

RESEARCH

Open Access

Characterization of a genetic mouse model of lung cancer: a promise to identify Non-Small Cell Lung Cancer therapeutic targets and biomarkers

Federica Riccardo¹, Maddalena Arigoni¹, Genny Buson², Elisa Zago², Manuela Iezzi³, Dario Livio Longo⁴, Matteo Carrara¹, Alessandra Fiore¹, Simona Nuzzo⁵, Silvio Biciato⁵, Patrizia Nanni⁶, Lorena Landuzzi⁷, Federica Cavallo¹, Raffaele Calogero^{1*}, Elena Quaglino^{1*}

From Tenth Annual Meeting of the Italian Society of Bioinformatics (BITS)
Udine, Italy. 21-23 May 2013

Background: Non-small cell lung cancer (NSCLC) accounts for 81% of all cases of lung cancer and they are often fatal because 60% of the patients are diagnosed at an advanced stage. Besides the need for earlier diagnosis, there is a high need for additional effective therapies. In this work, we investigated the feasibility of a lung cancer progression mouse model, mimicking features of human aggressive NSCLC, as biological reservoir for potential therapeutic targets and biomarkers.

Results: We performed RNA-seq profiling on total RNA extracted from lungs of a 30 week-old K-ras^{LA1}/p53^{R172HΔg} and wild type (WT) mice to detect fusion genes and gene/exon-level differential expression associated to the increase of tumor mass. Fusion events were not detected in K-ras^{LA1}/p53^{R172HΔg} tumors. Differential expression at exon-level detected 33 genes with differential exon usage. Among them nine, i.e. those secreted or expressed on the plasma membrane, were used for a meta-analysis of more than 500 NSCLC RNA-seq transcriptomes. None of the genes showed a significant correlation between exon-level expression and disease prognosis. Differential expression at gene-level allowed the identification of 1513 genes with a significant increase in expression associated to tumor mass increase. 74 genes, i.e. those secreted or expressed on the plasma membrane, were used for a meta-analysis of two transcriptomics datasets of human NSCLC samples, encompassing more than 900 samples. SPP1 was the only molecule whose over-expression resulted statistically related to poor outcome regarding both survival and metastasis formation. Two other molecules showed over-expression associated to poor outcome due to metastasis formation: GM-CSF and ADORA3. GM-CSF is a secreted protein, and we confirmed its expression in the supernatant of a cell line derived by a K-ras^{LA1}/p53^{R172HΔg} mouse tumor. ADORA3 is instead involved in the induction of p53-mediated apoptosis in lung cancer cell lines. Since in our model p53 is inactivated, ADORA3 does not negatively affect tumor growth but remains expressed on tumor cells. Thus, it could represent an interesting target for the development of antibody-targeted therapy on a subset of NSCLC, which are p53 null and ADORA3 positive.

Conclusions: Our study provided a complete transcription overview of the K-ras^{LA1}/p53^{R172HΔg} mouse NSCLC model. This approach allowed the detection of ADORA3 as a potential target for antibody-based therapy in p53 mutated tumors.

* Correspondence: raffaele.calogero@unito.it; elena.quaglino@unito.it

¹Molecular Biotechnology Center, University of Torino, 10126 Torino, Italy
Full list of author information is available at the end of the article

Background

Lung cancer is the most common cause of neoplasia-related death worldwide [1]. The vast majority of lung cancer cases (approximately 80%) are non-small cell lung cancers (NSCLC) and the remaining fraction is small cell lung cancers. Only a minority of NSCLC patients is suitable for radical treatment as curative care. Approximately two thirds of patients are diagnosed at an advanced stage, and of the remaining patients who undergo curative surgery, 30-50% have a recurrence with metastatic disease [2]. The 5-year relative survival rate among patients diagnosed with NSCLC is only 15%. Thus, the conventional treatments (i.e. surgery, radiotherapy and chemotherapy), have apparently reached a plateau of effectiveness in improving survival of advanced NSCLC patients [3]. Thus, the treatment of NSCLC is a major unmet need and new therapies focusing on the molecular mechanisms of lung tumorigenesis are urgently needed [4].

The discovery of new biomarkers for targeted therapies could greatly change the management and prognosis of many patients with NSCLC. Further, knowledge of the molecular pathways and mutational drivers of lung cancer will expand the use of targeted treatments. Hopefully, the identification of new therapeutic targets will provide personalized and precise treatments for lung cancer patients in the near future.

Indeed, considerable efforts were made to discover new molecular biomarkers associated to lung cancer, which could be used as early diagnostic markers or as new specific therapeutic targets to treat patients [5-7]. In our opinion, the identification of oncoantigens (i.e. tumor associated antigens that have a causal role in the promotion of tumor progression) [8,9] could provide new and more promising targets for personalized treatment in NSCLC.

In this study, we sought to identify new candidate biomarkers and/or potential oncoantigens involved in both initiation of lung cancer and/or its progression to an aggressive cancer phenotype. To this aim, we adapted to the lung cancer disease our consolidated pipeline for oncoantigen detection [8,10]. Thanks to the RNA-seq technology we also extended our pipeline to the detection of tumor specific transcript isoforms and fusion proteins [11]. Our pipeline requires the availability of an animal model for the cancer under study [8]. Thus, we used one of the models most closely simulating human metastatic lung cancer [12]. This model is based on the combination of a latent mutant K-ras allele at the endogenous locus (K-ras^{LA1}), which is spontaneously activated in vivo [13], and a particular mutation generated at the endogenous p53 allele containing an arginine-to-histidine substitution at codon 172 (p53^{R172HΔg}), corresponding to the hot spot mutation at human codon 175

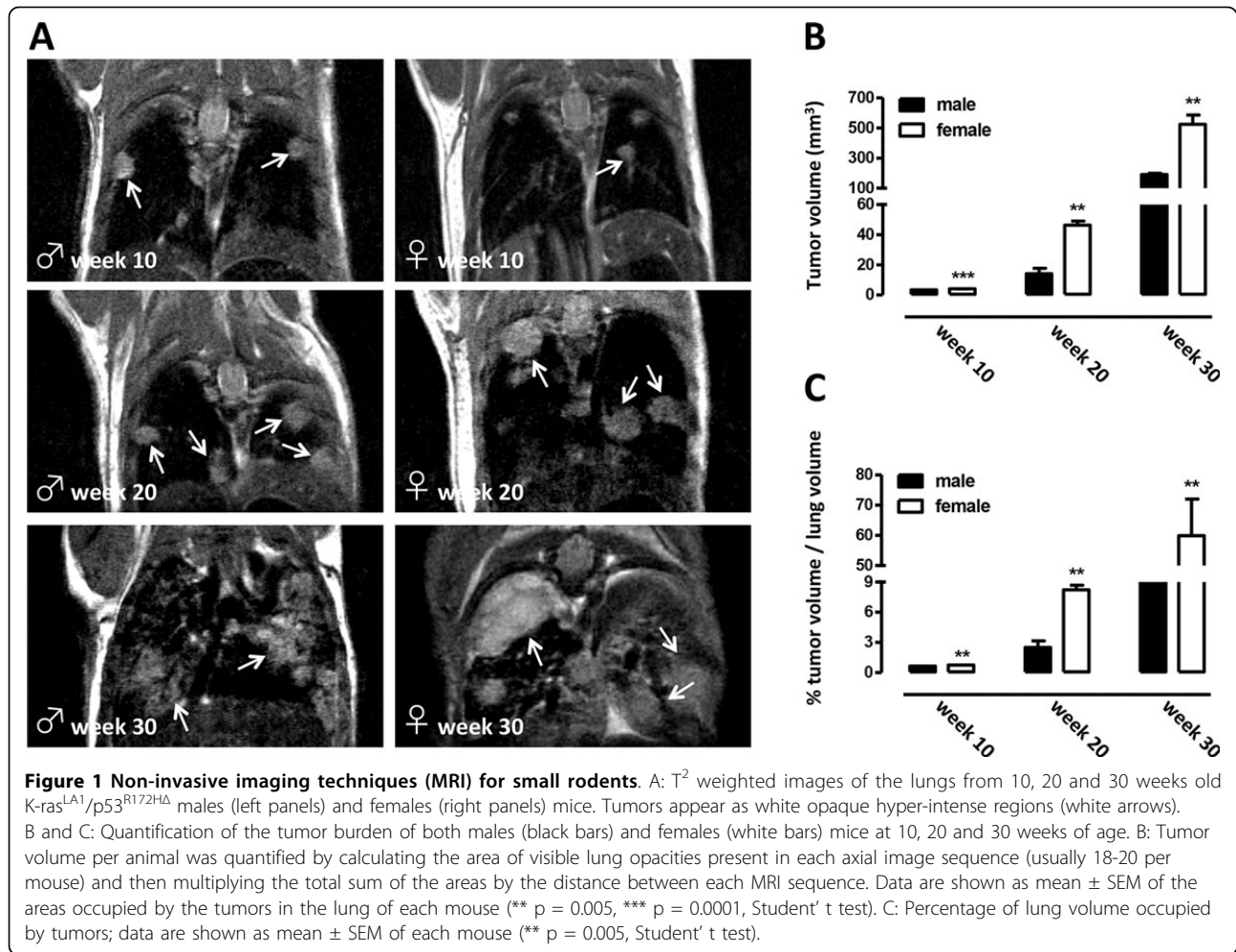
[14-16]. This mouse model develops lung adenocarcinomas with a high incidence of metastases and gender differences in cancer-related death. The use of our pipeline in the framework of metastatic lung cancer model, combined with the power of RNA-seq technology, allowed the identification of ADORA3 as new putative target for antibody-based therapy in mutant p53 tumors.

