科技部 Ministry of Science and Technology

The cost of publication in *Journal of Biomedical Science* is borne by the Ministry of Science and Technology, Taiwan.

JBS **JOURNAL OF BIOMEDICAL SCIENCE**

RESEARCH   **Open Access**

SNP rs10248565 in *HDAC9* as a novel genomic aberration biomarker of lung adenocarcinoma in non-smoking women

Liang-Chuan Lai[1,2*], Mong-Hsun Tsai[2,3], Pei-Chun Chen[7], Lee H Chen[2], Jen-Hao Hsiao[2], Shin-Kuang Chen[4], Tzu-Pin Lu[4], Jang-Ming Lee[8], Chung-Ping Hsu[9], Chuhsing K Hsiao[2,6] and Eric Y Chuang[2,4,5*]

Abstract

Background: Numerous efforts have been made to elucidate the etiology and improve the treatment of lung cancer, but the overall five-year survival rate is still only 15%. Although cigarette smoking is the primary risk factor for lung cancer, only 7% of female lung cancer patients in Taiwan have a history of smoking. Since cancer results from progressive accumulation of genetic aberrations, genomic rearrangements may be early events in carcinogenesis.

Results: In order to identify biomarkers of early-stage adenocarcinoma, the genome-wide DNA aberrations of 60 pairs of lung adenocarcinoma and adjacent normal lung tissue in non-smoking women were examined using Affymetrix Genome-Wide Human SNP 6.0 arrays. Common copy number variation (CNV) regions were identified by ≥30% of patients with copy number beyond 2 ± 0.5 of copy numbers for each single nucleotide polymorphism (SNP) and at least 100 continuous SNP variant loci. SNPs associated with lung adenocarcinoma were identified by McNemar's test. Loss of heterozygosity (LOH) SNPs were identified in ≥18% of patients with LOH in the locus. Aberration of SNP rs10248565 at *HDAC9* in chromosome 7p21.1 was identified from concurrent analyses of CNVs, SNPs, and LOH.

Conclusion: The results elucidate the genetic etiology of lung adenocarcinoma by demonstrating that SNP rs10248565 may be a potential biomarker of cancer susceptibility.

Keywords: Lung cancer, Microarray, rs10248565, *HDAC9*, Adenocarcinoma, Non-smoking

Background

One of most commonly diagnosed cancers is lung cancer, which accounts for nearly 18% of all cancer-related deaths worldwide [1]. In the United States and other Western countries, the 5-year survival rate of lung cancer is only 15% and has not improved over several decades. In Taiwan, lung cancer mortality rates have become the highest in the world [2,3]. Even though numerous research efforts have been devoted to the development of lung cancer treatment over the past few decades, the overall 5-year survival rate is still as low as 15% [4].

Smoking is the primary risk factor for lung cancer [5]. In Western countries, 70-90% of lung cancers are attributable to cigarette smoking, whereas in Taiwan, only 7%

of female lung cancer cases are associated with smoking [6,7]. Most non-smoking female patients with lung cancer have adenocarcinoma. However, the molecular mechanisms of lung adenocarcinoma in non-smoking women remain unclear.

Cancer appears to result from the progressive accumulation of genetic aberrations ranging from large, visible chromosome events to single nucleotide polymorphisms (SNPs). Genomic rearrangements that affect DNA sequences are called structural variants and include such things as insertions, deletions, duplications and inversions [8]. When the length of a structural variant is 1 kb or longer, it is defined as a copy number variation (CNV). CNVs have played important roles in recent cancer studies. Duplicated chromosomal regions may contain dominant oncogenes (e.g., *MYC* and *ERBB2* [9-11]), whereas deleted regions may harbor tumor suppressor genes (e.g., *RB1*, *CDKN2A*, and *PTEN* [9-15]). These genes play critical

* Correspondence: llai@ntu.edu.tw; chuangey@ntu.edu.tw
[1]Graduate Institute of Physiology, National Taiwan University, Taipei, Taiwan
[2]Bioinformatics and Biostatistics Core, Center of Genomic Medicine, National Taiwan University, Taipei, Taiwan
Full list of author information is available at the end of the article

roles in multiple processes including cell growth, proliferation, apoptosis, and metastasis. For example, genomic imbalances and losses at 16q were shown to be associated with more poorly differentiated subtypes of prostate cancer [16]. DNA CNVs explained about 12% of the gene expression variations in breast cancer [17]. Concordant changes between mRNA expression levels and CNVs were observed in several genes located in copy number variable regions in lung cancer [18,19]. Furthermore, gene CNVs have been shown to be useful in predicting patient survival outcomes in lung cancer [20,21]. For example, the amplification and overexpression of epidermal growth factor receptor (*EGFR*) and the under-expression of dual specificity phosphates 4 (*DUSP4*) served as effective prognostic biomarkers in lung cancer [19,22]. Therefore, it is important to characterize DNA copy number changes for both the basic understanding of cancer and its diagnosis.

In this study, the genomic aberrations of lung adenocarcinoma in non-smoking women were examined using genome-wide human SNP arrays. The primary advantage of SNP arrays for our purposes was that the probe intensity of both alleles at each SNP allows the detection of CNV breakpoints and the estimation of the associated number of copies. In addition, loss of heterozygosity (LOH), used for surveying segments of allelic losses, can be examined by analyzing the genotype of both normal and lung tissues [23,24].

Although SNP genotyping is often used for examining the associations between cancer and normal tissues, the main focus of this study was not to identify the association of SNPs with lung adenocarcinoma in non-smoking women. Instead, we took advantage of the ability of whole-genome SNP arrays to concurrently analyze CNVs, SNPs, and LOH in order to identify the novel focal loci of lung adenocarcinoma. All results indicated that the SNP rs10248565 in *HDAC9*, the gene encoding histone deacetylase, was related to lung carcinogenesis. In this study, we demonstrated that SNP rs10248565 in *HDAC9* may be a potential biomarker for identifying important genetic determinants of cancer susceptibility and elucidating the genetic etiology of lung cancer in non-smoking females.

