Published in final edited form as:

Sci Transl Med. 2013 May 22; 5(186): 186ra66. doi:10.1126/scitranslmed.3005723.

Ectopic Activation of Germline and Placental Genes Identifies Aggressive Metastasis-Prone Lung Cancers

Sophie Rousseaux[1,*], **Alexandra Debernardi**[1], **Baptiste Jacquiau**[1], **Anne-Laure Vitte**[1], **Aurélien Vesin**[1], **Hélène Nagy-Mignotte**[2], **Denis Moro-Sibilot**[2], **Pierre-Yves Brichon**[2], **Sylvie Lantuejoul**[2], **Pierre Hainaut**[3], **Julien Laffaire**[4], **Aurélien de Reyniès**[4], **David G. Beer**[5], **Jean-François Timsit**[1,2], **Christian Brambilla**[1,2], **Elisabeth Brambilla**[1,2], and **Saadi Khochbin**[1,*]

[1]INSERM, U823; Université Joseph Fourier, Grenoble 1; Institut Albert Bonniot, Grenoble F-38700, France

[2]Grenoble University Hospital (CHU), Grenoble F-38700, France

[*]Corresponding author. sophie.rousseaux@ujf-grenoble.fr (S.R.); khochbin@ujf-grenoble.fr (S.K.).

Author contributions: S.R. and S.K. conceived the whole project, designed the experiments and data analysis approaches, and coordinated authors' contributions. S.R., B.J., A.V., J.-F.T., and A.D. performed the statistical and data analyses. P.-Y.B. ensured surgical removal of tumor samples. C.B., D.M.-S., and H.N.-M. were involved in the recruitment, clinical assessment, and follow-up of the patients. E.B. and S.L. ensured the histopathological assessment of the tumors. P.H. performed the screen for P53 mutations; A.d.R. and J.L. double-checked all the statistical analyses and performed the lung cancer methylome analyses. D.G.B. contributed to the transcriptomic and survival analysis in the lung ADC validation study. C.B., D.M.-S., H.N.-M., E.B., and S.L. were involved in tumor bank constitution and management. A.-L.V. performed the validation experiments. S.R. and S.K. wrote the manuscript, which was read and approved by all co-authors.

Competing interests: The following patent applications include results presented in the paper: PCT/EP2009/053809, PCT/EP2011/068375, and PCT/EP2011/068377.

Data and materials availability: The transcriptomic data from our series of lung tumors have been deposited in the GEO database (http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE30219.

SUPPLEMENTARY MATERIALS

www.sciencetranslationalmedicine.org/cgi/content/full/5/186/186ra66/DC1

Materials and Methods

Fig. S1. Epigenetic "locking" of genes restricted to germline or placental expression.

Fig. S2. CpG methylation in the TSS regions of TS/PS genes.

Fig. S3. Prognostic value of ectopic expression of the 26 genes in lung cancer of specific histological subgroups.

Fig. S4. Prognostic value of ectopic expression of a subset of the 26 prognostic genes in an independent study of 443 lung ADCs.

Table S1. List of the normal tissue samples from GEO studies used for our analyses of gene expression profiles in normal tissues.

Table S2. List of genes with an expression restricted to male germ cells or placenta (TS/PS genes).

Table S3. Description and list of the 1776 cancer samples analyzed from the Exp0 cancer study (GSE2129).

Table S4. Percentage of positive expressions of TS/PS genes in the 14 different types of cancer analyzed from the Exp0 cancer study (GSE2129).

Table S5. List of tissue samples and cell lines analyzed on the dedicated microarray.

Table S6. List of primers used for qRT-PCR experiments.

Table S7. Key clinical and biological data for the 293 lung cancer patients in our study.

Table S8. Percentage of positive expressions of TS/PS genes in the different histological subtypes of our series of lung tumors.

Table S9. Methylation levels (β values) of the selected 347 CpG sites associated with 88 TS/PS genes in the 55 lung tumors in our series.

Table S10. List of 26 genes associated with poor survival in lung cancer and results of univariate analysis of association with prognosis in our lung cancer patients.

Table S11. HRs from the Cox regression model testing the association with prognosis in our set of 26 genes in a multivariate analysis, along with pathological data and TNM stages.

Table S12. List of the 3272 genes that are differentially expressed between patients classified by our 26 TS/PS genes in the good prognostic group (P1) and patients in the poor prognostic group (P3).

Table S13. List of relevant gene sets identified by GSEA from the Broad Institute Database corresponding to various GO terms (C5) and curated signatures (C2), including resistance to drugs or response to drugs, overlapping with our list of up- and down-regulated genes in P3 aggressive lung tumors.

[3]International Prevention Research Institute, Lyon F-69006, France

[4]Ligue Nationale Contre le Cancer, Cartes d'Identité des Tumeurs Program, Paris F-75013, France

[5]Thoracic Surgery, University of Michigan, 6304 Cancer Center, Ann Arbor, MI 48109, USA

Abstract

Activation of normally silent tissue-specific genes and the resulting cell "identity crisis" are the unexplored consequences of malignant epigenetic reprogramming. We designed a strategy for investigating this reprogramming, which consisted of identifying a large number of tissue-restricted genes that are epigenetically silenced in normal somatic cells and then detecting their expression in cancer. This approach led to the demonstration that large-scale "off-context" gene activations systematically occur in a variety of cancer types. In our series of 293 lung tumors, we identified an ectopic gene expression signature associated with a subset of highly aggressive tumors, which predicted poor prognosis independently of the TNM (tumor size, node positivity, and metastasis) stage or histological subtype. The ability to isolate these tumors allowed us to reveal their common molecular features characterized by the acquisition of embryonic stem cell/germ cell gene expression profiles and the down-regulation of immune response genes. The methodical recognition of ectopic gene activations in cancer cells could serve as a basis for gene signature–guided tumor stratification, as well as for the discovery of oncogenic mechanisms, and expand the understanding of the biology of very aggressive tumors.

