

# SCIENTIFIC REPORTS



OPEN

## An integrated genome-wide approach to discover deregulated microRNAs in non-small cell lung cancer: Clinical significance of *miR-23b-3p* deregulation

Received: 22 January 2015

Accepted: 18 May 2015

Published: 28 August 2015

Shahnaz Begum<sup>1,\*</sup>, Masamichi Hayashi<sup>2,\*</sup>, Takenori Ogawa<sup>2</sup>, Fayez J. Jabboure<sup>2</sup>, Mariana Brait<sup>2</sup>, Evgeny Izumchenko<sup>2</sup>, Sarit Tabak<sup>5</sup>, Steven A. Ahrendt<sup>6</sup>, William H. Westra<sup>1</sup>, Wayne Koch<sup>2</sup>, David Sidransky<sup>2</sup> & Mohammad O. Hoque<sup>2,3,4</sup>

In spite of significant technical advances, genesis and progression of non-small cell lung cancer (NSCLC) remain poorly understood. We undertook an integrated genetic approach to discover novel microRNAs that were deregulated in NSCLCs. A total 119 primary NSCLCs with matched normal were analyzed for genome-wide copy number changes. We also tested a subset of matched samples by microRNA expression array, and integrated them to identify microRNAs positioned in allelic imbalance area. Our findings support that most of the identified deregulated microRNAs (*miR-21*, *miR-23b*, *miR-31*, *miR-126*, *miR-150*, and *miR-205*) were positioned in allelic imbalance areas. Among microRNAs tested in independent 114 NSCLCs, overexpression of *miR-23b* was revealed to be a significantly poor prognostic factor of recurrence free survival (HR = 2.40, P = 0.005, 95%CI: 1.32–4.29) and overall survival (HR = 2.35, P = 0.005, 95%CI: 1.30–4.19) in multivariable analysis. In addition, overexpression of *miR-23b* in H1838 cell line significantly increased cell proliferation, while inhibition of *miR-23b* in H1437 and H1944 cell lines significantly decreased cell doubling time. In summary, integration of genomic analysis and microRNA expression profiling could identify novel cancer-related microRNAs, and *miR-23b* could be a potential prognostic marker for early stage NSCLCs. Further biological studies of *miR-23b* are warranted for the potential development of targeted therapy.

Surgical resection remains to be the major therapy for stage I and II cases of non-small cell lung cancer (NSCLC) and is considered as a first line of treatment for better survival<sup>1,2</sup>. A recent report<sup>3</sup> indicated that 5-year survival of 289 stage I NSCLC cases after surgical treatment was 63%. Many types of combination chemotherapy and targeted therapy have been investigated<sup>4</sup>. However, to date only limited response and survival benefit has been shown with any regimen<sup>5</sup>. The basis of these disappointing results stems from a lack of understanding of the basic biological mechanisms involved in the carcinogenesis of NSCLC.

<sup>1</sup>Department of Pathology, Johns Hopkins University, Baltimore, Maryland, 21231 USA. <sup>2</sup>Department of Otolaryngology-Head and Neck Surgery, Johns Hopkins University, Baltimore, Maryland, 21231 USA. <sup>3</sup>Department of Urology, Johns Hopkins University, Baltimore, Maryland, 21231 USA. <sup>4</sup>Department of Oncology, Johns Hopkins University, Baltimore, Maryland, 21231 USA. <sup>5</sup>Rosetta Genomics Ltd. 10 Plaut St., Rehovot, Israel, 76706. <sup>6</sup>Department of Surgery, Division of Surgical Oncology, University of Pittsburgh Medical Center, Pittsburgh, PA, 15213 USA. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to M.O.H. (email: mhoque1@jhmi.edu)

Cancer-specific molecular changes have utility not only as targets for therapy, but also as biomarkers for the determination of risk of recurrence for early-stage NSCLC<sup>6,7</sup>. Such prognostic capability may be due to the biologic significance of the alteration.

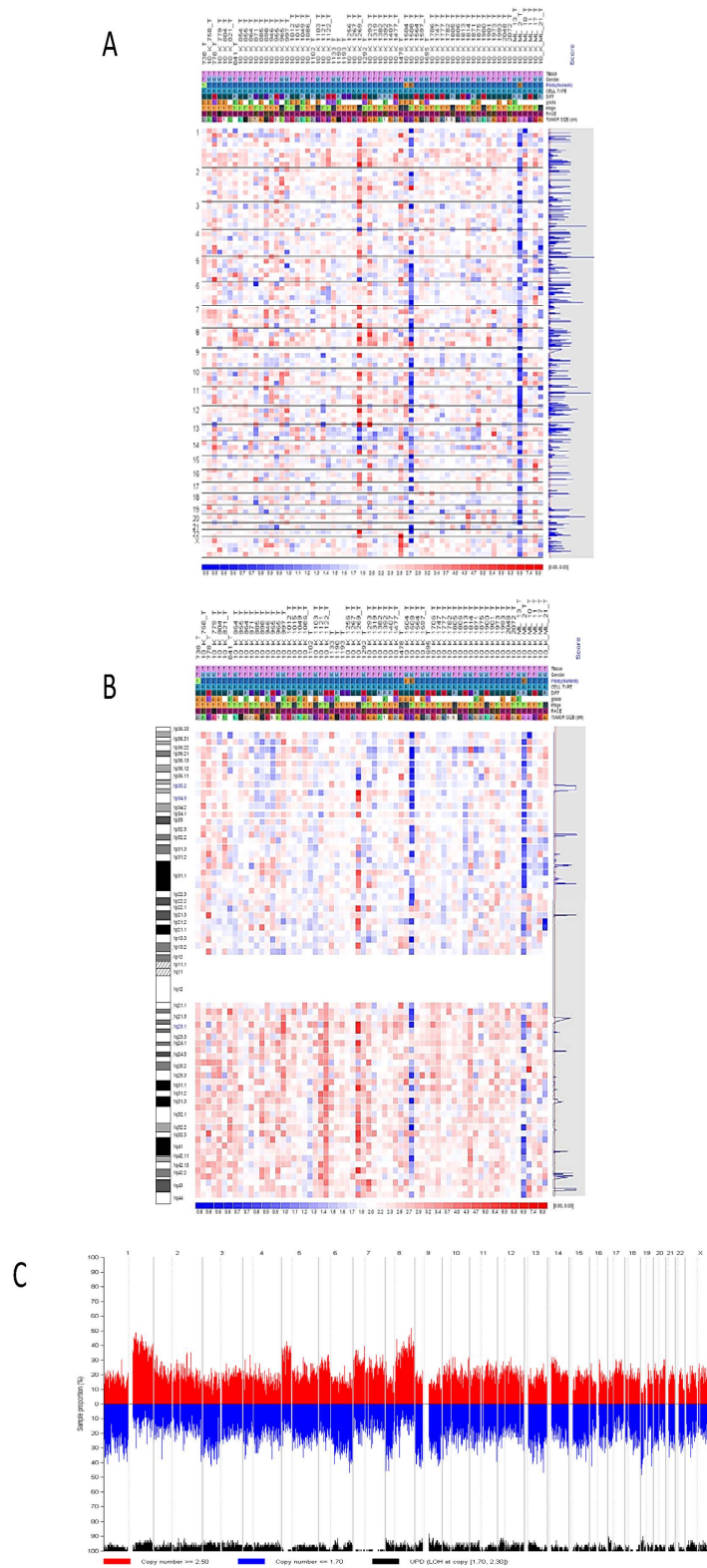
Soon after the identification of hundreds of new members of the microRNA family, it was shown that more than half of the known human microRNAs positioned in allelic imbalance areas of the cancer cells<sup>8</sup>. Such regions include minimal regions of loss of heterozygosity (LOH), which are thought to harbor tumor-suppressor genes (TSGs), minimal regions of amplification (MRA), which might contain oncogenes, common breakpoint regions in or near possible oncogenes or TSGs, and fragile sites (FRA). FRAs are preferential sites of sister-chromatid exchange, translocation, deletion, amplification or integration of plasmid DNA. The first evidence of involvement of microRNAs in cancer came from the finding that *miR-15a* and *miR-16-1* were down-regulated or deleted in most patients with chronic lymphocytic leukemia (CLL)<sup>9</sup>. *Mir-15a* and *miR-16-1* are located within the intron of a non-coding RNA gene of unknown function, called deleted in lymphocytic leukemia 2 (*DLEU2*). This gene resides in a 30 kb region at chromosome 13q14 that is deleted in >65% of CLL cases<sup>9</sup>, in 50% of mantle-cell lymphomas<sup>10</sup>, in 16–40% of multiple myeloma<sup>11</sup> and in 60% of prostate cancers<sup>12</sup>. This observation led researchers to investigate the association of microRNAs genomic locations and genomic regions that were involved in cancer. Strikingly, as noted above, 50% of the known microRNAs are located inside or close to fragile sites and in minimal regions of LOH, MRAs and common breakpoints are associated with cancer<sup>8</sup>. For example, the cluster 17–92 is located at 13q31, a region commonly amplified in lymphomas<sup>13</sup>; *miR-143* and *miR-145* are located at 5q33, which is frequently deleted in myelodysplastic syndromes, and *miR-142* is located 50 nucleotides from the t(8;17) breakpoint region, which involves chromosome 17 and *MYC*. This translocation juxtaposes the *MYC* gene to the *miR-142* promoter, and overexpresses an abnormal *MYC* that is associated with lymphomas and pro-lymphocytic leukemia<sup>14</sup>. The above information suggests that integration of allelic imbalance study using single-nucleotide polymorphisms (SNP) array and microRNA profiling in the same samples facilitates to identify deregulated microRNA (oncogene or tumor suppressor genes). Nymark *et al.* used microRNA expression profiling and array comparative genomic hybridization (CGH) data for detecting lung cancer-related microRNA markers and their targets. For instance, the gain at 12p13.31 was correlated with deregulated microRNAs and inversely correlated with target genes expressions<sup>15</sup>. Lazar *et al.* also integrated copy number alteration, mRNA and microRNA expression to identify NSCLC-related genes. They found *miR-944* expression was significantly correlated with copy number gains<sup>16</sup>. However, these studies did not demonstrate any *in vitro* assay of confirming the microRNA functions on living cancer cells.

In this study, we aimed to determine whether there were distinct patterns of copy number variations that correspond to the microRNA expression in NSCLCs. The discovery of such patterns would be useful for the identification of NSCLC-associated microRNA, focal genomic deletion or amplification and carcinogenesis pathway. We further explored the clinical significance of a panel of microRNAs in a well characterized cohort of NSCLC, and determined the biological significance of miR-23b using different *in vitro* assays.

## Results

We have employed an integrated genomic approach to identify altered genomic regions and deregulated microRNAs in NSCLC. Genome-wide copy number analysis in a well characterized cohort of NSCLC samples (Supplementary Figure S1) allow us to identify *de novo* allelic imbalance area in tumors that may underlie NSCLC tumorigenesis. Moreover, integration of these allelic imbalance areas with microRNA expression array data enabled us to select microRNA of interest for further studies.

