EBioMedicine 2 (2015) 831–840

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

Original Article

Gene Signatures Stratify Computed Tomography Screening Detected Lung Cancer in High-Risk Populations☆

Jiangting Hu [a,*], Mattia Boeri [b], Gabriella Sozzi [b], Dongxia Liu [a], Alfonso Marchianò [d,e], Luca Roz [b], Giuseppe Pelosi [f], Kevin Gatter [a], Ugo Pastorino [c], Francesco Pezzella [a]

[a] Nuffield Division of Clinical Laboratory Sciences, Radcliffe Department of Medicine, University of Oxford, John Radcliffe Hospital, United Kingdom
[b] Tumor Genomics Unit, Milan, Italy
[c] Division of Thoracic Surgery, Milan, Italy
[d] Division of Radiology, Milan, Italy
[e] Medical Statistics and Bioinformatics Unit, Milan, Italy
[f] Pathology Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy

ARTICLE INFO

Article history:
Received 16 April 2015
Received in revised form 29 June 2015
Accepted 1 July 2015
Available online 8 July 2015

Keywords:
Gene signature
CT screening
Lung cancer
Indolent
Aggressive
PI3K/PTEN/AKT signalling pathway

ABSTRACT

Background: Although screening programmes of smokers have detected resectable early lung cancers more frequently than expected, their efficacy in reducing mortality remains debatable. To elucidate the biological features of computed tomography (CT) screening detected lung cancer, we examined the mRNA signatures on tumours according to the year of detection, stage and survival.
Methods: Gene expression profiles were analysed on 28 patients (INT–IEO training cohort) and 24 patients of Multicentre Italian Lung Detection (MILD validation cohort). The gene signatures generated from the training set were validated on the MILD set and a public deposited DNA microarray data set (GSE11969). Expression of selected genes and proteins was validated by real-time RT-PCR and immunohistochemistry. Enriched core pathway and pathway networks were explored by GeneSpring GX10.
Findings: A 239-gene signature was identified according to the year of tumour detection in the training INT–IEO set and correlated with the patients' outcomes. These signatures divided the MILD patients into two distinct survival groups independently of tumour stage, size, histopathological type and screening year. The signatures can also predict survival in the clinically detected cancers (GSE11969). Pathway analyses revealed tumours detected in later years enrichment of the PI3K/PTEN/AKT pathway, with up-regulation of PDPK1, ITGB1 and down-regulation of FOXO1A. Analysis of normal lung tissue from INT–IEO cohort produced signatures distinguishing patients with early from late detected tumours.
Interpretation: The distinct pattern of "indolent" and "aggressive" tumour exists in CT-screening detected lung cancer according to the gene expression profiles. The early development of an aggressive phenotype may account for the lack of mortality reduction by screening observed in some cohorts.

1. Introduction

Lung cancer screening detects early cancers in a higher number than expected but nonetheless there is still no consensus on both efficacy in reducing mortality and safety (Melamed et al., 1984; Anon, 2014; Marcus et al., 2006; Bach et al., 2003).

Three European randomized CT screening clinical trials have so far failed to achieve a mortality reduction (Infante et al., 2009; Saghir et al., 2012; Pastorino et al., 2012).This is apparently in contrast to the results from two larger America based studies: the International ELCAP group published a positive report on the efficacy of CT screening, with an estimated 10-year survival of 80% overall and 92% in clinical stage I cancers undergoing surgery (Henschke et al., 2006)and the National Lung Screening Trial reported a 20% reduction in mortality in the LDCT group compared to chest X-rays group(The National Lung Screening Trial Research Team, 2011).

To explain these apparently contrasting findings, it has been raised a hypothesis that early-stage tumours found at baseline screening are mostly indolent tumours and removing them does little to reduce development of fatal, fast growing cancers which develop and are picked

☆ Funding: Italian Association for Cancer Research (AIRC) and Special program "Innovative Tools for Cancer Risk Assessment and early Diagnosis", Italian Ministry of Health (RF- 2010).
* Corresponding author at: Nuffield Division of Clinical Laboratory Sciences, Radcliffe Department of Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, United Kingdom.
E-mail address: jianting.hu@ndcls.ox.ac.uk (J. Hu).

http://dx.doi.org/10.1016/j.ebiom.2015.07.001

up later on in the screening programme (Bach, 2008; Pastorino, 2006). The evidences so far suggested that most advanced cancers do not reflect slow evolution of indolent carcinomas, but instead are fast-growing carcinomas with a de novo aggressive phenotype.

Remarkably, the year of detection has been associated with the clinical outcomes, as the tumours detected during the first two years had a better prognosis than those identified in later years of screening. Boeri et al. analysed microRNA expression of tissue and plasma of early detected lung cancer and found that specific signatures could distinguish tumours by the year of detection and predict their clinical outcome (Boeri et al., 2011; Sozzi et al., 2014).

In order to gain further insights on the biological features of CT screening detected tumours, in the present study we analysed the gene expression profile on two sets of spiral CT screening detected tumours: the training set from the pilot trial (INT–IEO) (Pastorino et al., 2003) and the validation set from the prospective randomized Multicentric Italian Lung Detection trial (MILD) (Bianchi et al., 2004). We compared the differential gene expression according to the year of the cancer detection on the training set then we validated the gene signatures on the validation set and on an independent public deposit data set (Takeuchi et al., 2006).

2. Methods

2.1. Sample

28 lung cancers and 23 non-adjacent normal lung tissues from INT–IEO cohort plus 24 tumours identified in the MILD trial were selected (Supplementary Table 1). Recruitment and diagnostic imaging workup have been previously described (Pastorino et al., 2003; Bianchi et al., 2004). Tumours detected during the first two years of screening are defined as CT1–2 those detected between the 3rd and the 5th year as CT3–5. For association analyses the following clinical parameters were considered: CT year, pathological stage and histopathology type. The χ^2 test was used to examine the associations between predictor variables. Further details were in Supplementary material and methods.

2.2. RNA Extraction, Sample Selection Criteria and mRNA Amplification

Total RNA was extracted using TRIzol (Invitrogen) and residual DNA removed by RNeasy (Qiagen). RNA quality was checked by Agilent bioanalyzer (Agilent Inc). The concentration was determined using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies). An RNA sample was further processed only if: (1) the ratio of A_{260}/A_{280} between 1.7–2.0; (2) concentration within the range of 0.5–10 μg/ml; (3) displaying two distinct peaks corresponding to the 28S and 18S ribosomal RNA bands at the ratio of 28S/18S > 0.5 with no degradation. One μg of total RNA was amplified with Amino Allyl MessageAmp™ aRNA Kit (Ambion, Austin, Texas) and indirectly labelled with Cy3 monofunctional dye (Amersham Biosciences UK Ltd., Bucks, United Kingdom) for the sample RNA and Cy5 monofunctional dye for the reference RNA (Stratagene®, Amsterdam, The Netherlands) and then co-hybridised onto the microarray.

2.3. cDNA Microarray Analysis

The Human Exonic Evidence-Based Oligonucleotide (HEEBO) array containing 44,544 of 70-mer probes (Stanford Functional Genomics Facility, Stanford University) was used. Microarray hybridization and processing were performed according to Stanford protocols (www.microarray.org/sfgf). After 20 h of incubation in a 42 °C hybridization oven, the microarray slides were washed with series of SSC and SDS and immediately scanned with a GenePix 4000B microarray scanner (Axon Instruments Inc., Union City, CA). The image QC (flag) set up as: SNR532 > 3, SNR635 > 3, mean of median background < 500, median PC > B + 1SD > 90%, feature variation <1, background variation <1, and feature with saturated pixels <1%. Data was then background subtracted and normalised by the globe intensity correction factor normalisation (LOWESS). The data were then imported into GeneSpring™ GX10 software (Agilent Inc, California) followed by log transformation and LOWESS normalisation. Quality control was performed considering values between 20th and 100th percentile. The data has been through the completed MIAME guideline checklist and is being deposited with Gene expression Omnibus (GEO accession no. GSE29827).

