microRNA classifiers are powerful diagnostic/prognostic tools in *ALK-*, *EGFR-*, and *KRAS*-driven lung cancers

Pierluigi Gasparini[a,1], Luciano Cascione[b,1], Lorenza Landi[c,1], Stefania Carasi[a], Francesca Lovat[a], Carmelo Tibaldi[c], Greta Alì[d], Armida D'Incecco[c], Gabriele Minuti[c], Antonio Chella[d], Gabriella Fontanini[d], Matteo Fassan[e], Federico Cappuzzo[c,2], and Carlo M. Croce[a,2]

[a]Comprehensive Cancer Center, Department of Molecular Virology, Immunology and Medical Genetics, The Ohio State University, Columbus, OH 43210; [b]Lymphoma and Genomics Research Program, Institute of Oncology Research, CH-6500 Bellinzona, Switzerland; [c]Department of Oncology, Tuscan Tumor Institute (ITT), 57100 Livorno, Italy; [d]University Hospital of Pisa, 56126 Pisa, Italy; and [e]Department of General Anatomic Pathology, University of Padua, 35100 Padova, Italy

microRNAs (miRNAs) can act as oncosuppressors or oncogenes, induce chemoresistance or chemosensitivity, and are major post-transcriptional gene regulators. Anaplastic lymphoma kinase (*ALK*), EGF receptor (*EGFR*), and V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (*KRAS*) are major drivers of non-small cell lung cancer (NSCLC). The aim of this study was to assess the miRNA profiles of NSCLCs driven by translocated *ALK*, mutant *EGFR*, or mutant *KRAS* to find driver-specific diagnostic and prognostic miRNA signatures. A total of 85 formalin-fixed, paraffin-embedded samples were considered: 67 primary NSCLCs and 18 matched normal lung tissues. Of the 67 primary NSCLCs, 17 were echinoderm microtubule-associated protein-like 4–*ALK* translocated (*ALK*[+]) lung cancers; the remaining 50 were not (*ALK*[−]). Of the 50 *ALK*[−] primary NSCLCs, 24 were *EGFR* and *KRAS* mutation-negative (i.e., WT; triple negative); 11 were mutant *EGFR* (*EGFR*[+]), and 15 were mutant *KRAS* (*KRAS*[+]). We developed a diagnostic classifier that shows how miR-1253, miR-504, and miR-26a-5p expression levels can classify NSCLCs as *ALK*-translocated, mutant *EGFR*, or mutant *KRAS* versus mutation-free. We also generated a prognostic classifier based on miR-769-5p and Let-7d-5p expression levels that can predict overall survival. This classifier showed better performance than the commonly used classifiers based on mutational status. Although it has several limitations, this study shows that miRNA signatures and classifiers have great potential as powerful, cost-effective next-generation tools to improve and complement current genetic tests. Further studies of these miRNAs can help define their roles in NSCLC biology and in identifying best-performing chemotherapy regimens.

microRNAs | lung cancer | EML4-*ALK* | *EGFR* | *KRAS*

Non-small cell lung carcinoma (NSCLC) includes adenocarcinomas and squamous cell carcinomas (1), which mainly are treated surgically, although chemotherapy is used both preoperatively (neoadjuvant chemotherapy) and postoperatively (adjuvant chemotherapy), with or without radiotherapy (2, 3). The best-characterized oncogenes that drive NSCLC are anaplastic lymphoma kinase (*ALK*), EGF receptor (*EGFR*), the V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (*KRAS*), the proto-oncogene receptor tyrosine kinase *MET*, and the recently identified receptor tyrosine kinase of the insulin receptor family *ROS1*. Knowledge of their mutational status is of key importance in choosing the most efficient chemotherapy regimen (4).

Recent clinical trials have shown that patients presenting with NSCLCs driven by *EGFR* and *ALK* gene rearrangements benefit from chemotherapies based on specific tyrosine kinase inhibitors (TKIs), such as EGFR-TKI and crizotinib, respectively. Despite the wide number of recently developed drugs, none targets mutant KRAS, although the inhibition of the downstream signaling component Ras-Raf-Erk seems to be a possible approach (2, 5).

microRNAs (miRNAs) are small, noncoding RNAs that are deregulated in malignancies (5) and can act as oncogenes or oncosuppressors (6); their deregulation also can induce chemoresistance or chemosensitivity in cancer cells (7, 8). Our group reported several tissue- and tumor-specific miRNA profiles and showed that different miRNA signatures can distinguish "hidden" subgroups within a single study class. Here we report the specific miRNA profiles of NSCLCs presenting with three well-studied driver mutations—*ALK*, *EGFR*, and *KRAS*—and the newly developed prognostic and diagnostic classifiers based on miRNA expression.