Methods

Sample collection

The study protocol was approved by the institutional review boards of National Taiwan University Hospital and Taichung Veterans General Hospital. The written consent form was approved by ethics committees, and all participants agreed with their written consents to participate in this study. In total, 120 pairs of cancer and normal lung tissue specimens were collected from non-smoking females. The selection criteria of clinical specimens depended on the pathology report, physical examination and cigarette smoking history. Surgical lung tissue specimens were immediately snap-frozen in liquid N_2 and stored at $-80°C$ until being further processed for DNA extraction. Only those paired samples passing quality checks (n = 61 pairs) were processed for SNP arrays.

Isolation of genomic DNA, DNA amplification, labeling and hybridization of SNP arrays

Genomic DNA was isolated by phenol/chloroform extraction following standard protocols with 0.5% SDS and 200 μg/ml proteinase K. Total genomic DNA (250 ng) was digested with a restriction enzyme (*Nsp*I or *Sty*I) and ligated to adaptors that recognize the cohesive four bp overhangs. All fragments resulting from restriction enzyme digestion were substrates for adaptor ligation. A generic primer that recognizes the adaptor sequence was used to amplify adaptor-ligated DNA fragments. PCR conditions had been optimized to preferentially amplify fragments in the 200 to 1,100 bp size range. The amplified DNA was then fragmented, labeled, and hybridized to Genome-Wide Human SNP 6.0 arrays (Affymetrix, Inc., Santa Clara, CA). After 16 hours of hybridization at 49°C, the arrays were washed by Fluidics Station 450 and scanned by GeneChip Scanner 3000. Microarray data of this study are MIAME compliant, and have been submitted to the MIAME compliant Gene Expression Omnibus (GEO) database (accession number GSE33355).

Copy number variation analysis

After scanning, the intensity data were analyzed by Partek® software (Partek®, St. Louis, MO, USA). Since both tumor and adjacent normal tissues were from the same individual, the reference baseline for each tumor tissue was its corresponding normal tissue. The criteria for searching for CNV regions in the whole genome were as follows: 1) copy number intensity ratio of tumor to normal tissue for each SNP was >2.5 or <1.5; 2) each individual had ≥100 continuous SNP variant loci; 3) the CNV regions existed in ≥30% of the study population. The overlapping genes located within the detected CNV regions were annotated using the documentation file version 30 provided by Affymetrix.

TaqMan® copy number assays

TaqMan® assays were used to validate the total copy number of CNV regions. Total genomic DNA (20 ng; 5 ng/μl) was used for TaqMan® Copy Number assays (Life Technologies, Carlsbad, CA, USA). All reactions were performed in duplicate, including the FAM™ dye label-based assay for the target of interest and the VIC® dye label-based TaqMan® Copy Number Reference Assay. The TaqMan® probes for the target of interest were labeled with FAM at the 5′ end and linked by a non-fluorescent quencher at the 3′ end. RNase P labeled with VIC dye (Life Technologies) was utilized as the reference gene,

which is known to exist in two copies in a diploid genome. All TaqMan® assays were performed following manufacturer's instructions and copy number calculation was conducted by the delta-delta threshold cycle ($\Delta\Delta Ct$) method. PCR was performed with an Applied Biosystems 7900HT Fast Real-Time PCR System (Applied Biosystems, Carlsbad, CA, USA). Results were analyzed by CopyCaller™ version 1.0. Tumor samples with a delta Ct value between target and reference sequences were measured, and then compared to their paired normal samples.

Single nucleotide polymorphism analysis

For SNP analysis, SNPs were obtained using Affymetrix® SNP Array 6.0 (each has more than 906,600 SNPs). After excluding SNPs with allele frequency <1% (157,703 SNPs) or call rate <90% (123 SNPs), 748,774 SNPs were further analyzed by McNemar-Bowker's test to examine the difference of genotypes between normal and tumor tissues from the same subject. SNPs were coded according to the number of minor alleles, i.e., AA, Aa and aa, denoted as 0, 1, 2, respectively. The nonparametric McNemar-Bowker's test was applied to examine the association between SNPs and tissues. The analyses were done in R version 2.9.0.

Loss of heterozygosity analysis

Loss of heterozygosity (LOH) was defined as heterozygosity in normal tissue and homozygosity in tumor tissue. The genotypes between tumor tissue and its normal counterpart from the same subject were compared using Genome-Wide Human SNP 6.0 arrays (Affymetrix, Inc., Santa Clara, CA). LOH SNPs were identified in ≥18% of patients with LOH in the locus.

Results

DNA genetic aberration analysis

In this study, pairs of adenocarcinoma and adjacent normal lung tissue specimens were collected from 61 non-smoking women for the purpose of examining genome-wide DNA aberrations. The majority (72%; n = 44) of women were in early stages (I + II) and the mean (SD) age was 59.4 (11). Their clinical characteristics are listed in Table 1. CNV, SNP, and LOH were concurrently analyzed

Table 1 Characteristics of 61 non-smoking female lung adenocarcinoma patients

Characteristics	Sample size	Age
Female	61	59.4 ± 11
Tumor types		
Adenocarcinoma	61	59.4 ± 11
Tumor stage		
I + II	44	60 ± 11
III + IV	17	58 ± 11

using Affymetrix Genome-Wide Human SNP 6.0 arrays. All chips' call rates were greater than 99%.

Copy number variation analysis

We first identified common CNV regions among these lung adenocarcinoma samples. The criteria for searching the CNV regions in whole genome were stated in the Methods section. In total, there were 424 CNV regions. Figure 1A shows the distribution of CNV for each chromosome among 61 paired samples. Each grey bar indicates the amplification or deletion regions in tumor tissue. Black bars indicate where >30% of patients (n > 18) had CNV. An expanded view of these results showed that one third or more of these patients had a genetic amplification at 7p21.3-7p21.1 and 7p11.2 (Figure 1B). In contrast, no common deletion regions were identified.

In order to validate the common amplification regions, four CNV regions in 7p21.1 were chosen for TaqMan® copy number assays. In the upper panel of Figure 1C, the positions of 4 CNV regions (black blocks) identified by SNP arrays and those examined by TaqMan® copy number assays (grey blocks) are shown. The TaqMan assays showed that the copy numbers in all 4 regions were greater than normal (lower panel of Figure 1C), indicating these regions are common amplification regions in non-smoking female lung cancer patients.