2.4. Statistical Analysis

28 INT–IEO and 24 MILD samples were analysed as training and validation set respectively. In the training set, the gene expression comparison between CT1–2 (n = 17)vs. CT3–5 (n = 11) and stage I (n = 19) vs. stages II–IV (n = 9) were carried out by parametric permutative (permutation times >1000) *t*-test (detailed description in supplementary method) to generate the significance threshold necessary for the recognition of genes differentially expressed in the two groups ($p < 0.01$). We then combined the cases according to the CT year and stage. 14 of them were CT1–2 at stage I, while six were CT3–5 at stages II–IV. The 14 of CT1–2/stage I and the 6 of CT3–5/stages II–IV were each compared to the respective normal samples, using the same statistical method with a number of permutations set >1000 and $p < 0.01$. Unsupervised hierarchical clustering by Pearson's distance measure on average linkage was followed to assess the distribution of each patient based on their similarities measured over the significantly expressed genes by the supervised filtered method.

Overall survival time for lung cancer patients was calculated from the diagnosis of the disease until death or by censoring at the last follow-up date. Survival curves were estimated using the product-limit method of Kaplan–Meier and were compared using the Log-rank test. Statistical analyses were carried out using SAS® (SAS Institute Inc., Cary, NC, USA) and R (URL: http://www.r-project.org/, last access Feb 8th 2010) software. Two-sided p values below 0.05 were considered statistically significant.

2.5. Biological Interpretation of the Microarray Data

To elucidate the gene signature as a pathway and its possible biological function, we employed an on-line method, DAVID bioinformatics resources (National Institute of Allergy and Infectious Diseases, NIH), a public Database for Annotation, Visualization and Integrated Discovery (http://david.abcc.ncifcrf.gov/) (Huang Da et al., 2009).

For pathway analysis, gene list from each comparison was imputed to GSGX10 which provides two approaches: enrichment analysis and network analysis. The former uses BioPAX (Biology Pathway Exchange) format that allows import pathway data from KEGG, Reactome. The latter uses the network database of biological associations extracted from up-to-date scientific literature (NLP) to construct the overall network and interaction such as activation, inhibition and binding to each other, and to identify the most enriched significant pathways in a given gene set.

2.6. Validation of Microarray Data

The 24-patient MILD cohort was kept as blind when doing microarray analysis. Validation of the signatures generated on the training set was performed on this data set as well as on an independent public dataset of clinically detected lung cancer (Takeuchi et al., 2006) (GEO accession number: GSE11969). Common features between different platforms (for the GSE11969 data set) were used for clustering analysis with centred correlation and complete linkage. Distribution of patients according to the signature were compared to clinical–pathological characteristics: year of screening (MILD set only), status, stage and histotype with a 2 × 2 contingency table and two-tailed Fisher's exact

test. Kaplan–Meier survival plots with Log-rank test (HR and 95% CI) were used to compare survival. Results were considered significant at p-value ≤ 0.05

2.7. TaqMan Real-Time Quantitative PCR

The microarray levels of expression of four selected genes ITGB1 (Hs01127543_m1), FOXO1A (Hs01054576_m1), SERPINA3 (Hs01038298_m1) and SELP (Hs00927900_m1) were validated using TaqMan qRT-PCR. From each samples, 0.5 μg of total RNA was converted to cDNA by the RetroScript kit (Applied Biosystems, Foster City, CA) using a random decamer as the primer in a 20-μl reaction according to the manufacturer's protocol. cDNA were then diluted 1:25 and five μl were used for qRT-PCR in a total volume of 20 μl containing 1× TaqMan gene expression master mix and one selected primer. Results of triplicate assays were log-transformed and mean expression values calculated. Relative expression for each gene was assessed based on real-time PCR data normalised to the control gene ACTB (Hs99999903_m1) by the ΔCt method. Relative fold-change was calculated by $2^{-\Delta\Delta Ct}$ and compared with log-transformed microarray data.

2.8. Immunohistochemistry

Paraffin-embedded, formalin-fixed tissues were sectioned (5-mm) onto glass slides. A monoclonal antibody against FOXO1A (Millipore clone 2H8.2) was used and immunostains and scoring (percentage of positive cells and intensity of staining) were performed as previously described (Leek et al., 2000).

3. Results

3.1. Comparison Between Patients Detected in the First Two (CT1–2) and the Last Three (CT3–5) Years of Screening

239 genes were differentially expressed between 17 CT1–2 and 11 CT3–5 tumours from the INT–IEO training set (Table 1 for selected genes and Supplementary Table 2 for the full list), 110 genes were up-regulated and 129 down-regulated in CT1–2 compared to CT3–5 by parametric permutative (permutation times > 1000) *t*-test ($p < 0.01$). 153 genes were differentially expressed in 19 stage I compared to 9 stages II–IV patients (selected genes in Table 2 and full list in Supplementary Table 3). Combined comparison of 14 CT1–2/stage I vs. 6 CT3–5/stages II–IV patients identified 218 differentially expressed genes (selected genes in Table 3 and full list in Supplementary Table 4). Tumours of CT3–5 or stages II–IV or combined CT3–5/stages II–IV had higher expression of genes associated with metastasis and high cell mobility.

Unsupervised hierarchical clustering of the 239-gene showed the separation of CT1–2 and CT3–5 (Fig. 1). Similarly the method applied to 153 genes distinguished stage I from stages II–IV and 218 genes of CT1–2/stage I vs. CT years3–5/stage II–IV tumours respectively (Supplementary Figs. 1 and 2).

We then analysed gene expression profiles in the normal lung from the same cohort using the same permutative *t*-test: 203 genes were differentially expressed between these histologically normal tissues of CT1–2 ($n = 15$) and CT3–5 ($n = 8$) patients ($p < 0.01$, Supplementary Table 5). Unsupervised hierarchical clustering separated the normal tissues according to the year of tumour detection (Supplementary Fig. 3).

We also compared 17 tumours with 15 normal tissues of CT1–2 and 11 tumours with 8 normal lung tissues of CT3–5 using the similar parametric permutative *t*-test ($p < 0.01$);a larger amount of differentially expressed genes were identified respectively (Supplementary Tables 6 &7).

GO pathway analysis was performed by importing the differentially expressed genes lists to the DAVID online database. Table 4 shows the top annotation clusters in the three lists of 239, 153 and 218 genes of tumour comparison according to the different criteria plus the 203 genes from normal tissue comparison (full lists of annotation in Supplementary Tables 8–11).

We also listed functional annotation clustering from the genes in the tumours vs. same period normal in Supplementary Tables 12 and 13. By focusing on a significant enrichment score ≥ 1, analysing the genes differentially expressed between tumour and normal of CT1–2, differences in cell adhesion, cell growth, ribosome activity, cell motility and regulation of kinase activity were identified. On the other hand, the same analysis on tumour vs. normal CT3–5, it is shown that different biological

Table 1
Selected genes of 239 differentially expressed by parametric permutative *t*-test between tumours detected in years 1–2 and in years 3–5 involved in metastases and tumour aggressiveness.