Results and discussion

Characterization of lung tumors of K-ras^{LA1}/p53^{R172HΔg} mice by non invasive MRI

A colony of K-ras^{LA1}/p53^{R172HΔg} double transgenic mice has been generated in our laboratory, by crossing one p53^{R172HΔg} male with one K-ras^{LA1} female, kindly provided us by Dr. Lozano (University of Texas, M.D. Anderson Cancer Center). These mice develop autochthonous lung adenocarcinomas with a high incidence of metastases and gender differences in cancer related death thus providing a realistic model of human metastatic lung cancer and an immunocompetent system for studying NSCLC and its prevention by novel agents [12]. By using non-invasive imaging techniques (MRI) for small rodents, a quantification of the number and the size of tumor lesions of K-ras^{LA1}/p53^{R172HΔg} mice during time was performed. The progression of lung tumors was monitored at 10, 20 and 30 weeks of age. Tumor lesions resulted as white opaque hyper-intense regions already evident in 10 week-old K-ras^{LA1}/p53^{R172HΔg} male and female mice (Figure 1A). The analysis of images collected at weeks 10, 20, and 30 weeks of age showed a significant increase in the total tumor volume in both K-ras^{LA1}/p53^{R172HΔg} males and females during cancer progression (Figures 1B and 1C). Moreover, starting from the 10th week of age, a significant increase in the number and size of lung lesions was observed between males and females, with females developing more lesions than males, as previously reported for survival [12]. These gender differences remain evident from early to advanced/late-stage of the disease (Figures 1B and 1C).

Histological analysis of lung sections from normal (Figure 2A) and 10 week-old K-ras^{LA1}/p53^{R172HΔg} male and female mice showed that white opacities revealed by the MRI analysis correspond to small foci of lung carcinoma growing with lepidic aspect (Figure 2B). These early lesions increase in number and dimensions and, at 20 weeks of age, become sub-pleural and intra-parenchymal tumors (Figure 2C and 2D, respectively), growing in masses with lepidic and solid growth aspects. Like in humans, in which the prevalence of adenocarcinomas of mixed subtypes led, in 2011, to a new WHO classification in which invasive adenocarcinomas are classified by predominant pattern and to the routinely definition of the percentage of histologic subtypes in clinical pathological



reports, at 30 weeks of age (Figure 2E,2F,2G,2H), lung adenocarcinomas of K-ras^{LA1}/p53^{R172HΔg} mice display, besides a predominance of zones with solid growth (Figure 2E), several types of differentiation, sometimes with prominent papillary growth pattern (Figure 2F), sometimes with less differentiated zones and aspects of large cell carcinoma (Figure 2G). Immunohistochemical analyses showed that these lesions are positive for TTF-1 (Thyroid Transcription Factor-1; Figure 2H), a typical marker of adenocarcinoma [17], and negative for p63, a marker of squamous tumors and for Synaptophysin, Chromogranin, and Neuron Specific Enolase (NSE; data not shown), markers of neuroendocrine tumors [18].

Transcription profiling

Microarray analysis

To estimate the importance of the gender effect on gene expression, we initially run a microarray experiment on lung tissues of 10, 20 and 30 week-old K-ras^{LA1}/p53^{R172HΔg} mice, using Affymetrix exon 1.0 arrays. The comparison did not show any significant difference at

the transcription level (not shown), suggesting that the differences in growth rate might be due to the endocrinological differences existing between male and female. Thus, we run a pair-end RNA-seq on two prototypical situations, WT and K-ras^{LA1}/p53^{R172HΔg} mice (MT), to detect genes/transcripts associated to the increase of tumor mass that might represent potential targets for precision medicine applications [19].

Fusion events detection

Direct sequencing of messenger RNA transcripts using the RNA-seq protocol [20] is rapidly becoming the standard method for detecting and quantifying expressed genes in a cell. One of the key features observed after cancer genomes analysis is a chromosomal abnormality. Genome rearrangements could result in aberrant gene fusions, and a number of them have been found to play important roles in carcinogenesis [21]. The discovery of novel gene fusions can lead to a better comprehension of cancer progression and development. Fusion events were detected in WT and MT samples using ChimeraScan