In order to understand the function of genes in common amplification regions, functional analysis was done using Ingenuity Pathway Analysis (IPA). The results revealed that the common amplification regions contain 29 genes. Network analysis showed that these genes were mainly involved in cellular development, cellular growth and proliferation, and cancer. Among these 29 genes, *EGFR* (encoding epidermal growth factor receptor) and *HDAC9* were previously reported to have an association with lung tumorigenesis.

Single nucleotide polymorphism analysis

Next, genotyping of SNPs was analyzed in normal and adenocarcinoma tissues from the same subject. SNPs with minor allele frequency <0.01 were excluded. After excluding SNPs with low minor allele frequency (181,503 SNPs) or SNPs with departing Hardy-Weinberg Equilibrium (*P*-value <0.0001; 2,816 SNPs), the remaining 684,877 SNPs on autosomal chromosomes were further analyzed by McNemar-Bowker's test to examine the differences in genotypes.

Since this study adopted a paired design, which provides less variation than general case–control studies and can achieve a higher statistical power, a strict criterion of *P*-values, Bonferroni correction, was not performed. As shown in Figure 2, a Manhattan plot showed that there were four SNPs with *P*-values smaller than 0.01. Each dot represents a SNP. The
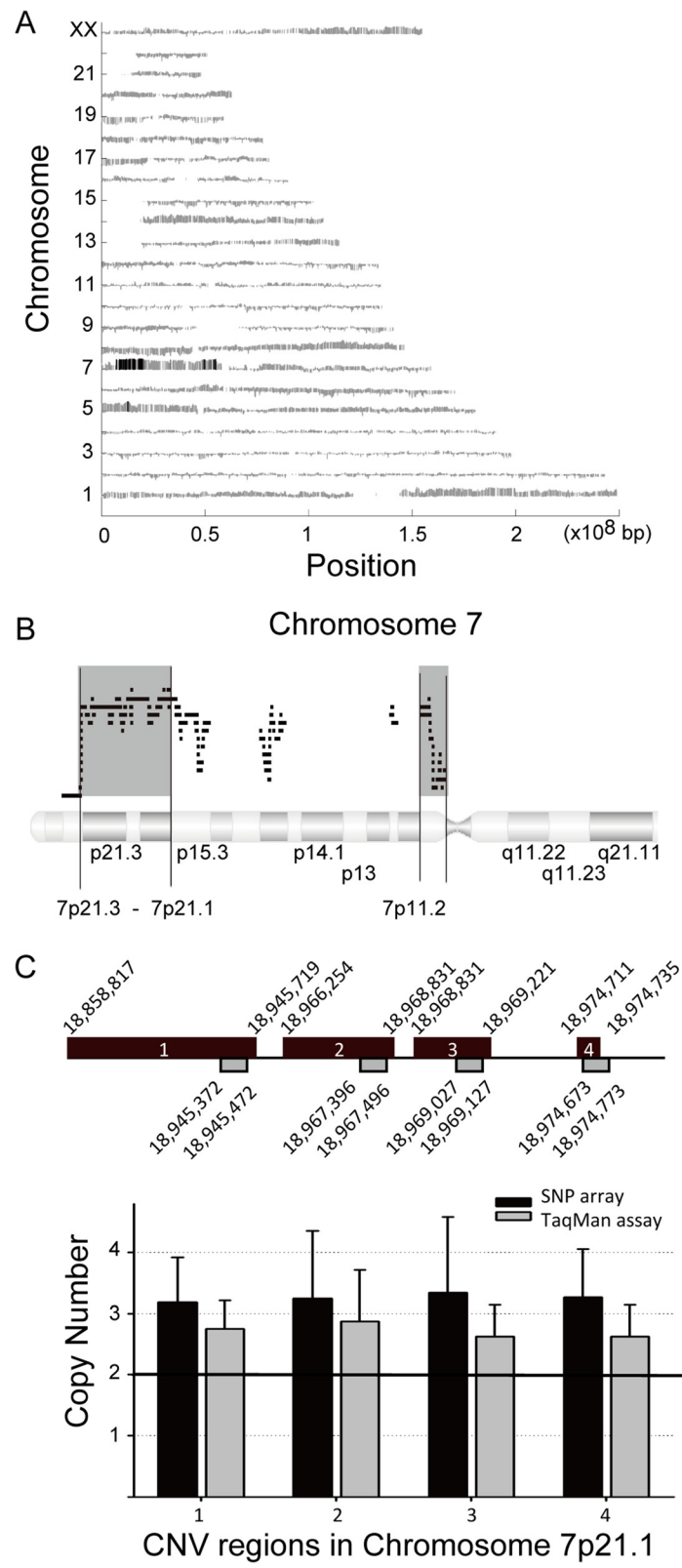
Figure 1 (See legend on next page.)

(See figure on previous page.)
Figure 1 Copy number variation analysis in non-smoking female lung adenocarcinoma patients. (A) Distribution of genome-wide CNV using Affymetrix GeneChip® SNP 6.0 analysis. The criteria for the CNV regions were that SNPs must have copy numbers >2.5 or <1.5 and there must be at least 100 continuous SNP variant loci. Grey bars indicate regions with gains (above) or losses (below) in copy number. Black bars indicate that ≥30% of the patients had a particular CNV. **(B)** Common gain regions (≥30% of patients) of CNVs (grey area) were identified in 7p21.3-7p21.1 and 7p11.2. Black lines of each row indicate regions with copy number amplifications at >100 continuous SNP loci for each patient. **(C)** TaqMan assay validation of CNVs in chromosome 7p21.1. Four CNV regions (black block 1–4) identified by SNP arrays were examined using TaqMan® copy number assays (grey block).

distribution of −log(P-value) of each SNP was plotted across chromosomes. Information on these four SNPs (rs1700874, rs10248565, rs11761619, and rs9316119) is listed in Table 2. Only rs1700874 was located in an intergenic region (between *TGFB2* and *LYPLAL1*); the rest of the SNPs were located in introns. Among them, the SNP with the lowest P-value, rs10248565, is located in an intron of *HDAC9*.