Gene ID	Fold	p Value	Direction	Biological functions
PCDHGB3	4.78	0.001	Up years 1–2	Neural cadherin-like adhesion proteins likely to play a role in the establishment of cell–cell connections in the brain
CLDN16	3.89	0.003	Up years 1–2	Tight junction. Associated with less aggressive, reduced tumour volume and lack of metastases in breast.
PLAC1	1.38	0.008	Up years 1–2	Associated with proliferation, motility, migration and invasion.
FOXO1A	1.4	0.009	Up years 1–2	Forkhead box O1A (rhabdomyosarcoma)
ITGB1	0.79	0.008	Down years 1–2	Controls invasion via regulation of MMP-2 collagenase expression through PI-3K, Akt, and Erk protein kinases and the c-Jun or via direct recruitment of MMP-2 to the cell surface
MFI2	0.77	0.008	Down years 1–2	Melanoma progression and metastases.
PECAM1	0.76	0.002	Down years 1–2	Endothelium. Associated with metastases.
PAPPA2	0.68	0.01	Down years 1–2	Metalloproteinase. Detected in some invasive extravillous trophoblasts in the first trimester
POSTN	0.63	0.008	Down years 1–2	Associated with aggressive metastatic tumours.
PTP4A3	0.61	0.006	Down years 1–2	Associated with cell proliferation and metastases
F8	0.59	0.004	Down years 1–2	Associated with cell proliferation and metastases
SERPINA3	0.58	0.009	Down years 1–2	Associated with cell proliferation and metastases
CCL25	0.57	0.002	Down years 1–2	Associated with metastatic melanoma.
OPHN1	0.53	0.007	Down years 1–2	Rho-GTPase-activating protein that promotes GTP hydrolysis of Rho subfamily members. Promotes cell migration.
AAMP	0.47	0.009	Down years 1–2	Heparin binding. Expressed strongly in metastatic colon adenocarcinoma cells in lymphatics.
CLCA2	0.47	0.01	Down years 1–2	It may serve as adhesion molecule for lung metastatic cancer cells, mediating vascular arrest and colonization.
MIZF	0.43	0.004	Down years 1–2	Transcription repressor. Possibly associated with invasiveness.
CTSB	0.34	0.002	Down yrs 1–2	Associated with metastases.
RLN2	0.32	0.005	Down yrs 1–2	Increase Cyclic AMP. Gs-adenylate cyclase and b-catenin pathway. Increases cell invasion and proliferation.
SELP	0.32	0.002	Down years 1–2	Associated with metastases.
MYO7B	0.3	0.001	Down years 1–2	Cell motility
SFSCN2	0.21	0.001	Down years 1–2	Associated with metastases.
CTSL1	0.18	0.001	Down years 1–2	Associated with metastases. Induced by hypoxia.
FZD5	0.18	0.002	Down years 1–2	Canonical WNT pathway. Lung oncogenesis, increases cell migration. Involves b-catenin.

Table 2
Selected genes of 153 genes differentially expressed by parametric permutative *t*-test between tumours stage 1 and stages 2–4 involving in metastases and tumour aggressiveness.

Gene ID	Fold	p Value	Direction	Biological functions
SUSD5	3.2	0.001	Up stage 1	Cell adhesion
LAMC3	1.94	0.003	Up stage 1	Stability of basement membranes and of cellular attachments to basement membranes,
VASP	1.39	0.002	Up stage 1	Associated with cell invasiveness
NPEPPS	1.32	0.003	Up stage 1	Associated with proliferation, migration, and invasion
PCLKC	0.74	0.006	Down stage 1	Cell adhesion. Tumour contact inhibition
DDC	0.5	0.009	Down stage 1	Associated with metastatic neuroblastoma.
GNE	0.47	0.01	Down stage 1	Induces sLex. Sialic acid — dependent processes in adhesion, signalling, differentiation, and metastasis.
NEXN	0.45	0.005	Down stage 1	F-actin associated. Induces cell migration and adhesion.
AAMP	0.38	0.001	Down stage 1	Heparin binding. Expressed strongly in metastatic colon adenocarcinoma cells in lymphatics.
CSPG3	0.35	0.003	Down stage 1	Thought to be involved in the modulation of cell adhesion and migration
EBAG9	0.34	0.005	Down stage 1	Associated with metastatic disease

activities are altered, for instance, in regulation of protein modification & metabolic process, protein folding/chaperone and oxidoreductase activity.

We tested the 239-gene signature differentially expressed between CT1–2 and CT3–5 on an independent cohort of 24 spiral-CT detected lung cancer from the MILD trial. Using complete linkage and centred correlation, the signature was able to divide the patients into two distinct groups independent from the tumour status, stage, histotype and the year of screening as shown in Table 5, with the overall reproducibility of R-index = 0.612 and D-index = 6.835 (Supplementary Tables 14). Survival analysis showed that despite a Hazard Ratio (HR) of 4.6 (95%CI 0.9–23.4), there was no significant association (Log-rank test p = 0.06) with the overall survival (Fig. 2a), but it became significant (p = 0.03), with 5.2 HR (95%CI 1.2–23.1), when considering disease free survival (Fig. 2b).

On the other hand, the signatures of stage generated comparing tumour stage I vs. II–IV in the training set, were ineffective in this validation sets (Supplementary Table 15).

We then tested the same signature on a deposited independent NSCLC data set with annotated clinical information (GSE11969) containing the expression profiles of 79 clinically detected lung adenocarcinomas and squamous cell carcinomas (Takeuchi et al., 2006). There were 118 features comparable with the dataset and able to divide the patients into two distinct survival groups independently from tumour stage and histotype (p = 0.013, Table 5). Moreover, the Kaplan–Meier curves showed clear differences in overall survival (Log-rank test p = 0.02) with 2.1 of HR (95%CI 1.1–3.9) as shown in Fig. 3a. When the analysis was restricted to 40 stage I tumours, the features also discriminated this early stage tumour with distinct clinical outcomes (p = 0.03 and HR = 2.9, 95%CI 1.1–7.8), suggesting that also clinically detected stage I tumours are a heterogeneous category comprising indolent and aggressive tumours (Fig. 3b). However, the stage generated signatures from the training set were ineffective in GSE11969 validation sets (Supplementary Table 15).

3.2. Building up Pathway-Based Gene Signatures

3.2.1. Pathway Network Model 1: Tumour of CT1–2 vs. CT3–5 and stage 1 vs. stages 2–4

We performed network modelling using GeneSpring GX10 on the 239-gene signature of differentially expressed by CT1–2 vs CT3–5. The results showed that the tumours detected in later years have an increased expression of the genes ITGB1, SELP, PECAM1, F8, SERPINA3 with loss of FOXO1A as hub nodes in the transcriptome regulatory network (Fig. 4). Further pathway enrichment analysis (Table 6) revealed in functional terms that these data would predict the aggressive tumours form later years more angiogenesis (ITGB1) and metastases (ITGB1, PECAM1, SELP), increased proliferation and diminished apoptosis following loss of FOXP1 transcription but the activation of the of PI3K/PTEN/AKT pathway(Fig. 5).

When analysing 153 genes of stage I vs. stages II–IV tumours, it is shown that SAPK/JNK and Alzheimer pathways as the most significantly related networks, both regulating AKT phosphorylation, which is known to be involved in the regulation of apoptosis and cell cycle

Table 3
Selected genes of 218 genes differentially expressed by parametric permutative *t*-test between tumours of CT1–2/stage 1 and CT3–5/stages 2–4 involved in metastases and tumour aggressiveness.