Numerous recent scientific reports show that the use of miRNAs as biomarkers has great potential because of their stability in tissues and bloodstream and also because their detection is relatively easy and can be performed by most laboratories, without particular expertise.

miRNA detection is feasible and cost-effective and thus represents an excellent diagnostic–prognostic tool to support more

Significance

microRNA profiles of anaplastic lymphoma kinase (*ALK*)-driven non-small cell lung cancers (NSCLCs) are currently not available in publically accessible databases. Identifying translocated *ALK*, mutant EGF receptor, and mutant V-Ki-ras2 Kirsten rat sarcoma cases in NSCLC is of value for determining which patients are more likely to benefit from a targeted therapy, to explicate mechanisms underlying chemotherapy survival, and ultimately in new drug development. microRNA-based classifiers are newly developed prognostic and diagnostic tools that can improve and complement the current gold-standard techniques. These classifiers also potentially represent an engine for boosting research on the role of these microRNAs in response to commonly used chemotherapy regimens in NSCLC to maximize patient outcomes.

established techniques (9). Generating miRNA signatures and classifiers can point basic researchers towards targets for investigations that can maximize the usefulness of information concerning diagnostic and chemotherapeutic outcomes in NSCLC patients.

Results

ALK-, EGFR-, KRAS-Driven NSCLCs Have Different miRNA Profiles.

The major aims of this project were to define the miRNA expression profiles of NSCLCs driven by three different oncogenes and to find predictors of mutational status and treatment outcome. The starting study cohort consisted of 85 RNAs extracted from formalin-fixed, paraffin-embedded (FFPE) samples: 67 primary NSCLC-derived RNAs from patients with a median age of 63 y and 18 lung tissues from healthy counterparts. Full demographic characteristics of the patient cohort are presented in Table 1 and in Table S1. Each of the 67 lung cancers was driven by translocation of ALK (ALK+), by mutant EGFR (EGFR+), or by mutant KRAS (KRAS+); 17 were echinoderm microtubule-associated protein-like 4 (EML4)-ALK translocated lung cancers (ALK+); the remaining 50 were not (i.e., ALK−). Within the ALK− subcohort, 24 were EGFR− and KRAS− (i.e., WT, also referred to as "triple negative"); 11 were EGFR+, and 15 were KRAS+. Further molecular characteristics of these subsets are presented in Table S2. All 85 samples passed the quality control and were run on the Human v2 miRNA Expression Assay from the NanoString nCounter system.

Table 1. Demographic characteristics of the NSCLC study group (n = 67 patients)

Patient characteristics	No. of patients	% of total
Median age in years (range)	63.5 (33–79)	
M/F	37/30	55.2/44.8
TNM stage		
I	13	19.4
IIA-B	18	26.9
IIIA-B	35	52.2
IV	1	1.5
Histology		
Adenocarcinoma	60	89.5
Adenosquamous	1	1.5
Squamous	4	6.0
Other*	2	3.0
Grading		
1/2/3	2/37/28	3.0/55.2/41.8
Adjuvant therapy		
Yes	24	35.8
No	43	64.2
Type of adjuvant therapy		
Chemotherapy alone	12	50.0
Radiotherapy alone	6	25.0
Chemo- and radiotherapy	6	25.0
Biological characteristics		
EGFR WT[†]	24	35.8
EGFR mutated	11	16.4
Exon 19	10	90.9
Exon 21	1	9.1
KRAS mutated[‡]	15	22.4
ALK+[§]	17	25.4
Overall survival in months (range)	34.4 (3.2–60.8)	

No data are available on smoking status. TNM, tumor/node/metastasis.
*Other histology included patients with clear cell carcinoma.
[†]EGFR WT included patients with EGFR WT and KRAS WT and ALK− (also called "triple negative") NSCLC.
[‡]Codon 12 exclusively.
[§]Defined by break-apart FISH assay.

Hierarchical clustering represented in the heat map in Fig. 1A shows 397 miRNAs differentially expressed in the whole set of 67 samples. These data show that miRNA expression profiles cluster separately into normal, EML4-ALK translocated (ALK+), and ALK− subgroups; furthermore within the ALK− group, the three subclasses WT, mutant EGFR (EGFR+), and mutant KRAS (KRAS+) also cluster differently. A full list of deregulated miRNAs associated with each heatmap and comparison is given in Dataset S1.

The Venn diagram shown in Fig. 1B clearly emphasizes the specificity of the deregulated miRNAs associated with each driver-mutation cancer subgroup: only two miRNAs are commonly deregulated in all three comparisons (center of the diagram), miR-570-3p and miR-376-3p. The high number of uniquely dysregulated miRNAs present in the ALK+ vs. mutant EGFR (n = 81) and ALK+ vs. mutant KRAS (n = 29) comparisons (blue and yellow sections, respectively) further underline the uniqueness of the ALK-driven miRNA profile. The comparison of mutant EGFR vs. mutant KRAS is distinguished by only eight miRNAs, making these two classes much more homogeneous than the ALK+ class (see Dataset S2 for a full list of the miRNAs in the different sections of the Venn diagram).

Fig. 1C shows differential clustering of 117 miRNAs in ALK+ (n = 17) vs. ALK− (n = 50) tumors. Further analysis of clustering by mutational status between these datasets is presented in Fig. 1 D–F. The comparison between the ALK+ and EGFR+ samples shows that 196 miRNAs are differentially expressed in the two groups (Fig. 1D). The heat map in Fig. 1E was generated by the clustering of 106 miRNAs from ALK+ and KRAS+ tumors. A relatively small pool of 78 miRNAs clearly divides mutant EGFR- and mutant KRAS-driven lung cancer, as shown in the heat map in Fig. 1F. In conclusion, this analysis shows that miRNA profiling can be a powerful predictor of mutational status in NSCLC. ALK+ cancers have an miRNA profile distinct from that of tumors driven by the other oncogenes and also appear to be a very heterogeneous group, judging by the large number of miRNAs differentially dysregulated in the ALK-driven cancers. This observation suggests that miRNA-definable subclasses within the ALK subgroup might be an interesting future challenge requiring a larger cohort of ALK-driven NSCLCs.