**Detection of allelic imbalance areas and associated genes from copy number analysis.** An overview of copy numbers of 66 adenocarcinoma (ADC) cases in whole genome were shown in Fig. 1A. Amplified regions were shown in red color and deleted regions were indicated by blue color. In each chromosome, copy number changes were analyzed more details and correlated with approximate locus (representative example for Chr.1 is shown Fig. 1B). As evident from Fig. 1B, amplified regions are more frequent in the chromosomal arm 1q (red color) in comparison with chromosomal arm 1p (blue color). An overview of whole genome copy number pattern was shown in Fig. 1C. In our analysis, copy numbers over 2.50 were defined as amplified (upper red bar) and those under 1.70 were considered as deleted (lower blue bar). Based on the later empirical cut-off of copy numbers, amplified regions and deleted regions were shown for ADC in Fig. 1B for chromosome 1. The same procedure was also performed for 53 squamous cell carcinoma (SCC) cases (data not shown). A summary of allelic imbalanced locus that showed over 15% amplification and/or deletion in ADC and SCC are available in Supplementary Table S2. Briefly, notable amplified area identified in our analysis were: 1q21.1-1q23.3 (ADC 22.7%, SCC 16.9%) that contained potential oncogene *S100A9*; 8q21.3-22.3 (ADC 30.3%, SCC 24.5%) that contained potential oncogene *MTDH*; and 8q24.11-8q24.21 (ADC 25.8%, SCC 20.7%) that contained well known oncogene *MYC*. Some examples of key deleted loci identified by our analysis were: 4q13.1 (ADC 15.2%, SCC 22.6%) that contained potential TSG *TSGEPA5*; 8p23.2 (ADC 18.2%, SCC 45.3%), that contained potential TSG *CSMD1* and 19q13.41-19q13.43 (ADC 18.2%, SCC 13.2%) that contained comparatively



**Figure 1.** Copy number analysis of 66 ADC cases: (A). Copy number signals in all the chromosomal arms are displayed. The headplot at the bottom depicted the log<sub>2</sub> ratio of copy number intensities; red color for amplifications and blue color for deletions. (B). More details of copy number alterations in Chromosomal arms 1p and 1q are displayed. The headplot at the bottom depicted the log<sub>2</sub> ratio of copy number intensities; red color for amplifications and blue color for deletions. (C). Genome-wide copy number frequencies are plotted according to their chromosomal locations. The light grey line within each chromosome denotes the centromere, separating both p and q arms. Red color denotes copy number amplification, blue color denotes copy number deletion and black color denotes copy neutral LOH regions (somatic uniparental disomy).

poorly characterized TSG *USP29*. In summary, in addition to confirm previously reported allelic imbalance loci such as LOH and copy number changes on chromosome 2q, 4q, 5q, 8p, 8q, 9p22 and 19q13, we identified novel LOH on 1q42 and copy number amplification on 1q21.1 and 1q32-q42 that contain some cancer associated genes that may have role in NSCLC tumorigenesis.

**Deregulated microRNAs determined by microRNA expression array.** Primary NSCLC and corresponding normal tissues were analyzed in 8 pairs of samples. Supplementary Table S3 summarized the clinicopathological information of all the 8 cases. The microRNA array we used here contained 688 mature microRNAs. Two types of analysis were carried out to identify probes which showed differential expression between tumor and adjacent normal tissues. Firstly, the expression ratios of individual microRNA between tumor and adjacent normal tissue were compared. Representative scatter plots are shown in Fig. 2A for 4 cases. Secondly, mean expression ratio profile across all patients was compared to the mean adjacent normal expression profile, and the differentially expressed microRNAs are shown in Fig. 2B a. In both analyses, differentially expressed microRNAs were expected to deviate from the bulk population. Four overexpressed (*miR-21*, *miR-193b*, *miR-205*, *miR-296*) and 4 under expressed microRNAs (*miR-126*, *miR-23b*, *miR-145* and *let-7b*) in tumor samples were identified by comparing mean expression values between tumors and normal (Fig. 2B a). Differential expression patterns were also observed in ADC or SCC specific manner (Fig. 2B b,c). Both ADCs and SCCs showed up-regulation of *miR-205* and down-regulation of *miR-126* and *miR-145* in tumors. Up-regulations of *miR-21*, *miR-150* and *miR-296* were only found in ADCs, whereas up-regulation of *miR-31* and down-regulation of *let-7b* were only seen in SCCs. In summary, considering mean expression values in all the 8 NSCLCs samples and group of ADC and SCC, 5 over-expressed (*miR-21*, *miR-31*, *miR-150*, *miR-205*, *miR-296*) and 3 under expressed microRNAs (*miR-126*, *miR-23b*, *miR-145*) in tumor samples were determined for technical validation.

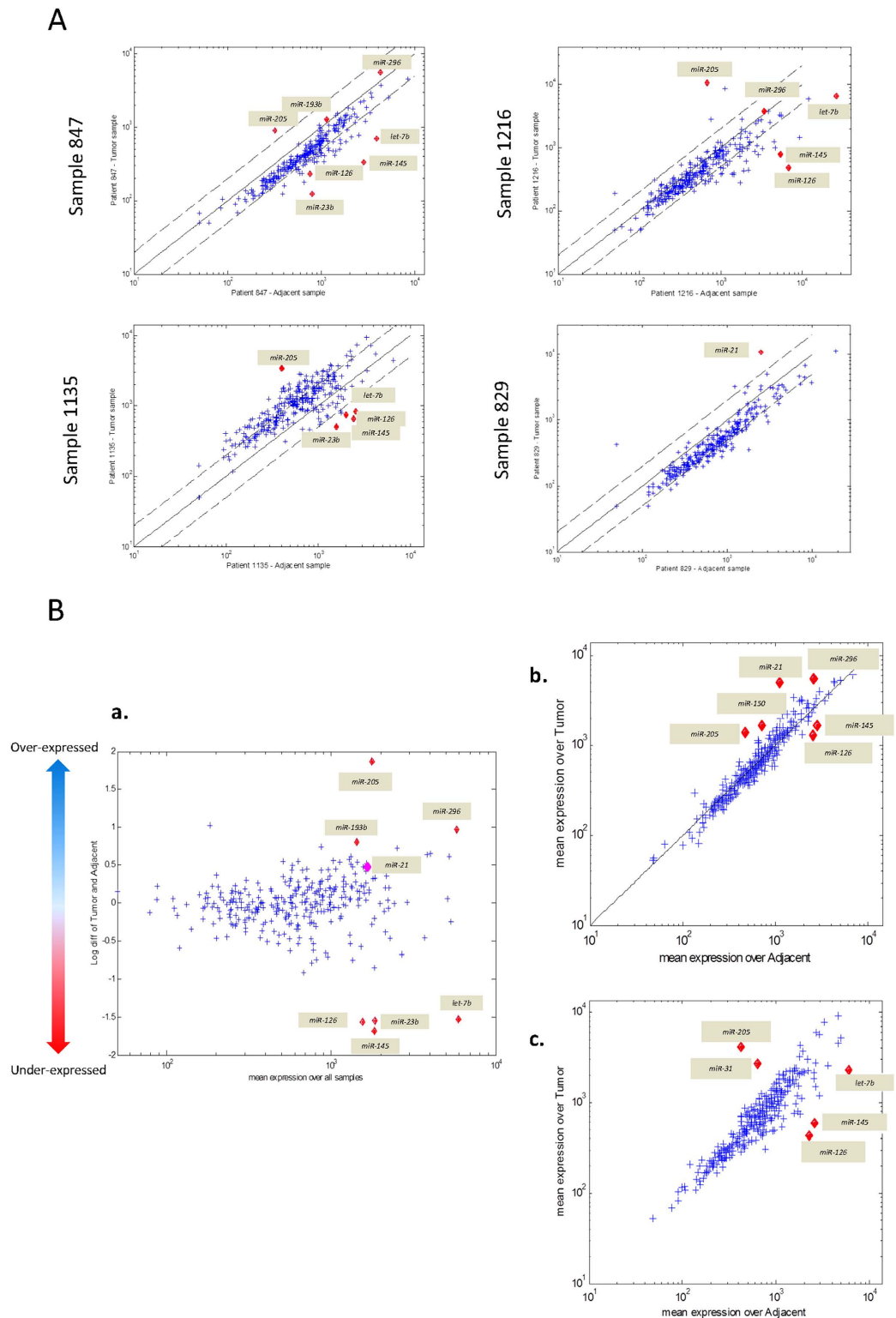
**Technical validation of microRNA expression data by quantitative reverse transcription PCR (Q-RT-PCR) and determination of cutoff values using a training set.** The comparative data between microRNA array results and Q-RT-PCR data are available in Supplementary Table S4 and S5. The individual consistency of the two techniques in representative 4 cases were high [*miR-205*:100% (3/3), *miR-296*:100% (2/2), *miR-21*:100% (1/1), *miR-23b*:50% (1/2), *miR-126*:100% (3/3), and *miR-145*:67% (2/3)] (Supplementary Table S4). Supplementary Table S5 summarizes the technical validation data of all the tested miRNAs in 8 tumor-normal paired samples.

Subsequently we tested additional 10 tumor and normal paired samples by Q-RT-PCR that results in matched 18 tumor-normal cohort as a training set ( $n = 18$ ). An overview of experimental design is shown in Fig. 3. Expression patterns of the training set (18 paired tumor-normal) are available in Supplementary Table S6, and scatter plots of tested miRNA by Q-RT-PCR are shown in Fig. 4. An optimal cut-off point with maximal sensitivity and specificity was determined by generating ROC curve for individual microRNA marker (solid line in Fig. 4). Using the cut-off value, the frequency of each deregulated microRNA in the tumor was as follows: 44% (8/18) overexpressed for *miR-205* ( $P = 0.018$ ), 56% (10/18) overexpressed for *miR-296* ( $P = 0.035$ ), 89% (16/18) overexpressed for *miR-21* ( $P < 0.001$ ), 22% (4/18) overexpressed for *miR-23b* ( $P = 0.104$ ), 89% (16/18) under expressed for *miR-126* ( $P < 0.001$ ), 78% (14/18) under expressed for *miR-145* ( $P < 0.001$ ), 33% (6/18) overexpressed for *miR-150* ( $P = 0.229$ ) and 83% (15/18) overexpressed for *miR-31* ( $P < 0.001$ ). A detailed summary of Q-RT-PCR analysis in training set is available in Table 1.