Loss of heterozygosity analysis

Lastly, we examined the distribution of LOH in each chromosome (Figure 3). Because the proportion of LOH loci ranged from 10% to over 50%, LOH for a SNP was defined as at least 18% of patients (≥11 patients) with LOH in the locus. As shown in Figure 3A, black bars indicate ≥18% of patients with the LOH SNP. In total, there were 30 SNPs indicating LOH. Most of these SNPs appeared in chromosome 7 (Figure 3B).

Among these LOH SNPs, we noticed that SNP rs10248565 was associated with lung adenocarcinoma and was located in the CNV region. Combining the results of CNV, SNP, and LOH analyses (Figure 4), we concluded that rs10248565 is a possible biomarker of lung adenocarcinoma in non-smoking females.

Discussion

It is well-known that there are many causative elements in cancer progression and tumorigenesis, such as sequence mutations, transcriptional alterations, and genomic changes. Among these complicated factors,

structural variations of DNA sequences have been widely reported to serve as a key driver to dysregulate the transcriptome during tumorigenesis [17]. Furthermore, since genes located within the variable regions are candidate oncogenes or tumor suppressors, an integrative analysis of CNV, SNP, and LOH may provide more information in dissecting the lung tumorigenic process. To help explain the relationship between copy number and gene expression, we performed an integrative analysis in paired lung adenocarcinoma tissue specimens to identify genomic alterations in tumor tissues.

Our CNV results showed that at least 30% of the samples had amplifications at chromosomes 7p21.3-7p21.1 and 7p11.2. However, no deletion regions were identified. This may due to the stringent selection criteria adopted here. Comparing with other studies, several aberrant regions have been detected using high resolution karyotyping techniques to scan lung cancer genome, such as amplifications of 3p25-27 and 5p13-14, and deletions of 3p21 and 9q21 [25]. Several studies reported that a deletion on chromosome 5q in small cell and squamous cell lung cancer subtypes may be associated with smoking history [26-29]. Conversely, amplifications of 5q have been detected in adenocarcinoma [30,31]. We did not observe any amplification regions in chromosome 5 in this study, which may be explained by differences in experimental design, selection criteria, and ethnicity of study populations.

Further investigations of the 29 genes located within these CNV regions identified several key players involved in the tumorigenic process. For instance, loss of
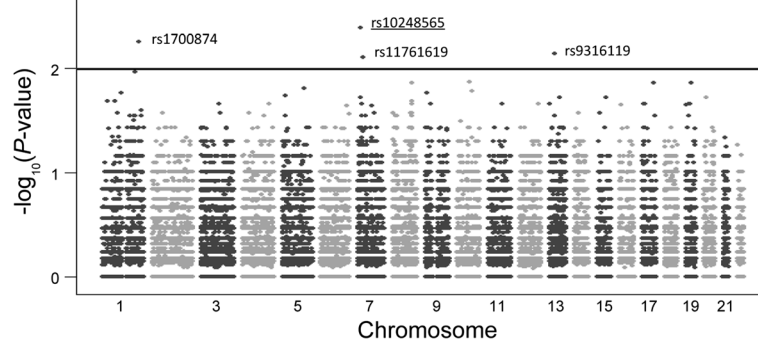

Figure 2 Single nucleotide polymorphism analysis in non-smoking female lung adenocarcinoma patients. Each dot denotes a SNP. The distribution of −log(P-value) of each SNP was plotted across chromosomes. The four SNPs with P-values smaller than 10^{-2} are labeled.

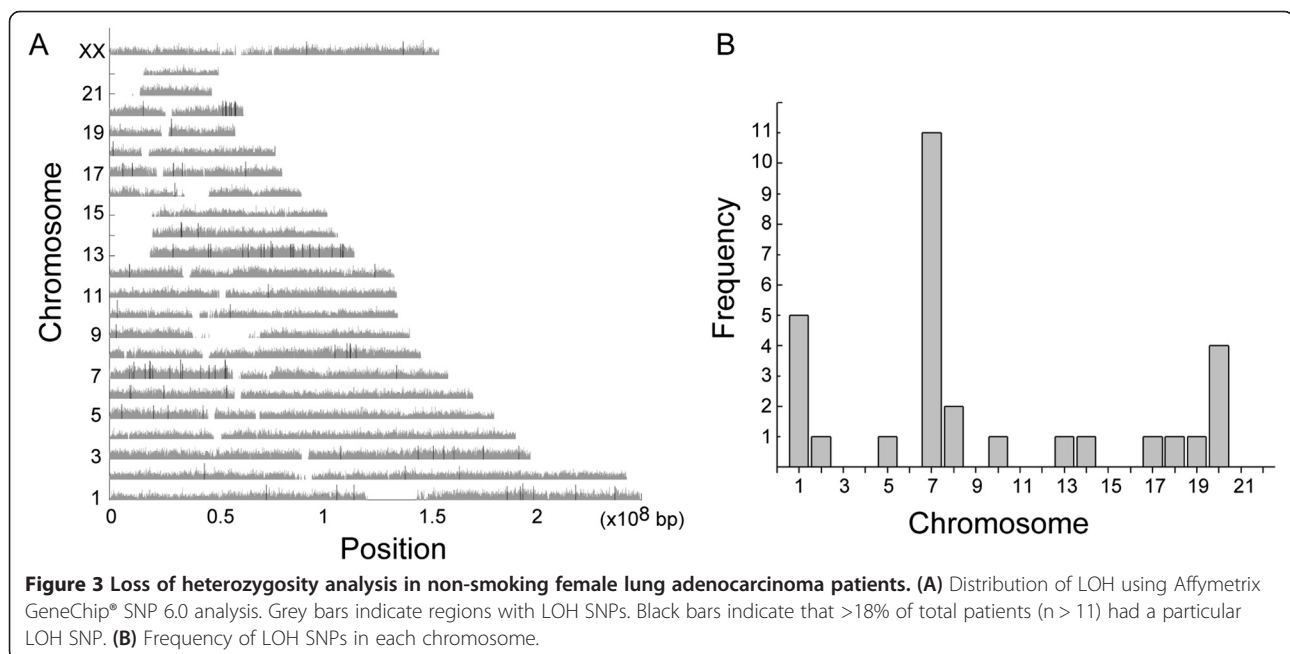**Table 2 SNPs significantly associated with lung tumors in non-smoking female patients**