Gene ID	Fold	p Value	Direction	Biological functions
S100A2	3.3	0.009	Up yr 1–2/s 1	Down regulated in metastatic Head Neck carcinoma
CD151	2.23	0.001	Up yr 1–2/s 1	Enhances cell motility, invasion and metastasis
MITF	1.89	0.009	Up yr 1–2/s 1	Associated with suppression of metastases.
CDK5	1.62	0.006	Up yr 1–2/s 1	Induces cell migration and apoptosis
TACSTD2	1.52	0.009	Up yr 1–2/s 1	Associated with metastases
PLAC1	1.51	0.01	Up yr 1–2/s 1	Associated with proliferation, motility, migration and invasion.
VASP	1.43	0.006	Up yr 1–2/s 1	Associated with cell invasiveness
ITGB1	0.70	0.003	Down yr 1–2/s 1	Controls invasion via regulation of MMP-2 collagenase expression through PI-3K, Akt, and Erk protein kinases and the c-Jun or via direct recruitment of MMP-2 to the cell surface
PCLKC	0.69	0.003	Down yr 1–2/s 1	Cell adhesion. Tumour contact inhibition
LAMA3	0.63	0.005	Down yr 1–2/s 1	Loss of expression associated with advanced stage.
GPR68	0.62	0.009	Down yr 1–2/s 1	Metastases suppressor gene. Inhibits cell migration.
CCL25	0.52	0.002	Down yr 1–2/s 1	Associated with metastatic melanoma.
ANXA9	0.39	0.006	Down yr 1–2/s 1	Cell–cell adhesion
NEXN	0.38	0.006	Down yr 1–2/s 1	F-actin associated. Induces cell migration and adhesion.
CLCA2	0.35	0.01	Down yr 1–2/s 1	It may serve as adhesion molecule for lung metastatic cancer cells, mediating vascular arrest and colonization.
RLN2	0.32	0.007	Down yr 1–2/s 1	Increase Cyclic AMP. Gs-adenylate cyclase and b-catenin pathway. Increases cell invasion and proliferation.
SELP	0.31	0.001	Down yr 1–2/s 1	Associated with metastases.
EBAG9	0.28	0.005	Down yr 1–2/s 1	Associated with metastatic disease
PDPK1	0.25	0.007	Down yr 1–2/s 1	Cell–matrix adhesion and oncogenesis
AAMP	0.24	0.001	Down yr 1–2/s 1	Heparin binding. Expressed strongly in metastatic colon adenocarcinoma cells in lymphatics.

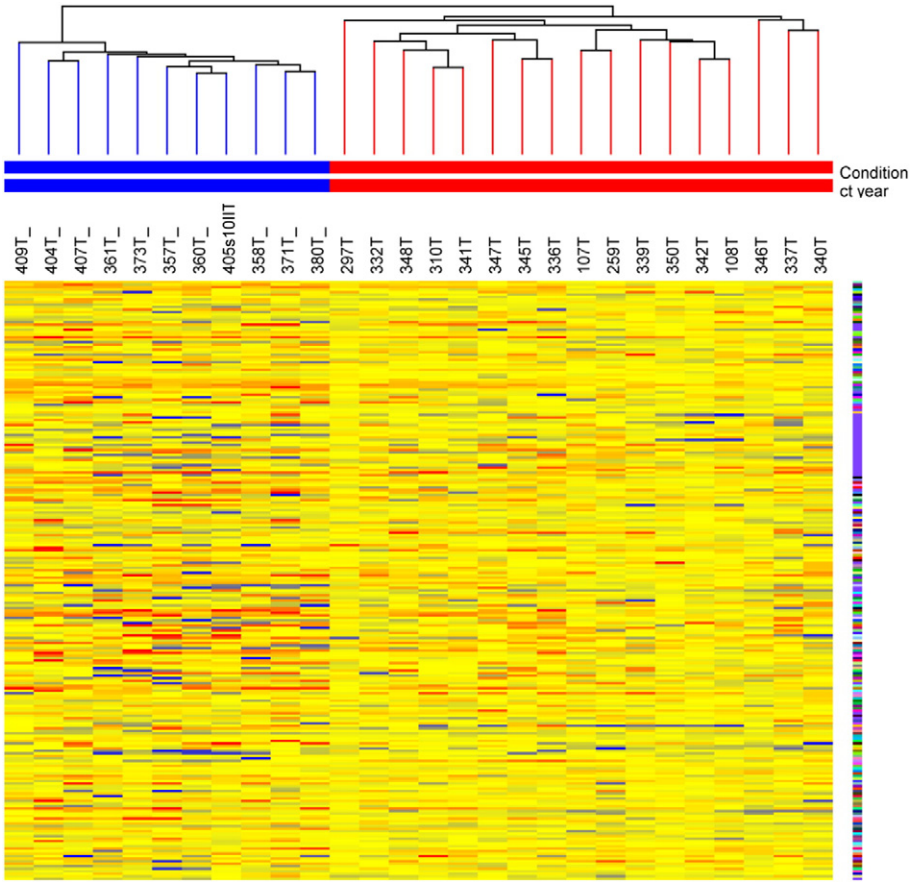
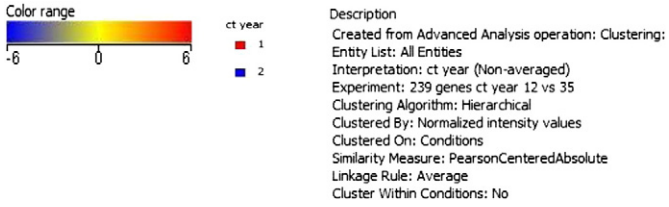


Fig. 1. Unsupervised hierarchical clustering of 17 tumours detected in years 1 and 2 (T) and 11 cases detected in years 3, 4 and 5 (T*) using 239 differentially expressed genes.



activity, mostly through ITGB1. Enrichment pathways on the genes differentially expressed between CT1–2/stage I and CT3–5/stages II–IV showed similar patterns (Table 6).

3.2.2. Pathway Network Model 2: Pathway Alteration Events in Normal Tissues

Similar pathway analysis approaches were applied to the gene list of 203 genes of normal tissues comparison according to the tumour detection time. The direct interaction of 203 genes revealed the direct connection of PSG1, PIK3R4, GRP, PSEN1, CEBPB and HIST1H2BK (supplementary Fig. 5). Further biological process analysis revealed that this network mainly functions in activation of MAPK activity, regulation of angiogenesis, keratinocyte proliferation, chromatin remodelling and assembly (supplementary Fig. 6).

3.3. Validation of Selected Genes

TaqMan QRT-PCR analysis of ITGB1, SERPINA3, FOXO1A and SELP expression in 28 cases showed similar expression patterns to those found in microarray analysis (supplementary Fig. 7).

3.4. Frequency and Expression Pattern of FOXO1A

Immunohistochemistry demonstrated nuclear expression of FOXO1A in 81% (14/17) of the INT–IEO CT1–2 and in 60% (6/10) of the CT3–5 cases and cytoplasmic staining in 76% (13/17) and 50% (5/10) tumours respectively (supplementary Fig. 8).