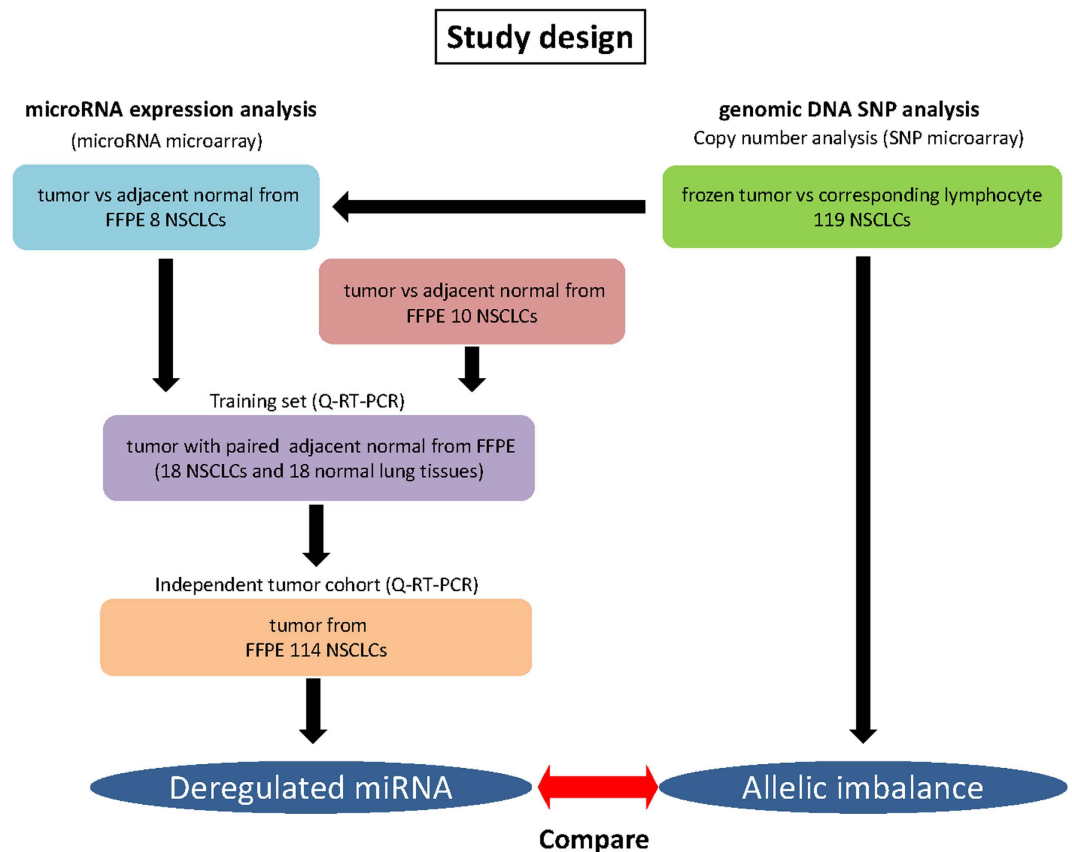
**Clinical validation of the microRNA panel in an independent cohort.** All the 8 microRNAs analyzed in training set, were also tested in an independent cohort of 114 formalin fixed paraffin embedded (FFPE) NSCLC samples. Detailed characteristics of all these 114 NSCLCs are available in Supplementary Table S7. Among all the clinicopathological factors, only TNM stage I cases was significantly dominant in this cohort (63/114, 55.3%,  $P = 0.011$ ) and the distribution of all other factors were almost similar. Delta Ct values (normalized by *miR-16*) of 114 NSCLCs were plotted in Fig. 4. Using the optimal cut off values determined in the training set, expression patterns of all the microRNAs were almost similar to those in tumors of the training set. Solid line shows empirical cut-off value of each microRNAs determined by ROC curve of training set ( $n = 18$ ), and this line divides the cohort into overexpressed group and under expressed group. The frequency of each deregulated microRNA in the tumor was as follows: 39% (44/114) overexpressed for *miR-205* ( $P < 0.001$ ), 95% (108/114) overexpressed for *miR-296* ( $P < 0.007$ ), 71% (81/114) overexpressed for *miR-21* ( $P < 0.001$ ), 23% (26/114) overexpressed for *miR-23b* ( $P < 0.001$ ), 64% (73/114) under expressed for *miR-126* ( $P < 0.001$ ), 78% (89/114) under expressed for *miR-145* ( $P < 0.001$ ), 67% (76/114) overexpressed for *miR-150* ( $P < 0.001$ ) and 67% (76/114) overexpressed for *miR-31* ( $P < 0.001$ ) (Table 1).

**Association of microRNA expression with clinicopathological factors.** We analyzed whether there is any association of each microRNA expression with all the available clinicopathological parameters (Supplementary Table S8). In the independent early stage dominant cohort ( $n = 114$ ), positive smoking history was significantly associated with low level *miR-126* ( $P = 0.030$ , student's t-test), N0 stage with high level *miR-23b* ( $P = 0.047$ ) and TNM stage I with high level *miR-205* ( $P = 0.034$ ).





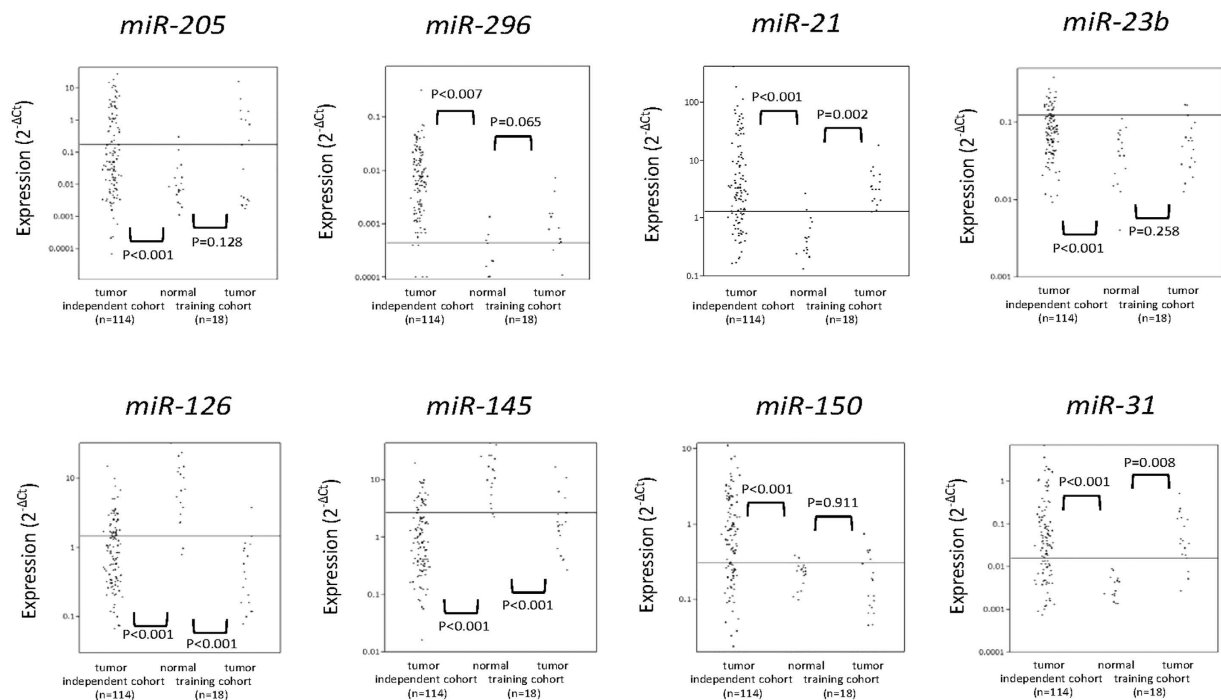
**Figure 2. MicroRNA expression array results:** (A). MicroRNA expression array results of representative 4 tumor-normal paired samples. Significantly differential expression between tumors and adjacent normal samples were found in some of microRNAs (red dots). (B). Scatter plots of promising microRNA that differentially expressed between tumors and normal by array analysis are shown: (a) Average microRNA expression ratios between tumor and adjacent normal tissue of all 8 cases (y axis) are plotted according to the mean expression (x axis). Differentially expressed microRNAs are expected to deviate from the bulk population (red dots). (b) Sub-group analyses scatter plot of mean tumor expression levels (y axis) and adjacent normal expression levels (x axis) of individual microRNAs for 3 Adenocarcinoma cases and (c) same for 5 squamous cell carcinoma cases. Probes with a large differential expression are identified as red dots.



**Figure 3. An overview of the study designs.** For microRNA expression analysis, we used 3 sample cohort, technical validation set ( $n = 8$ ), training set ( $n = 18$ ) and independent tumor set ( $n = 114$ ). For SNP analysis, we used 119 NSCLC cohorts that contain majority of samples we used for microRNA analysis. Integration of deregulated microRNAs and allelic imbalance results were performed in a subset of sample.

Significant expression difference between ADC and SCC was only seen for *miR-205* among 8 candidate microRNAs. SCC cell type was associated with *miR-205* overexpression ( $P < 0.001$ ) as previously reported<sup>17,18</sup>. Although microRNA profile has been reported to be different between African-American and Caucasian<sup>19</sup>, in our study no significant difference was found in any candidate microRNA expression among these two ethnic groups.

All the 114 patients of independent cohort were available for follow-up. The mean follow-up period was 46.3 months (range 1.0–204.0 months). Table 2 presents the results of the univariate and multivariate survival analyses considering all the available clinicopathological factors and the expression level of all the tested microRNAs. In the univariate analysis, positive alcohol history (HR = 1.77,  $P = 0.042$ , 95%CI:1.02–3.21), poor differentiation (HR = 2.71,  $P = 0.001$ , 95%CI:1.51–4.69), TNM stage II-IV (HR = 2.45,  $P < 0.001$ , 95%CI:1.44–4.21) and high *miR-23b* expression (HR = 1.87,  $P = 0.028$ , 95%CI:1.07–3.18) were significantly associated with poor recurrence free survival (RFS). The multivariate analysis confirmed that patients with SCC histology (HR = 1.92,  $P = 0.017$ , 95%CI:1.13–3.26), TNM stage II-IV (HR = 2.74,  $P < 0.001$ , 95% CI:1.60–4.75), low *miR-150* expression (HR = 1.92,  $P = 0.034$ , 95%CI:1.05–3.47) and patients with high *miR-23b* expression (HR = 2.40,  $P = 0.005$ , 95%CI:1.32–4.29) have independently high risk of recurrence. We then explored the association of clinicopathological factors with overall survival (OS) of patients. By univariate analysis, SCC histology (HR = 2.16,  $P = 0.006$ , 95%CI: 1.26–3.72), poor differentiation (HR = 2.42,  $P = 0.042$ , 95%CI: 1.34–4.21), TNM stage II-IV (HR = 2.11,  $P = 0.006$ , 95%CI: 1.24–3.63) and high *miR-23b* expression (HR = 1.93,  $P = 0.019$ , 95%CI: 1.12–3.30) were significantly associated with poor OS. In multivariable analysis, the significant factors were SCC (HR = 2.16,  $P = 0.006$ , 95%CI: 1.26–3.72), TNM stage II-IV (HR = 2.41,  $P = 0.002$ , 95% CI: 1.40–4.22), low *miR-150* expression (HR = 1.94,  $P = 0.038$ , 95%CI:1.04–3.56) and high *miR-23b* expression (HR = 2.35,  $P = 0.005$ , 95% CI: 1.30–4.19). We also examined these factors in TNM Stage I cases ( $n = 63$ ) (Table 3). Although many of factors turned out to be non-significant, histological SCC and high *miR-23b* expression were still significant prognostic factors in multivariable analysis of both recurrence free survival (RFS) and overall survival (OS). Especially, *miR-23b* was the solely significant miR factor of RFS (HR = 2.46,  $P = 0.041$ , 95% CI:1.04–5.62) and OS (HR = 2.64,  $P = 0.021$ , 95%CI:1.16–5.85) in this cohort of NSCLCs.