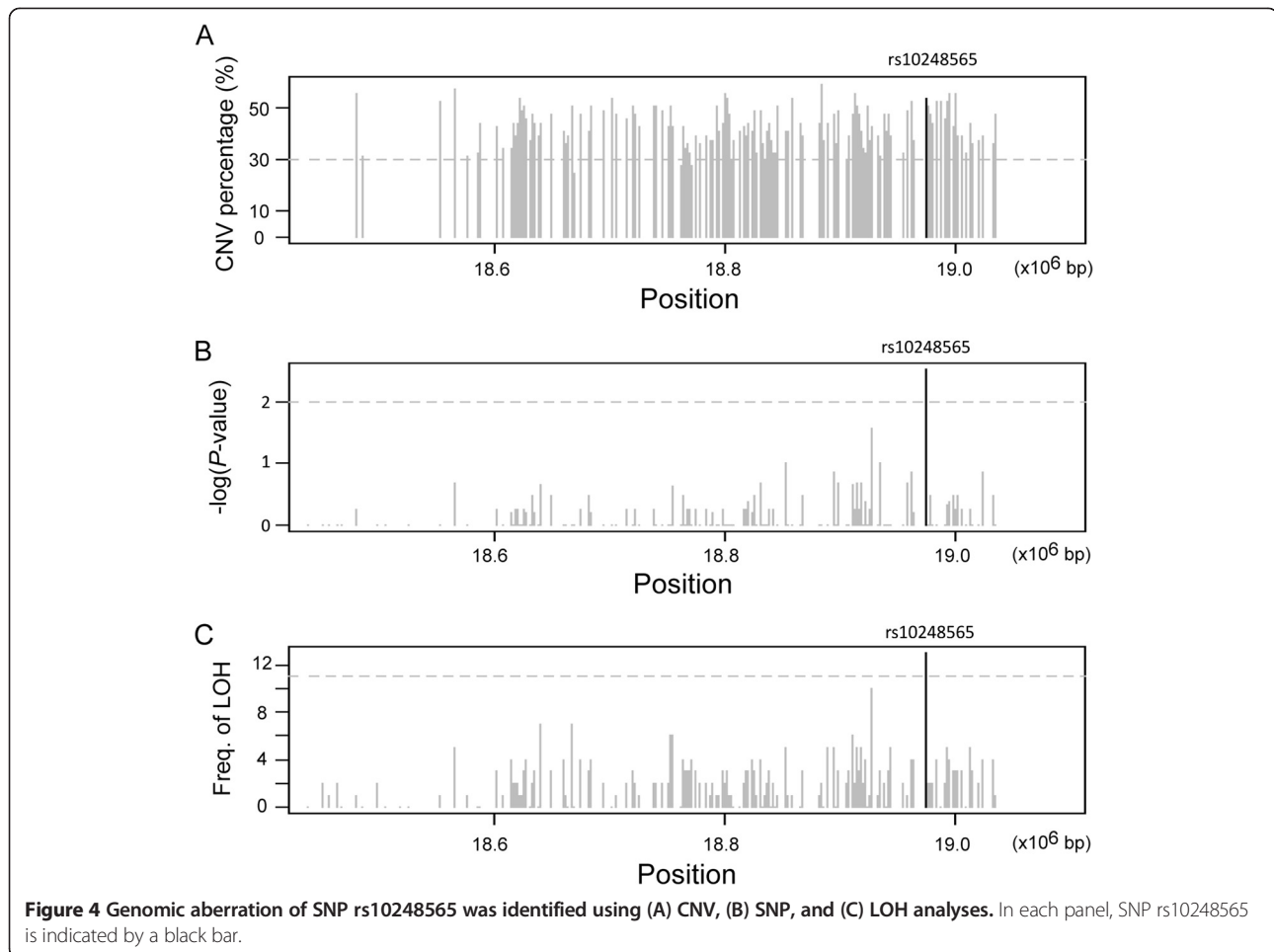| SNP	Chromosome	Position	P-value	Location
rs1700874	1	219,182,858	4×10^{-3}	Intergenic *TGFB2* & *LYPLAL1*
rs10248565	7	18,974,723	3×10^{-3}	*HDAC9* intron
rs11761619	7	33,549,392	8×10^{-3}	*BBS9* intron
rs9316119	13	45,696,862	6×10^{-3}	*GTF2F2* intron

docking protein 2 (*DOK2*) as well as expression of baculoviral IAP repeat-containing 2/3 (*BIRC2/3*) can facilitate lung cancer cell proliferation and contribute to lung tumor development [32,33]. EGFR is involved in the signal transduction pathways of cell proliferation, differentiation, adhesion, protection from apoptosis and survival. Numerous reports have shown that EGFR gene mutations are frequently detected in lung cancer, especially in adenocarcinoma, females, and non-smoking patients [34].

The gene encoding HDAC in chromosome 7p21.1 was identified in all CNV, SNP and LOH analyses and is worthy of mention here. HDAC is involved in deacetylation of lysine residues in the N-terminal tails of nucleosomal core histones [35], and it has also been implicated in the development of cancer [36]. The activity of several tumor suppressors is regulated in part by HDACs, such as p53 binding protein that regulates cell cycling in response to DNA damage [37]. HDAC inhibitors were developed as anti-cancer agents with a high degree of selectivity for killing cancer cells. In one study, inhibition of HDAC induced DNA damage which only normal cells, but not cancer cells, can repair, and resulted in cancer cell death [38].

Inhibition of *HDAC6* significantly enhanced cell death induced by the topoisomerase II inhibitors in transformed cells, but not in normal cells [21]. Inhibition of *HDAC1* and *HDAC2* enhanced the radiosensitivity of non-small cell lung cancer [39]. Unfortunately, the expression levels of *HDAC9* did not differ significantly between tumor tissue and adjacent normal tissue in our study (data not shown). This may be due to the location of the SNP in an intron of *HDAC9*, and further investigation of the mechanism of genomic aberration in *HDAC9* is warranted.

