Supporting Information

Boeri et al. 10.1073/pnas.1100048108

SI Materials and Methods

CT Screening Protocols. In the INT/IEO screening cohort of 1,035 high-risk heavy smokers, the median age was 58 y (range 50–84), 739 (71%) were men, average tobacco consumption was 26 cigarettes daily for 37 y (median pack/years = 40), and 14% were former smokers.

The following clinical parameters were evaluated: age, sex, pack/years index, forced expiratory ventilation in 1 s (FEV1%), CT year, pathological stage of detected cancers, histology, size, growth rate, standard uptake value (SUV) of PET. The χ^2 test was used to examine the associations between predictor variables. Overall survival (OS) curves of lung cancer patients were estimated with the Kaplan–Meier method and compared with the log-rank test, using time from lung cancer onset until death or by censoring at the last follow-up date. Statistical analyses were carried out using SAS (SAS Institute Inc., Cary, NC) and R (www.r-project.org, last accessed February 8, 2010) software. Two-sided P values < 0.05 were considered statistically significant.

The second trial was a prospective randomized trial named Multicentric Italian Lung Detection trial (MILD) launched in 2005 (MILD trial, validation set). Current or former smokers, at least 50 years old and without history of cancer within the prior 5 y, were randomized in two study groups: a control group undergoing a program of primary prevention with pulmonary function test evaluation and an early-detection group where periodic spiral-CT was associated with primary prevention and pulmonary function test evaluation. The early-detection group was further randomized in two arms: yearly low-dose spiral CT vs. spiral CT every 2 y. A total of 2,352 subjects were randomized in one of the two CT screening arms.

During enrollment and annual recall of all volunteers in both trials, whole blood was collected in EDTA vacuum tubes and plasma immediately separated by two centrifugation steps at 1,258 relative centrifugal force $\times g$ at 4 °C and stored in a biological bank, supported by a database recording all clinical and epidemiological information. Tissue samples from lung tumors and matching normal lung tissue (sampled at distance from the cancer lesion) were also collected when available from patients undergoing surgical resections. Tissue and plasma specimens were obtained according to the Internal Review and the Ethics Boards of the Istituto Nazionale Tumori of Milan.

miRNA Microarray Analysis in Tissue Samples. For expression analyses, we first used a set of 28 snap-frozen spiral-CT detected lung primary tumors and 24 paired normal lung tissues, collected during the INT/IEO trial. miRNA labeling and hybridization was performed using 5 μg of total TRIzol (Invitrogen) extracted RNA. The miRNA microarray (Ohio State University Comprehensive Cancer Center, version 2.0) used contained probes for 460 mature miRNAs spotted in quadruplicate (235 *Homo sapiens*, 222 *Mus musculus*, and three *Arabidopsis thaliana*) with annotated active sites selected for oligonucleotide design. Hybridization signals were detected with streptavidin–Alexa-647 conjugate, and scanned images (Perkin-Elmer ScanArray XL5K Scanner) were quantified using the GeneSpring software version 7.2 (Silicon Genetics, Redwood City, CA).

Statistical and Bioinformatics Analyses on Tissue Samples. On the microarray chips, after background subtraction and data transformation (to convert any negative value to 0.01), the average value of the four spots was normalized using a per-chip 50th percentile method that normalizes each chip on its median.

Class Comparison and Class Prediction Analyses. Statistical analyses were performed using BRB ArrayTools 3.8.1 software developed by Dr. Richard Simon at the National Cancer Institute (http://linus.nci.nih.gov/BRB-ArrayTools.html). MicroRNA differentially expressed between two classes were considered significant at the nominal 0.001–0.003 level of the univariate test based on 10,000 random permutations and were used for class prediction analyses with the multiple methods tool.

miRNA Profiling in Plasma Samples. miRNA expression profiling was performed in 40 plasma samples, collected 12–28 mo before and at time of the disease detection, from 19 patients in the training set and in 34 plasma samples from 22 patients from the validation set. Using mirVana PARISKit (Ambion), total RNA was extracted from 200-μl plasma samples, and miRNA expression was determined using the Megaplex Pools Protocol on microfluidic card type A (Applied Biosystems). The control groups were represented by 15 pools of 5–7 plasma samples each from disease-free individuals enrolled in the same trials and matched to the patients by sex, age, and smoking habit. For each microfluidic card (sample), the Ct of every miRNA was determined using the program SDS 2.2.2 (Applied Biosystems) and setting a threshold of 0.2 and a manual baseline from 3 to 18 cycles.

Quantitative Real-Time PCR. ***Tissues.*** Starting from 20 ng of total RNA in the reverse transcription (RT) step, TaqMan MicroRNA Assays (Applied Biosystems) were used for quantitative real-time PCR following their standard procedures. Relative quantification was performed using the $\Delta\Delta Ct$ method using as housekeeping the miRNA RNU-6B.

Plasma samples. Starting from 3 μl of the same plasma free-circulating RNA used for the Megaplex Pools Protocol (Applied Biosystems), selected miRNAs were validated with the Multiplex Pools Protocol (Applied Biosystems).

