

**Keywords:** lung cancer; miRNA; FFPE; molecular diagnosis and prognosis

# A microRNA-based prediction algorithm for diagnosis of non-small lung cell carcinoma in minimal biopsy material

N G Bediaga<sup>1,2,9</sup>, M P A Davies<sup>1,9</sup>, A Acha-Sagredo<sup>1,3,4</sup>, R Hyde<sup>1</sup>, O Y Raji<sup>1</sup>, R Page<sup>5</sup>, M Walshaw<sup>6</sup>, J Gosney<sup>7</sup>, A Alfirevic<sup>8</sup>, J K Field<sup>1</sup> and T Liloglou<sup>\*,1</sup>

<sup>1</sup>Roy Castle Lung Cancer Research programme, Department of Molecular and Clinical Cancer Medicine, Institute of Translational Medicine, University of Liverpool, Liverpool, UK; <sup>2</sup>BIOMICs Research Group, University of the Basque Country, Vitoria, Spain; <sup>3</sup>Oral Medicine and Pathology, Department of Stomatology II, UFI 11/25, University of the Basque Country, Leioa, Spain; <sup>4</sup>IKERBASQUE, Basque Foundation for Science, Bilbao, Spain; <sup>5</sup>Department of Thoracic Surgery, Liverpool Heart and Chest Hospital, Liverpool, UK; <sup>6</sup>Department of Respiratory Medicine, Liverpool Heart and Chest Hospital, Liverpool, UK; <sup>7</sup>Department of Pathology, Royal Liverpool and Broadgreen University Hospital Trust, Liverpool, UK and <sup>8</sup>Department of Molecular and Clinical Pharmacology, Institute of Translational Medicine, University of Liverpool, Liverpool, UK

**Background:** Diagnosis is jeopardised when limited biopsy material is available or histological quality compromised. Here we developed and validated a prediction algorithm based on microRNA (miRNA) expression that can assist clinical diagnosis of lung cancer in minimal biopsy material to improve clinical management.

**Methods:** Discovery utilised Taqman Low Density Arrays (754 miRNAs) in 20 non-small cell lung cancer (NSCLC) tumour/normal pairs. In an independent set of 40 NSCLC patients, 28 miRNA targets were validated using qRT-PCR. A prediction algorithm based on eight miRNA targets was validated blindly in a third independent set of 47 NSCLC patients. The panel was also tested in formalin-fixed paraffin-embedded (FFPE) specimens from 20 NSCLC patients. The genomic methylation status of highly deregulated miRNAs was investigated by pyrosequencing.

**Results:** In the final, frozen validation set the panel had very high sensitivity (97.5%), specificity (96.3%) and ROC-AUC (0.99,  $P = 10^{-15}$ ). The panel provided 100% sensitivity and 95% specificity in FFPE tissue (ROC-AUC = 0.97 ( $P = 10^{-6}$ )). DNA methylation abnormalities contribute little to the deregulation of the miRNAs tested.

**Conclusion:** The developed prediction algorithm is a valuable potential biomarker for assisting lung cancer diagnosis in minimal biopsy material. A prospective validation is required to measure the enhancement of diagnostic accuracy of our current clinical practice.

Lung cancer accounts for almost a quarter of cancer-related deaths and often presents at an advanced stage, which hinders effective treatment (Jemal *et al*, 2010; Jack *et al*, 2011). Early detection of lung cancer is a widely recognised unmet healthcare need and is expected to reduce mortality by allowing more effective therapies at earlier stages. However, implementation of early detection will

inevitably lead to testing smaller lesions, which poses diagnostic challenges. In addition, modern pathology is often called on to provide histo-molecular rather than histological diagnoses. Therefore, the availability of molecular biomarkers with high efficiency in minute biopsy material becomes increasingly important in aiding pathological diagnosis in confirming resection margins,

\*Correspondence: Dr T Liloglou; E-mail: tliloglo@liv.ac.uk

<sup>9</sup>These authors contributed equally to this work.

Received 1 June 2013; revised 12 September 2013; accepted 18 September 2013; published online 10 October 2013

© 2013 Cancer Research UK. All rights reserved 0007–0920/13

testing indeterminate nodules from CT scans and provide phenotypic information for therapeutic stratification purposes.

MicroRNAs (miRNAs or miRs) are short, non-coding RNAs that regulate expression of multiple genes by affecting stability and translational rate of messenger RNA (Garzon *et al*, 2009). MicroRNAs may offer great promise as biomarkers, for diagnostic, prognostic or therapeutic use (Liloglou *et al*, 2012). Their increased dynamic range compared with mRNA improves their diagnostic potential in samples with low percentage of cancer cells. In addition, their short (22nt) length and their association with nucleoproteins, confers greater stability (Jung *et al*, 2010) and thus extended life post sampling. Therefore, they are more tolerant to the standard clinical sample processing and storage conditions (Xi *et al*, 2007).

Lung-cancer-related miRNA expression patterns have been previously reported demonstrating a potential for detection in surrogate tissue (e.g. sputum, plasma) (Xing *et al*, 2010; Yu *et al*, 2010; Boeri *et al*, 2011). Besides the alterations of genomic miRNA copy numbers and modifications of the miRNA processing machinery, miRNA deregulation can also be due to epigenetic changes, such as the methylation status of miRNA loci (Yanai *et al*, 2006; Bandi *et al*, 2009; Farazi *et al*, 2011; Heller *et al*, 2012). Examples of methylated miRNA loci in non-small cell lung cancer (NSCLC) include members of the miR-34 family, hsa-miR-200c, hsa-miR-9, hsa-miR-124a and hsa-miR-126 (Lujambio *et al*, 2007; Gallardo *et al*, 2009; Ceppi *et al*, 2010; Wang *et al*, 2011; Heller *et al*, 2012; Watanabe *et al*, 2012).

In this study, we aimed to identify and validate a miRNA signature able to discriminate tumour from normal lung tissue with high sensitivity and specificity. We then further demonstrated the applicability of this panel in small fixed biopsy material. In addition, we investigated the role of DNA methylation in the deregulation of the miRNAs which demonstrated the most frequently aberrant expression patterns in this data set.

## MATERIALS AND METHODS

**Patients and samples.** Frozen tumour and/or normal samples from 112 NSCLC patients were obtained from Liverpool Heart and Chest Hospital. The demographic and clinical characteristics of the cancer patients are summarised in Table 1. Tissues were dissected within 30 min from surgical resection and were immediately stored at  $-80^{\circ}\text{C}$ . Formalin-fixed paraffin-embedded (FFPE) material was also available for the same tissues. The study protocol was approved by the Liverpool Research Ethics Committee and all patients provided written, informed consent in accordance with the Declaration of Helsinki and all local regulations.