The hypothesis underlying our SNP analysis was that if SNPs were associated with cancer, the proportions of different alleles would be different in cancer and normal groups. Previously, rs7086803 at 10q25.2, rs9387478 at 6q22.2 and rs2395185 at 6p21.32 were identified as lung cancer susceptibility loci in never-smoking women in Asia [40]. In this study, we identified another 4 SNPs (rs1700874, rs10248565, rs11761619, and rs9316119) that were significantly ($P < 0.01$) associated with lung cancer. SNP rs1700874 is located at an intergenic region in 1q41 between *TGFB2* and *LYPLAL1*. The transforming growth factor beta family plays an important role in cell cycling, cell growth, apoptosis, and protein synthesis, and is therefore involved in many pathological processes [41,42]. A previous study showed that *TGFB2* may correlate with heart disease and pulmonary function in mice [43]. The function of *LYPLAL1* is still unclear. SNP rs10248565 is located in 7p21.1 within *HDAC9*, the significance of which was discussed above. SNP rs11761619 is in 7p14.3 within *BBS9*. *BBS9* is associated with kidney and ovarian diseases [44,45], and may be a tumor suppressor gene for Wilms' tumor [46]. SNP rs9316119 is in 13q14.12 within *GTF2F2*,

Figure 3 Loss of heterozygosity analysis in non-smoking female lung adenocarcinoma patients. (A) Distribution of LOH using Affymetrix GeneChip® SNP 6.0 analysis. Grey bars indicate regions with LOH SNPs. Black bars indicate that >18% of total patients (n > 11) had a particular LOH SNP. **(B)** Frequency of LOH SNPs in each chromosome.

Figure 4 Genomic aberration of SNP rs10248565 was identified using (A) CNV, (B) SNP, and (C) LOH analyses. In each panel, SNP rs10248565 is indicated by a black bar.

which is known to affect the progression and survival of epithelial ovarian cancer [47].

In this study, we identified 30 SNPs with LOH. Most of these LOH SNPs were located in chromosome 7 (Figure 3B). LOH analysis has been used to identify genomic aberrations in previous studies. For instance, loss of heterozygosity at chromosomal regions 3p21.3 (site of *RASSF1A*, a member of the Ras association domain family, and *FUS1*), 3p14.2 (*FHIT*, a fragile histidine triad gene), 9p21 (*p16*), and 17p13 (*p53*) was identified as an early event in the development of non-small cell lung cancer [48].

It may seem contradictory that the SNP rs10248565 in *HDAC9* was increased in copy number but also showed loss of heterozygosity. The observed LOH accompanied by a gain in copy number may result from preferential amplification of one parental allele, because CNV analysis cannot identify situations in which the loss of one allele is followed by duplication of the remaining allele. Also, LOH cannot detect any amplification that might be involved in pathogenesis. Therefore, we conducted a concurrent LOH and CNV analysis with the expectation of more precisely defining the nature of genomic alternations observed in either analysis alone.

Conclusion

In conclusion, the high mortality of lung cancer worldwide is largely attributable to the difficulty of obtaining an early diagnosis and the lack of effective therapeutic methods. To improve survival rates in non-smoking lung cancer patients, a comprehensive analysis of the molecular signature of the carcinogenic processes in adenocarcinoma in non-smoking Taiwanese women was conducted to identify novel biomarkers for diagnosis and new molecular targets for drug development. Although more studies are still needed, SNP rs10248565 in *HDAC9* may be one of the potential biomarkers for lung adenocarcinoma in non-smoking women.

Microarray data from this study have been submitted to the Gene Expression Omnibus database (accession number GSE33355).

Authors' contributions
LCL, MHT, CKH, and EYC provided conception and design. LCL, CKH, and EYC provided financial support. JML and CPH provided study materials and patients. PCC and SKC collected and assembled data. LCL, MHT, PCC, LHC, JHH, TPL, and EYC analyzed and interpreted data. LCL, MHT, PCC, and EYC wrote manuscript. All authors read and approved the final manuscript.

Author details
[1]Graduate Institute of Physiology, National Taiwan University, Taipei, Taiwan. [2]Bioinformatics and Biostatistics Core, Center of Genomic Medicine, National Taiwan University, Taipei, Taiwan. [3]Institute of Biotechnology, National Taiwan University, Taipei, Taiwan. [4]YongLin Biomedical Engineering Center, National Taiwan University, Taipei, Taiwan. [5]Graduate Institute of Biomedical Electronics and Bioinformatics, Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan. [6]Department of Public Health, National Taiwan University, Taipei, Taiwan. [7]Department of Statistics and Informatics Science, Providence University, Taichung, Taiwan. [8]Department of Surgery, National Taiwan University Hospital, Taipei, Taiwan. [9]Division of Thoracic Surgery, Taichung Veterans General Hospital, Taichung, Taiwan.