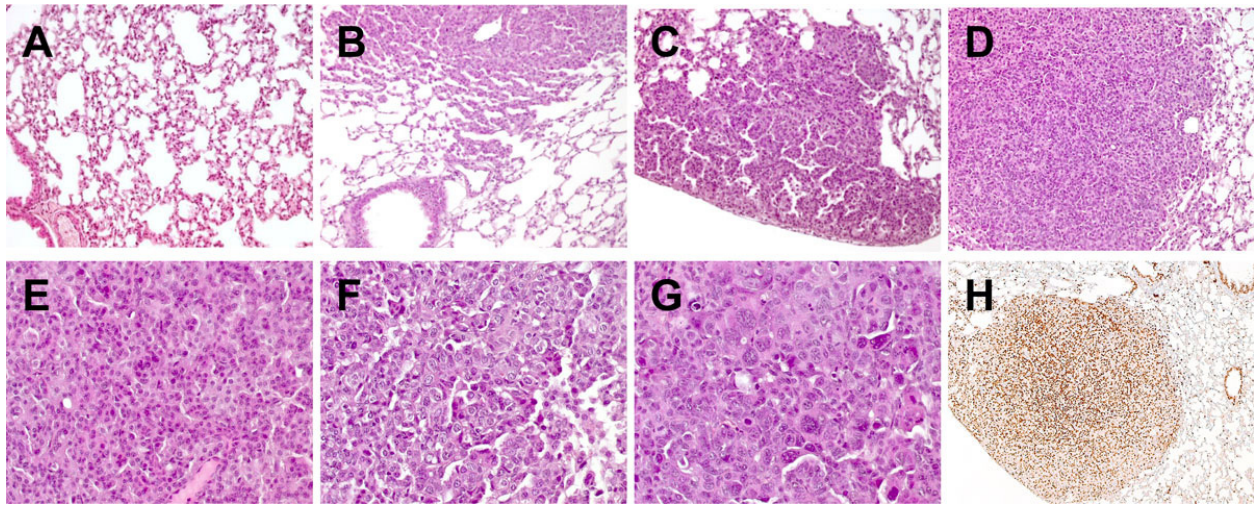


Figure 2 Morphological characterization of lung tumors from $K\text{-ras}^{\text{LA1}}/p53^{\text{R172HA}}$ mice. A-G: Hematoxylin-eosin evaluation of lung sections from a WT transgenic mouse (A), one representative 10- (B), 20- (C) and 30- (D-G) week-old $K\text{-ras}^{\text{LA1}}/p53^{\text{R172HA}}$ mice (A-D magnification $\times 200$; E-G magnification $\times 400$). A: normal lung tissue; B: initial lesions with aspects of lepidic growth; C: subpleural lesion with papillary and solid patterns; D: adenocarcinoma nodule with solid pattern of growth; E: tumor zone with a solid growth pattern composed of cohesive cell agglomerates in a nest-like configuration without acinar polarity; F: tumor zone with papillary growth. Papillae show fibrovascular cores lined by cells with large vesicular nuclei containing very prominent nucleoli; G: poorly differentiated tumor zone with highly polymorphic cells and cells with aberrant nuclei. H: Immunohistochemical staining for TTF-1 lung tumor lesions from one representative 30-week-old $K\text{-ras}^{\text{LA1}}/p53^{\text{R172HA}}$ mouse (magnification $\times 100$).

[22]. Since fusion detection tools are error prone [23], we filtered the putative fusions, reported by ChimeraScan, retaining only common events between the MT and not reported in the WT replicates. The detected fusions (AK029407:Ank3, Gimap1:Gimap5, Pisd-ps2:Pisd-ps) were subsequently discarded since they were all either read through events or fusions between homologue genes. Thus it seems that fusion products are not prominent events in tumors developing due to the presence of constitutively active K-ras and inactive p53.

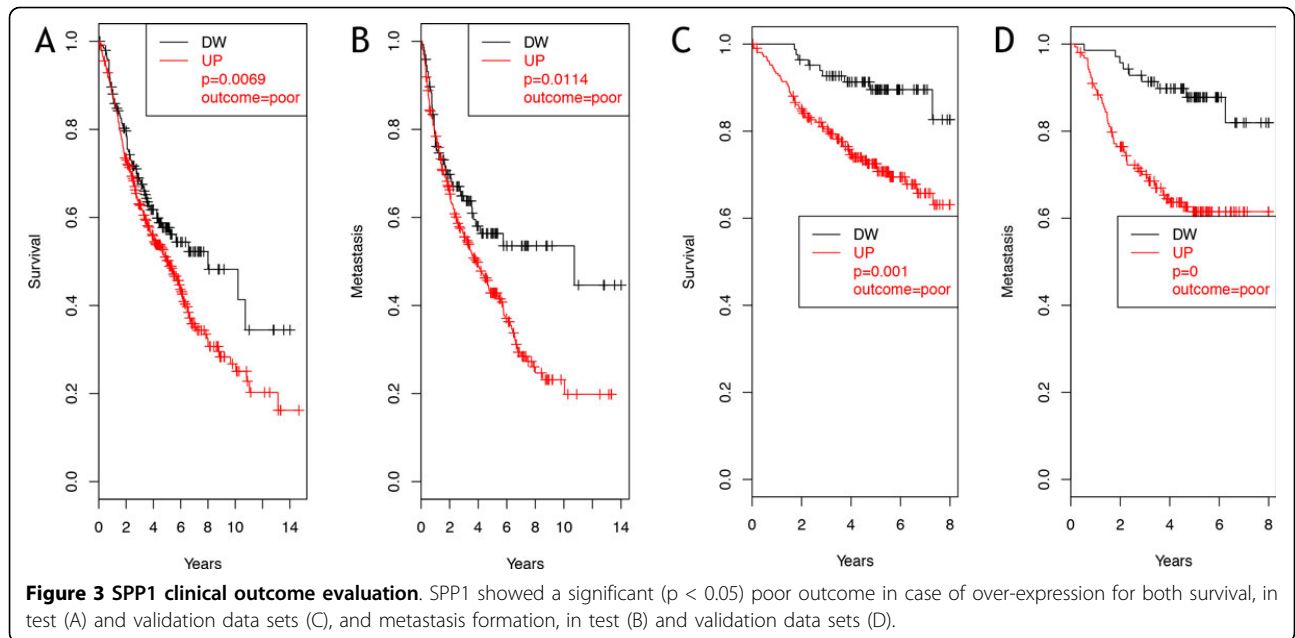
Exon-level analysis

Exon level analysis was run using DEXSeq Bioconductor package [24] and provided 33 genes with differential exon expression between WT and MT groups (FDR < 10%). Among them six (ITGAD, COL17A1, DCSTAMP, PTPRN, PTPRM and Klr1c) codify for proteins that were located on the plasma membrane and three (VWF, DMKN and TIMP3) for proteins secreted in the extracellular space. For 11 of the 33 detected genes, exon-level data for 509 tumors together with their clinical annotation were retrieved from the cancer genome atlas (<http://cancergenome.nih.gov/>). We scored the exons for their oncological power (see methods), which essentially represents the association between exon skipping/retention and poor outcome. Significant correlation between exon-level expression for the above-mentioned genes and poor prognosis could not be detected (not shown).

Gene-level analysis

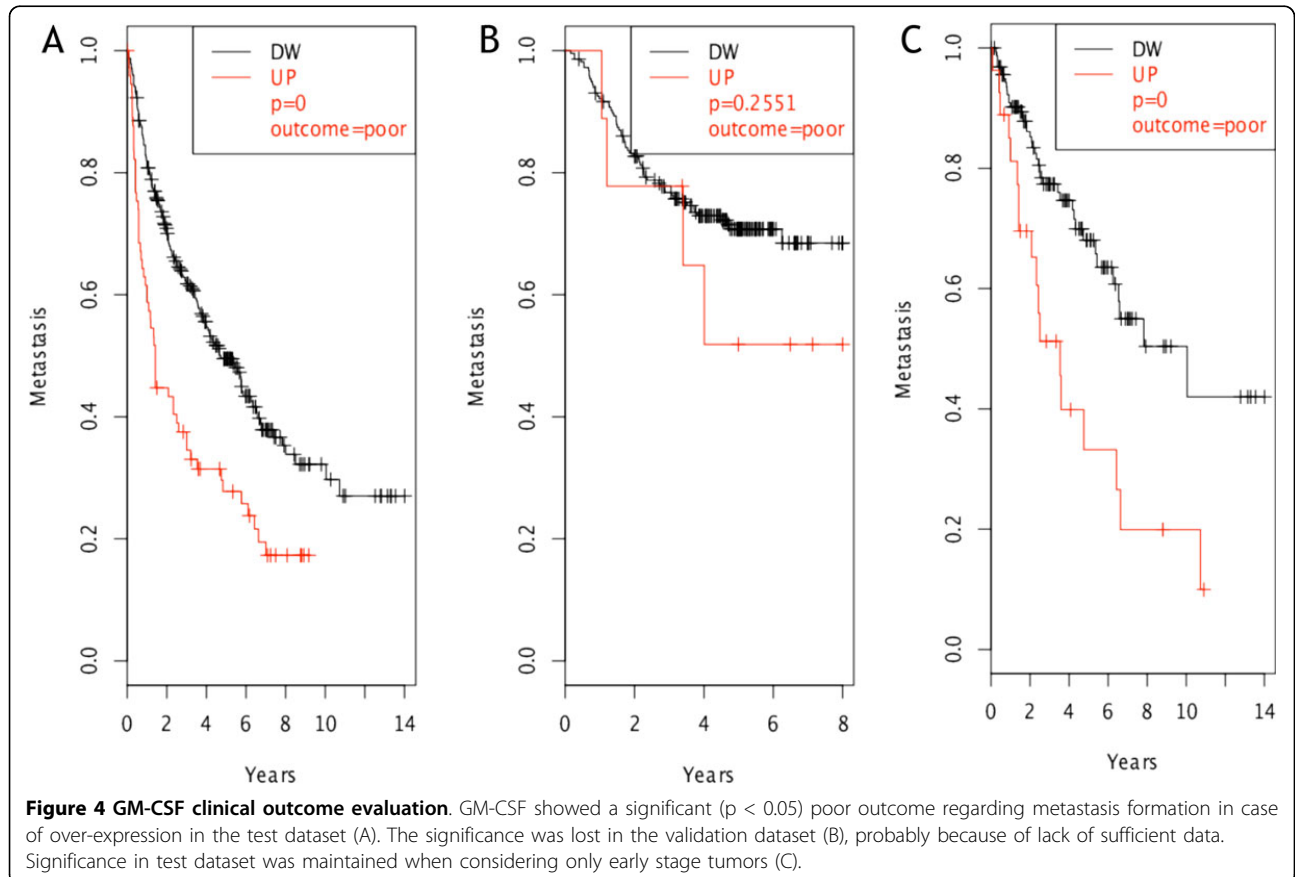
Gene-level analysis was run using DESeq Bioconductor package [25] and provided 1,513 genes with increased expression associated to tumor mass increment between WT and MT groups (FDR < 10%, $|\log_2\text{FC}| > 1$). We focused our analysis on 74 genes encoding for secreted and membrane bound proteins having a human ortholog (74). Thus, we run a meta-analysis on a set of public available transcriptomes of 989 NSCLC patients characterized by clinical outcome for survival and metastasis (see methods). The data set was divided in test and validation set, of 695 and 294 samples each, respectively. We scored the identified genes for their oncological power (CO score, see methods), which represents the association between up-modulation of a gene and poor clinical outcome.