**RNA isolation.** For frozen tissue, total RNA containing small RNA was extracted from five sections of  $10\ \mu\text{m}$  each for tumours and 10 sections for normal tissue, using a miRNeasy Mini Kit (Qiagen, Crawley, UK). For FFPE tissue, four sections of  $10\ \mu\text{m}$  each were first deparaffinized with xylene and then total RNA was extracted using the miRNeasy FFPE kit (Qiagen). The purity and concentration of RNA were determined by OD<sub>260/280/230</sub> readings using a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Hemel Hempstead, UK). The RNA integrity was determined by capillary electrophoresis using the RNA 6000 Nano Lab-on-a-Chip kit and the Bioanalyzer 2100 (Agilent Technologies, Warrington, UK).

**miRNA profiling.** MicroRNA profiling was performed using TaqMan Low Density Arrays (TLDA) (Applied Biosystems, Warrington, UK). The TaqMan Array Human microRNA Card Set v3.0 was used enabling accurate quantification of 754 human miRNAs. Briefly, 900 ng of total RNA was first reverse-transcribed using Megaplex Primer Pools, Human Pools Set v3.0 (Applied

**Table 1.** Clinical characteristics of non-small cell lung carcinoma patients in this study

	TLDA set	Validation set 1	Validation set 2	Total
N	20	45	47	112
<b>Age</b>				
mean (s.d.)	65.6 (1.5)	66.4 (1.2)	67.8 (1.3)	67.0 (8.0)
<b>Gender</b>				
Male:Female	14 : 6	20 : 25	27 : 20	61 : 51
<b>Histology</b>				
Adenocarcinoma	6	13	27	46
Squamous	14	32	20	66
<b>pT stage</b>				
1	0	4	12	16
2	18	37	28	73
3	2	4	6	12
4	0	0	1	1
<b>pN stage</b>				
0	8	26	31	65
1	7	16	9	32
2	5	3	7	15
<b>Clinical stage</b>				
IA	0	4	10	14
IB	8	21	18	47
IIA	0	0	2	2
IIB	6	15	9	30
IIIA	6	5	8	19

Abbreviation: TLDA = TaqMan Low Density Array.

Biosystems) and then quantified by PCR on TLDA cards using an Applied Biosystems 7900HT Real-Time PCR system.

Raw data for the analysis were extracted using the Sequence Detection System (SDS) Software v2.3 (Applied Biosystems). Each amplification curve was checked and optimal thresholds for each miRNA assay were defined using the RQ Manager Software (Applied Biosystems).  $C_t$  values  $>35$  were considered undetermined, whereas those  $<10$  were defined as unreliable. Valid data was imported into the HTqPCR package (Dvinge and Bertone, 2009) for visualisation, quality assessment, normalisation and testing for statistical significance in  $C_t$  values between different features. geNorm was used to determine the best set of normalisers, which were shown to be hsa-miR-26a, hsa-miR-140-5p, hsa-miR-195 and hsa-miR-30b.

**MicroRNA quantitative RT-PCR.** The identified miRNAs were evaluated in two independent validation sets from frozen tissue and a further test set from FFPE tissue by quantitative RT-PCR (qRT-PCR) using TaqMan miRNA assays (Applied Biosystems).  $C_t$  values of the target miRNAs were normalised in relation to that of the four reference identified by geNorm analysis. All assays were performed in duplicate, and no-template real-time controls were carried along in each experiment. MicroRNA expression levels (relative quantity, RQ) were calculated using the comparative  $C_t$  method ( $2^{-(C_t \text{ SAMPLE} - C_t \text{ MEAN CONTROL})}$ ) as previously described and fold change (FC) was computed as  $2^{-(\Delta C_t \text{ TUMOUR} - \Delta C_t \text{ NORMAL})}$  (Livak and Schmittgen, 2001).

**DNA methylation analysis.** DNA was isolated from frozen lung tumour and paired normal tissue using the DNeasy blood and tissue kit (Qiagen) and quantified using Picogreen (Invitrogen, Paisley, UK). One microgram DNA was bisulphite converted using the EZ-96 DNA Methylation-Gold Kit (ZymoResearch, Irvine, CA, USA). DNA methylation analysis was performed by pyrosequencing (Qiagen) as previously described (Shaw *et al.*, 2006). The primers and PCR annealing temperatures ( $T_a$ ) for pyrosequencing analysis are provided in Supplementary Table 1. The thermal profile for preparing the pyrosequencing templates was 95 °C for 5 min, followed by 40 cycles consisted of 94 °C for 30 s,  $T_a$  (Supplementary Table 1) for 30 s, 72 °C for 30 s and a final elongation at 72 °C for 10 min. Hyper/hypo-methylation cut-off points were calculated using 95% reference range ( $\text{mean} \pm 2 \times \text{s.d.}$ ) of the normal sample's methylation values. MicroRNAs -96, -182 and -183 are clustered sharing the same promoter thus a single pyrosequencing assays was designed for them.

### Statistical analysis

**Statistical analysis of the TLDA cards and further biological validations.** Statistical analysis of the TLDA cards and qRT-PCR data were conducted using statistical computing environment R (<http://www.r-project.org/>). For the analysis of the TLDA cards, significant miRNA differential expression between groups was classified utilising empirical Bayes moderated *t*-statistics with the limma package (<http://www.bioconductor.org/>). All reported *P*-values are adjusted by Benjamini-Hochberg procedure. We selected miRNAs with  $\geq 4$ -fold difference between lung tumour and normal and adjusted  $P < 0.05$  (HTqPCR package). Those markers showing the highest differential expression between tumour and normal tissues were selected for further biological validation. In addition, we validated two miRNAs (hsa-miR-27b and hsa-miR-34a), with a statistically significant  $> 3$ -FC, and which are currently under investigation in our laboratory owing to their involvement in the mitotic spindle checkpoint process. Both paired and non-paired *t*-test (two-tailed) were used to determine whether a statistically significant change was present in expression of validated miRNAs between the tumours and adjacent tissues.

**Data modelling.** A two-step classification scheme to discriminate tumour samples from normal tissue based on miRNA expression was implemented using the R packages e1071 (Dimitriadou *et al.*, 2009) and CMA (Slawski *et al.*, 2008). The scheme involved (i) feature selection and hyperparameter tuning using a combined Random Forest selection and support vector machine (SVM) classification approach with accuracy determined using five-fold cross-validation of validation set 1 (see the Results), and (ii) training an 8-miRNA linear SVM model on the validation set 1 using the package e1071. The latter model was validated in two independent test sets (based on the RNA from frozen and FFPE tissue, respectively) with performance assessed through receiver-operating characteristic (ROC) curve analysis using the R package epiR.

## RESULTS

**Discovery of aberrant miRNA signatures in NSCLC.** We used miRNA TLDA cards to profile expression signatures of mature miRNAs on NSCLC tissues and control normal lung tissue (taken distant to the tumour on the resected lobe). The tumour samples used in the discovery phase of this study (Table 1) included adenocarcinomas and squamous cell carcinomas, which are the most frequent types of NSCLC.

At a significance level of  $P < 0.05$  (adjusted for multiple testing) and a  $\geq 4$ -FC cut-off, 73 miRNAs were classified as over-expressed

and 8 were under-expressed in the cancer group. Two major clusters of samples were observed containing (i) all the normal tissues and three tumours and (ii) the remaining tumour samples (Supplementary Figure 1). Twenty-two over-expressed and six underexpressed miRNAs were selected for further validation, based on the high expression difference between tumour and normal tissues.

