as one group (n=16, 33.3%). The remaining 32 (66.7%) of the analyzed patients had squamous cell carcinoma. The age range in the study group was 41-80 years (mean age in males n=40, 66.2 years and in females n=8, 65.5 years) (Table S3). The patients did not undergo any preoperative chemo- or radiotherapy. At surgery, tissue specimens of appropriate size (1-2 cm^3) were cut from tumorous and morphologically tumor-free lung tissue. One half of each sample was fixed in formalin and used for pathological examination. The other half of each specimen was snap frozen and stored at -80°C until use. Control samples were obtained at a site distant from the removed tumor and confirmed to be tumor-free by the same pathologist.

DNA extraction and bisulfite modification

DNA was extracted from 50 mg of tumor and matching tumor-free lung tissue with the Dneasy® Blood & Tissue kit (Qiagen GmbH., Hilden, Germany) and with the Nucleospin® Tissue kit (Macherey-Nagel GmbH., Düren, Germany). DNA yield and purity were determined using the NanoDrop® ND1000 spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA). From each sample, 500 ng of genomic DNA was bisulfite modified using the EZ DNA Methylation™ Kit (Zymo Research, Orange, CA) according to the manufacturer’s recommendations.

Methylation validation by Sanger sequencing

For methylation chip validation 11 genes were chosen, five of these were genes from survival analysis and the remaining six were genes that distinguished between cancer and normal tissue (Primers used in study showed in Table S4). Primers for bisulfite-treated DNA PCR were designed using MethPrimer [36]. A 20 µl PCR was carried out in 80 mM Tris-HCl (pH 9,4-9,5), 20 mM (NH4)2SO4, 0,02% Tween-20 PCR buffer, 3 mM MgCl2, 1X Betaine, 0.25 mM dNTP mix, 2 units of Hot-start Taq polymerase, 50 pmol of the forward primer, 50 pmol of the reverse primer, and 20 ng of bisulfite- treated genomic DNA. PCR cycling conditions were 95°C 15 min for enzyme activation, 95°C 30 sec, 62°C 45 sec, 72°C 120 sec for 17 cycles, touchdown by -0.5°C for every cycle and 95°C 30 sec, 52°C 30 sec, 72°C 120 sec for 21 cycles. Sequencing was done as a service by the Core Facility of Estonian Biocenter.

Table 1
Distance of differentially methylated CpGs to the nearby transcription start sites (TSS, FDR corrected p<0.05, mean difference in methylation level in NSCLC vs tumor-free lung at least 13.6%). Distance to TSS of hyper- vs hypomethylated CpGs differed by p=0.0001 (Welch Two Sample t-test).

	Hypermethylated CpGs	**Hypomethylated CpGs**
Distance to TSS (median; 1st quartile; 3rd quartile)	**20.0**; -215; 229	**-63**; -469.2; 132.5
Located inside CpG island	86% (429)	23% (86)
Located outside CpG island	14% (67)	77% (287)

Table 2
Statistically significant Pearson’s correlations between differentially methylated CpG sites and gene expression values across 48 lung cancer samples and the control samples with available gene expression data. P-values were computed by permuting individuals and recalculating the median gene expression levels 1,000 times. All CpGs and genes represented by both methylation and gene expression arrays were included in permutations.

Symbol	Gene ID	Pearson R	Permuted p-value*
AGER	177	-0.8095317	0.0143
BRDT	676	-0.7664438	0.0477
CALML5	51806	-0.7729556	0.0419
ELAVL4	1996	-0.8363362	0.0027
GSTT1	2952	-0.7698949	0.0450
MAGEC1	9947	-0.7886890	0.0278
MB	4151	-0.7765738	0.0390
NR0B2	8431	-0.8705916	0.0002
P53AIP1	63970	-0.7665071	0.0477
PNLDC1	154197	-0.8677312	0.0002
PPP1R14D	54866	-0.8137714	0.0121

* - Permuted p-value - p-value computed by permuting individuals and recalculating the median gene expression levels 1,000 times.