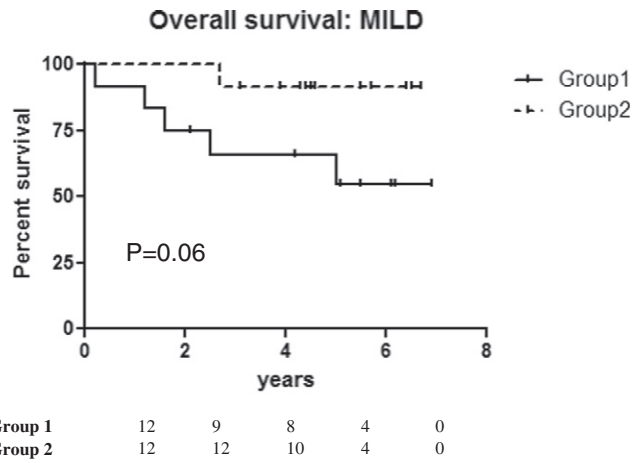
4. Discussion

Fifteen years ago, we launched a prospective early-detection trial with spiral CT, positron emission tomography and molecular markers in a cohort of 1035 heavy smokers (INT–IEO set). A second prospective randomized trial, MILD, including 4099 participants was launched in 2005. A study to compare the efficacy and cost-effectiveness of low-dose spiral CT screening in four published randomized trials concluded that there was no overall mortality difference in the CT arms compared with the control arms (Pastorino et al., 2012). The molecular findings including gene expression in the INT–IEO set and micoRNA signatures in both INT–IEO and MILD sets were previously reported (Boeri et al., 2011; Bianchi et al., 2004). Based on those findings, we proposed that the lung cancer natural history in early detection by CT screening can

Table 4
The top alteration in 4 genes lists of tumour & normal comparisons.

From tumour of CT1–2 vs. CT3–5				
Top annotation group	Annotation cluster terms	Enrichment score	Count %	FDR
1	Cysteine-type endopeptidase activity	1.82	6	0.0089
2	Endopeptidase activity	1.78	7	0.001
3	Glycosylation	1.51	42	0.0089
4	G-protein coupled receptor	0.89	11	0.009
5	Cell adhesion	0.87	10	0.0089
From tumour of stage I vs. stages II–IV				
Top annotation group	Annotation cluster terms	Enrichment score	Count %	FDR
1	Extracellular space	1.52	9	0.001
2	Jak-STAT signalling pathway	1.51	4	0.0064
3 Cell adhesion		0.92	7	0.001
From tumour of CT1–2/stage I vs. CT3–5/stages II–IV				
Top annotation group	Annotation cluster terms	Enrichment score	Count %	FDR
1	Cell motion	2.45	12	0.0077
2	Regulation of cell motion	1.55	12	0.0083
3	Cell adhesion	1.54	13	0.0092
4	Cell-substrate adherens junction	1.37	6	0.0066
5	Sodium ion binding	1.05	6	0.001
From normal of CT1–2 vs. CT3–5				
Top annotation group	Annotation cluster terms	Enrichment score	Count %	FDR
1	Chromatin assembly	1.36	5	0.098
2	DNA binding, high mobility group	1.30	3	0.0023
3	Regulation of cell morphogenesis involved in differentiation	1.29	4	0.0096
4	Endoplasmic reticulum	1.17	12	0.0067
5	RAS GTPase mediated signal transduction	0.97	12	0.0044

a) The overall survival according to the 239-gene signature of CT-year of screening



b) The disease free survival according to the 239-gene signature of CT-year of screening

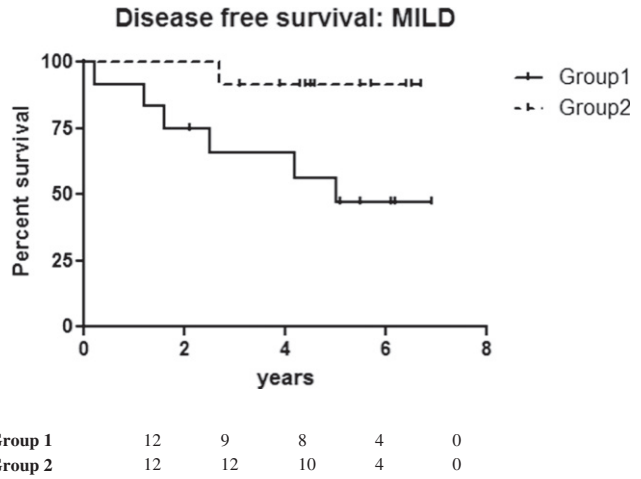


Fig. 2. Survival analysis of 24 MILD patients grouped according to the signature of CT-year of screening. These survival curves are based on (a) overall survival analysis and (b) disease free survival analysis. The (two-sided) *p* value is from by Log-rank (Mantel–Cox) test.

be stratified by their molecular signatures. Indeed, we find here that there are two clinically distinct types of early-detected lung cancer: "indolent" and "aggressive".

It is widely accepted that lung cancer progresses from pre-neoplastic to clinically detected diseases by accrual of genetic and epigenetic alterations, becoming metastatic in a later phase (Goldstraw et al., 2007).

Table 5
Patients' distribution of the MILD trial and the public GEO dataset GSE11969 by 239-gene signature.

	MILD*			GSE11969#		
	24 patients			79 patients		
	Group 1	Group 2	p-Value#	Group 1	Group 2	p-Value#
Alive	7	11	0.15	16	21	0.01
Dead	5	1		30	12	
Stage I	7	10	0.37	24	16	1.00
Stages II–IV	5	2		22	17	
ADK	6	7	1.00	29	16	0.25##
Other	6	5		17	17	
CT1–2	8	6	1.00			
CT3–5	4	6				

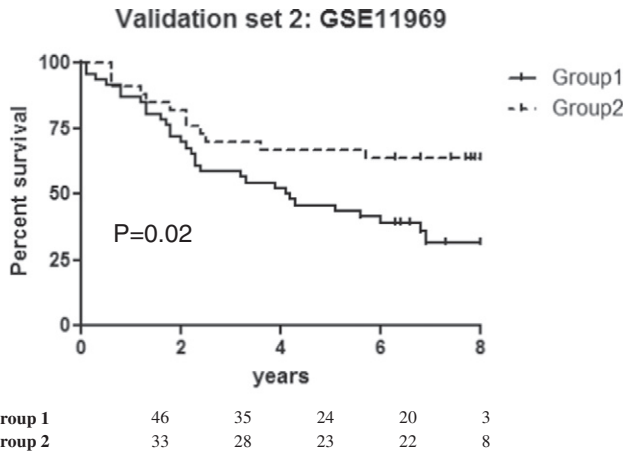
Clustering experiment using centred correlation, complete linkage and cutting dendrograms at 2 clusters.
* 231/239 features in common.
\# 118/239 features in common.
Two-tailed Fisher's exact test.

Strategies to reduce mortality have focused on early diagnosis to eradicate lesions before metastases occur. While the ability of these strategies to reduce overall mortality remains debatable, there is agreement that a higher number of tumours than predicted were found during screening programmes, with some tumours having a poor outcome while others behave in an indolent manner (Bach, 2008). A similar scenario is known to occur in other types of tumours e.g., non-Hodgkin's lymphomas which can develop as either indolent or de novo aggressive, with some indolent cases transforming into high-grade malignancies after several years (Horning and Rosenberg, 1984).

Boeri and colleagues analysed the role of microRNAs as biomarkers identifying a signature able to distinguish the indolent from the aggressive tumours (Boeri et al., 2011). In this study we have further demonstrated by mRNA profiling, that considerable differences between indolent and aggressive early detected tumours suggested that the latter accumulate unexpectedly fatal genetic/epigenetic aberrations over a rather short period of time.

