A DNA methylation microarray-based meta analysis identifies a panel of epigenetic biomarkers for NSCLC cancer diagnosis

Shicheng Guo1, Fengyang Yan, Yanyun Ma, Li Jin, Jiucun Wang

1, MOE Key Laboratory of Contemporary Anthropology and State Key Laboratory of Genetic Engineering, School of Life Sciences and Institutes of Biomedical Sciences, Fudan University, Shanghai, China

2, CMC Institute of Health Sciences, 1 Yaocheng Road, Taizhou, Jiangsu 225300, China

Key words: DNA methylation, non-small cell lung cancer, **biomarker**

Abbreviations: NSCLC, non-small cell lung cancer, microarray, Meta-analysis

MSP: methylation specific PCR

Abstract:[250]

[B]Aberrant DNA methylation is an important event in lung cancer initiation. So it is a promising biomarker for cancer early screening or diagnosis. Microarray analysis has yet to be widely accepted for molecular cancer diagnosis and classification of human cancers. There are exponential increase of microarray datasets in GEO,including methylation datasets,so we can make full use of such information to benefit clinical detection precision.

[Methos] Three independent microarray datasets were systematically quantified with serail of proecss: background-corrected, data intergate, batch effect emilination with Combat method and Differential ananlysis. External Validation was performed with Methylation-specific PCR(MSP) in 100 pairs of cancer and adajacent tissues.

[Result] From meta analysis we identified 3 there genes() which

[Conclusions] a panel of epigenetic biomarkers for NSCLC cancer diagnosis was established and

Introduction[400]

Lung cancer, a complex disease invovled both genetic and epigenetic changes, is the leading cause of cancer deaths worldwide[[1](#_ENREF_1)]. Among all types of cancer, the most commonly diagnosed, as well as the most common cause of cancer deaths, is lung cancer, with a mortality rate as high as 80-85% within 5 years[[1](#_ENREF_1)]. 80% of primary lung cancers are non-small cell lung carcinoma (NSCLC) which is characterized by a long asymptomatic latency and poor prognosis.

The most important things to resolve the predicament of mortality is to identify some high sensitivity and specificity diagnosis biomarkers. Evidence show that If early stage lung cancer is detected, the survival rate can increase dramatically[[2](#_ENREF_2)]. And this has been succeed in breast and prostate cancer, for which has repective early detection method. many imaging and cytology-based strategies have been employed in NSCLC diagnosis, however none have yet been proven effective.

Many studies have investigated the Identification of cancer biomarkers through analysis of genome-wide expression profiles has been popular severial years age.

Microarray technology is not only a new tool for the clinical lab but it can also improve the accuracy of the classical diagnostic techniques by suggesting novel tumor-specific markers.

Thus, it is essential to identify biomarkers for early prediction of lung cancer. Lung cancer diagnosis biomarker exploit is a much more challenging problem. Establishment of a panel of

We systematically quantified 3 independent microarray datasets from GEO[[3](#_ENREF_3)] and TCGA project(see Table 1)

**Materials and Methods**

**Datasets**

DNA methylation profiling data of 400 NSCLC associated samples: 294 NSCLC,99 cancer adjacent,7 normal lung tissue obtained from three publicly available independent datasets with two platform from illumina company, GoldGate and inf methylation27K respectively. As detailed in table 1. And 112 probe were shared by methylation 27K and goldenGate array.

**Reprocessing and Data integration**

All the datasets were background-corrected and normalised separately with recommended for each platform. Then they were merged and conducted quantile normaliziation collectively. As large number of evidence that there are still biases left with above adjustment. We use batch effect analysis tools, ComBat, which is believed to be the best programs[[4](#_ENREF_4)], to emilinate the batch effect exist in inpendent datasets.

Subsequent analyses were conducted in R[[5](#_ENREF_5)]. Hierarchical clustering was performed with the hclust function: Manhattan metric and average linkage for CpG loci with the highest variance. For inference, data were clustered using a recursively partitioned mixture model (RPMM) ([20](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2755616/?tool=pubmed#R20)). Associations between covariates and methylation at individual CpG loci were tested with generalized linear models, accounting for the beta-distribution of average beta as in Hsuing et al. ([21](http://www.ncbi.nlm.nih.gov/pubmed/17220338)). False discovery rate correction via Q–values were computed by the qvalue package ([22](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2755616/?tool=pubmed#R22)).