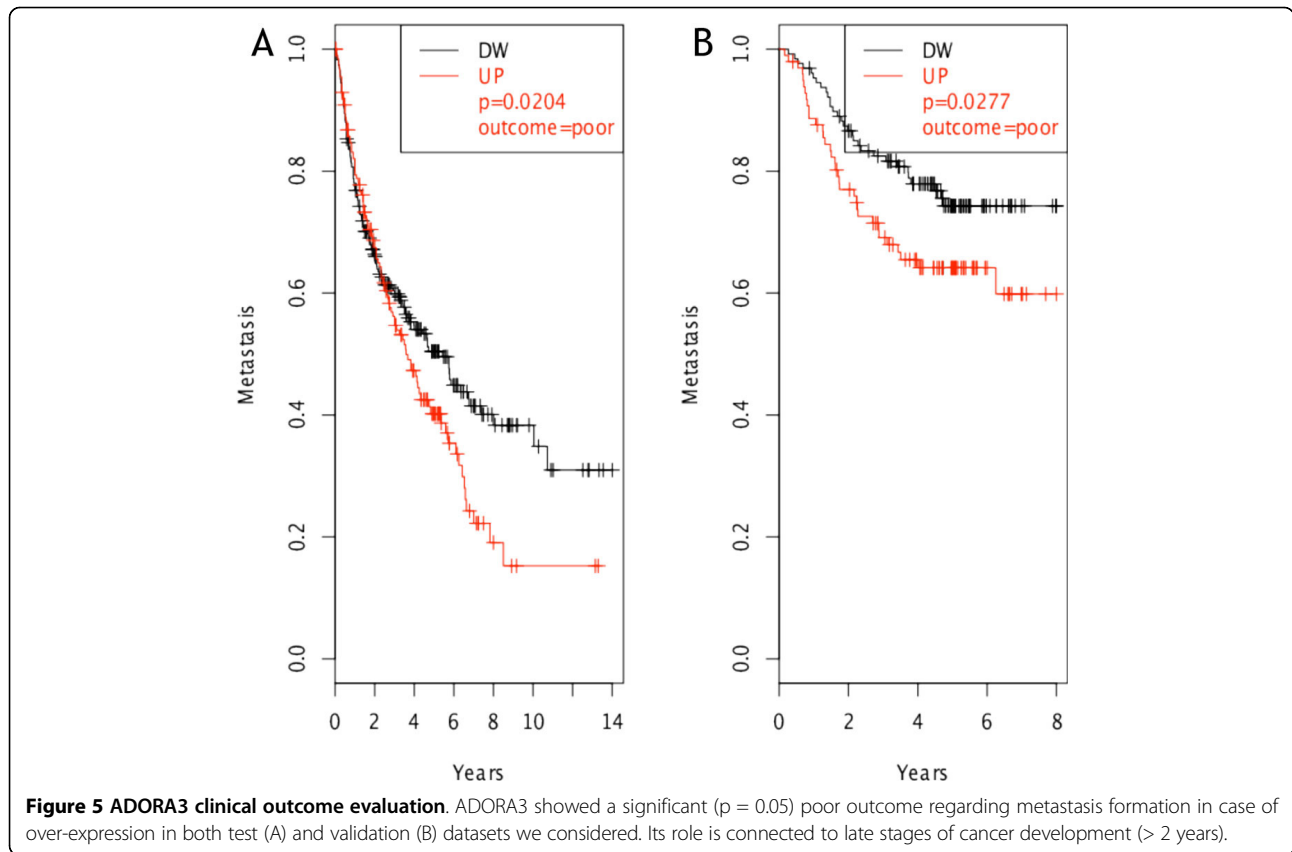
SPP1 (osteopontin) was the only molecule whose over-expression resulted statistically related to poor outcome regarding both survival and metastasis formation in NSCLC patients examined (Figure 3). This result was further maintained in both datasets evaluating only early tumor stage samples, i.e. category T1 based on the TNM staging system [26]. These results are in accordance with previous evidences that SPP1 is an early marker of tumor progression in NSCLC [27,28]. Among the identified genes, two additional molecules showed a significant over-expression in patients with poor outcome regarding metastasis formation: GM-CSF (Figure 4) and ADORA3 (Figure 5).



GM-CSF, the granulocyte and macrophage colony stimulating factor, is a monomeric, 4-helical, secreted cytokine known to inhibit inflammation and T-cell immunity [29]. It has been described to promote cancer

in pancreatic ductal neoplasia when over-expressed by a constitutively active form of K-ras [30], in accordance with our previously observed results in K-ras^{LA1}/p53^{R172HΔg} mice. The association of GM-CSF expression

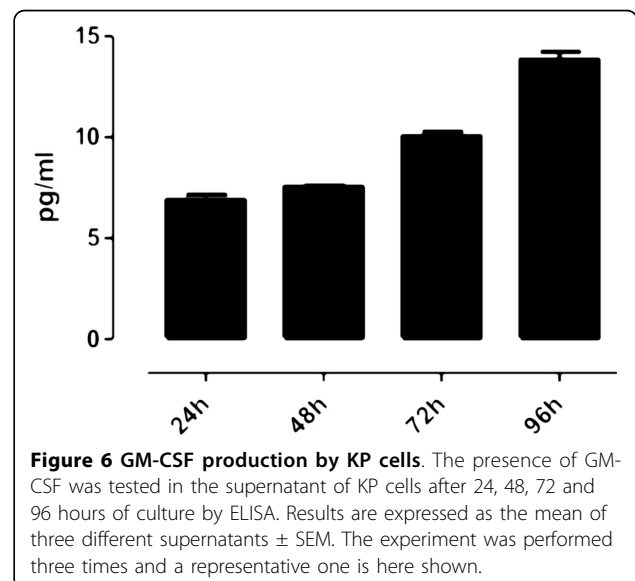




with poor outcome was obtained in the test dataset. The result could not be confirmed in the validation dataset probably due to the limited number of samples in high expression cluster (Figure 4B, red curve). Nevertheless, significance in the first dataset was maintained even only considering early stage T1 tumors (Figure 4C). Analysis of the supernatants from a cell line (KP cells) derived from a lung tumor of a 30 week-old $K\text{-ras}^{\text{LA1}}/p53^{\text{R172HAg}}$ mouse confirmed that they express GM-CSF (Figure 6). Taken together our data, with the observation that serum level of GM-CSF is significantly higher in colon adenocarcinoma patients [31], suggest that GM-CSF might represent a putative early marker in lung adenocarcinoma detection.