Both gene functional annotation and topological pathway network analyses indicated a significant overrepresentation of genes associated

a) The overall survival in all clinically detected NSCLC of validation set GSE11969 (N=79)



b) The overall survival in only stage I NSCLC of validation set GSE11969 (N=40)

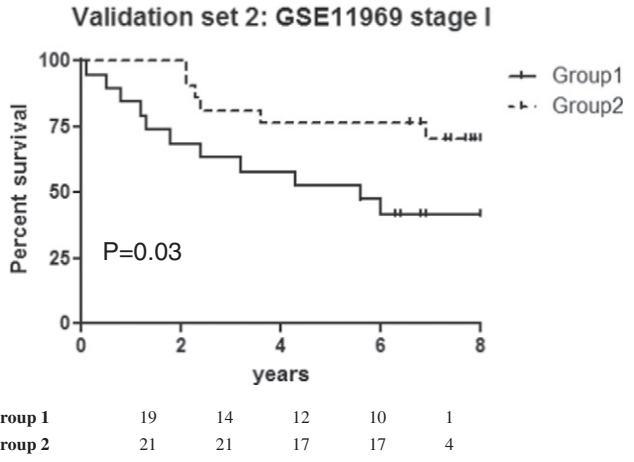


Fig. 3. Overall survival analysis of in silico data (GSE11969) considering (a) all the 79 patients or (b) the 40 stage I alone, grouped according to the signature of CT-year of screening. The (two-sided) *p* value is from by Log-rank (Mantel–Cox) test.

with peptidase activity, response to wound healing, cell adhesion, signal transducer activity and, ultimately, metastases in the aggressive early detected tumours. Further pathway enrichment analysis, whatever the classification criteria applied (year of detection, stage or both), reflected the activation of the PI3K/PDPK1/ITGB1/AKT pathway involving ITGB1 and FOXO1A in the most aggressive tumours (Testa and Bellacosa, 2001). The increase in ITGB1 levels is consistent with its known association with metastatic ability (Akiyama et al., 1995; Basson, 2008; Ritzenthaler et al., 2008). The findings overall are also consistent with the overexpression of mir-128a, which is predicted to target the 3′-untranslated region of FOXO1A eventually regulating AKT signalling (Sozzi et al., 2014), in the later year aggressive tumours of the same cohort (Boeri et al., 2011). Interestingly, mir-128 was also found overexpressed in endometrial cancer with concomitant repression of FOXO1A expression (Myatt et al., 2010). Thus, profiling studies on the INT/IEO training cohort by two approaches (miRNA and mRNA expression) in two different laboratories (Milan and Oxford) revealed FOXO1A/AKT pathway differential expression. Balsara et al. (2004) reported that phosphorylation of FOXO1A and AKT correlates in "in situ" lung lesions, possibly leading to invasiveness, while Maekawa et al. (2009) showed that the loss of expression of total FOXO1A is associated with advanced stage tumours. The association between phosphorylation or loss of FOXO1A and more aggressive disease has been also reported in prostate (Li et al., 2007), colorectal (Bravou et al., 2006) breast adenocarcinoma (Li et al., 2009) and acute myeloid leukaemia (Cheong et al., 2003). Note that PDPK1/AKT1 and the ITGB1 pathway plus FOXO1A has recently been identified as targets for treatment in light of their association with increased proliferation, metastasis and decreased apoptosis (Cen et al., 2007; Bloom and Calabro, 2009; Fang et al., 2008).

We found that not only the tumours, but also the normal lung tissue from patients identified in the first two years differ from the normal tissue of patients identified in the last three years, consistent with findings in microRNA expression analysis (Sozzi et al., 2014) and supporting the notion that the tumour aggressiveness is likely conditioned by the underlying "field cancerization" (Lochhead et al., 2015). These findings suggest the possibility of grouping patients as low or high risk by gene profiling the normal tissue.

We validated our signature on two independent validation sets: early detected tumours from the MILD trial and a cohort of patients with clinically detected lung cancer for which microarray data was available. In the MILD trial, the 239-gene signature from the training set was able to distinguish the patients by their outcome independent of tumour stage, histotype and year of screening, indicating that this signature is specific for discriminating between indolent and aggressive tumours. This finding was confirmed in a second validation set of a cohort of clinically detected lung cancers. The signature was not only effective in predicting the outcome when applied to all patients but was able to identify those patients who will survive more than 5 years and those who survived less than 2 years within stage I. This suggests that, as in the screening trial, stage 1 clinically detected tumours are also a mixture of already highly aggressive and indolent diseases.

This study conveys the translational significance in two aspects: firstly, CT screening detected early lung cancer is a pool of heterogeneous early tumours, not only from the histological type point of view, but a mixture of aggressive and indolent tumours. The gene signatures provided the molecular clue of disease outcomes independent from the cancer histopathology type. Our next step is to identify smaller, manageable signatures for personalised diagnostic purpose. It would be of value to be able to distinguish stage I tumours with aggressive characters which need targeted treatment to prevent metastatic relapse; secondly, the confirmed findings of "field cancerization" suggests the possibility to divide patients at low or high risk categories by gene profiling the normal mucosa, therefore to revise the protocol for recruiting the normal participants entering the screening trial.

In conclusion, we provide molecular evidence that both screening and early stage clinically detected non-small cell lung cancers can be divided into "indolent" and "aggressive" tumours suggesting that the metastatic phenotype appears much earlier than previously thought. The 239-gene signature may serve as risk stratification biomarkers to distinguish these clinically different tumours for personalised management, although it cannot be excluded that some indolent cancers will eventually become aggressive. Due to the limitation of this study, larger studies are needed to confirm that this categorization based on mRNA and miRNA expression signatures from screening detected tumour also fits with clinically detected cases, particularly in stage I lung cancer. Since these two groups of patients could be identified by non-invasive tools, such as PET/SUV (Pastorino et al., 2009) or circulating biomarkers, new prospective trials to improve the treatment of lung cancer may be conceived on this basis.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ebiom.2015.07.001.

Author Contributions

J.H., G.S., A.M. designed the study. U.P. initiated this work. J.H., G.S., D.L., and M.B. performed the research and obtained data. J.H., G.S.,