**Results**

**Batch effect elimination**

Fluorescent signals for methylated (Cy5) and unmethylated (Cy3) alleles give methylation level: β= (max(*Cy5*, 0))/(|*Cy3*| + |*Cy5*| + 100) with about 30 replicate bead measurements per locus.

Data integration from independent experiments is a main source of batch effect. Batch effect elimination was conducted using a empirical Bayes methods: ComBat[[6](#_ENREF_6)] after quantile normalization to three datasets. After principal component analysis, the first two principal components show the betch effect was fully decreased. Detail see figure 1. The data adjusted by the Combat was used to the following classficiation and differenation anaysis.

We obtained expression data for GBM samples at the TCGA web site (<http://tcga-data.nci.nih.gov/docs/publications/> gbm\_exp/).

Table 1 Datasets used for meta analysis

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Published | Acession | NSCLC | pNSCLC | Normal | Jaurnal |
| Karl T. Kelsey et al(2009)/USA | GSE16559 | 57 | 48 | 4 | *Cancer Res* |
| [Esteller M](http://www.ncbi.nlm.nih.gov/pubmed?term=%22Esteller%20M%22%5BAuthor%5D) et al(2012)/USA | GSE28094 | 33 | 0 | 3 | *Genome Res* |
| The Cancer Genome Atlas(2012)/USA | [TCGA](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE35341) | 262 | 51 | 0 | *TCGA* |

Datasets were downloaded either from websites or from public repositories (GEO, ArrayExpress).

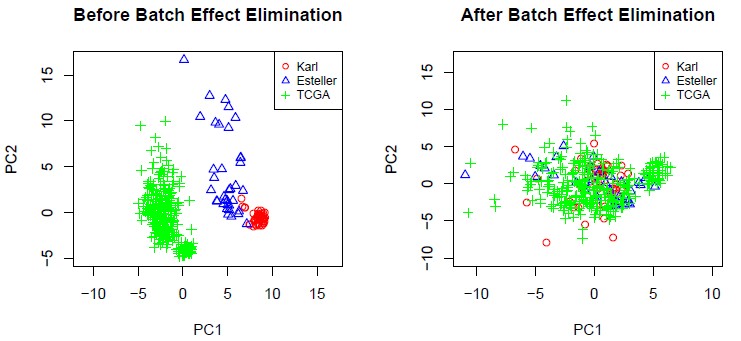


Figure 1 Fine **Batch effect elimination** ability of Combat

The effect of Combat on the first two components of the integret data.

**Classfication**

Four different method based on PCA, Decision tree, LDA,SVM were used to conducted classficiation. PCA differential analysis show there are significant larger number of principle components difference between adajacent cancer tissue and normal lung, which tell us we can’t merge such two kinds of data facilely.

Traditionally, Since the global DNA methylation data per case shows a bimodal distribution, Beta values < 0.3 defines the unmethylated CpGs and beta values > 0.8 the full methylated CpGs,beta values between 0.3-0.8 defines semi-methylated CpGs[[7](#_ENREF_7), [8](#_ENREF_8)]. It is tangly definition to the relationship between beta and Methylation status. Shinjo use beta > 0.4 as methylaiton status[[9](#_ENREF_9)].

Table 2 Four different method based on PCA, Decision tree, LDA,SVM

The classification model based on principal components

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Result | | | Leaving-one-out crossvalidation | | |
|  |  | Positive | Negative |  | Positive | Negative |  |
| 1,2,7,14 | Tumor | 292 | 2 | 99.32% | 292 | 2 | 99.32% |
|  | Para | 2 | 97 | 97.98% | 2 | 97 | 97.98% |
|  |  | 99.32% | 97.98% | 98.98% | 99.32% | 97.98% | 98.98% |
| 1,2,6,7,14 | Tumor | 292 | 2 | 99. 32% | 291 | 3 | 98.98% |
|  | Para | 1 | 98 | 98.99% | 0 | 99 | 100% |
|  |  | 99.66% | 98.00% | 99.24% | 100% | 97.06% | 99.24% |

Decision tree (C4.5)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Result | | | Leaving-one-out crossvalidation | | |
|  |  | Positive | Negative |  | Positive | Negative |  |
|  | Tumor | 293 | 1 | 99.66% | 282 | 12 | 95.92% |
|  | Para | 2 | 97 | 97.98% | 8 | 91 | 91.92% |
|  |  | 99.32% | 98.98% | 99.24% | 97.24% | 88.35% | 94.91% |