ADORA3 is a member of a family of 7-transmembrane G-protein-coupled receptor for adenosine. It has been reported to be involved in cell cycle regulation and tumor growth control both in vitro and in vivo [32]. It has been recently shown [33] that ADORA3 is involved in the induction of p53-mediated apoptosis in lung cancer cell lines. Since in our model p53 is inactivated, ADORA3 does not negatively affect tumor growth, but remain expressed on tumor cells. Although it does not represent a suitable oncoantigen, since its expression does not strictly affect tumor behavior; however, since it

is a tumor associated antigen it could represent an interesting target for the development of antibody-mediated therapy on the subset of NSCLC which are p53 null and ADORA3 positive.



Conclusions

The combination of powerful transcriptomics analysis, i.e. RNA-seq, genetically engineered mice models prone to develop tumors and large collection of human tumor transcriptomes offers new opportunities for the discovery and validation of therapeutic targets in the framework of personalized medicine. The identification of a known biomarker as osteopontin in the NSCLC mouse model confirmed the efficacy of our pipeline to detect targets in precision medicine. Moreover, our approach also allowed the identification of a new putative target, ADORA3, as well as a new putative biomarker, GM-CSF.

Methods

Mice

The heterozygous K-ras^{LA1} mice were crossed with heterozygous p53^{R172HΔg} mice (both kindly provided by Dr. G. Lozano, University of Texas, Houston, TX, USA) to generate K-ras^{LA1}/p53^{R172HΔg} and WT mice. The background of these mice was 129/Sv. Mice were maintained in the transgenic unit of the Molecular Biotechnology Center (University of Torino) under a 12 hour light-dark cycle and provided food and water ad libitum. Genotyped and individually tagged mice of the same age were treated in conformity with national and international laws and policies as approved by the Faculty Ethical Committee and all animal experiments were performed in accordance with European Union guidelines and national institutional regulations. Genotyping of K-ras^{LA1} mice was performed as previously described [13]. To determine p53^{R172HΔg} mouse genotypes, PCR analysis was performed on tail DNA using the following primer sets: BMGFD (covering part of intron 4 and of the exon 5; 5'- TCT CTT CCA GTA CTC TCC TC -3') and BMGRV (covering the end of exon 7 and part of intron 7; 5'- GCC TTC CTA CCT GGA GTC TT -3') (Invitrogen Corp., Carlsbad, CA) for the amplification of p53 allele. The resulting PCR product was then digested with HgaI restriction enzyme (Invitrogen) to discriminate p53 WT from p53^{R172HΔg} mutant alleles.

Cell line

KP is a cloned cell line established in vitro from a lung carcinoma that arose spontaneously in a K-ras^{LA1}/p53^{R172HΔg} mouse. KP cells were cultured in DMEM with Glutamax 1 (DMEM, Life Technologies) supplemented with 20% heat-inactivated fetal bovine serum (Invitrogen).

Magnetic Resonance Imaging (MRI)

MR images were acquired on a Bruker Avance 300 (Bruker, Ettlingen, Germany) operating at 7T using a 30 mm insert birdcage. Mice at different weeks of age (i.e. 10, 20 and 30 weeks, n = 3 each group) were anesthetized by

injecting intramuscularly a mixture of tiletamine/zolazepam 20 mg/kg (Zoletil 100; Virbac, Milperra, Australia) and 5 mg/kg xylazine (Rompun; Bayer, Milano, Italy). Breath rate was monitored throughout in vivo MRI experiments using a respiratory air pillow (SA Instruments, Stony Brook, NY).

T₂w axial, coronal and sagittal MR images with an in-plane resolution of 100 μm were acquired with a breath-triggered sequence respiratory gating to reduce lung movement artefacts using a RARE sequence (typical setting TR/TE/NEX/RARE factor = 6.0 s/4.14 ms/2/16) preceded by a fat-suppression module. A 256 × 256 acquisition matrix was used with a field of view of 25 × 25 mm². The slice thickness was 1 mm, and the number of slices was 18 to 20, which was sufficient to cover the entire lung so that tumor volume could be measured. The T₂w sequence can display the tumor location, size, and shape in both left and right lungs, providing clear boundaries with normal lung tissue.

Tumor Volume Measurements

Data analysis of MR images was performed by using an open source application, ITK-Snap (<http://www.itksnap.org>), for segmentation of the lung nodules in three-dimensions, calculating both the number and the size of tumor lesions [34]. Tumor volume per animal was quantified by calculating the area of visible lung opacities hyper intense regions present in each axial or coronal image slice sequence (usually 18-20 per mouse) and then multiplying the sum of the areas by the distance between each MRI sequence slice. The post-processing of the segmented data provides the voxel counts and the volume (mm³) and displays the shape of the segmented structure. Tumor volumes were normalized relative to the total lung volumes at the indicated times and expressed as percentage of lung volume occupied by tumors.

Lung tumor collection

Normal lung tissues and primary lung adenocarcinomas were collected from WT and K-ras^{LA1}/p53^{R172HΔg} mice, at different stages of cancer progression (corresponding to 10, 20 and 30 weeks of age). Groups of three to six WT and K-ras^{LA1}/p53^{R172HΔg} mice were sacrificed by cervical dislocation at the indicated times. Specimens for RNA extraction and gene expression profile analysis were stored in RNA later (Sigma-Aldrich, Milano, Italy) at 4° C for 24 h and then snap-frozen in liquid nitrogen and stored at -80° C until use. Tissues for histological and immunohistochemical studies were fixed in 10% neutral-buffered formalin and embedded in paraffin.

Histopathological and immunohistochemical analysis

Tumors and tissues collected from K-ras^{LA1}/p53^{R172HΔg} mice were fixed in formalin or PLP (Paraformaldehyde/Lysine/Periodate) and embedded in paraffin or frozen in

OCT, respectively. Sections were stained with hematoxylin and eosin (H&E) for histological evaluation. Immunohistochemical staining was performed with the following primary antibodies: anti-TTF-1 (Thyroid Transcription Factor-1), anti-p63, anti-Synaptophysin and anti-Neuron Specific Enolase (NSE). Slides were then incubated with the appropriate biotinylated secondary antibody. Immunoreactive antigens were detected using NeutrAvidin™ Alkaline Phosphatase Conjugated (Thermo Scientific-Pierce Biotechnology, Rockford, USA) and Vulcan Fast Red (Biocare Medical, Concord, CA) or DAB Chromogen System (Dako Corporation, Carpinteria, CA, USA).

RNA extraction

Total RNA was isolated from lung specimens by using an IKA-Ultra-Turrax® T8 homogenizer (IKA-Werke, Staufen, Germany) and TRIzol® reagent (Invitrogen), according to the manufacturer's instructions. Genomic DNA contaminations were removed from total RNA by using DNA-free kit (Ambion, Warrington, England) as per manufacturer's instructions. Total RNA concentration and purity were assessed using NanoVue Plus Spectrophotometer (GE Healthcare, Milano, Italy); RNA quality was evaluated on an Agilent 2100 Bioanalyzer following the manufacturer's recommendations (Agilent Technologies, Milano, Italy), with a RNA integrity number (RIN) greater than 8.0 considered acceptable for expression profiling by microarray.