Conclusion:

Methylation analysis was performed using Infinium® HumanMethylation27 RevB BeadChips (Illumina Inc.). The assay covers 27,578 CpGs in 14,495 genes located predominantly in CpG islands within proximal promoter regions, between 1.5 kb upstream and 1 kb downstream of the transcription start sites (TSS). A CpG island in this assay is defined as a nucleotide sequence of (1) 200 bp or greater in length, (2) 50% or greater in GC-percent, and (3) 0.60 or greater in the ratio of observed CpG sites over expected CpG sites in that region [10]. The HumanMethylation27 beadchips also cover CpG sites in the regulatory regions of 1,000 well-known cancer genes, 150 differentially methylated genes in various cancers and 110 miRNA genes. The chips were processed according to the manufacturer's standard protocols. We performed cluster analysis of the methylation profiles by using the Limma program of Bioconductor package in R statistical computing software (www.bioconductor.org). Prior to analysis we quantile normalized the methylation data to eliminate systematic differences between the chips. The analysis was performed using t-tests with an empirical Bayes' correction from the Bioconductor Limma package [37]. The differentially methylated genes were clustered hierarchically and visualized using a heatmap. All the methylation differences (differences between the Beta-values representing the calculated level of methylation from 0 to 1, alternatively 0% to 100% for each analyzed CpG) were identified using a false discovery rate (FDR) corrected $p<0.05$ and ≥ 0.136 mean methylation level difference that was previously shown to detect differences with at least 95% confidence [38]. For Kaplan-Meier survival analysis, we divided the Beta values into low (0-0.25), medium (0.25-0.75) and high (0.75-1) methylation group and performed a log-rank test to assess the difference in survival between the groups. We corrected these p-values using FDR and used 0.05 as the significance level.