LDA

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Result | | | Leaving-one-out crossvalidation | | |
|  |  | Positive | Negative |  | Positive | Negative |  |
|  | Tumor | 0 | 294 | 0.00% | 22 | 272 | 7.48% |
|  | Para | 0 | 99 | 100% | 1 | 98 | 98.99% |
|  |  | 0.00% | 25.19% | 25.19% | 95.65% | 26.49% | 30.53% |

SVM

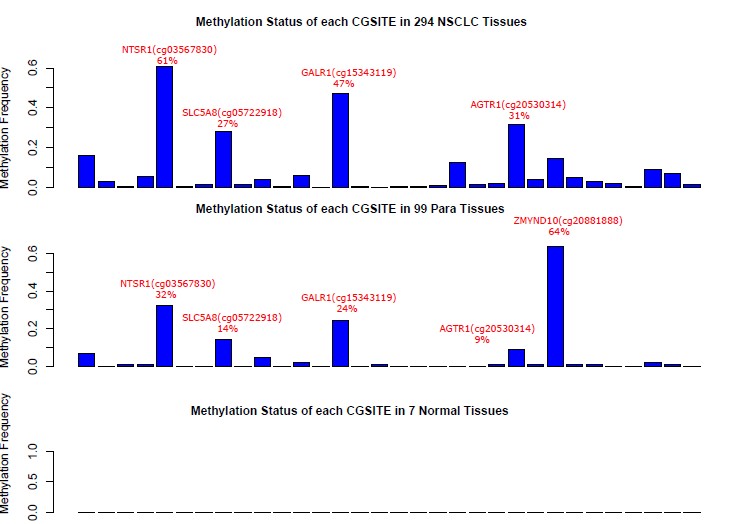
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Result | | | Leaving-one-out crossvalidation | | |
|  |  | Positive | Negative |  | Positive | Negative |  |
|  | Tumor | 293 | 1 | 99.66% | 291 | 3 | 98.98% |
|  | Para | 0 | 99 | 100% | 1 | 98 | 98.99% |
|  |  | 100% | 99.00% | 99.75% | 99.66% | 97.03% | 98.98% |

**Methylation-Specfic PCR Validation**

Methylation arrays despite its similarity to microarray measurement, RT-PCR could give slightly different results.

**Differential Methylation Analysis and Gene Ontology Enrichment**

There are 8 cgsite are all hypermethylation in 7 normal lung tissue, meanwhile, they are all hypermethylated in almost (at least 90%) all tumor tissue and para-tumor tissues.



Previous studies on DNA methylation in PCNSL were limited by their restriction to a low number of selected genes.Our data are in line with Chu et al. and onzales-Gomez et al. [5,8] analyzing the DNA methylation of CDKN2B, DAPK1, GSTP1, MGMT, MLH1, RARB, THBS1, TIMP2, and TIMP3 in PCNSL by MSP (Additional file 8) further supporting the validity of our analysis. This is of special

interest, since our data on PCNSL are based on a minor number of five cases, because of the limited availability of sample material.

Binary-Classification

Discussion

neurotensin receptor 1 (high affinity); Receptor for the tridecapeptide neurotensin. It is associated with G proteins that activate a phosphatidylinositol- calcium second messenger system

Neurotensin (NT) **[**[**9**](#_ENREF_9)**]** is a brain and gastrointestinal peptide that acts on three subtypes of neurotensin receptors; the G-protein-coupled receptors NTS1 (high affinity) and NTS2 (low affinity) and the single transmembrane receptor NTS3 (Sortilin 1). These receptors have many central and peripheral actions including modulation of dopaminergic transmission in nigro-striatal and mesocorticolimbic pathways, modulation of the digestive tract and hypothermic and analgesic effects. NTS1 is distributed widely throughout the brain and gastrointestinal tract, whilst NTS2 is confined to areas of the brain involved in the descending control of nociception, including the periaqueductal gray and dorsal raphe. The human genes encoding NTS1 and NTS2 receptors are located on chromosomes 20q13 and 2p25.1 respectively.

Gene introduction involved in this paper.

Neurotensin receptor-1 **(NTSR-1)** is a G-protein coupled receptor (GPCR) that has been recently identified as a mediator of cancer progression[[10](#_ENREF_10)].Neurotensin receptor 1 determines the outcome of non-small cell lung cancer[[11](#_ENREF_11)]. Inhibition of neurotensin receptor 1 selectively sensitizes prostate cancer to ionizing radiation[[12](#_ENREF_12)].The neurotensin receptor-1 promotes tumor development in a sporadic but not an inflammation-associated mouse model of colon cancer[[13](#_ENREF_13)]. The potential use of the neurotensin high affinity receptor 1 as a biomarker for cancer progression and as a component of personalized medicine in selective cancers[[14](#_ENREF_14)].