**Figure 4. Scatter plots of 8 microRNA expression by quantitative-RT-PCR: Individual microRNA expression of independent tumor cohort (n = 114), 18 tumor-adjacent normal paired samples (training set) are shown as delta Ct values (reference gene: *miR-16*). The solid line shows empiric cut-off value which is generated by maximizing sensitivity and specificity of each of microRNA on individual ROC curves. Each tumor cohort of training set (n = 18) and independent set (n = 114) was compared with normal sample set (n = 18) by student's t-test, two-tailed.  $P < 0.05$  were considered as significant.**

microRNA (Cut-off value)	Expression	A:Normal in training set (n = 18)	B:Cancer in training set (n = 18)	P value (A vs B)	C:Cancer in independent cohort (n = 114)	P value (A vs C)
<i>miR-205</i> (0.173)	Overexpressed Under expressed	6% (1/18) 94% (17/18)	44% (8/18) 56% (10/18)	<b>0.018</b>	39% (44/114) 61% (70/114)	<b>0.006</b>
<i>miR-296</i> (0.00044)	Overexpressed Under expressed	17% (3/18) 83% (15/18)	56% (10/18) 44% (8/18)	<b>0.035</b>	95% (108/114) 5% (6/114)	<b>&lt;0.001</b>
<i>miR-21</i> (1.297)	Overexpressed Under expressed	11% (2/18) 89% (16/18)	89% (16/18) 11% (2/18)	<b>&lt;0.001</b>	71% (81/114) 29% (33/114)	<b>&lt;0.001</b>
<i>miR-23b</i> (0.124)	Overexpressed Under expressed	0% (0/18) 100% (18/18)	22% (4/18) 78% (14/18)	0.104	23% (26/114) 77% (88/114)	<b>0.023</b>
<i>miR-126</i> (1.470)	Overexpressed Under expressed	89% (16/18) 11% (2/18)	11% (2/18) 89% (16/18)	<b>&lt;0.001</b>	36% (41/114) 64% (73/114)	<b>&lt;0.001</b>
<i>miR-145</i> (2.680)	Overexpressed Under expressed	89% (16/18) 11% (2/18)	22% (4/18) 78% (14/18)	<b>&lt;0.001</b>	22% (25/114) 78% (89/114)	<b>&lt;0.001</b>
<i>miR-150</i> (0.302)	Overexpressed Under expressed	11% (2/18) 89% (16/18)	33% (6/18) 67% (12/18)	0.229	67% (76/114) 33% (38/114)	<b>&lt;0.001</b>
<i>miR-31</i> (0.016)	Overexpressed Under expressed	0% (0/18) 100% (18/18)	83% (15/18) 17% (3/18)	<b>&lt;0.001</b>	67% (76/114) 33% (38/114)	<b>&lt;0.001</b>

**Table 1. Frequency of high or low expression of each microRNA in the training set (n = 18) and in independent set (n = 114). P value was calculated by Fisher's exact test, two-tailed. If significant displayed bold.**

As *miR-23b* overexpression was correlated significantly with poor RFS and OS, we further performed Kaplan-Meier curves analysis for *miR-23b* and associated clinical factors. Kaplan-Meier curves of RFS and OS in relation with *miR-23b* expression were shown in Fig. 5A. In log-rank test, overexpression of *miR-23b* correlated significantly with poor RFS ( $P = 0.020$ ) and poor OS ( $P = 0.013$ ) (Fig. 5A a). These findings remained statistically significant in subgroup analysis adjusting for T1 stage (RFS:  $P = 0.021$ ,

	RFS	Univariate	Multivariable	OS	Univariate	Multivariable
Age (>60 vs ≤ 60)	>60	HR = 1.02, P = 0.935 95%CI:0.59–1.83		>60	HR = 1.17, P = 0.590 95%CI:0.67–2.15	
Race (AA vs non-AA)	AA	HR = 1.06, P = 0.852 95%CI:0.54–1.94		AA	HR = 1.14, P = 0.692 95%CI:0.58–2.09	
Gender (Female vs Male)	Male	HR = 1.20, P = 0.498 95%CI:0.71–2.06		Male	HR = 1.21, P = 0.479 95%CI:0.71–2.09	
Smoking history (Yes vs No)	Yes	HR = 1.29, P = 0.344 95%CI:0.76–2.23		Yes	HR = 1.33, P = 0.291 95%CI:0.78–2.31	
Alcohol history (Yes vs No)	Yes	<b>HR = 1.77, P = 0.042</b> <b>95%CI:1.02–3.21</b>		Yes	HR = 1.50, P = 0.141 95%CI:0.88–2.65	
Histology (SCC vs ADC)	SCC	HR = 1.61, P = 0.075 95%CI:0.95–2.72	<b>HR = 1.92,</b> <b>P = 0.017</b> <b>95%CI:1.13–3.26</b>	SCC	<b>HR = 1.71, P = 0.048</b> <b>95%CI:1.01–2.90</b>	<b>HR = 2.16,</b> <b>P = 0.006</b> <b>95%CI:1.26–3.72</b>
Differentiation (Poor vs Well/Mod)	Poor	<b>HR = 2.71, P = 0.001</b> <b>95%CI:1.51–4.69</b>		Poor	<b>HR = 2.42, P = 0.042</b> <b>95%CI:1.34–4.21</b>	
TNM stage (II/III/IV vs I)	II–IV	<b>HR = 2.45, P &lt; 0.001</b> <b>95%CI:1.44–4.21</b>	<b>HR = 2.74,</b> <b>P &lt; 0.001</b> <b>95%CI:1.60–4.75</b>	II–IV	<b>HR = 2.11, P = 0.006</b> <b>95%CI:1.24–3.63</b>	<b>HR = 2.41,</b> <b>P = 0.002</b> <b>95%CI:1.40–4.22</b>
<i>miR-21</i> (High vs Low)	High	HR = 1.65, P = 0.093 95%CI:0.92–3.14		High	HR = 1.61, P = 0.110 95%CI:0.90–3.07	
<i>miR-23b</i> (High vs Low)	High	<b>HR = 1.87, P = 0.028</b> <b>95%CI:1.07–3.18</b>	<b>HR = 2.40,</b> <b>P = 0.005</b> <b>95%CI:1.32–4.29</b>	High	<b>HR = 1.93, P = 0.019</b> <b>95%CI:1.12–3.30</b>	<b>HR = 2.35,</b> <b>P = 0.005</b> <b>95%CI:1.30–4.19</b>
<i>miR-31</i> (High vs Low)	High	HR = 1.22, P = 0.485 95%CI:0.70–2.21		High	HR = 1.70, P = 0.081 95%CI:0.94–3.29	
<i>miR-126</i> (High vs Low)	Low	HR = 1.42, P = 0.203 95%CI:0.83–2.52		Low	HR = 1.31, P = 0.329 95%CI:0.77–2.31	
<i>miR-145</i> (High vs Low)	Low	HR = 1.07, P = 0.844 95%CI:0.58–2.11		Low	HR = 1.02, P = 0.961 95%CI:0.56–1.97	
<i>miR-150</i> (High vs Low)	Low	HR = 1.58, P = 0.104 95%CI:0.91–2.70	<b>HR = 1.92,</b> <b>P = 0.034</b> <b>95%CI:1.05–3.47</b>	Low	HR = 1.49, P = 0.173 95%CI:0.83–2.56	<b>HR = 1.94,</b> <b>P = 0.038</b> <b>95%CI:1.04–3.56</b>
<i>miR-205</i> (High vs Low)	High	HR = 1.22, P = 0.459 95%CI:0.72–2.06		High	HR = 1.39, P = 0.228 95%CI:0.81–2.35	
<i>miR-296</i> (High vs Low)	High	HR = 1.13, P = 0.833 95%CI:0.42–4.65		High	HR = 1.22, P = 0.725 95%CI:0.45–5.03	

**Table 2. Correlation of clinicopathological and miR factors with clinical outcome in validation cohort (n = 114).** AA: African-American, ADC: adenocarcinoma, SCC: squamous cell carcinoma, N.S.: not significant, Hazard ratios, 95% confidence intervals and P values were obtained using Cox proportional hazards models for RFS and OS. If significant displayed bold.

OS: P = 0.012) (Fig. 5A b), N0 stage (RFS: P = 0.032, OS: P = 0.023) (Fig. 5A c) and TNM stage I (OS: P = 0.038) (Fig. 5A d).

**MicroRNA expression and allelic imbalance.** We integrated allelic imbalance area from copy number analysis cohort (n = 119) (Fig. 3) with loci of all the 8 microRNAs we tested in this study. We considered allelic imbalance loci if a SNP locus is amplified or deleted at least in 15% of our tested samples. The integrated data of copy number variation (CNV) determined by SNP array analysis and deregulated candidate microRNAs were shown in Table 4. Four out of 8 deregulated microRNAs were located in gene locus with over 15% allelic imbalance (amplification or deletion). Among 119 NSCLCs, the locus of *miR-205* showed amplifications in 29/119 (24.4%) cases and deletions in 13/119 (10.9%) cases. The locus of *miR-126* showed amplifications in 11/119 (9.2%) cases and deletions in 22/119 (18.5%) cases. The locus of *miR-150* showed amplifications in 10/119 (8.4%) and deletions in 19/119 (16.0%) cases. The locus of *miR-31* showed amplifications in 6/119 (5.0%) and deletions in 20/119 (16.8%) cases. The locus of rest of the microRNAs (*miR-296*, *miR-21*, *miR-23b* and *miR-145*) had no or low level of allelic imbalance.

**Overexpression and inhibition of miR-23b in lung cancer cell lines.** To determine the biological consequences of *miR-23b* deregulation, we first analyzed the expression level of *miR-23b* in a panel of NSCLC cell lines and one immortalized normal broncho-epithelial cell line (Fig. 5B a). We decided to perform functional analysis using cell lines that were derived from ADCs because of the wide variability of expression levels. We selected H1838 cell line for *miR-23b* overexpression (mimic), and H1437 & H1944 cell lines for inhibition of *miR-23b*. We performed [3-(4,5-dimethylthiazol-2-yl)-2,5 diphenyl tetrazolium bromide] (MTT) assay using these three cells either overexpressing or inhibiting *miR-23b*. In