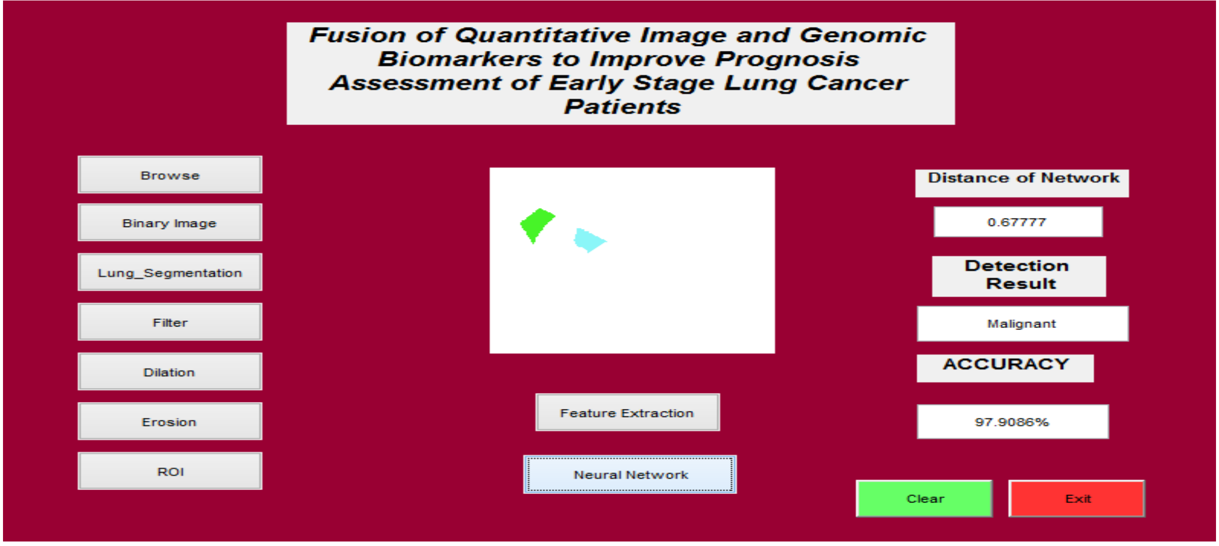
References

1. Baylin SB, Ohm JE (2006) Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction? Nat Rev Cancer 6: 107-116.
2. Feinberg AP, Ohlsson R, Henikoff S (2006) The epigenetic progenitor origin of human cancer. Nat Rev Genet 7: 21-33.
3. Uribe-Lewis S, Woodfine K, Stojic L, Murrell A (2011) Molecular mechanisms of genomic imprinting and clinical implications for cancer. Expert Rev Mol Med 13: e2.
4. Foley DL, Craig JM, Morley R, Olsson CA, Olsson CJ, et al. (2009) Prospects for epigenetic epidemiology. Am J Epidemiol 169: 389-400.
5. Brambilla E, Travis WD, Colby TV, Corrin B, Shimosato Y (2001) The new World Health Organization classification of lung tumours. Eur Respir J 18: 1059-1068.
6. Chu XY, Hou XB, Song WA, Xue ZQ, Wang B, et al. (2011) Diagnostic values of SCC, CEA, Cyfra21-1 and NSE for lung cancer in patients with suspicious pulmonary masses: A single center analysis. Cancer Biol Ther 11.
7. Bird A (2002) DNA methylation patterns and epigenetic memory. Genes Dev 16: 6-21.
8. Rollins RA, Haghighi F, Edwards JR, Das R, Zhang MQ, et al. (2006) Large-scale structure of genomic methylation patterns. Genome Res 16: 157-163.
9. Grønbaek K, Hother C, Jones PA (2007) Epigenetic changes in cancer. Apmis 115: 1039-1059.
10. Takai D, Jones PA (2002) Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc Natl Acad Sci U S A 99: 3740-3745.
11. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, et al. (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet 41: 178-186.
12. Kalari S, Pfeifer GP (2010) Identification of driver and passenger DNA methylation in cancer by epigenomic analysis. Adv Genet 70: 277-308.
13. Bock C, Tomazou EM, Brinkman AB, Muller F, Simmer F, et al. (2010) Quantitative comparison of genome-wide DNA methylation mapping technologies. Nat Biotechnol 28: 1106-1114.
14. Valk K, Vooder T, Kolde R, Reintam MA, Petzold C, et al. (2011) Gene Expression Profiles of Non-Small Cell Lung Cancer: Survival Prediction and New Biomarkers. Oncology 79: 283-292.
15. Yang RY, Hsu DK, Yu L, Ni J, Liu FT (2001) Cell cycle regulation by galectin-12, a new member of the galectin superfamily. J Biol Chem 276: 20252-20260.
16. Potapenko IO, Haakensen VD, Luders T, Helland A, Bukholm I, et al. (2010) Glycan gene expression signatures in normal and malignant breast tissue; possible role in diagnosis and progression. Mol Oncol 4: 98-118.
17. Ehrich M, Field JK, Liloglou T, Xinarianos G, Oeth P, et al. (2006) Cytosine methylation profiles as a molecular marker in non-small cell lung cancer. Cancer Res 66: 10911-10918.
18. Bartling B, Hofmann HS, Weigle B, Silber RE, Simm A (2005) Down-regulation of the receptor for advanced glycation end-products (RAGE) supports non-small cell lung carcinoma. Carcinogenesis 26: 293-301.
19. Flonta SE, Arena S, Pisacane A, Michieli P, Bardelli A (2009) Expression and functional regulation of myoglobin in epithelial cancers. Am J Pathol 175: 201-206.
20. Kim H, Lapointe J, Kaygusuz G, Ong DE, Li C, et al. (2005) The retinoic acid synthesis gene ALDH1a2 is a candidate tumor suppressor in prostate cancer. Cancer Res 65: 8118-8124.
21. Yoo KH, Park YK, Kim HS, Jung WW, Chang SG (2010) Epigenetic inactivation of HOXA5 and MSH2 gene in clear cell renal cell carcinoma. Pathol Int 60: 661-666.
22. Faller WJ, Rafferty M, Hegarty S, Gremel G, Ryan D, et al. (2010) Metallothionein 1E is methylated in malignant melanoma and increases sensitivity to cisplatin-induced apoptosis. Melanoma Res 20: 392-400.
23. Ye YW, Wu JH, Wang CM, Zhou Y, Du CY, et al. (2011) Sox17 regulates proliferation and cell cycle during gastric cancer progression. Cancer Lett.
24. Dammann R, Li C, Yoon JH, Chin PL, Bates S, et al. (2000) Epigenetic inactivation of a RAS association domain family protein from the lung tumour suppressor locus 3p21.3. Nat Genet 25: 315-319.
25. Hirai K, Koizumi K, Haraguchi S, Hirata T, Mikami I, et al. (2005) Prognostic significance of the tumor suppressor gene maspin in non-small cell lung cancer. Ann Thorac Surg 79: 248-253.
26. Khattar NH, Lele SM, Kaetzel CS (2005) Down-regulation of the polymeric immunoglobulin receptor in non-small cell lung carcinoma: correlation with