**GALR1,** galanin receptor subtype 2 suppresses cell proliferation in several cancers such as head and neck[[15](#_ENREF_15), [16](#_ENREF_16)],oral squamous cell carcinoma[[17](#_ENREF_17)]. And its inactivation can be caused by promoter hypermethyaltion[[15](#_ENREF_15)].

**AGTR1,** angiotensin II receptor type 1, is associated with cell proliferation[[18](#_ENREF_18)] and the cancer classfication in breast cancer[[19](#_ENREF_19)].

**SLC5A8,** a tumor suppressor gene,is usally suppressed in cancers[[20-22](#_ENREF_20)],Functional identification of SLC5A8, a tumor suppressor down-regulated in colon cancer, as a Na(+)-coupled transporter for short-chain fatty acids[[22](#_ENREF_22)]. Aberrant methylation and histone deacetylation associated with silencing of SLC5A8 in gastric cancer[[21](#_ENREF_21)]. Protein expressions and genetic variations of SLC5A8 in prostate cancer risk and aggressiveness[[23](#_ENREF_23)]

**ZMYND10 (alias:BLU)**: Frequent inactivation of RASSF1A, **BLU,** and SEMA3B on 3p21.3 by promoter hypermethylation and allele loss in non-small cell lung cancer[[24](#_ENREF_24)].Characterization of a multiple epigenetic marker panel for lung cancer detection and risk assessment in plasma[[25](#_ENREF_25)].

Smoking can cause down regulation of BLU and low expression will persist in following cancer process[[26](#_ENREF_26)].3p21.3 tumor suppressor gene cluster flanking RASSF1 (i.e., SEMA3B, HYAL3, HYAL2, HYAL1, TUSC2, RASSF1, ZMYND10, NPRL2, TMEM115, and CACNA2D2). epigenetic repression is involved in the down-regulation of multiple genes at 3p21.3 in breast cancer cells[[27](#_ENREF_27)].Hypermethylation of RASSF1A and BLU tumor suppressor genes in non-small cell lung cancer: implications for tobacco smoking during adolescence[[28](#_ENREF_28)].

Acknowledgment

Grant Support

Note: GSE19434 and GSE16559 are share almost same set samples,so we just only use GSE16559 datasets.

"cg07285167" "CSF3R" "chr1:36948981-36949018"

"cg09313705" "HOXB2" "chr17:46622443-46622480"

cg27650175 cancer hypomethylation DAB2IP

cg20881888 cancer hypermethylation ZMYND10

It is demonstrated that DAB2IP is a tumor suppressor gene inactivated by methylation in several cancers

First, they can be measured by relatively cheap and widely used clinical methods such as RT-PCR, ELISA and immunohistochemistry. Second, they can be detected in serum or other body fluids permitting the establishment of noninvasive diagnostic test, which is very important especially in the cases of cancers with more difficult access for diagnostic biopsy (e.g., lung, ovary, pancreas).

Reference[100]

1. Siegel, R., D. Naishadham, and A. Jemal, *Cancer statistics, 2012.* CA Cancer J Clin, 2012. **62**(1): p. 10-29.

2. Henschke, C.I., et al., *Survival of patients with stage I lung cancer detected on CT screening.* N Engl J Med, 2006. **355**(17): p. 1763-71.

3. Edgar, R., M. Domrachev, and A.E. Lash, *Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.* Nucleic Acids Research, 2002. **30**(1): p. 207-10.

4. Chen, C., et al., *Removing batch effects in analysis of expression microarray data: an evaluation of six batch adjustment methods.* Plos One, 2011. **6**(2): p. e17238.

5. Dessau, R.B. and C.B. Pipper, *[''R"--project for statistical computing].* Ugeskr Laeger, 2008. **170**(5): p. 328-30.

6. Johnson, W.E., C. Li, and A. Rabinovic, *Adjusting batch effects in microarray expression data using empirical Bayes methods.* Biostatistics, 2007. **8**(1): p. 118-27.

7. Ron-Bigger, S., et al., *Aberrant epigenetic silencing of tumor suppressor genes is reversed by direct reprogramming.* Stem Cells, 2010. **28**(8): p. 1349-54.

8. Richter, J., et al., *Array-based DNA methylation profiling of primary lymphomas of the central nervous system.* Bmc Cancer, 2009. **9**: p. 455.