	RFS	Univariate	Multivariable	OS	Univariate	Multivariable
Age (>60 vs ≤60)	≤60	HR = 1.01, P = 0.973 95%CI:0.41–2.28		>60	HR = 1.19, P = 0.690 95%CI:0.52–3.06	
Race (AA vs non-AA)	Non-AA	HR = 1.20, P = 0.706 95%CI:0.49–3.63		Non-AA	HR = 1.22, P = 0.682 95%CI:0.50–3.67	
Gender (Female vs Male)	Male	HR = 1.09, P = 0.831 95%CI:0.49–2.46		Male	HR = 1.08, P = 0.850 95%CI:0.50–2.38	
Smoking history (Yes vs No)	Yes	HR = 1.59, P = 0.250 95%CI:0.72–3.67		Yes	HR = 1.46, P = 0.336 95%CI:0.68–3.27	
Alcohol history (Yes vs No)	Yes	HR = 1.37, P = 0.436 95%CI:0.62–3.16		Yes	HR = 1.17, P = 0.698 95%CI:0.54–2.58	
Histology (SCC vs ADC)	SCC	HR = 2.13, P = 0.062 95%CI:0.96–4.93	<b>HR = 2.54, P = 0.026 95%CI:1.12–6.05</b>	SCC	<b>HR = 2.28, P = 0.039 95%CI:1.04–5.23</b>	<b>HR = 2.66, P = 0.016 95%CI:1.20–6.21</b>
Differentiation (Poor vs Well/Mod)	Poor	HR = 2.50, P = 0.141 95%CI:0.71–7.00		Poor	<b>HR = 3.90, P = 0.023 95%CI:1.23–10.54</b>	
<i>miR-21</i> (High vs Low)	High	HR = 2.06, P = 0.104 95%CI:0.87–5.68		High	HR = 2.19, P = 0.075 95%CI:0.93–5.99	
<i>miR-23b</i> (High vs Low)	High	HR = 2.00, P = 0.101 95%CI:0.87–4.43	<b>HR = 2.46, P = 0.041 95%CI:1.04–5.62</b>	High	<b>HR = 2.23, P = 0.051 95%CI:0.99–4.85</b>	<b>HR = 2.64, P = 0.021 95%CI:1.16–5.85</b>
<i>miR-31</i> (High vs Low)	High	HR = 1.19, P = 0.680 95%CI:0.53–2.81		High	HR = 1.58, P = 0.269 95%CI:0.71–3.87	
<i>miR-126</i> (High vs Low)	Low	HR = 1.50, P = 0.326 95%CI:0.67–3.54		Low	HR = 1.36, P = 0.439 95%CI:0.63–3.12	
<i>miR-145</i> (High vs Low)	Low	HR = 1.12, P = 0.805 95%CI:0.47–3.08		High	HR = 1.08, P = 0.864 95%CI:0.42–2.46	
<i>miR-150</i> (High vs Low)	Low	HR = 1.24, P = 0.626 95%CI:0.48–2.86		Low	HR = 1.16, P = 0.740 95%CI:0.45–2.64	
<i>miR-205</i> (High vs Low)	High	HR = 1.75, P = 0.167 95%CI:0.79–3.95		High	<b>HR = 2.20, P = 0.048 95%CI:1.01–5.03</b>	
<i>miR-296</i> (High vs Low)	High	HR = 1.17, P = 0.877 95%CI:0.25–20.85		High	HR = 1.35, P = 0.757 95%CI:0.29–24.16	

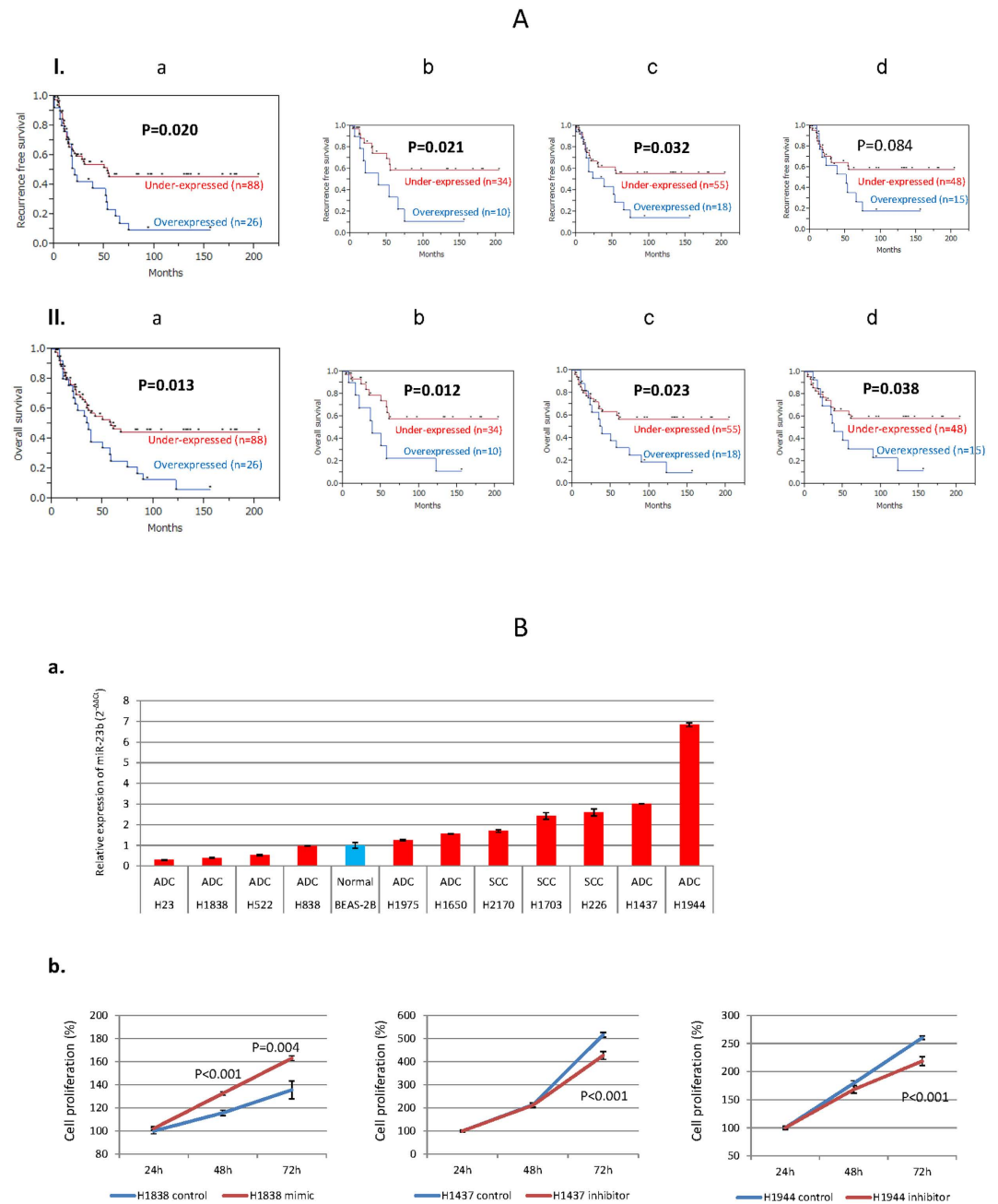
**Table 3. Correlation of clinicopathological and miR factors with clinical outcome in TNM Stage I case (n = 63).** AA: African-American, ADC: adenocarcinoma, SCC: squamous cell carcinoma, N.S.: not significant, Hazard ratios, 95% confidence intervals and P values were obtained using Cox proportional hazards models for RFS and OS. If significant displayed bold.

the mimic study, cell viability was significantly increased in *miR-23b* transfected H1838 cells than control at 72 hours after transfection ( $p < 0.001$ , student t-test). While in the inhibition study, *miR-23b* siRNA transfected H1437 and H1944 cells showed significant decrease in cells viability compared to the control at 72 hours after transfection ( $p < 0.001$ , student t-test) (Fig. 5B b). These phenotypic characteristics due to forced alterations of *miR-23b* were consistent with primary NSCLC data that *miR-23b* is a potential oncogene.

## Discussion

Generally, recurrent losses and gains of broad chromosomal regions in cancer suggest that multiple genes located in the same chromosomal regions may be concurrently function as a tumor suppressor gene (TSG) and oncogene. In addition to the identification of novel allelic imbalance area, our data provided confirmation of previously reported allelic imbalance area that may contain novel NSCLC related cancer genes. Furthermore, the integrated approach we employed in this study support the previous assumption that almost half of microRNA are located in the allelic imbalance areas.

NSCLC is an extensively heterogeneous disease at the molecular level which in part may be related to easy exposure of lung to different kinds of environmental stimuli. Various environmental stimuli and



**Figure 5. Association of clinical outcome with miR-23b expression in NSCLC and its *in vitro* oncogenic potential: (A). I:** Correlation of *miR-23b* expression with recurrence free survival (RFS): Total 114 NSCLCs are divided into *miR-23b* overexpressed ( $n = 26$ ) and under expressed group ( $n = 88$ ): (a) The association of *miR-23b* expression with RFS of NSCLC patients was analyzed by Kaplan-Meier curve and log-rank test. Subgroup analyses were also performed in (b) T1 stage cases ( $n = 44$ ), (c) N0 stage cases ( $n = 73$ ) and (d) TNM stage I cases ( $n = 63$ ). **II:** Overall survival (OS) analysis based on *miR-23b* expression by Kaplan-Meier curve and log-rank test: (a) The association of *miR-23b* expression with OS of NSCLC patients was examined. Subgroup analyses were also performed in (b) T1 stage cases ( $n = 44$ ); (c) N0 stage cases ( $n = 73$ ) and (d) TNM stage I cases ( $n = 63$ ).  $P < 0.05$  were considered as significant. (B). Modulation of *miR-23b* in lung cancer cell lines to determine its *in vitro* cell growth potential: (a) Relative expression (mean  $\pm$  standard error) of *miR-23b* in 11 lung cancer cell lines (red bars). BEAS-2B, a normal lung epithelial cell line (blue bar) was used as a reference. ADC: Adenocarcinoma, SCC: Squamous cell carcinoma. Three adenocarcinoma cell lines were chosen to modulate *miR-23b* to understand the biological effect of this microRNA. (b) *MiR-23b* mimic was transfected to H1838 which showed low level *miR-23b* expression among the cell lines, while *mir-23b* inhibitor was transfected to H1437 and H1944 cell lines that showed high level *miR-23b* expression among the cell lines. Average cell proliferation ratio  $\pm$  standard error was shown at each time point. P values were shown if there was significance between transfected cells and controls (student t-test, two-tailed).

microRNA	Locus	ADC (n = 66)		SCC (n = 53)		Total (n = 119)	
		Amplification	Deletion	Amplification	Deletion	Amplification	Deletion
<i>miR-205</i>	Chr1 q32.2	23/66 (34.8%)	4/66 (6.1%)	6/53 (11.3%)	9/53 (16.7%)	29/119 (24.4%)	13/119 (10.9%)
<i>miR-296</i>	Chr20 q13.32	No allelic imbalance				No allelic imbalance	
<i>miR-21</i>	Chr17 q23.2	No allelic imbalance				No allelic imbalance	
<i>miR-23b</i>	Chr9 q22.32	No allelic imbalance				No allelic imbalance	
<i>miR-126</i>	Chr9 q34	4/66 (6.1%)	10/66 (15.2%)	7/53 (13.2%)	12/53 (22.6%)	11/119 (9.2%)	22/119 (18.5%)
<i>miR-145</i>	Chr5 q32	No allelic imbalance				No allelic imbalance	
<i>miR-150</i>	Chr19 q13.33	8/66 (12.1%)	12/66 (18.2%)	2/53 (3.7%)	7/53 (13.2%)	10/119 (8.4%)	19/119 (16.0%)
<i>miR-31</i>	Chr9 p21.3	4/66 (6.1%)	8/66 (12.1%)	2/53 (3.7%)	12/53 (22.6%)	6/119 (5.0%)	20/119 (16.8%)

**Table 4. Association of microRNA expression and allelic imbalance.**

individual response to a particular stimuli leads to differential molecular alterations that include mutations, translocations, copy number alterations etc. All these alterations may have an impact on NSCLC initiation and/or progression and may be targeted for therapy and the development of preventive strategies. Comprehensive investigation of focal copy number alterations in NSCLC has led to the identification of multiple cancer-driving genes with potential therapeutic implications<sup>20–22</sup>. However, all these approaches for identification of genetic alterations have limitations and therefore development of novel analytical approaches may generate new, previously not discovered targets for further studies. By our analysis, we have identified novel regions of allelic imbalance that contain several cancer related genes such as *CHD1L* and *S100A9* at locus 1q21.1; and *CENPF* and *ESRRG* at locus 1q32–q42. However, since the main focus of this study was to identify NSCLC related microRNAs, we yet not further validated and functionally characterized any of these genes for potential clinical implications.