Diagnostic Classifier.

Aiming to identify a lung miRNA expression signature predictive of ALK, EGFR, and KRAS mutational status, we used the Conditional Inference classification Trees (CTree) implemented in the Bioconductor package party (10); the prediction accuracy of the classification algorithm was estimated using 10-fold cross-validation (11).

We identified a signature comprising three miRNAs: miR-1253, miR-504, and miR-26a-5p. This classifier based on miRNA expression levels can distinguish mutation-free (WT) NSCLCs from translocated ALK-, mutant EGFR-, or mutant KRAS-driven NSCLCs with an accuracy of 0.79 [95% confidence interval (CI) 0.67–0.88] and a multiclass area under the curve (AUC) of 0.692. The highest performances are reached in the classification of the three main cancer drivers as shown by the CTree in Fig. 2A. The histograms represent the driver-specific positive prediction rates of the model: 87.5%, 87.5%, 66.7%, and 100% of the observations in that node are classified as ALK+, mutant EGFR, WT, or mutant KRAS, respectively (also see the confusion matrix in Dataset S3A). The predictor's performance in the 10-fold cross-validation is summarized in Dataset S3B.

Fig. 2B shows the Kaplan–Meier plot of overall survival (OS) considering only the oncogenic driving alterations as stratification criteria, the Concordance Probability Estimate (CPE) is 0.69 (95% CI 0.68–0.70), and the Akaike's Information Criteria (AIC) is 88.75. The stratification represented in Fig. 2B is well known to clinicians who see daily that mutant EGFR and ALK+ cancer patients have a better response to chemotherapy (mostly

GENETICS
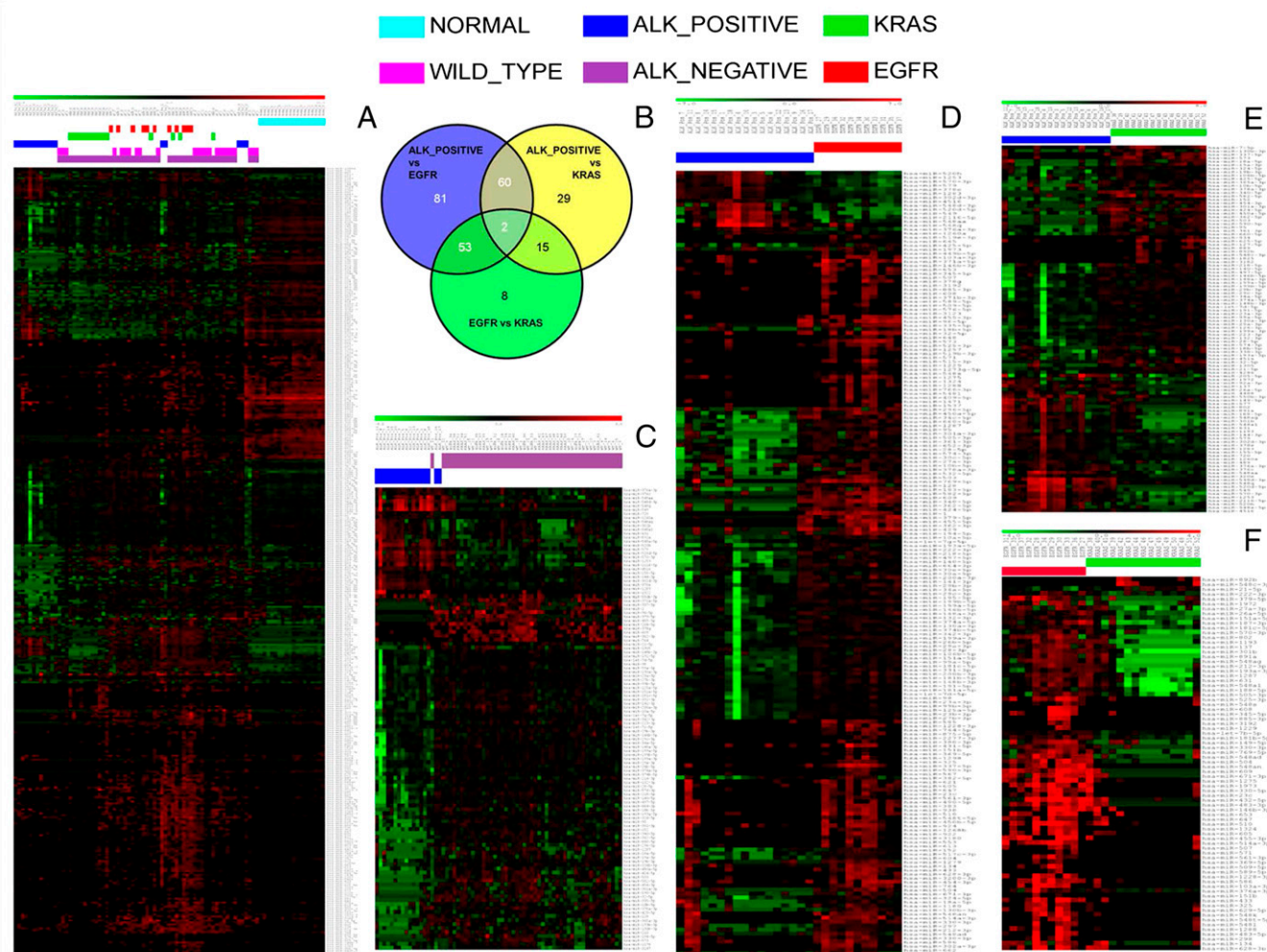
Fig. 1. Identification of miRNAs differentially expressed in different driver mutations. The heat maps represent relative miRNA expression as indicated in the green-to-red key bar at the top. Samples are shown in columns; miRNAs are shown in rows. Color-coded bars above the dendrograms identify the different study subgroups. (*A*) Heat map showing miRNA profiling of the whole cohort (*n* = 85) and their different clustering. (*B*) Venn diagram showing miRNAs commonly and specifically deregulated in drivers. (*C*) Heat map showing the miRNAs differentially deregulated in *ALK*⁺ and *ALK*⁻ samples. (*D*) Heat map showing the miRNAs differentially deregulated in *ALK*⁺ and mutant *EGFR* samples. (*E*) Heat map showing the miRNAs differentially deregulated in *ALK*⁺ and mutant *KRAS* samples. (*F*) Heat map showing the miRNAs differentially deregulated in mutant *EGFR* and mutant *KRAS* samples.