Microarray data generation and analysis

Total RNA was then used to create the biotin-labelled cDNA probes to be hybridized on GeneChips Exon 1.0 ST mouse microarrays following the procedure described by the manufacturer (Affymetrix, Santa Clara, CA). Arrays were scanned on Affymetrix Gene Chip Scanner 3000 7G and the CEL files were analysed as follows.

The CEL files resulting from the analysis of image files were analysed using oneChannelGUI 1.6.5 [35]. Gene-level expression was calculated using RMA method (Robust Multichip Average) [36] and normalized by quantile sketch method [37].

The gender effect was modelled to evaluate if any gene was associated to the difference in tumor growth observed between males and females.

The maSigPro Bioconductor library was used to assess differential expression at gene level [38]. maSigPro statistics follows a two-step regression strategy. It first adjusts the model by the least squared technique to identify differentially expressed genes and selects significant genes applying false discovery rate control procedures ($FDR \leq 0.05$). Secondly, backward stepwise regression is applied to study differences between experimental groups ($p \leq 0.05$). The final list of significant differentially expressed

genes is defined using the R^2 values ($R^2 \geq 0.6$) of this second step.

Data were deposited on GEO database: GSE30878

RNA-seq and transcriptome analysis

RNA libraries were sequenced using (HiSeq2000, Illumina, CA, USA). Two pools of total RNA extracted from 30 week-old mice ($n = 3$) were generated for WT and MT. Each pool was sequenced twice to increase the coverage. A total of 51,756,477 and 70,406,984 paired-end (PE) reads were obtained for the first and the second MT replicates, respectively. In the case of the WT replicates 79,079,459 and 63,675,355 PE reads were observed, respectively. Data were deposited on GEO database: GSE51144

Fusion detection

De-novo discovery of chimeric transcripts was done by ChimeraScan [22] with default parameters. For the first and the second MT datasets 5066 and 4543 putative events were measured, respectively. 4533 and 4351 putative events were found for the first and second WT dataset, respectively. Gene fusions were annotated using chimera Bioconductor package. Only the fusion events in common between replicates were retained.

Gene/Exon-level analysis

Reads were mapped on mouse reference genome mm9 using TopHat version 2.0.4, using default parameters and UCSC annotation (<http://genome.ucsc.edu/>).

Mapped reads were counted for each replicate of WT and MT using DEXSeq package [24]. Briefly, dexseq_count.py script was used to associate reads to exons and differentially expressed exons were detected using $FDR < 0.1$ and $|\log_2 \text{Fold Change}| > 1$.

Then, geneCountTable function was used to collapse exon-level in gene-level counts. Differential expression was subsequently evaluated using DESeq package [25] ($FDR < 0.1$, $|\log_2 \text{Fold Change}| > 1$).

Collection and processing of lung cancer expression data

Microarrays

Seven datasets containing microarray data of lung cancer samples (adenocarcinoma and squamous cell carcinoma) and annotations on patients' clinical outcome were collected. All data were measured on different Affymetrix arrays and have been downloaded from NCBI Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>), caArray (<https://array.nci.nih.gov/caarray/home.action>), and the Computational Biology Center of the Memorial Sloan-Kettering Cancer Center (<http://cbio.mskcc.org/>). The complete list of datasets is provided in Table 1.

Prior to analysis, the datasets were reorganized by eliminating duplicate samples and samples without outcome information. Briefly, the original studies have been

Table 1 Original lung cancer datasets

Source	Affymetrix platform	Samples	References
GEO GSE3141	HG-U133 Plus 2.0	111	Bild et al., 2006 [51]
GEO GSE19188	HG-U133 Plus 2.0	156	Hou J et al., 2010 [40]
caArray jacob-00182	HG-U133A	468	Shedden et al., 2008 [52]
http://cbio.mskcc.org/Public/lung_array_data/	HG-U133Av2	129	Nguyen et al., 2009 [42]; Chitale et al., 2009 [43]
GEO GSE10245	HG-U133 Plus 2.0	58	Kuner et al., 2009 [44]
GEO GSE31210	HG-U133 Plus 2.0	226	Okayama H et al., 2012 [45,53]
GEO GSE14814	HG-U133A	90	Zhu CQ et al., 2010 [54]

modified as follows: GSE3141 [39] has been re-named as Duke (Duke University) and used as it is; GSE19188 [40] has been re-named EMC and used after removal of samples lacking the patient outcome information; Shedden [41] has been split into MI (187 samples from the University of Michigan Cancer Center), DFCI (82 samples of the Dana-Farber Cancer Institute); HLM (92 samples collected at the Moffitt Cancer Center), and MSKCC_1 (107 samples from the Memorial Sloan-Kettering Cancer Center); Ladanyi-Gerald [42,43] has been re-named as MSKCC_2 (Memorial Sloan-Kettering Cancer Center) and used as it is; GSE10245 [44] has been re-named DKFZ (German Cancer Research Center) and used as it is; GSE31210 [45] re-named NCCRI (National Cancer Center Research Institute, Japan) and used as it is; GSE14814 [45] re-named OCI-PMH (Ontario Cancer Institute, Princess Margaret Hospital) and used after removal of large cell undifferentiated carcinoma samples. This re-organization resulted in a compendium (meta-dataset) comprising 989 unique adenocarcinoma samples from seven independent cohorts. The type and content of clinical and pathological annotations of the meta-dataset samples have been derived from the original cohorts.

According to Cordenonsi et al., [46] clinical information among the various datasets was standardized redefining the outcome descriptions based on the clinical annotations of each individual study. Specifically, we defined two major types of events, i.e., metastasis and survival.

Raw expression data (i.e., CEL files) obtained from different platforms was integrated using an approach inspired by geometry and probe content of HG-U133 Affymetrix arrays [47]. Briefly, probes with the same oligonucleotide sequence, but located at different coordinates on different type of arrays, have been arranged in a virtual platform grid. As for any other microarray geometry, this virtual grid has been used as a reference to create a virtual Chip Definition File (virtual-CDF), containing probes shared among the various HG-U133 platforms and their coordinates on the virtual platform, and a virtual-CEL file containing the fluorescence intensities of the original CEL files properly re-mapped on the

virtual grid. Expression values for 21981 meta-probesets were generated from the transformed virtual-CEL files using a virtual-CDF obtained merging HG-U133A, HG-U133Av2, and HG-U133 Plus2 original CDFs. Fluorescence signals were background adjusted, normalized using quantile normalization, and gene expression levels calculated using median polish summarization (RMA; [48]). The entire procedure was implemented as an R script. The meta-dataset is available upon request to the authors.

RNAseq

Public RNA sequencing human lung adenocarcinoma data and related clinical metadata were downloaded from The Cancer Genome Atlas repositories (<http://cancergenome.nih.gov/>). Two datasets were available at the day of the download, containing respectively a total of 162 (RNASeq) and 452 (RNASeqV2) samples with exon-level expressions. After filtering the transcriptomes on the basis of the available clinical annotations we obtained a dataset of 509 NSCLC adenocarcinoma transcriptomes (Additional file 1). The entire procedure was implemented as an R script.

Clinical Outcome score evaluation

The microarray meta-dataset was split in two separate groups containing respectively 695 (from cohorts published between 2005 and 2009) and 294 samples (from cohorts published between 2011 and 2012).

Exon-level analysis was done on 137 and 372 samples derived from Cancer Genome Atlas RNASeq dataset and from RNASeqV2 dataset, respectively

Expression levels of each putative target (gene/exon) discovered by the analysis of RNA-seq data were divided in two clusters using a k-means clustering ($k = 2$). Median expression for each cluster was calculated. The label "UP" was associated to the cluster characterized by the higher median expression, while the other cluster was labelled "DOWN".