We have identified several differentially expressed microRNAs in cancer and matching normal samples of NSCLC. Some of our findings are consistent with previously reported data while others are inconsistent. As for example, *miR-31* was reported to be overexpressed in colorectal cancer<sup>23</sup> which is consistent with our finding; however, it was also reported to be under expressed due to promoter hypermethylation in prostate cancer<sup>24</sup>. These contradictions could be due to tumor context, technology used for analysis and different endogenous and exogenous stimuli for the genesis of a given tumor in a given population. Other microRNAs that were overexpressed in our study such as *miR-21*, *miR-205* and *miR-296* were also reported to be deregulated in other cancer types. As for example, *miR-205* was reported to be under expressed in non-muscle invasive urothelial cancer, however it was found to be overexpressed in muscle-invasive urothelial cancer<sup>25</sup>. Similarly in consistent with our results, *miR-126* and *miR-145* were also reported to be under expressed in NSCLC<sup>26,27</sup>, and functionally *miR-145* was revealed to inhibit NSCLC cell proliferation by targeting a well-known oncogene, *c-Myc*<sup>28</sup>. However, contradictory findings were also reported for *miR-145*<sup>29</sup>. In the later study, although *miR-145* was under expressed in lung cancer, overexpression of *miR-145* was related to poor survival. Thus, detail biological studies are necessary for understanding the exact role of *miR-145* in the genesis of NSCLC.

Our array analysis data indicate that *miR-23b* is under expressed in NSCLC (Fig. 2B a). However, in our independent cohort study by Q-RT-PCR (Fig. 4), we found that *miR-23b* was overexpressed in NSCLC. These differences may derive from the fact that the array result only showed the total average of 8 samples. Total average is not necessarily indicates whether the overexpressed or under expressed cases are dominant. Actually, each of subgroup analysis (3ADCs and 5 SCCs) of *miR-23b* did not show either overexpression or under expression (Supplementary Table S5) which is also supported by Q-RT-PCR results of training set (n = 18) (Supplementary Table S6). Another explanation is the existence of dual role of this microRNA as described above for *miR-205*. If the function of a given microRNA differs in each tumor stage, the expression pattern will depend on the characteristics of the cohort. Actually, *miR-23b* expression was significantly different between pathological N0 stage and N1/N2 stages (P = 0.047, student t-test, Supplementary Table S8). Both overexpression and under expression of *miR-23b* were reported previously in various solid tumors<sup>30</sup>. Our independent tumor cohort (n = 114) mainly consisted of TNM stage I tumors (63/114, 55.2%), and it is slightly different from array cohort (3/8, 37.5%) (Supplementary Table S7).

The integration of allelic imbalance determined by SNP array and our limited microRNA expression data support that half of the microRNAs are positioned in allelic imbalance areas. Although loci imbalance status of *miR-126* and *miR-150* (Table 4) showed good correlation with microRNA expression by Q-RT-PCR (Supplementary Table S5, S6), the expression pattern of *miR-31* and *miR-205* didn't match their loci imbalance status. Possible explanations underlying these inconsistencies are existence of another deregulating mechanism, inadequate threshold setting or array noises. Meanwhile, in order to find low level allelic imbalance, we manually analyzed SNP data in the locus of *miR-21* and *miR-23b* using the same threshold (amplification >2.5, deletion <1.7). Expressions of both of these microRNAs were overexpressed in tumors by Q-RT-PCR analysis (Fig. 4). We then extended our analysis in all the

SNP data set ( $n = 119$ ) and as expected observed one of previously well characterized microRNA, *miR-21* locus showed significant amplification in tumor (Supplementary Table S9), and our focus *miR-23b* locus was also showed significant amplification in tumor (Supplementary Table S10). Overexpression of *miR-21* and *miR-23b* in NSCLCs seemed to be partially due to allelic amplification.

While it is evident that *miR-23b* is a potential oncogene for the genesis and progression of NSCLC by our initial functional analysis data and testing of NSCLC samples, *miR-23b* has been found to have dual role in carcinogenesis, both as a tumor suppressor and as an oncomir. A detailed review of functional consequences by *miR-23b* has been published recently<sup>30,31</sup>. As a tumor suppressor, Majid *et al.*<sup>31</sup> reported that *miR-23b* was frequently silenced in prostate cancer by methylation, and it had anti-proliferative and anti-invasive properties through repressing Src kinase/Akt pathway. Promoter hypermethylation of *miR-23b* was also found in gliomas<sup>32</sup>. Another recent report suggested that radio-resistant pancreatic cancer cell lines showed the reduced levels of *miR-23b*, and overexpression of *miR-23b* sensitized the cells to radiation by targeting ATG12, a known autophagy-related protein<sup>33</sup>. On the contrary, Jin *et al.* indicated the oncogenic character of *miR-23b* in breast cancer, showing that ERBB2, EGF and TNF- $\alpha$  promote its expression through AKT/NF- $\kappa$ B pathway<sup>34</sup>. Chen *et al.* revealed that down-regulation of *miR-23b* was followed by the inhibition of  $\beta$ -catenin/Tcf-4 and HIF-1 $\alpha$ /VEGF signaling pathways in glioma<sup>35</sup>. Recently, Zaman *et al.* found that PTEN was overexpressed after *miR-23b-3p* knock-down in renal cancer cells and also found inverse correlation of *miR-23b* with PTEN expression in human samples<sup>36</sup>. PTEN inactivation was reported as a poor prognostic factor<sup>37,38</sup> or a chemotherapy resistant factor<sup>39,40</sup> for NSCLC. As a future study, we have to elucidate the gene targets of *miR-23b* in NSCLCs including PTEN<sup>36</sup>, and find mechanism of involvement of *miR-23b* in lung carcinogenesis.

There are several reports about the prognostic significance of miRNAs in early stage NSCLCs<sup>41–45</sup>. Yanaihara *et al* tested a panel of miRNAs in 104 NSCLCs including 65 stage I tumors (62.5%)<sup>46</sup>. They found *miR-205*, *miR-21* and *miR-150* were overexpressed, while *miR-126* and *miR-145* were under expressed in NSCLCs. Other groups reported overexpression of *miR-31* in NSCLCs<sup>47,48</sup>. Patnaik *et al* examined miRNA expression profiles that might predict recurrence of localized stage I NSCLC after surgery<sup>49</sup>. They compared 37 recurrent and 40 non-recurrent cases. Although they did not focus on *miR-23b* in their study, it was listed as one of the significantly overexpressed microRNAs in recurrent cases ( $P = 0.003$ , Fold change: 2.51). All these reports may partially support our results. In this study, we evaluated prognostic significance of miRNAs expression in early stage NSCLCs using formalin fixed paraffin embedded (FFPE) samples. If reliable prognostic markers would be available in early stage NSCLC, patients could be handled in a more appropriate way to increase survival time. In addition, miRNA marker also has potential for monitoring of disease as miRNAs are more stable than messenger RNAs due to its small size<sup>50</sup>, and can be tested in bodily fluids such as in sputum<sup>51</sup> and serum<sup>52</sup>. Several studies have already been published using miRNA microarray data from FFPE samples<sup>53–55</sup>. As they are often the only available tissue source with comprehensive clinical data and long-term follow up, it is meaningful to prove their qualities. Hall *et al* showed that microRNAs were not subjected to the same deterioration seen in other RNA types<sup>53</sup>. Based on these reports, this study also showed that micro array results from FFPE samples were almost consistent with Q-RT-PCR results.

In conclusion, we have identified novel allelic imbalance regions that could harbor potential NSCLC related genes. Our integrated analysis revealed that a substantial numbers of microRNA were located in allelic imbalance area. More interestingly, from the clinical context, increased *miR-23b* expression in the tumor is a novel candidate biomarker of significant for poor survival of NSCLC patients. However, further validation of *miR-23b* in a multi-centered prospective study is needed before any potential clinical implementation; and detail *miR-23b* deregulation mechanisms including downstream pathway should be evaluated not only for the potential as a prognostic marker but also for suitability of early detection marker and targeted therapy of NSCLC.

## Methods

**Clinical samples.** Different set of NSCLC clinical samples were tested in this study. An overview of study design was shown in Fig. 3. A total 119 patients undergoing surgical resection of a primary NSCLC at The Johns Hopkins Hospital (Baltimore, MD, USA), the Johns Hopkins Bayview Medical Center (Baltimore, MD, USA) or the Medical College of Wisconsin, Froedtert Memorial Hospital (Milwaukee, WI, USA) were included for SNP array analysis. Among these specimens, 66 were adenocarcinoma (ADC) (including those with bronchoalveolar components), and 53 were squamous cell carcinoma (SCC). Details criteria of these samples are available in Supplementary Table S1. A subset (8 pairs) of the cohort (Supplementary Table S3) was analyzed for microRNA expression array analysis, and 8 candidate microRNAs were selected for further validation based of microRNA array data bioinformatics analysis. Ten additional sample pairs were added to these 8 samples, and tested by Q-RT-PCR as a training set ( $n = 18$ ) (Fig. 3). As an independent set, 114 NSCLC tumors were used for testing 8 candidate microRNAs. The clinicopathological characteristics of this independent cohort of 114 NSCLC samples were summarized in Supplementary Table S7. All samples were obtained as anonymized materials in accordance with the guidelines which were approved by the Johns Hopkins University Institutional Review Board. Thus, this study was qualified for exemption under the U.S. Department of Health and Human Services policy for protection of human subjects [45 CFR 46.101(b)].

**DNA extraction.** Hematoxylin-Eosin stained sections were histologically examined at every 20 sections for the presence or absence of tumor cells, as well as for tumor density. Only sections that showed more than 70% of tumor cells were used for DNA extraction. Microdissected tissues and lymphocytes were digested with 1% SDS and 50  $\mu\text{g/ml}$  proteinase K (Boehringer, Mannheim, Germany) at 48°C overnight followed by phenol/chloroform extraction and ethanol precipitation of DNA as previously described<sup>56</sup>.

**SNP microarray analysis.** Genomic DNA from microdissected frozen tumor tissues and corresponding lymphocyte were analyzed in parallel. Briefly 250 ng DNA was digested with *XbaI* (New England Biolabs Inc., Ipswich, MA, USA), ligated to the adaptor, and amplified by polymerase chain reaction (PCR) using a single primer. After purification of PCR products with the MinElute 96 UF PCR purification kit (Qiagen, Valencia, CA, USA), amplicons were quantified, fragmented, labeled and subsequently hybridized on Affymetrix GeneChip1 Mapping 10K 2.0 SNP microarrays following the manufacturer's instructions (Affymetrix Inc., Santa Clara, CA, USA). After washing and staining, the arrays were scanned for data analysis.