using pemetrexed and the TKI crizotinib, respectively) than do patients presenting with mutant *KRAS* cancers, for which, at present, there are no targeted drugs.

Prognostic Classifier. We then applied the same CTree method to identify an miRNA expression signature predictive of outcome and evaluated its potential.

The CTree in Fig. 2*C* shows the prognostic signature consisting of two miRNAs, miR-769-5p and let7d-5p, whose expression was significantly correlated with survival (Fig. 2*D*). miR-769-5p alone, when its expression is ≥2.825, distinguishes samples belonging to the prognostic category group A with best survival probability. To distinguish the prognostic category group B, with a medium survival rate, from group C, with a low survival rate, the expression level of miR-769-5p must be <2.825, and that of let-7d-5p must be <9.495 and ≥9.495. The Kaplan–Meier curves in Fig. 2*D* show statistically different OS for the three groups (Mantel–Cox test, *P* < 0.0001). The Cox model built using this classification has a CPE equal to 0.82 (95% CI 0.76–0.88), suggesting that our miRNA predictor has better discriminatory power than the mutational status-based predictor represented in Fig. 2*B*. Notably, 35% of patients were classified as high risk (group C), and the 2-y Kaplan–Meier estimates

of OS using two miRNAs as classifiers of prognostic and mutational status differed substantially in this group (23% vs. 40%, respectively) as well as in the low-risk group A (100% vs. 90%).

The absolute difference in OS between the low- and high-risk groups was 77% with the two-miRNAs signature stratification compared with 41% with mutational stratification. The AUC ranged from 0.79 to 0.91 across time for the prognostic classifier based on two miRNAs and from 0.61–0.71 across time for the mutational status model (red and green lines, respectively, in Fig. S1). The higher AUC of the prognostic model based on two miRNAs suggests its better performance. Similar conclusions were reached using the global model fit criterion (AIC) and discrimination measure (CPE), with the prognostic classifier based on two miRNAs achieving a better global model fit, lower AIC (68.02 vs. 88.75), and higher CPE in discrimination (0.82 vs. 0.69) for this population. Both the AIC and CPE show that the prognostic classifier based on two miRNAs achieved the best fit and discrimination for this population (Dataset S3*C*).