**Validation of differentially expressed miRNA biomarkers for NSCLC detection.** Validation of the 28 miRNAs identified in the discovery phase, was undertaken by qRT-PCR (Supplementary Table 2) in an independent cohort of frozen surgical tumour and normal paired samples from 45 NSCLC patients (validation set 1, Table 1). As in the discovery phase, internal normalisation was performed using the average of the four control miRNAs (hsa-miR-140, hsa-miR26a, hsa-miR-195 and hsa-miR-30b) which were also assayed individually. All the targets tested were significantly differentially expressed in lung tumours compared with paired normal lung (Supplementary Table 2).

**Construction of a diagnostic miRNA signature.** The data set consisting of normalised  $\Delta\text{CT}$  values for 28 miRNA markers analysed in validation set 1 was used to construct a miRNA-based diagnostic classifier. Five-fold cross-validation was used to select the number of miRNA variables, the cost hyperparameter and the specific miRNAs used in an SVM-based classifier. Random-Forest-based variable selection implemented in the R package CMA, revealed that the value of having additional miRNAs in the classifier decreases significantly after including the top 10 miRNAs. Consequently, SVM models were constructed containing 4–10 miRNAs, with Random Forest feature selection used to select the most informative miRNA variables. Sensitivity and specificity for the resulting models (using the above cross-validation scheme) was highest with eight miRNAs, the selected diagnostic variables being: hsa-miR-96, hsa-miR-450a, hsa-miR-183, hsa-miR-9, hsa-miR-577, hsa-Let-7i, hsa-miR-27b and hsa-miR-34a (sensitivity, 1; specificity, 0.978; misclassification rate, 0.011). Together with the four control miRNAs, this constitutes a 12-miRNA panel. The final SVM model, trained using the eight diagnostic miRNAs and the whole of validation set 1 was constructed for use in a second round of validation.

**Second validation of the eight miRNA signature.** The diagnostic performance of the constructed algorithm was assessed on a second independent set of tumour-normal sample pairs, composed of 40 tumours, 47 normal lung samples taken distal to the tumour site and 7 samples of histologically normal lung immediately adjacent to the tumour (validation set 2, Table 1). For this validation step, tumour and normal samples were entered blindly into the discriminatory model in order to determine sensitivity and specificity.

The resulting ROC curve (Figure 1A) has an AUC of 0.989 (95% CI 97.4–100%,  $P = 10^{-15}$ ). Using a prediction score cut-off that maximised the sum of sensitivity and specificity on the data (0.75), the panel of the markers produced 97.5% sensitivity and 96.3% specificity for NSCLC in the independent validation cases. The corresponding positive predictive value was 95.1% and negative predictive value was 98.1%. The distribution of prediction scores and the optimum cut-point defined are shown in Figure 1B; the prediction scores fall largely below 0.32 for normal and above 0.75 for tumour. Thus only 1/40 tumour samples and 2/54 normal samples were misclassified. The distance between the two groups of scores is indicative of the veracity of the prediction score (ranging from 0.51–0.75 all giving the same predictive values).

All the miRNAs in this panel showed significantly higher expression levels in the NSCLC tumours compared with the normal (Table 2, Figure 2). Notably, despite a clear separation

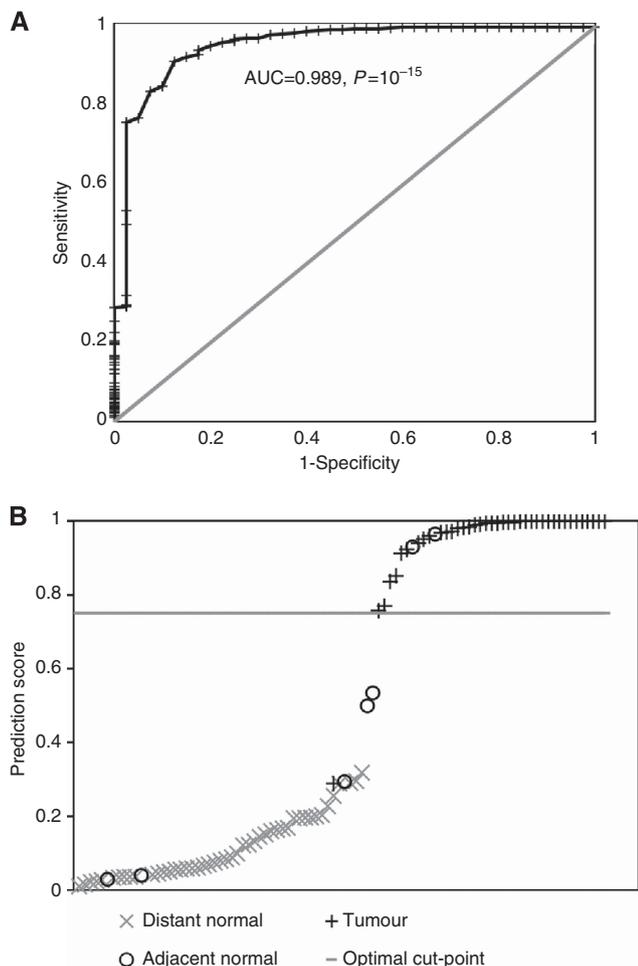


Figure 1. Evaluation of the prediction score for tumour/normal status utilising the validation set 1 data. (A) ROC curve. (B) Distribution of prediction scores for tumour samples (+), normal samples distant to tumour (x) and normal samples adjacent to tumour (circles).

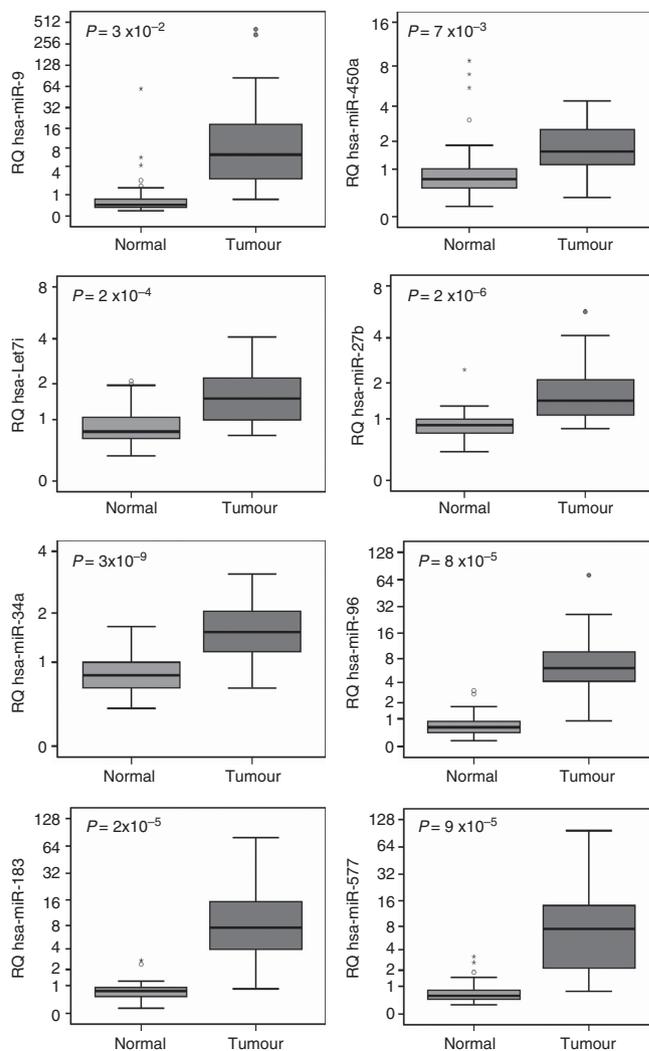


Figure 2. Boxplot of expression levels (RQ values) of the eight miRNAs in the model in normal and tumour samples of the validation set 2.