Exponential survival models [49] from the survival R package, were fitted for the UP and DOWN clusters and the significance (P_{true}) of the differences between the models were tested [50]. Then, we performed a random assignment of UP and DOWN labels to the samples and

we tested the significance (P*) of the difference between these null models. This procedure was repeated n times (n = 10000), randomly removing, at each repetition step, 10% of the samples.

Clinical Outcome score (CO) was then calculated with the following formula:

$$CO = \text{on}$$

FR and EQ generated the animal model and prepared samples for histological and microarray analysis, MA and AF prepared samples for RNA-seq, GB and EZ sequenced the RNA-seq libraries, MI did the histological analyses, DLL run the NMR analyses. MC and RAC did transcriptome data analysis; SN and SB prepared the lung transcriptome dataset. PN and LL generate the KP cell line. RAC, FC and EQ conceived, designed and supervised the study, and wrote the paper.

Competing interests

The authors declare that they have no competing interests.

Declaration

The publication costs for this article were funded by grants from the Italian Association for Cancer Research; the Epigenomics Flagship Project EPIGEN. This article has been published as part of BMC Bioinformatics Volume XX Supplement X, 2014: Italian Society of Bioinformatics (BITS): Annual Meeting 2013.

This article has been published as part of *BMC Genomics* Volume 15 Supplement 5, 2014: Italian Society of Bioinformatics (BITS): Annual Meeting 2013: Genomics. The full contents of the supplement are available online at <http://www.biomedcentral.com/bmcgenomics/supplements/15/S3>

Authors' details

¹Molecular Biotechnology Center, University of Torino, 10126 Torino, Italy.

²Department of Biotechnology, University of Verona, 37134 Verona, Italy.

³Department of Oncology and Neurosciences, G. d'Annunzio University, 66100 Chieti, Italy. ⁴Department of Chemistry IFM and Center for Molecular Imaging, University of Turin, 10126 Turin, Italy. ⁵Center for Genome Research, Dept. of Life Sciences, University of Modena and Reggio Emilia, 41100 Modena, Italy. ⁶Department of Medicina Specialistica, Diagnostica e Sperimentale, University of Bologna, 40126 Bologna, Italy. ⁷Laboratory of Experimental Oncology, Rizzoli Orthopaedic Institute, 40136 Bologna Italy.

Published: 6 May 2014

References

- Lovly CM, Carbone DP: **Lung cancer in 2010: One size does not fit all.** *Nat Rev Clin Oncol* 2011, **8**(2):68-70.
- Gibbons DL, Lin W, Creighton CJ, Zheng S, Berel D, Yang Y, Raso MG, Liu DD, Lozano G, et al: **Expression signatures of metastatic capacity in a genetic mouse model of lung adenocarcinoma.** *PLoS One* 2009, **4**(4):e5401.
- Pallis AG, Serfass L, Dziadzisko R, van Meerbeeck JP, Fennell D, Lacombe D, Welch J, Gridelli C: **Targeted therapies in the treatment of advanced/metastatic NSCLC.** *Eur J Cancer* 2009, **45**(14):2473-2487.
- Dempke WC, Suto T, Reck M: **Targeted therapies for non-small cell lung cancer.** *Lung Cancer* 2010, **67**(3):257-274.
- Greenberg AK, Lee MS: **Biomarkers for lung cancer: clinical uses.** *Curr Opin Pulm Med* 2007, **13**(4):249-255.
- Sung HJ, Cho JY: **Biomarkers for the lung cancer diagnosis and their advances in proteomics.** *BMB Rep* 2008, **41**(9):615-625.
- Sudhindra A, Ochoa R, Santos ES: **Biomarkers, Prediction, and Prognosis in Non-Small-Cell Lung Cancer: A Platform for Personalized Treatment.** *Clin Lung Cancer* 2011.
- Cavallo F, Calogero RA, Forni G: **Are oncoantigens suitable targets for anti-tumour therapy?** *Nat Rev Cancer* 2007, **7**(9):707-713.
- Cavallo F, De Giovanni C, Nanni P, Forni G, Lollini PL: **2011: the immune hallmarks of cancer.** *Cancer Immunol Immunother* 2011, **60**(3):319-326.
- Calogero RA, Quaglino E, Saviozzi S, Forni G, Cavallo F: **Oncoantigens as anti-tumor vaccination targets: the chance of a lucky strike?** *Cancer Immunol Immunother* 2008.
- Maher CA, Palanisamy N, Brenner JC, Cao X, Kalyana-Sundaram S, Luo S, Khrebtukova I, Barrette TR, Grasso C, Yu J, et al: **Chimeric transcript discovery by paired-end transcriptome sequencing.** *Proceedings of the National Academy of Sciences of the United States of America* 2009, **106**(30):12353-12358.
- Zheng S, El-Naggar AK, Kim ES, Kurie JM, Lozano G: **A genetic mouse model for metastatic lung cancer with gender differences in survival.** *Oncogene* 2007, **26**(48):6896-6904.
- Johnson L, Mercer K, Greenbaum D, Bronson RT, Crowley D, Tuveson DA, Jacks T: **Somatic activation of the K-ras oncogene causes early onset lung cancer in mice.** *Nature* 2001, **410**(6832):1111-1116.
- Liu G, McDonnell TJ, Montes de Oca Luna R, Kapoor M, Mims B, El-Naggar AK, Lozano G: **High metastatic potential in mice inheriting a targeted p53 missense mutation.** *Proc Natl Acad Sci USA* 2000, **97**(8):4174-4179.
- Lang W, Wang H, Ding L, Xiao L: **Cooperation between PKC-alpha and PKC-epsilon in the regulation of JNK activation in human lung cancer cells.** *Cell Signal* 2004, **16**(4):457-467.
- Olive KP, Tuveson DA, Ruhe ZC, Yin B, Willis NA, Bronson RT, Crowley D, Jacks T: **Mutant p53 gain of function in two mouse models of Li-Fraumeni syndrome.** *Cell* 2004, **119**(6):847-860.
- Ueno T, Linder S, Elmberger G: **Aspartic proteinase napsin is a useful marker for diagnosis of primary lung adenocarcinoma.** *Br J Cancer* 2003, **88**(8):1229-1233.
- Kontic M, Stojisic J, Kacar-Kukric V, Jekic B, Bunjevacki V: **Multidisciplinary approach in diagnosis of lung carcinoma.** *Experimental oncology* 2010, **32**(2):111-113.
- Gonzalez de Castro D, Clarke PA, Al-Lazikani B, Workman P: **Personalized cancer medicine: molecular diagnostics, predictive biomarkers, and drug resistance.** *Clinical pharmacology and therapeutics* 2013, **93**(3):252-259.
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: **Mapping and quantifying mammalian transcriptomes by RNA-Seq.** *Nature methods* 2008, **5**(7):621-628.
- Maher CA, Kumar-Sinha C, Cao X, Kalyana-Sundaram S, Han B, Jing X, Sam L, Barrette T, Palanisamy N, Chinnaiyan AM: **Transcriptome sequencing to detect gene fusions in cancer.** *Nature* 2009, **458**(7234):97-101.
- Iyer MK, Chinnaiyan AM, Maher CA: **ChimeraScan: a tool for identifying chimeric transcription in sequencing data.** *Bioinformatics* 2011, **27**(20):2903-2904.
- Carrara M, Beccuti M, Lazzarato F, Cavallo F, Cordero F, Donatelli S, Calogero RA: **State-of-the-art fusion-finder algorithms sensitivity and specificity.** *BioMed research international* 2013, **2013**:340620.
- Anders S, Reyes A, Huber W: **Detecting differential usage of exons from RNA-seq data.** *Genome research* 2012, **22**(10):2008-2017.
- Anders S, Huber W: **Differential expression analysis for sequence count data.** *Genome biology* 2010, **11**(10):R106.
- Wittekind C: **[2010 TNM system: on the 7th edition of TNM classification of malignant tumors].** *Der Pathologe* 2010, **31**(5):331-332.
- Shojaei F, Scott N, Kang X, Lappin PB, Fitzgerald AA, Karlcek S, Simmons BH, Wu A, Lee JH, Bergqvist S, et al: **Osteopontin induces growth of metastatic tumors in a preclinical model of non-small lung cancer.** *Journal of experimental & clinical cancer research: CR* 2012, **31**:26.
- Chambers AF, Wilson SM, Kerkviet N, O'Malley FP, Harris JF, Casson AG: **Osteopontin expression in lung cancer.** *Lung cancer* 1996, **15**(3):311-323.
- Bayne LJ, Beatty GL, Jhala N, Clark CE, Rhim AD, Stanger BZ, Vonderheide RH: **Tumor-derived granulocyte-macrophage colony-stimulating factor regulates myeloid inflammation and T cell immunity in pancreatic cancer.** *Cancer cell* 2012, **21**(6):822-835.
- Pylayeva-Gupta Y, Lee KE, Hajdu CH, Miller G, Bar-Sagi D: **Oncogenic Kras-induced GM-CSF production promotes the development of pancreatic neoplasia.** *Cancer cell* 2012, **21**(6):836-847.
- Mroczo B, Szmietkowski M, Wereszczynska-Siemiatkowska U, Okulczyk B, Kedra B: **Pretreatment serum levels of hematopoietic cytokines in**