References
1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D: **Global cancer statistics.** *CA Cancer J Clin* 2011, **61**:69–90.
2. Chen CJ, Wu HY, Chuang YC, Chang AS, Luh KT, Chao HH, Chen KY, Chen SG, Lai GM, Huang HH: **Epidemiologic characteristics and multiple risk factors of lung cancer in Taiwan.** *Anticancer Res* 1990, **10**:971–976.
3. Yang SP, Luh KT, Kuo SH, Lin CC: **Chronological observation of epidemiological characteristics of lung cancer in Taiwan with etiological consideration–a 30-year consecutive study.** *Jpn J Clin Oncol* 1984, **14**:7–19.
4. Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ: **Cancer statistics, 2007.** *CA Cancer J Clin* 2007, **57**:43–66.
5. Wood ME, Kelly K, Mullineaux LG, Bunn PAJ: **The inherited nature of lung cancer: a pilot study.** *Lung Cancer* 2000, **30**:135–144.
6. Ger LP, Liou SH, Shen CY, Kao SJ, Chen KT: **Risk factors of lung cancer.** *J Formos Med Assoc* 1992, **91**(Suppl 3):S222–S231.
7. Hirayasu T, Iwamasa T, Kamada Y, Koyanagi Y, Usuda H, Genka K: **Human papillomavirus DNA in squamous cell carcinoma of the lung.** *J Clin Pathol* 1996, **49**:810–817.
8. Alkan C, Coe BP, Eichler EE: **Genome structural variation discovery and genotyping.** *Nat Rev Genet* 2011, **12**:363–376.
9. Attiyeh EF, London WB, Mosse YP, Wang Q, Winter C, Khazi D, McGrady PW, Seeger RC, Look AT, Shimada H, Brodeur GM, Cohn SL, Matthay KK, Maris JM: **Chromosome 1p and 11q deletions and outcome in neuroblastoma.** *N Engl J Med* 2005, **353**:2243–2253.
10. Seeger RC, Brodeur GM, Sather H, Dalton A, Siegel SE, Wong KY, Hammond D: **Association of multiple copies of the N-myc oncogene with rapid progression of neuroblastomas.** *N Engl J Med* 1985, **313**:1111–1116.
11. Little CD, Nau MM, Carney DN, Gazdar AF, Minna JD: **Amplification and expression of the c-myc oncogene in human lung cancer cell lines.** *Nature* 1983, **306**:194–196.
12. Kamb A, Gruis NA, Weaver-Feldhaus J, Liu Q, Harshman K, Tavtigian SV, Stockert E, Day RS 3rd, Johnson BE, Skolnick MH: **A cell cycle regulator potentially involved in genesis of many tumor types.** *Science* 1994, **264**:436–440.
13. Friend SH, Bernards R, Rogelj S, Weinberg RA, Rapaport JM, Albert DM, Dryja TP: **A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma.** *Nature* 1986, **323**:643–646.
14. Mosse YP, Laudenslager M, Longo L, Cole KA, Wood A, Attiyeh EF, Laquaglia MJ, Sennett R, Lynch JE, Perri P, Laureys G, Speleman F, Kim C, Hou C, Hakonarson H, Torkamani A, Schork NJ, Brodeur GM, Tonini GP, Rappaport E, Devoto M, Maris JM: **Identification of ALK as a major familial neuroblastoma predisposition gene.** *Nature* 2008, **455**:930–935.
15. Li J, Yen C, Liaw D, Podsypanina K, Bose S, Wang SI, Puc J, Miliaresis C, Rodgers L, McCombie R, Bigner SH, Giovanella BC, Ittmann M, Tycko B, Hibshoosh H, Wigler MH, Parsons R: **PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer.** *Science* 1997, **275**:1943–1947.
16. Teixeira MR, Ribeiro FR, Eknaes M, Waehre H, Stenwig AE, Giercksky KE, Heim S, Lothe RA: **Genomic analysis of prostate carcinoma specimens obtained via ultrasound-guided needle biopsy may be of use in pre-operative decision-making.** *Cancer* 2004, **101**:1786–1793.
17. Pollack JR, Sorlie T, Perou CM, Rees CA, Jeffrey SS, Lonning PE, Tibshirani R, Botstein D, Borresen-Dale AL, Brown PO: **Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors.** *Proc Natl Acad Sci U S A* 2002, **99**:12963–12968.
18. Hirsch FR, Varella-Garcia M, Cappuzzo F: **Predictive value of EGFR and HER2 overexpression in advanced non-small-cell lung cancer.** *Oncogene* 2009, **28**(Suppl 1):S32–S37.
19. Chitale D, Gong Y, Taylor BS, Broderick S, Brennan C, Somwar R, Golas B, Wang L, Motoi N, Szoke J, Reinersman JM, Major J, Sander C, Seshan VE, Zakowski MF, Rusch V, Pao W, Gerald W, Ladanyi M: **An integrated genomic analysis of lung cancer reveals loss of DUSP4 in EGFR-mutant tumors.** *Oncogene* 2009, **28**:2773–2783.
20. Kawano O, Sasaki H, Okuda K, Yukiue H, Yokoyama T, Yano M, Fujii Y: **PIK3CA gene amplification in Japanese non-small cell lung cancer.** *Lung Cancer* 2007, **58**:159–160.
21. Namdar M, Perez G, Ngo L, Marks PA: **Selective inhibition of histone deacetylase 6 (HDAC6) induces DNA damage and sensitizes transformed cells to anticancer agents.** *Proc Natl Acad Sci U S A* 2010, **107**:20003–20008.
22. Li AR, Chitale D, Riely GJ, Pao W, Miller VA, Zakowski MF, Rusch V, Kris MG, Ladanyi M: **EGFR mutations in lung adenocarcinomas: clinical testing experience and relationship to EGFR gene copy number and immunohistochemical expression.** *J Mol Diagn* 2008, **10**:242–248.
23. Huang J, Wei W, Zhang J, Liu G, Bignell GR, Stratton MR, Futreal PA, Wooster R, Jones KW, Shapero MH: **Whole genome DNA copy number changes identified by high density oligonucleotide arrays.** *Hum Genomics* 2004, **1**:287–299.
24. Nannya Y, Sanada M, Nakazaki K, Hosoya N, Wang L, Hangaishi A, Kurokawa M, Chiba S, Bailey DK, Kennedy GC, Ogawa S: **A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays.** *Cancer Res* 2005, **65**:6071–6079.
25. Balsara BR, Testa JR: **Chromosomal imbalances in human lung cancer.** *Oncogene* 2002, **21**:6877–6883.
26. Boelens MC, Gustafson AM, Postma DS, Kok K, van der Vries G, van der Vlies P, Spira A, Lenburg ME, Geerlings M, Sietsma H, Timens W, van den Berg A, Groen HJ: **A chronic obstructive pulmonary disease related signature in squamous cell lung cancer.** *Lung Cancer* 2011, **72**:177–183.
27. Boelens MC, Kok K, van der Vlies P, van der Vries G, Sietsma H, Timens W, Postma DS, Groen HJ, van den Berg A: **Genomic aberrations in squamous cell lung carcinoma related to lymph node or distant metastasis.** *Lung Cancer* 2009, **66**:372–378.
28. Miura I, Graziano SL, Cheng JQ, Doyle LA, Testa JR: **Chromosome alterations in human small cell lung cancer: frequent involvement of 5q.** *Cancer Res* 1992, **52**:1322–1328.
29. Hartel PH, Shackelford AL, Hartel JV, Wenger SL: **Del(5q) is associated with clinical and histological parameters in small cell neuroendocrine lung carcinoma.** *Int J Surg Pathol* 2008, **16**:419–423.
30. Yen CC, Liang SC, Jong YJ, Chen YJ, Lin CH, Chen YM, Wu YC, Su WC, Huang CY, Tseng SW, Whang-Peng J: **Chromosomal aberrations of malignant pleural effusions of lung adenocarcinoma: different cytogenetic changes are correlated with genders and smoking habits.** *Lung Cancer* 2007, **57**:292–301.
31. Shen H, Gao W, Wu YJ, Qiu HR, Shu YQ: **Multicolor fluorescence in situ hybridization and comparative genomic hybridization reveal molecular events in lung adenocarcinomas and squamous cell lung carcinomas.** *Biomed Pharmacother* 2009, **63**:396–403.
32. Dai Z, Zhu WG, Morrison CD, Brena RM, Smiraglia DJ, Raval A, Wu YZ, Rush LJ, Ross P, Molina JR, Otterson GA, Plass C: **A comprehensive search for**