Table 2. Comparison of expression levels (FC) of the validated eight miRNAs between frozen tumours and normal samples (validation set 2) and FFPE tissues

	Frozen samples (validation set 2)			FFPE samples		
	FC	t	P	FC	t	P
hsa-Let-7i	2.4	-3.9	<0.001	2.67	-4.59	<0.001
hsa-miR-183	16.4	-4.89	<0.001	6.98	-8.32	<0.001
hsa-miR-27b	2.2	-5.41	<0.001	1.67	-5.62	<0.001
hsa-miR-34a	0.61	-6.59	<0.001	2.25	-3.83	<0.001
hsa-miR-450a	3.55	-2.76	<0.01	4.59	-7.16	<0.001
hsa-miR-577	22.2	-4.37	<0.001	12.51	-5.83	<0.001
hsa-miR-9	39.6	-2.3	<0.05	10.76	-5.18	<0.001
hsa-miR-96	16.3	-4.41	<0.001	9.17	-9.237	<0.001

Abbreviations: FC = fold change; FFPE = formalin-fixed paraffin-embedded.

between normal and tumour for prediction score, individually each miRNA displayed some overlap.

In order to experimentally establish the lowest proportion of cancer cells for positive identification with the current algorithm,

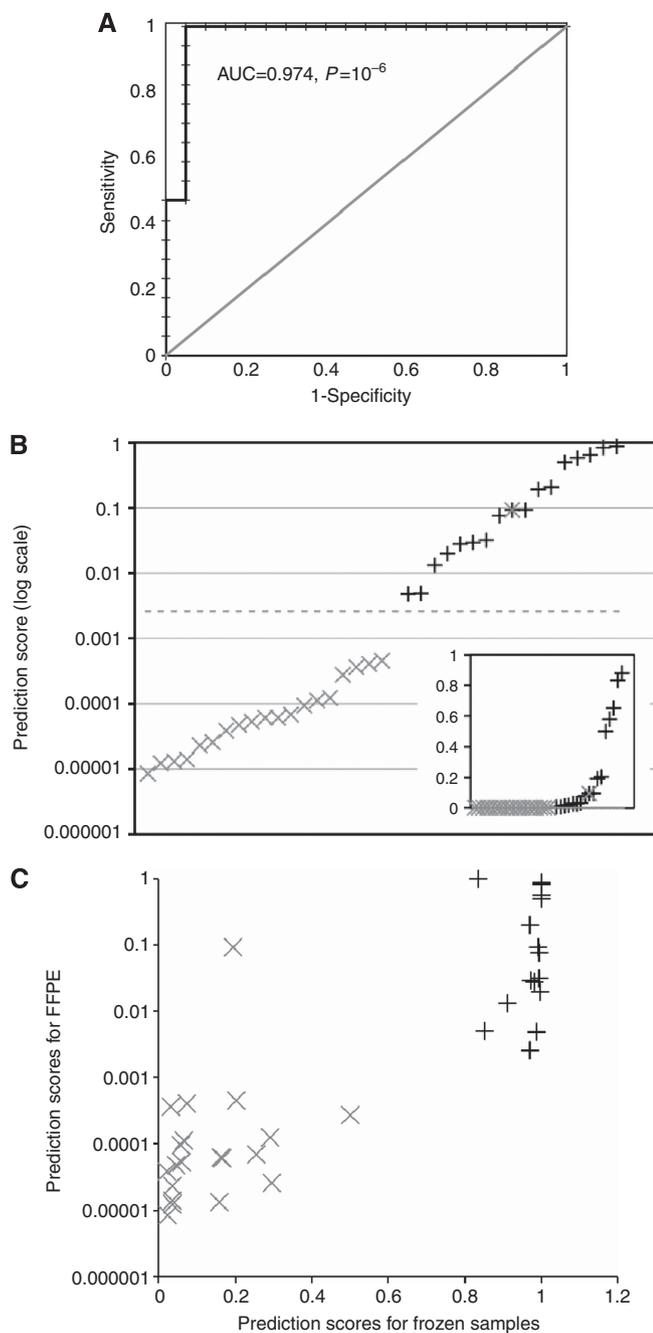
we applied the method to a two-fold dilution series of tumour in normal RNA. We used two tumour samples with 75% tumour content as estimated by histopathology, and a pool of three normal tissue RNAs. In both dilution series, the diagnostic algorithm produced a sigmoidal response being able to correctly classify the dilutions with  $\geq 18.8\%$  of tumour RNA (Supplementary Figure 1).

**Validation of marker panel in FFPE tissue samples.** In order to test the discriminatory efficiency of the prediction score model in FFPE tissue, normal and tumour FFPE samples from 22 patients in discovery or validation sets were also assessed (Supplementary Table 3). The prediction score threshold was lower (0.0026) for FFPE in comparison to frozen tissue (0.75). Although the prediction score distributions differed, there was still a clear separation between tumour and normal FFPE samples (Figure 3B) and using the optimal FFPE cut-point all tumour samples were correctly identified and only 1 of the 22 normal samples was misclassified. The ROC analysis showed that the panel performed well, giving an AUC of 0.974, ( $P=10^{-6}$ , Figure 3A), with a sensitivity of 100% and specificity of 95%.

It is of note that the percentage of tumour cells in some of these samples was as low as 18%, as determined by histopathological analysis, and that the prediction score was independent of the percentage of tumour cells in tumour samples. Prediction scores for both frozen and FFPE tissue were highly correlated ( $r=0.83$ ,  $P=10^{-10}$ ) (Figure 3C).

**Correlation between clinical data and the expression of the miRNAs in the model.** The expression levels of the miRNAs in the model are independent of the age, gender and tumour stage. Only the tumour size demonstrated a statistically significant correlation with RQ-hsa-miR-96 ( $P < 0.008$ ) and RQ-hsa-miR-450a ( $P < 0.018$ ).