dysregulated expression of the transcription factors USF and AP2. J Biomed Sci 12: 65-77.

27. Muller-Hagen G, Beinert T, Sommer A (2004) Aspects of lung cancer gene expression profiling. Curr Opin Drug Discov Devel 7: 290-303.
28. Jin M, Kawakami K, Fukui Y, Tsukioka S, Oda M, et al. (2009) Different histological types of non-small cell lung cancer have distinct folate and DNA methylation levels. Cancer Sci 100: 2325-2330.
29. Bertazza L, Mocellin S (2010) The dual role of tumor necrosis factor (TNF) in cancer biology. Curr Med Chem 17: 3337-3352.
30. Strassburg CP, Strassburg A, Nguyen N, Li Q, Manns MP, et al. (1999) Regulation and function of family 1 and family 2 UDP-glucuronosyltransferase genes (UGT1A, UGT2B) in human oesophagus. Biochem J 338 (Pt 2): 489-498.
31. Araki J, Kobayashi Y, Iwasa M, Urawa N, Gabazza EC, et al. (2005) Polymorphism of UDP-glucuronosyltransferase 1A7 gene: a possible new risk factor for lung cancer. Eur J Cancer 41: 2360-2365.
32. Tekpli X, Zienolddiny S, Skaug V, Stangeland L, Haugen A, et al. (2011) DNA methylation of the CYP1A1 enhancer is associated with smoking-induced genetic alterations in human lung. Int J Cancer.
33. Di Pietro E, Wang XL, MacKenzie RE (2004) The expression of mitochondrial methylenetetrahydrofolate dehydrogenase-cyclohydrolase supports a role in rapid cell growth. Biochim Biophys Acta 1674: 78-84.
34. Hillion J, Wood LJ, Mukherjee M, Bhattacharya R, Di Cello F, et al. (2009) Upregulation of MMP-2 by HMGA1 promotes transformation in undifferentiated, large-cell lung cancer. Mol Cancer Res 7: 1803-1812.
35. Mountain CF (2000) The international system for staging lung cancer. Semin Surg Oncol 18: 106-115.
36. Li LC, Dahiya R (2002) MethPrimer: designing primers for methylation PCRs. Bioinformatics 18: 1427-1431.
37. Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 3: Article3.
38. Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, et al. (2009) Genome-wide DNA methylation profiling using Infinium® assay. Epigenomics 1: 177-200.

Output:



Fusion of Quantitative Image and Genomic Biomarkers to Improve Prognosis Assessment of Early Stage Lung Cancer Patients

Distance of Network
0.67777

Detection Result
Malignant

ACCURACY
97.9086%