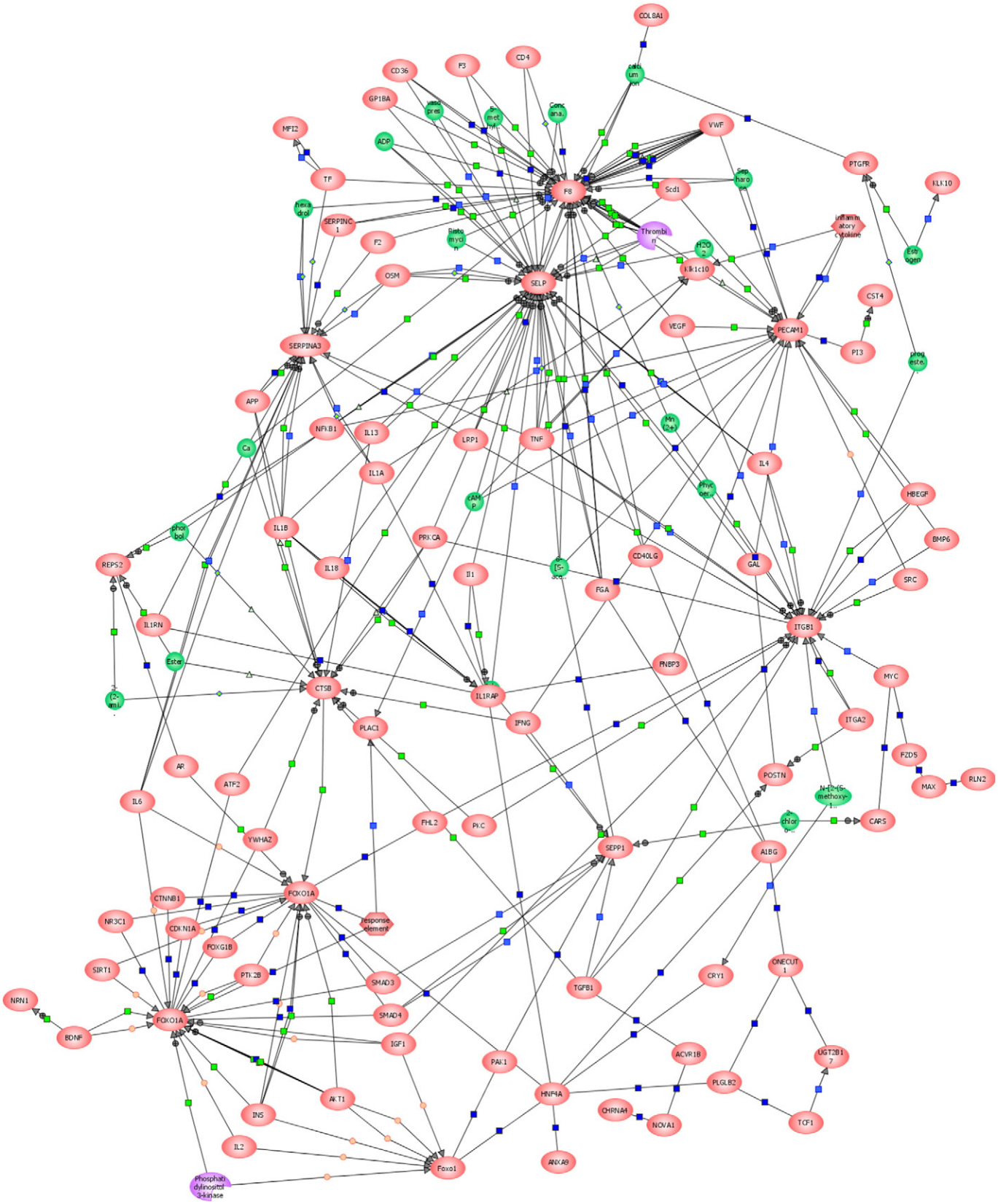


Fig. 4. The 239-gene of differentially expressed genes according to the CT screening year were imported into the GeneSpring GX10 for searching the common regulators of these genes. The connection between these genes was built up and unlinked nodes (genes) were removed. Blue lines and squares signify that a defined regulatory relationship exits between genes. Grey lines and squares signify that a putative regulatory relationship between genes has been identified but not biochemically defined. +, positive regulation; −, negative regulation.

Table 6
Pathway enrichment analysis.

239 genes tumours detected in years 1–2 vs years 3–5				
	Object identifier	Common object	Size	Name
0	1625598	1	73	Long-term potentiation
1	1625611	2	82	ERK-PI3K (collagen) signalling
2	1625626	2	86	Integrin signalling
3	1625629	2	44	PTEN signalling
4	1625630	2	63	AKT signalling
5	1625632	2	32	ACH-R apoptosis signalling
6	1625637	1	144	Apoptosis
7	1625640	1	24	Interferon-alpha signalling
153 genes stage 1 vs stages 2–4				
	Object identifier	Common object	Size	Name
0	1625597	1	130	SAPK–JNK signalling
1	1625605	1	72	Alzheimer's disease
218 genes combined ct/stage t1 vs t2				
	Object identifier	Common object	Size	Name
0	1625598	1	73	Long-term potentiation
1	1625605	1	72	Alzheimer's disease
2	1625611	1	82	ERK–PI3K (collagen) signalling
3	1625626	1	86	Integrin signalling
4	1625629	2	44	PTEN signalling
5	1625630	1	63	AKT signalling
6	1625632	1	32	ACH-R apoptosis signalling
7	1625633	1	94	Wnt signalling (Calcium)

M.B., A.M., F.P., L.R., and G.P. analysed and interpreted the results. J.H., F.P., K.G., U.P. wrote the paper. All authors read, gave comments and approved the final version of the manuscript. All authors had full access to the data in the study and take responsibility for the integrity of the results.

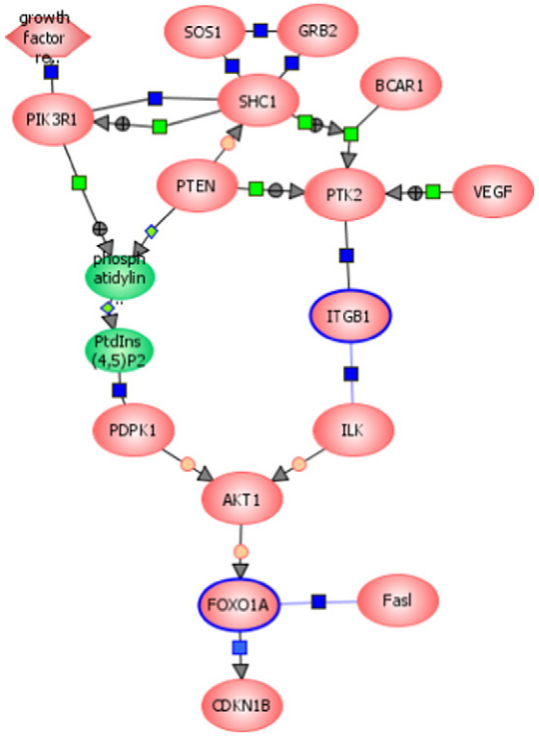


Fig. 5. Pathway enrichment analysis reveals PI3K/PTEN/AKT signalling and apoptosis pathway as the most significantly related pathway. ITGB1 has higher level of expression in the late year tumours while FOXO1A has higher levels in the tumours detected at first two years.

Declaration

The authors have no conflict of interests related to this article.

Acknowledgements

The authors thank Prof. Massimo Bellomi, MD, Dr. Elvio De Fiori, MD, and Dr. Cristiano Rampinelli, MD, from the Division of Radiology of the European Institute of Oncology (IEO), for the revision of spiral CT data and Dr. Xu Yu Jin, John Radcliffe Hospital, Oxford, for helping data analysis and critically reading the manuscript and Dr. Danielle Fletcher from Agilent for the statistical advices.

This work was supported by grants from Italian Association for Cancer Research (AIRC): IG research Grant No.14318, No.11991; the Special program "Innovative Tools for Cancer Risk Assessment and early Diagnosis", 5 × 1000, No. 12162; and the Italian Ministry of Health (RF-2010). The funding agency has no role in the actual experimental design, analysis, or writing of this manuscript.