We also found that the expression of hsa-miR-9, hsa-miR-96 and hsa-miR-577 was significantly higher in squamous carcinomas than in adenocarcinomas while the opposite relationship was found for hsa-miR-34a. Nonetheless it should be noted that the discriminatory power of the model is good regardless the histology of the tumour as each of the miRNAs in the model is highly overexpressed in all the tumours compared with normal tissue.



**Figure 3.** Validation of prediction score for tumour/normal status in FFPE samples. (A) ROC curve. (B) Distribution of prediction score for tumour (+) and normal (x) samples and (C) correlation of the prediction scores from matched frozen and FFPE samples.

Kaplan–Meier analysis demonstrated no statistically significant association between the expression of the miRNAs in the model and the patients’ prognosis.

**Methylation of miRNA loci in NSCLC patients.** Eleven out of the 28 miRNAs found to be deregulated in NSCLC in this study harbour a high/medium CpG content in their promoter (Table 3). We therefore examined their methylation status by pyrosequencing in 39 pairs of primary NSCLC tumours and corresponding normal lung tissues from the first validation set. Furthermore, we have evaluated LINE-1 hypomethylation, a surrogate marker of global DNA hypomethylation (Daskalos *et al*, 2009).

Hypermethylation in NSCLC tissue was observed for hsa-miR-34a (83%), hsa-miR-9.2 (45%) and hsa-miR-9.3 (46%), while hypomethylation was detected in the tumours for hsa-miR-27b (51%), hsa-miR-182 (31%), hsa-miR-24.1 (39%), hsa-miR-23b (28%), hsa-miR200b (23%), hsa-miR-486 (44%) and hsa-miR-338 (74%) (Table 3). Hypomethylation in the tumours was also detected for LINE-1 (77%) while hsa-Let-7i was not methylated in either the tumour or normal tissue. LINE-1 methylation levels correlated with those of hsa-miR-27b ( $r = 0.52$ ,  $P = 1 \times 10^{-6}$ ), hsa-miR-338 ( $r = 0.65$ ,  $P = 2 \times 10^{-10}$ ) and hsa-miR-486 ( $r = 0.53$ ,  $P = 7 \times 10^{-7}$ ), while it inversely correlated to that of hsa-miR-34a ( $r = -0.66$ ,  $P = 3 \times 10^{-10}$ ). The criteria for these Spearman’s correlations were set to  $r > 0.5$ ,  $P < 10^{-4}$  to adjust for multiple testing. In addition, the methylation levels of hsa-miR-27b correlated to that of hsa-miR-23 ( $r = 0.56$ ,  $P = 1.2 \times 10^{-7}$ ) and hsa-miR-24.1 ( $r = 0.50$ ,  $P = 1.9 \times 10^{-5}$ ), which belong to the same cluster.

Upon comparing the miRNA expression in relation to the methylation levels of their corresponding loci, we found that the expression of hsa-miR-27b was higher 1.8-fold (RQ:  $0.093 \pm 0.06$ ) in the hypomethylated tumours compared with the non-hypomethylated ones (RQ:  $0.053 \pm 0.026$ , Mann–Whitney,  $P = 0.026$ ). For the remaining miRNAs, the trends noted between methylation and expression did not reach statistical significance.

On the other hand, hypomethylation in miR-96/182/183 locus was associated with lower expression of hsa-miR-96 (hypomethylated: RQ =  $0.021 \pm 0.019$ ; non-hypomethylated: RQ =  $0.041 \pm 0.025$ ,  $P = 0.019$ ) and hsa-miR-183 (hypomethylated:

**Table 3.** Methylation levels in tumour and normal tissues

miRNA locus	Methylation level (mean $\pm$ s.d.)		T-test	
	Tumour	Normal	P-value	% hyper/hypomethylation
hsa-Let-7i	0 $\pm$ 0	0.1 $\pm$ 0.4	0.2	0
hsa-miR-96/182/183	89.4 $\pm$ 2.5	90.4 $\pm$ 0.9	<b>0.04</b>	31
hsa-miR-23b	75.8 $\pm$ 11.7	81.6 $\pm$ 2.3	<b>0.002</b>	28
hsa-miR-27b	84.8 $\pm$ 13.2	93.9 $\pm$ 2.2	<b><math>7 \times 10^{-5}</math></b>	51
hsa-miR-24.1	80.7 $\pm$ 15	87.2 $\pm$ 3.9	<b>0.005</b>	39
hsa-miR-200b	83.9 $\pm$ 11.2	88 $\pm$ 2	<b>0.03</b>	23
hsa-miR-338	48.4 $\pm$ 13.9	67.6 $\pm$ 3.3	<b><math>1 \times 10^{-9}</math></b>	74
hsa-miR-34a	76.2 $\pm$ 14.6	57.4 $\pm$ 5.3	<b><math>5 \times 10^{-8}</math></b>	83
hsa-miR-486	89.1 $\pm$ 7.4	93.4 $\pm$ 1	<b>0.0007</b>	44
hsa-miR-9.2	10.8 $\pm$ 13.5	1.8 $\pm$ 1.6	<b>0.0002</b>	45
hsa-miR-9.3	12.7 $\pm$ 10.2	7.4 $\pm$ 2.2	<b>0.001</b>	46
LINE-1	55.4 $\pm$ 8.2	68.6 $\pm$ 2.1	<b><math>2 \times 10^{-12}</math></b>	77

Significant P-values are highlighted in bold.

RQ =  $0.06 \pm 0.04$ ; non-hypomethylated: RQ =  $0.04 \pm 0.04$ ,  $P = 0.028$ ). Notably, in this case the relationship of decreased miRNA expression with genomic DNA hypomethylation was only true for tumours, as normal tissue still had lower expression than tumours with non-hypomethylated miR-96/182/183 locus (Supplementary Figure 2).

## DISCUSSION

Histology is the cornerstone of pathological analysis for lung cancer diagnosis, but can be challenging for a number of reasons including limited sample availability (as for needle biopsies of CT-detected indeterminate nodules) and ambiguous microscopic appearance (close to resection margins, or when quality is compromised by technical artefacts). There is a growing reliance on molecular diagnosis, especially for identification of clinically relevant histo-molecular subgroups. Cancer-specific miRNA expression patterns have been previously reported in lung cancer raising the question of clinical utility (Xing *et al*, 2010; Yu *et al*, 2010; Boeri *et al*, 2011). Here we examine miRNA suitability for molecular diagnosis in NSCLC.

In this study, we developed a miRNA panel based on an unbiased, systematic way, and evaluated its diagnostic efficiency through a multi-stage independent validation series. Furthermore, we confirmed the applicability of this panel in FFPE tissue, demonstrating the potential to use in CT-assisted diagnostic biopsies. We applied a systematic approach by using different technical platforms in three independent patient groups to discover and validate miRNA biomarkers for lung squamous cell carcinoma and adenocarcinoma, which comprise the large majority of NSCLCs. The similarity in the principle of TLDA cards (discovery) and qRT-PCR technology (validation) certainly contributed to successful downstream validation of all the miRNA candidates selected in discovery phase. A biomarker panel of eight discriminatory miRNAs and four normalisation control miRNAs has been assembled, demonstrating high sensitivity and specificity of delineation of normal lung and NSCLC tumour samples.