**RNA extraction for microRNA expression analysis.** Total RNA was extracted from the 10- $\mu\text{m}$ -thick FFPE tissue sections as previously described<sup>57</sup> using the Ambion RecoverAll Total Nucleic Acid Isolation Kit for FFPE tissues (Applied Biosystems/Ambion, Austin, TX, USA) according to the manufacturer's instructions. Only sections that showed more than 70% of tumor cells were used for RNA extraction. The quantity and quality of the total RNA was verified with the NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Twelve pairs of tumor-adjacent samples were hybridized. Of these, 8 pairs were considered highest quality and were used for the following experiments.

**miRdicator™ array platform.** Custom microarrays were produced by printing DNA oligonucleotide probes representing 688 microRNAs (Sanger database, version 9 and additional Rosetta validated and predicted microRNAs). Each probe, printed in triplicate, was carried up to 22-nt linker at the 3' end of the microRNA's complement sequence in addition to an amine group used to couple the probes to coated glass slides. 10/20  $\mu\text{M}$  of each probe were dissolved in 2x SSC + 0.0035% SDS and spotted in triplicate on Schott Nexterion® Slide E (Applied Microarrays Inc., Tempe, AZ, USA) coated microarray slides using a Genomic Solutions® BioRobotics MicroGrid II (Genomic solutions, Beverly, MA, USA) according to the manufacturer's directions. Sixty four negative control probes were designed using the sense sequences of different microRNAs. Two groups of positive control probes were designed to hybridize for miRdicator™ array (Rosetta Genomics Inc., Philadelphia, PA, USA) synthetic spikes. Small RNA was added to the RNA before labeling to verify the labeling efficiency. Probes for abundant small RNA [e.g. small nuclear RNAs (U43, U49, U24, Z30, U6, U48, U44), 5.8 s and 5 s ribosomal RNA] were spotted on the array to verify RNA quality. The slides were blocked in a solution containing 50 mM ethanolamine, 1 M Tris (pH9.0) and 0.1% SDS for 20 min at 50°C, then thoroughly rinsed with water and spun dry.

**Cy-dye labeling of microRNA for miRdicator™ array.** Total RNA (3–5  $\mu\text{g}$ ) was labeled by ligation of a RNA-linker, p-rCrU-Cy/dye (Dharmacon, Lafayette, CO, USA)<sup>58</sup> to the 3'-end with Cy3 or Cy5. The labeling reaction contained total RNA, spikes (20–0.1 fmoles), 300 ng RNA-linker-dye (Dharmacon), 15% DMSO, 1x ligase buffer and 20 units of T4 RNA ligase (New England Biolabs Inc.) and proceeded at 40°C for 1 hr followed by 1 hr at 37°C. The labeled RNA was mixed with 3x hybridization buffer (Ambion, Austin, TX, USA), heated to 95°C for 3 min and then added on top of the miRdicator™ array. Slides were hybridize 12–16 hr in 42°C, followed by two washes in room temperature with 1xSSC and 0.2% SDS and a final wash with 0.1xSSC. The array was scanned using an Agilent Microarray Scanner Bundle G2565BA resolution of 10  $\mu\text{m}$  at 100% power) (Agilent Technologies, Santa Clara, CA, USA). The data was analyzed using SpotReader software (Niles Scientific, Seattle, WA). Standard bioinformatics and statistical analysis were performed.

**Real Time reverse transcriptase (RT) polymerase chain reaction (Q-RT-PCR) for Quantification of microRNAs.** Briefly, a total of 10 ng RNA isolated from primary tissues was reverse transcribed using TaqMan reverse transcription kit (Applied Biosystems, Foster City, CA, USA) and microRNA-specific primers provided with TaqMan microRNA assays (Applied Biosystems) in 15  $\mu\text{L}$  reaction volume that contains 3  $\mu\text{L}$  of RT Primer Mix, 0.15  $\mu\text{L}$  of 100 mM dNTPs, 1  $\mu\text{L}$  of Reverse Transcriptase enzyme 50 U/ $\mu\text{L}$ , 0.19  $\mu\text{L}$  of RNase inhibitor 20 U/ $\mu\text{L}$ , 4.16  $\mu\text{L}$  of Nuclease Free water and 5  $\mu\text{L}$  of RNA (10 ng). RT reaction was carried out with annealing at 16°C for 30 min followed by extension at 42°C for 30 min. 1.3  $\mu\text{L}$  of the RT reaction was then used with 1  $\mu\text{L}$  of specific primers for each microRNAs (Applied Biosystems) in triplicate wells for 45-cycles PCR on a 7900HT thermocycler (Applied Biosystems). The thermal cycling parameters were as follows: 50°C for 2 min, 95°C for 10 min, followed by a third step for denaturation at 95°C for 15 s and annealing/extension at 60°C for 1 min repeated for 40 cycles. SDS v2.4 software (Applied Biosystems) was used to determine cycle threshold (Ct) values of the fluorescence measured during PCR. All experiments were done in triplicate. Two normalization steps were considered: loading the same quantity of template RNA in each well and normalizing the data against endogenous genes (*hsa-miR-16*, *RNU6*). As expression of *hsa-miR-16* was evenly distributed across the samples, we decided



to use *hsa-miR-16* for normalization in this study (data not shown). The ABI TaqMan SDS v 2.4 software was utilized to obtain raw Ct values. Relative quantification of microRNA expression was calculated with the  $2^{-(\Delta\Delta Ct)}$  method (Applied Biosystems User Bulletin N 2) (P/N 10303859).

**Lung cancer cell lines.** Lung cancer cell lines H23, H226, H522, H838, H1437, H1650, H1703, H1838, H1944, H1975, H2170 and SV40-Immortalized normal human bronchial epithelium cell line BEAS-2B were obtained from and propagated according to the recommendations of American Type Culture Collection (ATCC). Mediums and antibiotics were purchased from Mediatech (Manassas, VA, USA) and supplemented with fetal bovine serum (10%) (Hyclone, Logan, UT, USA), 100 µg/ml streptomycin and 100 I.U/ml penicillin (both from Life technologies). Cells were grown at 37 °C in a humidified atmosphere composed of 95% air and 5% CO<sub>2</sub> in a monolayer culture. All cancer cell lines were maintained in RPMI 1640, and BEAS-2B was grown in BEGM (Lonza, Walkersville, MD, USA) medium.

**Transfection of miR-23b.** We first determined the expression level of *miR-23b* in different lung cancer cell line (H23, H226, H522, H838, H1437, H1650, H1703, H1838, H1944, H1975, H2170) and one SV40-immortalized normal alveolar cell line (BEAS-2B). Based on expression patterns, we selected 3 cell lines (one expressing low level and two expressing high level of *miR-23b*) to determine the biological effect on cell lines due to *miR-23b* modulation. H1838, H1437 and H1944 cells ( $5 \times 10^3$ – $10 \times 10^3$  cells/well) were plated on 96-well plates with 100 µl of growth medium without antibiotics. The cells were transfected with RNAiMAX (Life technologies, Carlsbad, CA, USA) according to the manufacturer's protocol. Both mirVana® miRNA mimic Negative Control #1 and mirVana® miRNA inhibitor Negative Control #1 (Life technologies) were used as controls. Transfection efficiency was maximized at 50 nM for mimic (H1838: 2,328, 672%) and 3.3 nM for inhibition (H1437: 0.007%, H1944: 0.010%).

**Cell Proliferation assay (MTT assay).** Transfected lung cancer cell lines were plated on 96-well plates at a density of  $5 \times 10^3$  to  $1 \times 10^4$  per well. Cellular viability was measured by the MTT proliferation assay kit (ATCC, Manassas, VA, USA) according to the manufacturer's instructions as described in the previous paper from our group<sup>59</sup>. Each assay was performed in triplicate, and each experiment was repeated at least three times. The extent of cellular survival was represented as a percentage of the first measurement day.

**Data Analysis and Statistical Consideration.** Allelic calls for tumor DNA and corresponding normal genomic DNA were obtained from normalized SNP array data using the GDAS genotyping software supplied by the array manufacturer (Affymetrix). A Hidden Markov Model was applied to infer the probability of allelic imbalance for each SNP in tumor DNA compared to corresponding normal DNA using the dChip software (Cheng Li Lab). Further details of identified allelic imbalance area are described in our previous study<sup>60</sup>.

We determined an optimal cut off value for each tested microRNA using a training set that consist of 18 tumors with paired normal. ROC curve was generated for each microRNA and the empiric cut off value was selected by maximizing sensitivity and specificity. Based on this cut off value, we divided another set of 114 tumor samples into overexpressed and under expressed groups and the differences of clinical outcomes were compared between overexpressed and under expressed groups. For clinical outcomes, RFS was defined as the time from surgery to the time of first documentation of any disease recurrence. OS was defined as the time from surgery to the time of death of disease. Those who remained alive were censored at the last date the subject was known to be alive. Associations of microRNA expressions with RFS and OS were evaluated using Cox proportional hazards model with hazard ratios and 95% confidence intervals estimated for multivariable analysis. For other statistics, continuous variables were analyzed by Student's t-test and categorical variables were analyzed by Fisher's exact test. All statistical analyses were performed using JMP 9 software (SAS institute, Cary, NC, USA). The level of statistical significance was set at  $P < 0.05$  in two-tailed.

## References

1. Ettinger, D. S. & Kris, M. G. Panel NN-SCLCPG. NCCN: Non-small cell lung cancer. *Cancer control : journal of the Moffitt Cancer Center* **8**, 22–31 (2001).
2. Harless, W. & Qiu, Y. Cancer: A medical emergency. *Medical hypotheses* **67**, 1054–1059 (2006).
3. Harpole, D. H., Jr., Herndon, J. E., 2nd, Young, W. G., Jr., Wolfe, W. G. & Sabiston, D. C., Jr. Stage I nonsmall cell lung cancer. A multivariate analysis of treatment methods and patterns of recurrence. *Cancer* **76**, 787–796 (1995).
4. Pallis, A. G. *et al.* Management of elderly patients with NSCLC; updated expert's opinion paper: EORTC Elderly Task Force, Lung Cancer Group and International Society for Geriatric Oncology. *Annals of oncology : official journal of the European Society for Medical Oncology/ESMO* **25**, 1270–1283 (2014).
5. Cardenal, F., Nadal, E., Jove, M. & Faivre-Finn, C. Concurrent systemic therapy with radiotherapy for the treatment of poor-risk patients with unresectable stage III non-small cell lung cancer - a review of the literature. *Annals of oncology : official journal of the European Society for Medical Oncology/ESMO*. (2014).
6. Brock, M. V. *et al.* DNA methylation markers and early recurrence in stage I lung cancer. *The New England journal of medicine* **358**, 1118–1128 (2008).
7. Beer, D. G. *et al.* Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nature medicine* **8**, 816–824 (2002).