DNA amplification in lung cancer identifies inhibitors of apoptosis cIAP1 and cIAP2 as candidate oncogenes. *Hum Mol Genet* 2003, **12**:791–801.

33. Berger AH, Niki M, Morotti A, Taylor BS, Socci ND, Viale A, Brennan C, Szoke J, Motoi N, Rothman PB, Teruya-Feldstein J, Gerald WL, Ladanyi M, Pandolfi PP: Identification of DOK genes as lung tumor suppressors. *Nat Genet* 2010, **42**:216–223.

34. Shigematsu H, Lin L, Takahashi T, Nomura M, Suzuki M, Wistuba II, Fong KM, Lee H, Toyooka S, Shimizu N, Fujisawa T, Feng Z, Roth JA, Herz J, Minna JD, Gazdar AF: Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *J Natl Cancer Inst* 2005, **97**:339–346.

35. Gray SG, Ekstrom TJ: The human histone deacetylase family. *Exp Cell Res* 2001, **262**:75–83.

36. Wade PA: Transcriptional control at regulatory checkpoints by histone deacetylases: molecular connections between cancer and chromatin. *Hum Mol Genet* 2001, **10**:693–698.

37. Yang XJ, Gregoire S: Class II histone deacetylases: from sequence to function, regulation, and clinical implication. *Mol Cell Biol* 2005, **25**:2873–2884.

38. Lee JH, Choy ML, Ngo L, Foster SS, Marks PA: Histone deacetylase inhibitor induces DNA damage, which normal but not transformed cells can repair. *Proc Natl Acad Sci U S A* 2010, **107**:14639–14644.

39. Han Y, Zhang Y, Yang LH, Mi XY, Dai SD, Li QC, Xu HT, Yu JH, Li G, Zhao J, Han C, Yuan XM, Wang EH: X-radiation inhibits histone deacetylase 1 and 2, upregulates Axin expression and induces apoptosis in non-small cell lung cancer. *Radiat Oncol* 2012, **7**:183.

40. Lan Q, Hsiung CA, Matsuo K, Hong YC, Seow A, Wang Z, Hosgood HD 3rd, Chen K, Wang JC, Chatterjee N, Hu W, Wong MP, Zheng W, Caporaso N, Park JY, Chen CJ, Kim YH, Kim YT, Landi MT, Shen H, Lawrence C, Burdett L, Yeager M, Yuenger J, Jacobs KB, Chang IS, Mitsudomi T, Kim HN, Chang GC, Bassig BA, *et al*: Genome-wide association analysis identifies new lung cancer susceptibility loci in never-smoking women in Asia. *Nat Genet* 2012, **44**:1330–1335.

41. Jenkins G: The role of proteases in transforming growth factor-beta activation. *Int J Biochem Cell Biol* 2008, **40**:1068–1078.

42. Bonsi L, Marchionni C, Alviano F, Lanzoni G, Franchina M, Costa R, Grossi A, Bagnara GP: Thrombocytopenia with absent radii (TAR) syndrome: from hemopoietic progenitor to mesenchymal stromal cell disease? *Exp Hematol* 2009, **37**:1–7.

43. Sanford LP, Ormsby I, Gittenberger-de Groot AC, Sariola H, Friedman R, Boivin GP, Cardell EL, Doetschman T: TGFbeta2 knockout mice have multiple developmental defects that are non-overlapping with other TGFbeta knockout phenotypes. *Development* 1997, **124**:2659–2670.

44. Kang H, Lee SK, Kim MH, Song J, Bae SJ, Kim NK, Lee SH, Kwack K: Parathyroid hormone-responsive B1 gene is associated with premature ovarian failure. *Hum Reprod* 2008, **23**:1457–1465.

45. Swolin-Eide D, Hansson S, Larsson L, Magnusson P: The novel bone alkaline phosphatase B1x isoform in children with kidney disease. *Pediatr Nephrol* 2006, **21**:1723–1729.

46. Vernon EG, Malik K, Reynolds P, Powlesland R, Dallosso AR, Jackson S, Henthorn K, Green ED, Brown KW: The parathyroid hormone-responsive B1 gene is interrupted by a t(1;7)(q42;p15) breakpoint associated with Wilms' tumour. *Oncogene* 2003, **22**:1371–1380.

47. Le Page C, Ouellet V, Quinn MC, Tonin PN, Provencher DM, Mes-Masson AM: BTF4/BTNA3.2 and GCS as candidate mRNA prognostic markers in epithelial ovarian cancer. *Cancer Epidemiol Biomarkers Prev* 2008, **17**:913–920.

48. Wistuba II, Mao L, Gazdar AF: Smoking molecular damage in bronchial epithelium. *Oncogene* 2002, **21**:7298–7306.