Validation

This study has two major limitations: the difficulty of collecting an external validation cohort and the impossibility of performing
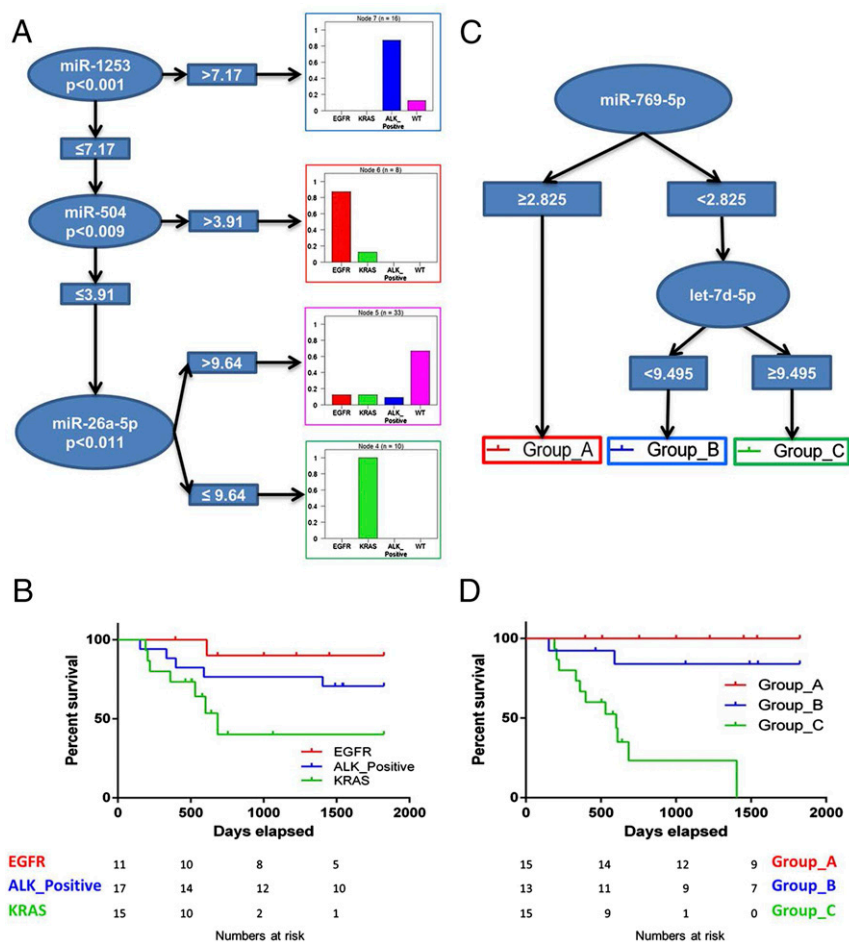
Fig. 2. Prognostic and diagnostic miRNA signatures. (*A*) CTree representation of the diagnostic classifier based on three miRNAs that is able to distinguish the three different drivers and the WT status. (*B*) Kaplan–Meier plot of OS considering only the oncogenic driving alterations as stratification criteria. (*C*) CTree representation of the prognostic signature based on two miRNAs. (*D*) Kaplan–Meier plot of OS identified by the expression levels of two miRNAs as shown in C.

a full in silico validation because of the absence of publicly available data of NSCLCs with proved mutational status profiled for miRNAs.

The Cancer Genome Atlas (TCGA) has incomplete information on both mutational status and miRNA profiling of NSCLCs [lung adenocarcinoma (LUAD) dataset] (12), which we summarize in Dataset S4, *Tab1*.

TCGA has only three *ALK*-translocated, 28 mutant *EGFR*, and 66 mutant *KRAS* samples profiled for miRNAs (Dataset S4, *Tab2*); these numbers represent a relatively small validation dataset for our signatures. Further analysis of the TCGA miRNA profile showed that miR-1253 expression levels are absent in these sample subsets (Dataset S4, *Tab3*), probably because they were excluded by the TCGA analysis cut off. The absence of hsa-mir-1253 TCGA expression data precludes the prognostic validation of the classifier we generated, because this miRNA represents its first decisional node. The value of the miRNA signatures is based on the pool of miRNAs taken together; considered singly, they might not be significant enough to be selected as a feature in the predictor. With these limitations in mind, we proceeded with various feasible in silico validation approaches.

Diagnostic Validation. We ran the diagnostic classifier (without miR-1253 and the corresponding discrimination for *ALK*⁺ samples) on 105 TCGA samples with unknown mutation status profiled for miRNA (Dataset S4, *Tab4*). The diagnostic miRNA

signature classified the cohort as 65% WT, 15% mutant *KRAS*, and 20% mutant *EGFR*, slightly higher than the percentages in recent epidemiological studies (40%, 15%, and 10%, respectively) (Dataset S4, *Tab5*) (13).

We then validated the miR-504 and miR-26a-5P expression pattern in the TCGA mutant *EGFR* and mutant *KRAS* subsets. The χ^2 test was used to investigate the relationship between dichotomized miRNA expression and mutational status in the TCGA LUAD dataset. The association between miR-504 (but not miR-26a expression) and mutant *EGFR* was confirmed: a high level of miR-504 expression was associated with *EGFR* mutations [odds ratio 2.86 (95% CI 1.07–7.71), $P = 0.04$], although low levels of miR-26a-5p were not significantly associated with the mutant *KRAS* or WT phenotype.

Prognostic Validation. The validation of the prognostic signature on the TCGA dataset presented some other difficulties because of the heterogeneity of the two cohorts in terms of OS: in the TCGA dataset, 20 of 76 patients (26%) died with a median survival time of 261 mo, whereas in the cohort we collected 13 of 43 patients (30%) died with a median survival of time 1,002 mo. Furthermore, in the LUAD dataset the only three patients with ALK-EML4 fusions had a limited follow-up period (141 and 162 mo), and all three were alive at last follow-up.

The miRNA signature predicting outcome was tested on a set of 89 samples (with unknown mutational status, miRNA profiling,

and follow-up; see Dataset S4, *Tab4*), and the predictor was able to distinguish a group with worst prognosis (as group C), even though the trend shown by the curves seems similar to that seen for groups A and B in Fig. 2*D* (the difference between group A and group B is not statistically significant).

In the NSCLC study cohort we used, the univariate Cox proportional hazards regression model correlated 57 miRNAs to the OS of patients; for each we dichotomized the cohort using the median miRNA expression as cutoff (all results are presented in Dataset S5). We then generated Kaplan–Meier survival curves, finding 14 miRNAs that significantly (log-rank test < 0.05) split the cohort into high-risk vs. low-risk subpopulations.