8. Calin, G. A. *et al.* Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci USA* **101**, 2999–3004 (2004).
9. Calin, G. A. *et al.* Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci USA* **99**, 15524–15529 (2002).
10. Stilgenbauer, S. *et al.* Expressed sequences as candidates for a novel tumor suppressor gene at band 13q14 in B-cell chronic lymphocytic leukemia and mantle cell lymphoma. *Oncogene* **16**, 1891–1897 (1998).
11. Elnenaï, M. O. *et al.* Delineation of the minimal region of loss at 13q14 in multiple myeloma. *Genes Chromosomes Cancer* **36**, 99–106 (2003).
12. Dong, J. T., Boyd, J. C. & Frierson, H. F., Jr. Loss of heterozygosity at 13q14 and 13q21 in high grade, high stage prostate cancer. *Prostate* **49**, 166–171 (2001).
13. Ota, A. *et al.* Identification and characterization of a novel gene, C13orf25, as a target for 13q31–q32 amplification in malignant lymphoma. *Cancer Res* **64**, 3087–3095 (2004).
14. Tsujimoto, Y., Finger, L. R., Yunis, J., Nowell, P. C. & Croce, C. M. Cloning of the chromosome breakpoint of neoplastic B cells with the t(14;18) chromosome translocation. *Science* **226**, 1097–1099 (1984).
15. Nymark, P. *et al.* Integrative analysis of microRNA, mRNA and aCGH data reveals asbestos- and histology-related changes in lung cancer. *Genes, chromosomes & cancer* **50**, 585–597 (2011).
16. Lazar, V. *et al.* Integrated molecular portrait of non-small cell lung cancers. *BMC medical genomics* **6**, 53 (2013).
17. Molina-Pinelo, S. *et al.* MicroRNA-dependent regulation of transcription in non-small cell lung cancer. *PLoS one* **9**, e90524 (2014).
18. Cao, P. *et al.* Comprehensive expression profiling of microRNAs in laryngeal squamous cell carcinoma. *Head & neck* **35**, 720–728 (2013).
19. Theodore, S. C. *et al.* MicroRNA profiling of novel African American and Caucasian Prostate Cancer cell lines reveals a reciprocal regulatory relationship of miR-152 and DNA methyltransferase 1. *Oncotarget* **5**, 3512–3525 (2014).
20. Weir, B. A. *et al.* Characterizing the cancer genome in lung adenocarcinoma. *Nature* **450**, 893–898 (2007).
21. Weiss, J. *et al.* Frequent and focal FGFR1 amplification associates with therapeutically tractable FGFR1 dependency in squamous cell lung cancer. *Science translational medicine* **2**, 62ra93 (2010).
22. Chitale, D. *et al.* An integrated genomic analysis of lung cancer reveals loss of DUSP4 in EGFR-mutant tumors. *Oncogene* **28**, 2773–2783 (2009).
23. Manceau, G. *et al.* Hsa-miR-31-3p expression is linked to progression-free survival in patients with KRAS wild-type metastatic colorectal cancer treated with anti-EGFR therapy. *Clinical cancer research: an official journal of the American Association for Cancer Research* **20**, 3338–3347 (2014).
24. Lin, P. C. *et al.* Epigenetic repression of miR-31 disrupts androgen receptor homeostasis and contributes to prostate cancer progression. *Cancer research* **73**, 1232–1244 (2013).
25. Wiklund, E. D. *et al.* Coordinated epigenetic repression of the miR-200 family and miR-205 in invasive bladder cancer. *International journal of cancer Journal international du cancer* **128**, 1327–1334 (2011).
26. Jusufovic, E. *et al.* let-7b and miR-126 are down-regulated in tumor tissue and correlate with microvessel density and survival outcomes in non-small-cell lung cancer. *PLoS one* **7**, e45577 (2012).
27. Cho, W. C., Chow, A. S. & Au, J. S. Restoration of tumour suppressor hsa-miR-145 inhibits cancer cell growth in lung adenocarcinoma patients with epidermal growth factor receptor mutation. *European journal of cancer* **45**, 2197–2206 (2009).
28. Chen, Z. *et al.* miRNA-145 inhibits non-small cell lung cancer cell proliferation by targeting c-Myc. *Journal of experimental & clinical cancer research: CR* **29**, 151 (2010).
29. Campayo, M. *et al.* Low miR-145 and high miR-367 are associated with unfavourable prognosis in resected nonsmall cell lung cancer. *The European respiratory journal* **41**, 1172–1178 (2013).
30. Donadelli, M., Dando, I., Fiorini, C. & Palmieri, M. Regulation of miR-23b expression and its dual role on ROS production and tumour development. *Cancer letters* **349**, 107–113 (2014).
31. Majid, S. *et al.* miR-23b represses proto-oncogene Src kinase and functions as methylation-silenced tumor suppressor with diagnostic and prognostic significance in prostate cancer. *Cancer research* **72**, 6435–6446 (2012).
32. Geng, J. *et al.* Methylation mediated silencing of miR-23b expression and its role in glioma stem cells. *Neuroscience letters* **528**, 185–189 (2012).
33. Wang, P. *et al.* MicroRNA 23b regulates autophagy associated with radioresistance of pancreatic cancer cells. *Gastroenterology* **145**, 1133–1143 e1112 (2013).
34. Jin, L. *et al.* Prooncogenic factors miR-23b and miR-27b are regulated by Her2/Neu, EGF, and TNF-alpha in breast cancer. *Cancer research* **73**, 2884–2896 (2013).
35. Chen, L. *et al.* VHL regulates the effects of miR-23b on glioma survival and invasion via suppression of HIF-1alpha/VEGF and beta-catenin/Tcf-4 signaling. *Neuro-oncology* **14**, 1026–1036 (2012).
36. Zaman, M. S. *et al.* Inhibition of PTEN gene expression by oncogenic miR-23b-3p in renal cancer. *PLoS one* **7**, e50203 (2012).
37. Fidler, M. J. *et al.* PTEN and PIK3CA gene copy numbers and poor outcomes in non-small cell lung cancer patients with gefitinib therapy. *Br J Cancer* **105**, 1920–1926 (2011).
38. Buckingham, L. *et al.* PTEN, RASSF1 and DAPK site-specific hypermethylation and outcome in surgically treated stage I and II nonsmall cell lung cancer patients. *International journal of cancer Journal international du cancer* **126**, 1630–1639 (2010).
39. Kim, S. M. *et al.* Acquired resistance to cetuximab is mediated by increased PTEN instability and leads cross-resistance to gefitinib in HCC827 NSCLC cells. *Cancer letters* **296**, 150–159 (2010).
40. Bidkhori, G., Moeini, A. & Masoudi-Nejad, A. Modeling of tumor progression in NSCLC and intrinsic resistance to TKI in loss of PTEN expression. *PLoS one* **7**, e48004 (2012).
41. Heegaard, N. H. *et al.* Circulating micro-RNA expression profiles in early stage nonsmall cell lung cancer. *International journal of cancer Journal international du cancer* **130**, 1378–1386 (2012).
42. Edmonds, M. D. & Eischen, C. M. Differences in miRNA expression in early stage lung adenocarcinomas that did and did not relapse. *PLoS one* **9**, e101802 (2014).
43. Geng, Q. *et al.* Five microRNAs in plasma as novel biomarkers for screening of early-stage non-small cell lung cancer. *Respiratory research* **15**, 149 (2014).
44. Saito, M. *et al.* The association of microRNA expression with prognosis and progression in early-stage, non-small cell lung adenocarcinoma: a retrospective analysis of three cohorts. *Clinical cancer research: an official journal of the American Association for Cancer Research* **17**, 1875–1882 (2011).
45. Lu, Y. *et al.* MicroRNA profiling and prediction of recurrence/relapse-free survival in stage I lung cancer. *Carcinogenesis* **33**, 1046–1054 (2012).
46. Yanaihara, N. *et al.* Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer cell* **9**, 189–198 (2006).
47. Meng, W. *et al.* MicroRNA-31 predicts the presence of lymph node metastases and survival in patients with lung adenocarcinoma. *Clinical cancer research: an official journal of the American Association for Cancer Research* **19**, 5423–5433 (2013).

48. Dong, Z., Zhong, Z., Yang, L., Wang, S. & Gong, Z. MicroRNA-31 inhibits cisplatin-induced apoptosis in non-small cell lung cancer cells by regulating the drug transporter ABCB9. *Cancer letters* **343**, 249–257 (2014).
49. Patnaik, S. K., Kannisto, E., Knudsen, S. & Yendamuri, S. Evaluation of microRNA expression profiles that may predict recurrence of localized stage I non-small cell lung cancer after surgical resection. *Cancer research* **70**, 36–45 (2010).
50. Xi, Y. *et al.* Systematic analysis of microRNA expression of RNA extracted from fresh frozen and formalin-fixed paraffin-embedded samples. *Rna* **13**, 1668–1674 (2007).
51. Yu, L. *et al.* Early detection of lung adenocarcinoma in sputum by a panel of microRNA markers. *International journal of cancer Journal international du cancer* **127**, 2870–2878 (2010).
52. Zhu, W. *et al.* Expression of miR-29c, miR-93, and miR-429 as potential biomarkers for detection of early stage non-small lung cancer. *PloS one* **9**, e87780 (2014).
53. Hall, J. S. *et al.* Enhanced stability of microRNA expression facilitates classification of FFPE tumour samples exhibiting near total mRNA degradation. *Br J Cancer* **107**, 684–694 (2012).
54. Hui, A. B. *et al.* Comprehensive MicroRNA profiling for head and neck squamous cell carcinomas. *Clinical cancer research: an official journal of the American Association for Cancer Research* **16**, 1129–1139 (2010).
55. Lee, T. S. *et al.* Aberrant microRNA expression in endometrial carcinoma using formalin-fixed paraffin-embedded (FFPE) tissues. *PloS one* **8**, e81421 (2013).
56. Hoque, M. O., Lee, C. C., Cairns, P., Schoenberg, M. & Sidransky, D. Genome-wide genetic characterization of bladder cancer: a comparison of high-density single-nucleotide polymorphism arrays and PCR-based microsatellite analysis. *Cancer research* **63**, 2216–2222 (2003).
57. Rosenfeld, N. *et al.* MicroRNAs accurately identify cancer tissue origin. *Nat Biotechnol* **26**, 462–469 (2008).
58. Thomson, J. M., Parker, J., Perou, C. M. & Hammond, S. M. A custom microarray platform for analysis of microRNA gene expression. *Nature methods* **1**, 47–53 (2004).
59. Hayashi, M. *et al.* Epigenetic inactivation of VGF associated with Urothelial Cell Carcinoma and its potential as a non-invasive biomarker using urine. *Oncotarget* **5**, 3350–3361 (2014).
60. LeBron, C. *et al.* Genome-wide analysis of genetic alterations in testicular primary seminoma using high resolution single nucleotide polymorphism arrays. *Genomics* **97**, 341–349 (2011).

### Acknowledgements

We thank Dr. Prodipto Pal, for helping us in bioinformatic analysis of SNP array data. Grant Support: S. Begum is supported by a Young Clinical Scientist Award(YCSA) from the Flight Attendant Medical Research Institute (FAMRI, Miami, FL, USA), and M.O. Hoque is supported by the Young Investigator Award from the International Association for the Study of Lung Cancer and Clinical Innovative Award from FAMRI (Miami, FL, USA). This work was also supported by National Cancer Institute grant U01-CA84986.

### Author Contributions

S.B., M. Hayashi and M. Hoque wrote the main manuscript and design the study. S.B., M. Hayashi., T.O., F.J., S.T. and S.A. performed all experiments. M.B., E.I., W.W., W.K. and D.S. reviewed the manuscript and provide some reagents for the experiments.

### Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Begum, S. *et al.* An integrated genome-wide approach to discover deregulated microRNAs in non-small cell lung cancer: Clinical significance of *miR-23b-3p* deregulation. *Sci. Rep.* **5**, 13236; doi: 10.1038/srep13236 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>