It should be noted that none of the miRNAs identified as significantly up-regulated in tumour tissue would be individually used as diagnostic biomarkers; Figure 2 clearly indicates that despite significant differential expression, there is overlap in expression levels. This is due to the significant variability between tumours in terms of expression levels and the admixture of cell types. However, by modelling the eight miRNA markers, a robust discriminatory tumour prediction score is achieved, which adds further support to the use of multi-locus biomarker panels for optimum sensitivity and specificity, as has been seen for other molecular alterations such as DNA methylation (Nikolaidis *et al*, 2012).

The prediction model built on the first validation data set was subsequently blindly tested in a second independent set of frozen tissue samples, resulting in a sensitivity and specificity of >96%. It is important to note that the efficiency of this panel proved to be uniform across pathological and clinical stages of lung cancer in our data set, indicating that it is very efficient from the early stages of the disease. However, having in mind its ultimate clinical use, it was imperative that we tested this panel in small FFPE samples as these are the standard pathology specimens to be utilised in CT-screening studies. The prediction model was still valid in this alternative sample type (sensitivity of 100%, specificity of 95%, ROC-AUC value of 0.97). It is of note the 275 ng of total RNA used in these assays represented on average only 8% of a single 10 micron FFPE section. This underlines the feasibility of using this panel in NSCLC diagnostic applications

when limited amounts of material are available, for example, FNA or biopsy samples taken to identify indiscriminate nodules detected by CT screening.

The lowest proportion of tumour material in the clinical samples assayed in this study was ~18%. Moreover, we determined experimentally that the lowest fraction of tumour cells required for correct classification by this algorithm using two dilution series of frozen tumour into normal tissue RNA was 18.8%. This demonstrates that the potential analytical sensitivity of this method is just below 20%, thus being tolerant to significant contamination by normal tissue material. Of course, certain diagnostic biopsies may contain even lower fractions of tumour cells. In such cases, diagnosis in NSCLC biopsies can be also facilitated by specific immunohistochemical markers (Kayser *et al*, 2013). It is likely that a combination of the two methods may significantly complement each other, resulting in a highly accurate diagnostic test. The ability to detect significant miRNA patterns in small samples, containing a variety of proportions of tumour tissue, as exemplified by the FFPE results, may be partly due to a proportion of this miRNA expression resulting from tumour-reactive stroma. The relatively high prediction scores, of two of the normal adjacent sections compared with tumour-distant normal samples supports this suggestion. An additional question that has to be addressed following this study is the predictive accuracy of this panel in preneoplastic lesions, in other words the ability to predict lung preneoplastic associated with malignant transformation. This will significantly boost the early detection potential of the method.

Our miRNA profiling revealed significant increase of hsa-miR-183 and hsa-miR-96 expression in lung cancer tissue compared with normal tissue, being in agreement to previous reports (Cho *et al*, 2009; Miko *et al*, 2009; Ma *et al*, 2011; Zhu *et al*, 2011). The miR-183 family members (hsa-miR-96, hsa-miR-182 and hsa-miR-183) are involved in the regulation of a wide range of cellular processes including cell proliferation (Segura *et al*, 2009; Lin *et al*, 2010), senescence (Li *et al*, 2009), cell migration (Lowery *et al*, 2010; Sarver *et al*, 2010) and metastasis (Segura *et al*, 2009). Hsa-miR-183 expression has demonstrated a potential oncogenic role by targeting EGR1 and PTEN (Sarver *et al*, 2010), but opposite results have also been found in a highly metastatic human pulmonary giant cell line, in which up-regulation of hsa-miR-183 repressed invasion and migration through targeting Ezrin, thus indicating that it also has a metastasis suppressor role (Wang *et al*, 2008). Hsa-miR-96 down-regulates both RAD51 and REV1 (involved in the homologous recombination and repair of DNA) and it might have a critical role in inhibition of DNA repair and chemosensitivity (Wang *et al*, 2012).

In our study, hsa-miR-27b was found to be markedly up-regulated in lung cancer tissues. The miR-27b/miR24.1/miR-23b cluster was significantly hypomethylated, and this hypomethylation significantly correlated with the overexpression of hsa-miR-27b, thus indicating the potential epigenetic deregulation of this miRNA in lung cancer. Down-regulation of hsa-miR-27b has been reported in lung cancer tissue (Yanaijara *et al*, 2006; Hirota *et al*, 2012) and in serum from NSCLC patients (Hennessey *et al*, 2012). Hsa-miR-27b expression levels have also been previously correlated with invasiveness of breast cancer (Wang *et al*, 2009) and with regulation of angiogenesis (Kuehbachner *et al*, 2008).

Hsa-miR-9, which represses the expression of E-cadherin promoting cell motility and invasiveness (Lu *et al*, 2012), was found to be over-expressed in NSCLC in our study. miR-9.2 and miR-9.3 loci were hypermethylated in 44.7% and 46.2% of NSCLC tumours examined. The hypermethylation of miR-9.3 has been previously reported (Heller *et al*, 2012) (Kitano *et al*, 2011). In our study, hypermethylation of miR-9.2/9.3 did not correlate with expression of miR-9, thus indicating that miR-9 might be regulated through other mechanisms; for example, regulated by PROX1

(a homeobox transcription factor) as has recently been reported (Lu *et al*, 2012).

miR-34a was overexpressed in most of the tumours in this study and hypermethylation of miR34a was a frequent alteration (83.3% of the tumours), but this aberrant methylation was not related to miR-34a expression. This is in agreement with previous reports showing no direct correlation between miR-34 methylation status and miR-34 expression levels (Corney *et al*, 2010; Tanaka *et al*, 2012) but contrasts one study reporting that methylation of this promoter region suppresses miR-34a expression in lung cancer (Gallardo *et al*, 2009). The miR-34 family includes three processed miRNAs that are encoded by two different genes: miR-34a is encoded by its own transcript, whereas miR-34b and miR-34c share a common primary transcript. The promoter regions of both loci include a p53-binding site (He *et al*, 2007), and it has been shown that the miR-34 gene family members are downstream transcriptional targets of p53.

In this study, we have observed a very diverse degree of deregulation (up- and down-) of many miRNAs between normal lung and tumour tissue. The reduction of Dicer expression in a fraction of NSCLCs (Karube *et al*, 2005) might imply an overall reduction of miRNA levels which has not seen in our data set. This is not surprising as miRNA expression regulation is a complicated process and Dicer is only a part of this. Nevertheless the inclusion of multiple endogenous controls in our reactions would normalise values for such an overall reduction and thus would not affect the algorithm.

Overall, our results confirm that DNA methylation has a minor role in regulating the expression of these miRNAs in the diagnostic panel validated in this study. DNA hypomethylation may be responsible for the up-regulation of mature hsa-miR-27b, but conversely may have a role in the decreased expression of hsa-miR-96 and hsa-miR-183 in some tumours. Moreover, the correlations observed in methylation levels of the miR-27b cluster stresses the importance of cis-acting expression regulation whereby deregulation of one member of the cluster is accompanied by similar deregulation of other cluster members.