We determined whether miRNAs are independent prognostic factors by performing a multivariate Cox proportional hazards regression analysis to the OS and mutational status (with mutational status and miRNA expression as covariates). We found 43 miRNAs associated with OS independently of the mutational status; 10 of those miRNAs improved risk stratification beyond the information provided by traditional classificator based on mutational status only.

We then tested the prognostic value of the aforementioned miRNAs in the TCGA LUAD dataset (Dataset S5). We validated the prognostic value of six miRNAs in the univariate model (hsa-miR-1287, hsa-miR-181c-5p, hsa-miR-200a-3p, hsa-miR-200b-3p, hsa-miR-29c-3p, and hsa-miR-9-5p), but only hsa-miR-181c-5p identified high-risk and low-risk subpopulations in both cohorts. In addition, we also validated hsa-miR-141-3p, hsa-miR-200a-3p, and hsa-miR-200b-3p as independent prognostic factors and possibly as prognostic biomarkers in NSCLC.

miRNA Expression Validation. Last, considering that we generated the miRNA profiling starting from FFPE samples, opening the possibility of applying our classifiers to the large numbers of archival specimens present in pathology laboratories, we wanted to validate the miRNA expression data. We aimed to validate the expression of a pool of miRNAs that are well expressed across all the subgroups of the dataset by performing real-time quantitative PCR (TaqMan qRT-PCR assay) on the RNAs used for profiling. In a subset of samples randomly chosen based on RNA availability, we tried to validate six miRNAs (hsa-miR-518-5P, hsa-mir-520-5P, hsa-mir-520h, hsa-mir-548d-3P, hsa-mir-548q, and hsa-mir-549, plus two normalizers). Five of the six miRNAs presented an expression level across the subtypes reflecting the one detected by NanoString technology; box plots representing this qRT-PCR–based validation are shown in Fig. S2.

Discussion

miRNAs are well-known key players in downstream oncogenic pathways, behaving as oncogenes or oncosuppressors in many types of cancer. They also have been clearly identified as modulators of chemosensitivity and chemoresistance in several cancer models (7, 8, 14). Their central role in tumorigenesis and their stable and long-lasting presence in tissues and body fluids underlie the increased efforts and interest in defining their roles as possible next-generation biomarkers.

Here we show that the miRNA expression profiles of the cohort representing translocated *ALK*-, mutant *EGFR*-, and mutant *KRAS*-driven NSCLCs are clearly different; the expression clusters of the different cancer groups show that each driver subgroup is recognizable by a specific miRNA subset. The heterogeneity of the miRNA profile in the ALK$^+$ study group also suggests the possible presence of miRNA-definable subclasses within this subclass, representing opportunities both for understanding *ALK*-driven tumor biology and for new insights for chemotherapeutic development.

We have generated a three-miRNA classifier based on miR-1253, miR-504, and miR-26a-5p expression levels that can categorize samples as *ALK$^+$*, mutant *EGFR*, or mutant *KRAS*.

Because the presence of a targetable driver mutation (*EGFR* or *ALK*) is of key importance in selecting the best targeted treatment (versus chemotherapy), genetic tests are crucial. To perform a genetic test to detect these mutations, a team of highly trained personnel is needed: the procedure is delicate, expensive, and time consuming, and results must be read by a pathologist. We believe that the development of fast prescreening tests to improve and complement current genomic tests is of great interest; miRNAs detection is easy, affordable, and does not require particular expertise. For these reasons miRNAs represent diagnostic biomarkers. Although we are aware that our three-miRNAs predictor has only 79% accuracy, we believe it can be improved significantly by testing our signature in other independent study cohorts; unfortunately we were unable to collect such cohorts or to find them available online, mainly, but not only, because of difficulties in collecting *ALK*-translocated samples.

Considering the OS of the patients, we built another classifier based on the expression levels of only two miRNAs: miR-769-5p and Let-7d-5p. Knowing the expression levels of these two prognostic miRNAs, we can predict good, medium, or poor survival. This newly developed miRNA-based prognostic classifier has shown an improved ability to discriminate both high-risk (with a 2-y OS <50%) and low-risk groups as compared with the commonly used predictor based on mutational status only. In conclusion, this classifier could represent a useful tool in clinical settings for selecting the optimal chemotherapy regimen.

Further studies of these miRNAs should be pursued to define their roles in these driver-specific cancers, in particular in mutant *KRAS*-driven NSCLCs that still lack specifically designed chemotherapies.

In NSCLC, the chromosome inversion inv (2) (p21;p23) generates the EML4-ALK fusion protein (the N-terminal regions of the EML4 protein fused with the 3′ end of ALK kinase) and results in the constitutive activation of this kinase in cancer cells (15–17). Because of the very limited presence of *ALK$^+$* mutations in NSCLCs (4–7%), we were not able to find a validation set for this study. Ongoing clinical trials recruiting *ALK$^+$* patients (i.e., the Alchemist Lung Cancer Trials; www.cancer.gov/researchandfunding/areas/clinical-trials/nctn/alchemist) may be valuable sources of samples in the future, but such samples are not yet available.

Currently, publically accessible data repositories have no NSCLC miRNA data associated with mutational status in a number that could allow us any complete in silico validation.