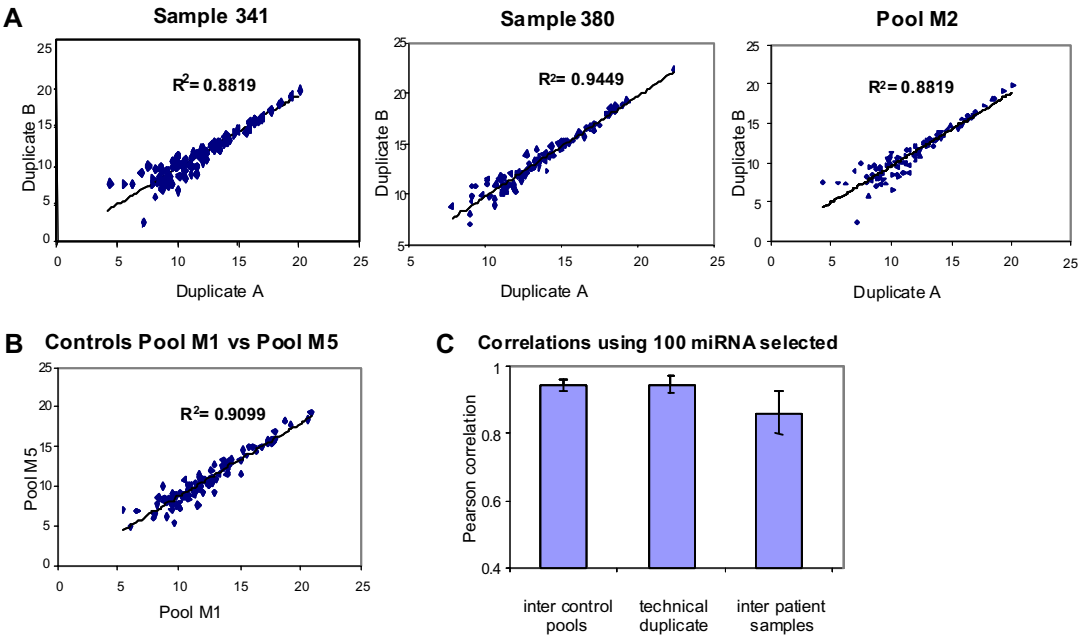

Fig. S1. Consistency of miRNA expression measurement in plasma samples by quantitative real-ttime PCR considering only the 100 miRNAs selected for class comparison analysis. (*A*) Technical duplicates were performed for two patient samples (341 and 380) and for a control pool (M2). The graphical representation was performed plotting the first miRNA values obtained on abscissa (duplicate A) and the values obtained in the second evaluation in ordinate (duplicate B). The linear regression value shows a good reproducibility of measurements. (*B*) Correlation between two different control pools. (*C*) Graphical representation of average values of all Pearson correlation coefficients between control pools, technical duplicates, and between all patient samples (before and at time of disease).

Table S1. Clinical–pathological characteristics of patients with CT-detected tumor in INT-IEO screening trial according to CT year of detection

	CT year		
Characteristic	1–2	3–5	Total
Lung cancer	22	16	38
Resected	21 (95)	12 (75)	33* (87)
Stage I	17 (77)	7 (44)	24 (63)
Stage II–IV	5 (23)	9 (56)	14 (37)
Adeno	17 (95)	10 (63)	27 (71)

The number in parentheses is the percent of all detected lung cancers.
*Twenty-eight tumor tissue and 24 normal lung samples were available for miRNA expression analysis.

Table S2. Top 10 pathways identified in pathway enrichment analysis, using miRNA signatures discriminating tumor vs. normal samples

KEGG pathway ($P < 0.001$)	No. of genes
MAPK signaling pathway	159
Regulation of actin cytoskeleton	133
Focal adhesion	130
Wnt signaling pathway	102
Axon guidance	93
Insulin signaling pathway	92
TGF-beta signaling pathway	69
ErbB signaling pathway	64
Adherens junction	62
Ribosome	3

Table S3. Clinical–pathological characteristics of patients from training and validation sets selected for miRNA expression analysis in plasma samples

Characteristic	Training set, Trial INT-IEO, *n* = 19	Validation set, MILD trial, *n* = 22
Sex		
Male	12 (63.2%)	16 (72.7%)
Female	7 (36.8%)	6 (27.3%)
Age (y)	57.5 ± 5.6 (SD)	61.9 ± 7 (SD)
Smoking habit (pack/years index)	60.3 ± 23.8 (SD)	55 ± 21 (SD)
Screening year of disease detection		
First year	1 (5.3%)	
Second year	7 (36.8%)	5 (22.7%)
Third to fifth year	11 (57.9%)	14 (63.6%)
Interval cancers		3 (13.7%)
Histotype		
Adenocarcinoma (ADC)	14 (73.7%)	14 (63.6%)
Squamous cell carcinoma (SCC)	3 (15.8%)	4 (18.2%)
Other	2 (10.5%)	4 (18.2%)
Stage		
Ia–Ib	12 (63.2%)	15 (68.2%)
II–III–IV	7 (36.8%)	7 (31.8%)
Median follow-up (mo)	67*	14 (min = 4, max = 46)
Prognosis		
Disease free	11 (57.9%)	16 (72.7%)
Alive with disease		1 (4.6%)
Dead	8 (42.1%)	5[†] (22.7%)
Control pools[‡]	5	10

Numbers in parentheses represent the percent of all selected cases.

*An outlier has a follow-up of 35 mo.

[†]A subject died of clinical complications.

[‡]Disease-free individuals pooled by sex, age and smoking habit to best match with patient's characteristics. Each pool contained 5-7 samples.