- patients with colorectal adenomas and cancer. *International journal of colorectal disease* 2007, **22**(1):33-38.
32. Yaer R, Jones MR, Chen JF, Ravid K: **Animal models for the study of adenosine receptor function.** *Journal of cellular physiology* 2005, **202**(1):9-20.
 33. Otsuki T, Kanno T, Fujita Y, Tabata C, Fukuoka K, Nakano T, Gotoh A, Nishizaki T: **A3 adenosine receptor-mediated p53-dependent apoptosis in Lu-65 human lung cancer cells.** *Cellular physiology and biochemistry: international journal of experimental cellular physiology, biochemistry, and pharmacology* 2012, **30**(1):210-220.
 34. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G: **User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability.** *Neuroimage* 2006, **31**(3):1116-1128.
 35. Sanges R, Cordero F, Calogero RA: **oneChannelGUI: a graphical interface to Bioconductor tools, designed for life scientists who are not familiar with R language.** *Bioinformatics* 2007, **23**(24):3406-3408.
 36. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP: **Exploration, normalization, and summaries of high density oligonucleotide array probe level data.** *Biostatistics* 2003, **4**(2):249-264.
 37. Bolstad BM, Irizarry RA, Astrand M, Speed TP: **A comparison of normalization methods for high density oligonucleotide array data based on variance and bias.** *Bioinformatics* 2003, **19**(2):185-193.
 38. Conesa A, Nueda MJ, Ferrer A, Talon M: **maSigPro: a method to identify significantly differential expression profiles in time-course microarray experiments.** *Bioinformatics* 2006, **22**(9):1096-1102.
 39. Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D, Joshi MB, Harpole D, Lancaster JM, Berchuck A, et al: **Oncogenic pathway signatures in human cancers as a guide to targeted therapies.** *Nature* 2006, **439**(7074):353-357.
 40. Hou J, Aerts J, den Hamer B, van Ijcken W, den Bakker M, Riegman P, van der Leest C, van der Spek P, Foekens JA, Hoogsteden HC, et al: **Gene expression-based classification of non-small cell lung carcinomas and survival prediction.** *PLoS one* 2010, **5**(4):e10312.
 41. Shedden K, Chen W, Kuick R, Ghosh D, Macdonald J, Cho KR, Giordano TJ, Gruber SB, Fearon ER, Taylor JM, et al: **Comparison of seven methods for producing Affymetrix expression scores based on False Discovery Rates in disease profiling data.** *BMC bioinformatics* 2005, **6**:26.
 42. Nguyen DX, Chiang AC, Zhang XH, Kim JY, Kris MG, Ladanyi M, Gerald WL, Massague J: **WNT/TCF signaling through LEF1 and HOXB9 mediates lung adenocarcinoma metastasis.** *Cell* 2009, **138**(1):51-62.
 43. Chitale D, Gong Y, Taylor BS, Broderick S, Brennan C, Somwar R, Golas B, Wang L, Motoi N, Szoke J, et al: **An integrated genomic analysis of lung cancer reveals loss of DUSP4 in EGFR-mutant tumors.** *Oncogene* 2009, **28**(31):2773-2783.
 44. Kuner R, Muley T, Meister M, Ruschhaupt M, Buness A, Xu EC, Schnabel P, Warth A, Poustka A, Sultmann H, et al: **Global gene expression analysis reveals specific patterns of cell junctions in non-small cell lung cancer subtypes.** *Lung Cancer* 2009, **63**(1):32-38.
 45. Okayama H, Kohno T, Ishii Y, Shimada Y, Shiraishi K, Iwakawa R, Furuta K, Tsuta K, Shibata T, Yamamoto S, et al: **Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas.** *Cancer research* 2012, **72**(1):100-111.
 46. Cordenonsi M, Zanconato F, Azzolin L, Forcato M, Rosato A, Frasson C, Inui M, Montagner M, Parenti AR, Poletti A, et al: **The Hippo transducer TAZ confers cancer stem cell-related traits on breast cancer cells.** *Cell* 2011, **147**(4):759-772.
 47. Fallarino F, Volpi C, Fazio F, Notartomaso S, Vacca C, Busceti C, Biciato S, Battaglia G, Bruno V, Puccetti P, et al: **Metabotropic glutamate receptor-4 modulates adaptive immunity and restrains neuroinflammation.** *Nature medicine* 2010, **16**(8):897-902.
 48. Irizarry RA, Ooi SL, Wu Z, Boeke JD: **Use of mixture models in a microarray-based screening procedure for detecting differentially represented yeast mutants.** *Statistical applications in genetics and molecular biology* 2003, **2**:Article1.
 49. Andersen PK, Borch-Johnsen K, Deckert T, Green A, Hougaard P, Keiding N, Kreiner S: **A Cox regression model for the relative mortality and its application to diabetes mellitus survival data.** *Biometrics* 1985, **41**(4):921-932.
 50. Harrington DP FT: **A class of rank test procedures for censored survival data.** *Biometrika* 1982, **69**:553-566.
 51. Thomsen HS, Dorph S: **Interventional urology today.** *Annals of medicine* 1992, **24**(3):167-169.
 52. Shedden K, Taylor JM, Enkemann SA, Tsao MS, Yeatman TJ, Gerald WL, Eschrich S, Jurisica I, Giordano TJ, Misek DE, et al: **Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study.** *Nature medicine* 2008, **14**(8):822-827.
 53. Yamauchi M, Yamaguchi R, Nakata A, Kohno T, Nagasaki M, Shimamura T, Imoto S, Saito A, Ueno K, Hatanaka Y, et al: **Epidermal growth factor receptor tyrosine kinase defines critical prognostic genes of stage I lung adenocarcinoma.** *PLoS one* 2012, **7**(9):e43923.
 54. Zhu CQ, Ding K, Strumpf D, Weir BA, Meyerson M, Pennell N, Thomas RK, Naoki K, Ladd-Acosta C, Liu N, et al: **Prognostic and predictive gene signature for adjuvant chemotherapy in resected non-small-cell lung cancer.** *Journal of clinical oncology: official journal of the American Society of Clinical Oncology* 2010, **28**(29):4417-4424.

doi:10.1186/1471-2164-15-S3-S1

Cite this article as: Riccardo et al.: Characterization of a genetic mouse model of lung cancer: a promise to identify Non-Small Cell Lung Cancer therapeutic targets and biomarkers. *BMC Genomics* 2014 **15**(Suppl 3):S1.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