To our knowledge, this research is the first complete miRNA profiling of three well-known NSCLC-driver mutations; the data will be freely available for the researchers (Gene Expression Omnibus accession no. GSE72526), integrating the patchy NSCLC data available on TCGA database.

We are aware of the limitations of our study, and we hope that the publication of this research containing several interesting findings will encourage high-quality data sharing and the publishing of reports replicating and validating our data. The predictors/signatures generated represent a previously unreported, useful, cost-effective way to complement the gold-standard techniques such as FISH to maximize patient outcome and also provide an engine to boost research on the role of these miRNAs in responses to commonly used chemotherapy regimens in NSCLC.

Materials and Methods

Sample Inclusion Criteria. A total of 88 samples were collected. Three samples (two WT and one translocated *ALK*) did not pass the RNA quality control and were eliminated from further analysis. Of the remaining 85 samples, 67 primary NSCLC-derived RNAs and 18 normal counterpart tissues (see Table 1 and Tables S1 and S2) were collected at the Tuscan Tumor Institute (ITT), Livorno, Italy and University Hospital of Pisa, Pisa, Italy. The investigation was conducted in accordance with the ethical standards, the Declaration of Helsinki, and national and international guidelines on research with human subjects, The Ohio State University IRB protocol no. 2005C0014 (C.M.C.), and study 94 protocol LIVONCO2013-03 (F.C.). Each of the 67 lung cancers was

driven by translocated *ALK*, mutant *EGFR*, or mutant *KRAS*; 17 were EML4-*ALK* translocated lung cancers (*ALK*+); the remaining 50 were not (i.e., were *ALK*−). Within the *ALK*− subcohort, 24 were *EGFR*− and *KRAS*− (WT; i.e., triple negative); 11 were mutant *EGFR* (*EGFR*+); 15 were mutant *KRAS* (*KRAS*+). In the present study all patients had resectable disease, and all patients underwent surgery with radical intent. Thirty-two patients (48%) had pathological N2-stage disease, and only one patient presented with stage IV disease because of the presence of a single adrenal metastasis, a condition suitable for surgery. The seventh edition of the lung cancer TNM classification and staging system was used. Adjuvant chemotherapy was offered according to international guidelines and was delivered to the patients (44%) who were candidates for adjuvant treatment. The remaining patients were considered unsuitable for adjuvant therapy.

RNA Extraction and NanoString nCounter Assay. Total RNAs were isolated from FFPE tissues using the RecoverAll kit (Ambion) following the manufacturer's protocol. About 100 ng of total RNA per sample was processed with the Human v2 miRNA Expression Assay from the nCounter system (NanoString) based on MiRBase v. 18 in the Nucleic Acid Shared Resource of The Ohio State University.

Data Analysis. The NanoString miRNA panel detects 800 endogenous miRNAs, five housekeeping transcripts [β-actin (NM_001101.2), β-2 microglobulin (NM_004048.2), GAPDH (NM_002046.3), RPL19 (NM_000981.3), and RPLP0 (NM_001002.3)], and six positive and eight negative proprietary spike-in controls. Unlike traditional hybridization microarrays, NanoString does not associate targets with spatial coordinates; instead, the system generates copy numbers of target-specific molecular barcodes attached to detection probes, theoretically eliminating position-dependent effects. Raw data, which are proportional to copy number, were log-transformed and normalized by the quantile method after application of a manufacturer-supplied correction factor for several miRNAs. Data were filtered to exclude features below the detection threshold (defined for each sample by a cutoff corresponding to approximately twice the SD of negative control probes plus their mean) in at least 20% of the samples.

Using R/Bioconductor and the filtered dataset, linear models for microarray data analysis (Limma) was used with a contrast matrix for the studied comparisons. *P* values were used to rank miRNAs of interest, and correction for multiple comparisons was done by the Benjamini–Hochberg method. Raw data that were above background and the corresponding quantile-normalized data also were imported into Multi-Experiment Viewer. The samples were clustered hierarchically by Pearson correlation distance and average linkage.

The miRNA microarray expression data were submitted to the Gene Expression Omnibus dataset (accession no. GSE72526). All fold-changes associated with these analyses were represented in log₂ scale (logFC), and only data with an adjusted *P* value <0.05 were considered statistically significant.

Results of statistical analysis are expressed as mean ± SD unless indicated otherwise. GraphPad Prism version 5.0 was used for graphic purposes.

OS curves were done according to Kaplan–Meier method. Censoring occurred at the date of death from any cause or at the time of the last known follow-up. Comparisons of outcomes between subgroups were performed using the log-rank test. The Cox proportional hazard model was used for univariate and multivariate analyses of prognostic factors. The performances of the "driving alteration" model and miRNA signature for OS were compared by a measure of global fit (AIC) and by a measure of discrimination (CPE) along with its 95% CI (18–20). Low AIC values indicate better fit, and high CPE values indicate better discrimination. The area under the receiver operator characteristic curve (ROC) over time is shown as the AUC (21).

To develop an miRNA signature of OS, we used the Recursive Partitioning algorithm implemented in the Bioconductor package party. Recursive partitioning is a fundamental tool with good performance in data mining. It helps us explore the structure of a set of data while developing easy-to-visualize and simple decision rules for predicting a categorical outcome (classification tree). To assess the performance of the CTree predictor, we used the caret, CPE, and risksetROC packages.