References

Akiyama, S.K., Olden, K., Yamada, K.M., 1995. Fibronectin and integrins in invasion and metastasis. Cancer Metastasis Rev. 14 (3), 173–189 (Epub 1995/09/01).
Anon., 2014. Screening for lung cancer. Too many uncertainties, even for smokers. Prescrire Int. 23 (145), 19–23 (Jan).
Bach, P.B., 2008. Is our natural-history model of lung cancer wrong? Lancet Oncol. 9 (7), 693–697 (Epub 2008/07/05).
Bach, P.B., Kelley, M.J., Tate, R.C., McCrory, D.C., 2003. Screening for lung cancer: a review of the current literature. Chest 123 (1 Suppl.), 72S–82S.
Balsara, B.R., Pei, J., Mitsuuchi, Y., et al., 2004. Frequent activation of AKT in non-small cell lung carcinomas and preneoplastic bronchial lesions. Carcinogenesis 25 (11), 2053–2059 (Epub 2004/07/09).
Basson, M.D., 2008. An intracellular signal pathway that regulates cancer cell adhesion in response to extracellular forces. Cancer Res. 68 (1), 2–4 (Epub 2008/01/04).
Bianchi, F., Hu, J., Pelosi, G., et al., 2004. Lung cancers detected by screening with spiral computed tomography have a malignant phenotype when analyzed by cDNA microarray. Clin. Cancer Res. 10, 6023–6028.
Bloom, L., Calabro, V., 2009. FN3: a new protein scaffold reaches the clinic. Drug Discov. Today 14 (19–20), 949–955 (Epub 2009/07/07).
Boeri, M., Verri, C., Conte, D., Roz, L., et al., 2011. MicroRNA signatures in tissues and plasma predict development and prognosis of computed tomography detected lung cancer. Proc. Natl. Acad. Sci. U. S. A. 108 (9), 3713–3718 (Epub 2011/02/09).
Bravou, V., Klironomos, G., Papadaki, E., et al., 2006. ILK over-expression in human colon cancer progression correlates with activation of beta-catenin, down-regulation of E-cadherin and activation of the Akt-FKHR pathway. J. Pathol. 208 (1), 91–99 (Epub 2005/11/10).
Cen, L., Hsieh, F.C., Lin, H.J., et al., 2007. PDK-1/AKT pathway as a novel therapeutic target in rhabdomyosarcoma cells using OSU-03012 compound. Br. J. Cancer 97 (6), 785–791 (Epub 2007/09/13).
Cheong, J.W., Eom, J.I., Maeng, H.Y., et al., 2003. Constitutive phosphorylation of FKHR transcription factor as a prognostic variable in acute myeloid leukemia. Leuk. Res. 27 (12), 1159–1162 (Epub 2003/08/19).
Fang, W., Li, X., Jiang, Q., et al., 2008. Transcriptional patterns, biomarkers and pathways characterizing nasopharyngeal carcinoma of Southern China. J. Transl. Med. 6, 32.
Goldstraw, P., Crowley, J., Chansky, K., et al., 2007. The IASLC lung cancer staging project: proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM classification of malignant tumours. J. Thorac. Oncol. 2 (8), 706–714 (Epub 2007/09/01).
Henschke, C.I., Yankelevitz, D.F., Libby, D.M., et al., 2006. Survival of patients with stage I lung cancer detected on CT screening. N. Engl. J. Med. 355 (17), 1763–1771 (Epub 2006/10/27).
Horning, S.J., Rosenberg, S.A., 1984. The natural history of initially untreated low-grade non-Hodgkin's lymphomas. N. Engl. J. Med. 311, 1471–1475.
Huang Da, W., Sherman, B.T., Lempicki, R.A., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4 (1), 44–57 (Epub 2009/01/10).
Infante, M., Cavuto, S., Lutman, F.R., et al., 2009. A randomized study of lung cancer screening with spiral computed tomography: three-year results from the DANTE trial. Am. J. Respir. Crit. Care Med. 180 (5), 445–453 (Epub 2009/06/13).
Leek, R.D., Hunt, N.C., Landers, R.J., et al., 2000. Macrophage infiltration is associated with VEGF and EGFR expression in breast cancer. J. Pathol. 190 (4), 430–436.
Li, R., Erdamar, S., Dai, H., et al., 2007. Forkhead protein FKHR and its phosphorylated form p-FKHR in human prostate cancer. Hum. Pathol. 38 (10), 1501–1507 (Epub 2007/06/29).

Li, J., Yang, L., Song, L., et al., 2009. Astrocyte elevated gene-1 is a proliferation promoter in breast cancer via suppressing transcriptional factor FOXO1. Oncogene 28 (36), 3188–3196 (Epub 2009/07/28).
Lochhead, P., Chan, A., Nishihara, R., et al., 2015. Etiologic field effect: reappraisal of the field effect concept in cancer predisposition and progression. Mod. Pathol. 28, 14–29.
Maekawa, T., Maniwa, Y., Doi, T., et al., 2009. Expression and localization of FOXO1 in non-small cell lung cancer. Oncol. Rep. 22 (1), 57–64 (Epub 2009/06/11).
Marcus, P.M., Bergstralh, E.J., Zweig, M.H., et al., 2006. Extended lung cancer incidence follow-up in the Mayo Lung Project and overdiagnosis. J. Natl. Cancer Inst. 98 (11), 748–756 (Epub 2006/06/08).
Melamed, M.R., Flehinger, B.J., Zaman, M.B., et al., 1984. Screening for early lung cancer. Results of the Memorial Sloan–Kettering study in New York. Chest 86 (1), 44–53 (Epub 1984/07/01).
Myatt, S.S., Wang, J., Monteiro, L.J., et al., 2010. Definition of microRNAs that repress expression of the tumor suppressor gene FOXO1 in endometrial cancer. Cancer Res. 70 (1), 367–377 (Epub 2009/12/24).
Pastorino, U., 2006. Early detection of lung cancer. Respiration 73 (1), 5–13 (Epub 2006/02/25).
Pastorino, U., Bellomi, M., Landoni, C., et al., 2003. Early lung-cancer detection with spiral CT and positron emission tomography in heavy smokers: 2-year results. Lancet 362 (9384), 593–597.
Pastorino, U., Landoni, C., Marchiano, A., et al., 2009. Fluorodeoxyglucose uptake measured by positron emission tomography and standardized uptake value predicts long-term survival of CT screening detected lung cancer in heavy smokers. J. Thorac. Oncol. 4 (11), 1352–1356 (Epub 2009/10/29).
Pastorino, U., Rossi, M., Rosato, V., et al., 2012. Annual or biennial CT screening versus observation in heavy smokers: 5-year results of the MILD trial. Eur. J. Cancer Prev. 21 (3), 308–315 (May).
Ritzenthaler, J.D., Han, S., Roman, J., 2008. Stimulation of lung carcinoma cell growth by fibronectin–integrin signalling. Mol. BioSyst. 4 (12), 1160–1169 (Epub 2009/04/28).
Saghir, Z., Dirksen, A., Ashraf, H., et al., 2012. CT screening for lung cancer brings forward early disease. The randomised Danish lung cancer screening trial: status after five annual screening rounds with low-dose CT. Thorax 67, 296e301.
Sozzi, G., Boeri, M., Rossi, M., et al., 2014. Clinical utility of a plasma-based miRNA signature classifier within computed tomography lung cancer screening: a correlative MILD trial study. J. Clin. Oncol. 32, 768–773.
Takeuchi, T., Tomida, S., Yatabe, Y., et al., 2006. Expression profile-defined classification of lung adenocarcinoma shows close relationship with underlying major genetic changes and clinicopathologic behaviors. J. Clin. Oncol. 24 (11), 1679–1688 (Epub 2006/03/22).
Testa, J.R., Bellacosa, A., 2001. AKT plays a central role in tumorigenesis. Proc. Natl. Acad. Sci. U. S. A. 98 (20), 10983–10985 (Epub 2001/09/27).
The National Lung Screening Trial Research Team, 2011. Reduced lung-cancer mortality with low-dose computed tomographic screening. N. Engl. J. Med. 365, 395–409.