In conclusion, we have developed a panel of miRNAs that can be reliably utilised in both frozen and FFPE lung tissue to assist clinical diagnosis of NSCLC with high sensitivity and specificity. This panel could enhance the diagnostic efficiency by either complementing histological diagnosis or assessing tumour margins; it certainly seems to be most relevant to the emerging importance of CT-detected lesions, where it might be applied to minimal biopsies of indeterminate nodules. The full clinical potential for this miRNA biomarker panel needs to be tested in a prospective setting, expanding the lung cancer types (small cell, large cell, etc.) to be included and also testing preneoplastic lesions.

## ACKNOWLEDGEMENTS

Supported by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. HEALTH-F2-2010-258677 (CURELUNG project) and grant agreement no. 258868 (LCAOS project). Additional funding from the Roy Castle Lung Cancer Foundation; Naiara G Bediaga and Amelia Acha-Sagredo are recipients of postdoctoral fellowships from the Basque Government.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## REFERENCES

- Bandi N, Zbinden S, Gugger M, Arnold M, Kocher V, Hasan L, Kappeler A, Brunner T, Vassella E (2009) miR-15a and miR-16 are implicated in cell cycle regulation in a Rb-dependent manner and are frequently deleted or down-regulated in non-small cell lung cancer. *Cancer Res* **69**(13): 5553–5559.
- Boeri M, Verri C, Conte D, Roz L, Modena P, Facchinetti F, Calabro E, Croce CM, Pastorino U, Sozzi G (2011) MicroRNA signatures in tissues and plasma predict development and prognosis of computed tomography detected lung cancer. *Proc Natl Acad Sci USA* **108**(9): 3713–3718.
- Ceppi P, Mudduluru G, Kumarswamy R, Rapa I, Scagliotti GV, Papotti M, Allgayer H (2010) Loss of miR-200c expression induces an aggressive, invasive, and chemoresistant phenotype in non-small cell lung cancer. *Mol Cancer Res* **8**(9): 1207–1216.
- Cho WC, Chow AS, Au JS (2009) Restoration of tumour suppressor hsa-miR-145 inhibits cancer cell growth in lung adenocarcinoma patients with epidermal growth factor receptor mutation. *Eur J Cancer* **45**(12): 2197–2206.
- Corney DC, Hwang CI, Matoso A, Vogt M, Flesken-Nikitin A, Godwin AK, Kamat AA, Sood AK, Ellenson LH, Hermeking H, Nikitin AY (2010) Frequent downregulation of miR-34 family in human ovarian cancers. *Clin Cancer Res* **16**(4): 1119–1128.
- Daskalos A, Nikolaidis G, Xinarianos G, Savvari P, Cassidy A, Zakopoulou R, Kotsinas A, Gorgoulis V, Field JK, Liloglou T (2009) Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer. *Int J Cancer* **124**(1): 81–87.
- Dimitriadou EHK, Leisch F, Meyer D, Weingessel A (2009) [http://CRAN.R-project.org/package=e1071] website.
- Dvinge H, Bertone P (2009) HTqPCR: high-throughput analysis and visualization of quantitative real-time PCR data in R. *Bioinformatics* **25**(24): 3325–3326.
- Farazi TA, Spitzer JI, Morozov P, Tuschl T (2011) miRNAs in human cancer. *J Pathol* **223**(2): 102–115.
- Gallardo E, Navarro A, Vinolas N, Marrades RM, Diaz T, Gel B, Quera A, Bandres E, Garcia-Foncillas J, Ramirez J, Monzo M (2009) miR-34a as a prognostic marker of relapse in surgically resected non-small-cell lung cancer. *Carcinogenesis* **30**(11): 1903–1909.
- Garzon R, Calin GA, Croce CM (2009) MicroRNAs in Cancer. *Annu Rev Med* **60**: 167–179.
- He L, He X, Lim LP, de Stanchina E, Xuan Z, Liang Y, Xue W, Zender L, Magnus J, Ridzon D, Jackson AL, Linsley PS, Chen C, Lowe SW, Cleary MA, Hannon GJ (2007) A microRNA component of the p53 tumour suppressor network. *Nature* **447**(7148): 1130–1134.
- Heller G, Weinzierl M, Noll C, Babinsky V, Ziegler B, Altenberger C, Minichsdorfer C, Lang G, Dome B, End-Pfutzenreuter A, Arns BM, Grin Y, Klepetko W, Zielinski CC, Zochbauer-Muller S (2012) Genome-wide miRNA expression profiling identifies miR-9-3 and miR-193a as targets for DNA methylation in non-small cell lung cancers. *Clin Cancer Res* **18**(6): 1619–1629.
- Hennessey PT, Sanford T, Choudhary A, Mydlarz WW, Brown D, Adai AT, Ochs MF, Ahrendt SA, Mambo E, Califano JA (2012) Serum microRNA biomarkers for detection of non-small cell lung cancer. *PLoS One* **7**(2): e32307.
- Hirota T, Date Y, Nishibatake Y, Takane H, Fukuoka Y, Taniguchi Y, Burioka N, Shimizu E, Nakamura H, Otsubo K, Ieiri I (2012) Dihydropyrimidine dehydrogenase (DPD) expression is negatively regulated by certain microRNAs in human lung tissues. *Lung Cancer* **77**(1): 16–23.
- Jack RH, Davies EA, Moller H (2011) Lung cancer incidence and survival in different ethnic groups in South East England. *Br J Cancer* **105**(7): 1049–1053.
- Jemal A, Siegel R, Xu J, Ward E (2010) Cancer statistics, 2010. *CA: Cancer J Clin* **60**(5): 277–300.
- Jung M, Schaefer A, Steiner I, Kempkensteffen C, Stephan C, Erbersdobler A, Jung K (2010) Robust microRNA stability in degraded RNA preparations from human tissue and cell samples. *Clin Chem* **56**(6): 998–1006.
- Karube Y, Tanaka H, Osada H, Tomida S, Tatematsu Y, Yanagisawa K, Yatabe Y, Takamizawa J, Miyoshi S, Mitsudomi T, Takahashi T (2005) Reduced expression of Dicer associated with poor prognosis in lung cancer patients. *Cancer Sci* **96**(2): 111–115.