9. Haase, C., et al., *Neurotensin receptors in adeno- and squamous cell carcinoma.* Anticancer Research, 2006. **26**(5A): p. 3527-33.

10. Heakal, Y., et al., *Neurotensin receptor-1 inducible palmitoylation is required for efficient receptor-mediated mitogenic-signaling within structured membrane microdomains.* Cancer Biology & Therapy, 2011. **12**(5): p. 427-35.

11. Alifano, M., et al., *Neurotensin receptor 1 determines the outcome of non-small cell lung cancer.* Clinical Cancer Research, 2010. **16**(17): p. 4401-10.

12. Valerie, N.C., et al., *Inhibition of neurotensin receptor 1 selectively sensitizes prostate cancer to ionizing radiation.* Cancer Research, 2011. **71**(21): p. 6817-26.

13. Bugni, J.M., et al., *The neurotensin receptor-1 promotes tumor development in a sporadic but not an inflammation-associated mouse model of colon cancer.* International Journal of Cancer, 2012. **130**(8): p. 1798-805.

14. Dupouy, S., et al., *The potential use of the neurotensin high affinity receptor 1 as a biomarker for cancer progression and as a component of personalized medicine in selective cancers.* Biochimie, 2011. **93**(9): p. 1369-78.

15. Misawa, K., et al., *Epigenetic inactivation of galanin receptor 1 in head and neck cancer.* Clinical Cancer Research, 2008. **14**(23): p. 7604-13.

16. Kanazawa, T., et al., *Galanin receptor subtype 2 suppresses cell proliferation and induces apoptosis in p53 mutant head and neck cancer cells.* Clinical Cancer Research, 2009. **15**(7): p. 2222-30.

17. Henson, B.S., et al., *Galanin receptor 1 has anti-proliferative effects in oral squamous cell carcinoma.* Journal of Biological Chemistry, 2005. **280**(24): p. 22564-71.

18. Hernandez Schulman, I., M.S. Zhou, and L. Raij, *Cross-talk between angiotensin II receptor types 1 and 2: potential role in vascular remodeling in humans.* Hypertension, 2007. **49**(2): p. 270-1.

19. Rhodes, D.R., et al., *AGTR1 overexpression defines a subset of breast cancer and confers sensitivity to losartan, an AGTR1 antagonist.* Proc Natl Acad Sci U S A, 2009. **106**(25): p. 10284-9.

20. Park, J.Y., et al., *Silencing of the candidate tumor suppressor gene solute carrier family 5 member 8 (SLC5A8) in human pancreatic cancer.* Pancreas, 2008. **36**(4): p. e32-9.

21. Ueno, M., et al., *Aberrant methylation and histone deacetylation associated with silencing of SLC5A8 in gastric cancer.* Tumour Biology, 2004. **25**(3): p. 134-40.

22. Miyauchi, S., et al., *Functional identification of SLC5A8, a tumor suppressor down-regulated in colon cancer, as a Na(+)-coupled transporter for short-chain fatty acids.* Journal of Biological Chemistry, 2004. **279**(14): p. 13293-6.

23. Lin, H.Y., et al., *Protein expressions and genetic variations of SLC5A8 in prostate cancer risk and aggressiveness.* Urology, 2011. **78**(4): p. 971 e1-9.

24. Ito, M., et al., *Frequent inactivation of RASSF1A, BLU, and SEMA3B on 3p21.3 by promoter hypermethylation and allele loss in non-small cell lung cancer.* Cancer Letters, 2005. **225**(1): p. 131-9.

25. Hsu, H.S., et al., *Characterization of a multiple epigenetic marker panel for lung cancer detection and risk assessment in plasma.* Cancer, 2007. **110**(9): p. 2019-26.

26. Boelens, M.C., et al., *Current smoking-specific gene expression signature in normal bronchial epithelium is enhanced in squamous cell lung cancer.* Journal of Pathology, 2009. **218**(2): p. 182-91.

27. da Costa Prando, E., L.R. Cavalli, and C.A. Rainho, *Evidence of epigenetic regulation of the tumor suppressor gene cluster flanking RASSF1 in breast cancer cell lines.* Epigenetics, 2011. **6**(12): p. 1413-24.

28. Marsit, C.J., et al., *Hypermethylation of RASSF1A and BLU tumor suppressor genes in non-small cell lung cancer: implications for tobacco smoking during adolescence.* International Journal of Cancer, 2005. **114**(2): p. 219-23.