Mutations Status Detection. *ALK* translocation was detected by FISH on tumor sections obtained from paraffin-embedded tumor blocks with the use of a commercially available break-apart probe specific to the *ALK* locus (Vysis LSI *ALK* Dual Color, Break Apart Rearrangement Probe; Abbott Molecular) according to the manufacturer's instructions on all NSCLC samples. Tumor samples were scored by three independent investigators (including G.A. and G.F.) blinded to the clinicopathological characteristics of the patients and to immunohistochemical results, according to the score proposed by Kwak et al. (22). Mutational profiling of *EGFR* (exons 18–21) was performed as previously reported (23). Pyrosequencing assays were performed for sequence analysis of *KRAS* (codons 12 and 13) (24). See *SI Materials and Methods* for the full details of the protocols for detecting mutational status. Full details of the mutations detected in each case are reported in Table S2.

qRT-PCR. cDNA was reverse transcribed from 10 ng of total RNA of each sample using specific miRNA primers from the TaqMan MicroRNA Assays and reagents from the TaqMan MicroRNA Reverse Transcription Kit (Life Technologies). Subsequently, in the PCR step, PCR products were amplified from cDNA samples using the TaqMan MicroRNA Assays together with the TaqMan Universal PCR Master Mix. All assays were performed in triplicate according to the manufacturer's instructions.

1. Davidson MR, Gazdar AF, Clarke BE (2013) The pivotal role of pathology in the management of lung cancer. *J Thorac Dis* 5(Suppl 5):S463–S478.
2. Chen Z, Fillmore CM, Hammerman PS, Kim CF, Wong KK (2014) Non-small-cell lung cancers: A heterogeneous set of diseases. *Nat Rev Cancer* 14(8):535–546.
3. Langer CJ, Besse B, Gualberto A, Brambilla E, Soria JC (2010) The evolving role of histology in the management of advanced non-small-cell lung cancer. *J Clin Oncol* 28(36):5311–5320.
4. Li T, Kung HJ, Mack PC, Gandara DR (2013) Genotyping and genomic profiling of non-small-cell lung cancer: Implications for current and future therapies. *J Clin Oncol* 31(8): 1039–1049.
5. Volinia S, et al. (2006) A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc Natl Acad Sci USA* 103(7):2257–2261.
6. Croce CM (2009) Causes and consequences of microRNA dysregulation in cancer. *Nat Rev Genet* 10(10):704–714.
7. Garofalo M, et al. (2009) miR-221&222 regulate TRAIL resistance and enhance tumorigenicity through PTEN and TIMP3 downregulation. *Cancer Cell* 16(6):498–509.
8. Garofalo M, et al. (2012) EGFR and MET receptor tyrosine kinase-altered microRNA expression induces tumorigenesis and gefitinib resistance in lung cancers. *Nat Med* 18(1):74–82.
9. Heneghan HM, Miller N, Kerin MJ (2010) MiRNAs as biomarkers and therapeutic targets in cancer. *Curr Opin Pharmacol* 10(5):543–550.
10. Hothorn T, Hornik K, Zeileis A (2006) Unbiased recursive partitioning: A conditional inference framework. *J Comput Graph Stat* 15(3):651–674.
11. Zelleis A, Hothorn T, Hornik K (2008) Model-based recursive partitioning. *J Comput Graph Stat* 17(2):492–514.
12. Cancer Genome Atlas Research Network (2014) Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 511(7511):543–550.
13. Levy MA, Lovly CM, Pao W (2012) Translating genomic information into clinical medicine: Lung cancer as a paradigm. *Genome Res* 22(11):2101–2108.
14. Acunzo M, et al. (2012) miR-130a targets MET and induces TRAIL-sensitivity in NSCLC by downregulating miR-221 and 222. *Oncogene* 31(5):634–642.
15. Camidge DR, et al. (2012) Activity and safety of crizotinib in patients with ALK-positive non-small-cell lung cancer: Updated results from a phase 1 study. *Lancet Oncol* 13(10): 1011–1019.
16. Gerber DE, Minna JD (2010) ALK inhibition for non-small cell lung cancer: From discovery to therapy in record time. *Cancer Cell* 18(6):548–551.
17. Soda M, et al. (2007) Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature* 448(7153):561–566.
18. Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Automat Contr* 19:716–723.
19. Gönen MHG (2005) Concordance probability and discriminatory power in proportional hazards regression. *Biometrika* 92(4):965–970.
20. Harrell FE (2001) *Regression Modeling Strategies with Applications to Linear Models, Logistic Regression, and Survival Analysis.* (Springer, New York).
21. Heagerty PJ, Zheng Y (2005) Survival model predictive accuracy and ROC curves. *Biometrics* 61(1):92–105.
22. Kwak EL, et al. (2010) Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N Engl J Med* 363(18):1693–1703.
23. Boldrini L, et al. (2009) Epidermal growth factor receptor and K-RAS mutations in 411 lung adenocarcinoma: A population-based prospective study. *Oncol Rep* 22(4): 683–691.
24. Querings S, et al. (2011) Benchmarking of mutation diagnostics in clinical lung cancer specimens. *PLoS One* 6(5):e19601.

GENETICS

Gasparini et al.