- Kayser G, Csanadi A, Otto C, Plönes T, Bittermann N, Rawluk J, Passlick B, Werner M (2013) Simultaneous multi-antibody staining in non-small cell lung cancer strengthens diagnostic accuracy especially in small tissue samples. *PLoS One* **8**(2): e56333.
- Kitano K, Watanabe K, Emoto N, Kage H, Hamano E, Nagase T, Sano A, Murakawa T, Nakajima J, Goto A, Fukayama M, Yatomi Y, Ohishi N, Takai D (2011) CpG island methylation of microRNAs is associated with tumor size and recurrence of non-small-cell lung cancer. *Cancer Sci* **102**(12): 2126–2131.
- Kuehbach A, Urbich C, Dimmeler S (2008) Targeting microRNA expression to regulate angiogenesis. *Trends Pharmacol Sci* **29**(1): 12–15.
- Li G, Luna C, Qiu J, Epstein DL, Gonzalez P (2009) Alterations in microRNA expression in stress-induced cellular senescence. *Mech Ageing Dev* **130**(11–12): 731–741.
- Liloglou T, Bediaga NG, Brown BR, Field JK, Davies MP (2012) Epigenetic biomarkers in lung cancer. *Cancer Lett*; e-pub ahead of print 27 April 2012; PMID: 22546286.
- Lin H, Dai T, Xiong H, Zhao X, Chen X, Yu C, Li J, Wang X, Song L (2010) Unregulated miR-96 induces cell proliferation in human breast cancer by downregulating transcriptional factor FOXO3a. *PLoS One* **5**(12): e15797.
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta C(T)}$  method. *Methods* **25**(4): 402–408.
- Lowery AJ, Miller N, Dwyer RM, Kerin MJ (2010) Dysregulated miR-183 inhibits migration in breast cancer cells. *BMC Cancer* **10**: 502.
- Lu MH, Huang CC, Pan MR, Chen HH, Hung WC (2012) Prospero Homeobox 1 promotes epithelial-mesenchymal transition in colon cancer cells by inhibiting E-cadherin via miR-9. *Clin Cancer Res* **18**: 6416–6425.
- Lujambio A, Ropero S, Ballestar E, Fraga MF, Cerrato C, Setien F, Casado S, Suarez-Gauthier A, Sanchez-Cespedes M, Git A, Spiteri I, Das PP, Caldas C, Miska E, Esteller M (2007) Genetic unmasking of an epigenetically silenced microRNA in human cancer cells. *Cancer Res* **67**(4): 1424–1429.
- Ma L, Huang Y, Zhu W, Zhou S, Zhou J, Zeng F, Liu X, Zhang Y, Yu J (2011) An integrated analysis of miRNA and mRNA expressions in non-small cell lung cancers. *PLoS One* **6**(10): e26502.
- Miko E, Czinner Z, Csanky E, Boros G, Buslig J, Dezso B, Scholtz B (2009) Differentially expressed microRNAs in small cell lung cancer. *Exp Lung Res* **35**(8): 646–664.
- Nikolaidis G, Raji OY, Markopoulou S, Gosney JR, Bryan J, Warburton C, Walshaw M, Sheard J, Field JK, Liloglou T (2012) DNA methylation biomarkers offer improved diagnostic efficiency in lung cancer. *Cancer Res* **72**(22): 5692–5701.
- Sarver AL, Li L, Subramanian S (2010) MicroRNA miR-183 functions as an oncogene by targeting the transcription factor EGR1 and promoting tumor cell migration. *Cancer Res* **70**(23): 9570–9580.
- Segura MF, Hanniford D, Menendez S, Reavie L, Zou X, Alvarez-Diaz S, Zakrzewski J, Blochin E, Rose A, Bogunovic D, Polsky D, Wei J, Lee P, Belitskaya-Levy I, Bhardwaj N, Osman I, Hernando E (2009) Aberrant miR-182 expression promotes melanoma metastasis by repressing FOXO3 and microphthalmia-associated transcription factor. *Proc Natl Acad Sci USA* **106**(6): 1814–1819.
- Shaw RJ, Liloglou T, Rogers SN, Brown JS, Vaughan ED, Lowe D, Field JK, Risk JM (2006) Promoter methylation of P16, RARbeta, E-cadherin, cyclin A1 and cytoglobin in oral cancer: quantitative evaluation using pyrosequencing. *Br J Cancer* **94**(4): 561–568.
- Slawski M, Daumer M, Boulesteix AL (2008) CMA: a comprehensive bioconductor package for supervised classification with high dimensional data. *BMC Bioinform* **9**: 439.
- Tanaka N, Toyooka S, Soh J, Kubo T, Yamamoto H, Maki Y, Muraoka T, Shien K, Furukawa M, Ueno T, Asano H, Tsukuda K, Aoe K, Miyoshi S (2012) Frequent methylation and oncogenic role of microRNA-34b/c in small-cell lung cancer. *Lung Cancer* **76**(1): 32–38.
- Wang G, Mao W, Zheng S (2008) MicroRNA-183 regulates Ezrin expression in lung cancer cells. *FEBS Lett* **582**(25–26): 3663–3668.
- Wang Y, Huang JW, Calses P, Kemp CJ, Taniguchi T (2012) MiR-96 downregulates REV1 and RAD51 to promote cellular sensitivity to cisplatin and PARP inhibition. *Cancer Res* **72**(16): 4037–4046.
- Wang Y, Rathinam R, Walch A, Alahari SK (2009) ST14 (suppression of tumorigenicity 14) gene is a target for miR-27b, and the inhibitory effect of ST14 on cell growth is independent of miR-27b regulation. *J Biol Chem* **284**(34): 23094–23106.
- Wang Z, Chen Z, Gao Y, Li N, Li B, Tan F, Tan X, Lu N, Sun Y, Sun J, Sun N, He J (2011) DNA hypermethylation of microRNA-34b/c has prognostic value for stage non-small cell lung cancer. *Cancer Biol Ther* **11**(5): 490–496.
- Watanabe K, Emoto N, Hamano E, Sunohara M, Kawakami M, Kage H, Kitano K, Nakajima J, Goto A, Fukayama M, Nagase T, Yatomi Y, Ohishi N, Takai D (2012) Genome structure-based screening identified epigenetically silenced microRNA associated with invasiveness in non-small-cell lung cancer. *Int J Cancer* **130**(11): 2580–2590.
- Xi Y, Nakajima G, Gavin E, Morris CG, Kudo K, Hayashi K, Ju J (2007) Systematic analysis of microRNA expression of RNA extracted from fresh frozen and formalin-fixed paraffin-embedded samples. *RNA* **13**(10): 1668–1674.
- Xing L, Todd NW, Yu L, Fang H, Jiang F (2010) Early detection of squamous cell lung cancer in sputum by a panel of microRNA markers. *Mod Pathol* **23**(8): 1157–1164.
- Yanaihara N, Caplen N, Bowman E, Seike M, Kumamoto K, Yi M, Stephens RM, Okamoto A, Yokota J, Tanaka T, Calin GA, Liu CG, Croce CM, Harris CC (2006) Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell* **9**(3): 189–198.
- Yu L, Todd NW, Xing L, Xie Y, Zhang H, Liu Z, Fang H, Zhang J, Katz RL, Jiang F (2010) Early detection of lung adenocarcinoma in sputum by a panel of microRNA markers. *Int J Cancer* **127**(12): 2870–2878.
- Zhu W, Liu X, He J, Chen D, Hunag Y, Zhang YK (2011) Overexpression of members of the microRNA-183 family is a risk factor for lung cancer: a case control study. *BMC Cancer* **11**: 393.

This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported License.

Supplementary Information accompanies this paper on British Journal of Cancer website (<http://www.nature.com/bjc>)