INTRODUCTION

Cell differentiation is associated with the establishment of specific patterns of cell type– and tissue-specific gene expression, which largely rely on the cell's "epigenetic landscape," mainly shaped by chemical modifications of the genome and the associated histones. In differentiated cells, these epigenetic mechanisms not only help activate and maintain specific gene expression patterns but also control a genome-wide repression of tissue-specific genes (1, 2).

Recent investigations have demonstrated that a global deregulation of epigenetic signaling is an early and recurrent event that occurs during oncogenic cell transformation. Aberrant gene activity is a direct consequence of these anomalies. DNA methylation–associated repression of tumor suppressor genes in cancer cells is well documented and now recognized as an important oncogenic event (3, 4). A less studied consequence of epigenetic deregulations in transformed cells is the ectopic activation of various cell- and tissue-specific genes (5, 6). The aberrant activation of genes in cancer represents a promising source of cancer biomarkers, as exemplified by the discovery of an almost universal cancer marker, the follicle-stimulating hormone (FSH) receptor, abnormally expressed in many types of cancers (7). However, these unprogrammed gene activation events have to date been only sporadically discovered.

Among these genes, male germ cell–specific genes are of particular interest. Indeed, these genes, normally exclusively expressed in testis, have been found to be sporadically

derepressed in several somatic cancers and have hence been called cancer/testis (C/T) genes (8), and proposed as candidate cancer markers with additional potential for exploitation as novel therapeutic targets (C/T gene products are highly immunogenic). However, the sporadic nature of their ectopic expression hinders their use as reliable cancer indicators when detected individually.

The identification of genes with a highly specific pattern of expression, whose silencing was associated with a specific epigenetic signature in normal adult somatic tissues, served as a basis for a genome-wide inventory of all tissue-specific genes with aberrant activations ("off context") in cancer cells. We show here that this approach provides us with access to a large number of candidate cancer biomarkers and sheds new insight upon the biology of cancer with applications in translational medicine.

RESULTS

Most tissue-restricted genes are epigenetically marked germline genes

To evaluate the extent of ectopic gene activations in cancer cells, it was necessary to first identify the human genes whose physiological pattern of expression is normally restricted to one tissue.

To this end, we combined two strategies, exploiting human ESTs (expressed sequence tags) and transcriptomic data sets. A list of genes whose ESTs were specifically found in a given tissue was established. In parallel, human tissue transcriptomic data (normal tissue samples listed in table S1) were analyzed to isolate a second list of genes with a clearly predominant expression in one tissue (meaning, that gene expression in this tissue was more than 3 SDs above the mean of the values of expression in all tissues). The two gene lists were then compared, and genes that were common to both lists were selected and defined as tissue-specific (fig. S1A, step 1).

Both EST and transcriptomic approaches revealed that testis and germline tissues show the highest number of genes satisfying our selection criteria for strict tissue specificity (fig. S1, B and C, respectively), which also explains why most of the already known off-context expressed genes are testis genes (8). Only germline- and placenta-specific genes were selected as potential biomarkers, because they are the only ones that are never expressed in the healthy somatic tissues of an adult organism and remain unknown to the immune system. A total of 506 tissue-restricted genes, including 439 germline- and 67 placenta-restricted genes, were identified by the present approach (hereinafter named TS and PS genes, respectively; listed in table S2).

To test our hypothesis that the single tissue-restricted expression of testis-specific/placenta-specific (TS/PS) genes and their consistently repressed state in adult somatic tissues could be linked to a particular epigenetic status in normal cells, we interrogated several genome-wide epigenetic mapping data sets to characterize the status of these genes (fig. S1A, step 2). Using methylated DNA immunoprecipitation and chromatin immunoprecipitation data from published genome-wide studies (9–11), we extracted epigenomic data corresponding to our list of TS/PS genes. This analysis demonstrated that most of the TS/PS genes show

particular promoter sequence features and are associated with a specific epigenetic status in somatic cells, where their expression is tightly locked. In particular, most TS/PS genes are associated with either CpG-poor promoters (46%) or CpG-rich, but hypermethylated, promoters (35%), whereas when all human genes are considered, the proportions of CpG-poor and hypermethylated CpG-rich promoters drop to 22 and 6%, respectively (fig. S1D). We also screened our TS/PS genes for specific transcriptional/epigenetic features in different somatic tissues from various genome-wide studies (10, 11). We found that TS/PS genes are nearly all devoid of active histone marks, including acetylation, H3K4 methylation, and binding of RNA polymerase II, but are relatively enriched in repressive histone modification H3K9 and H3K27 methylation (fig. S1E). The epigenetic status of these TS/PS genes is therefore consistent with our expression data and confirms that we have isolated a specific fraction of human genes. The other observation from these analyses is the homogeneous and consistent nature of the TS/PS gene epigenetic signature, which seems independent of the tissue type, because it was found both in embryonic stem cells and in multiple types of differentiated adult cells, and was reproducible between multiple independent studies (fig. S1E).

Large-scale activations of TS/PS genes occur in all cancer cells

We systematically investigated the expression status of TS/PS genes in human cancers. The analysis of a large set of transcriptomic data from 1776 solid tumor samples derived from 14 different cancer types (described in table S3) showed that hundreds of these genes are aberrantly activated across all cancer types tested (Fig. 1A and table S4), therefore providing access to an invaluable source of cancer biomarkers.

Before going further with the analyses of ectopic expressions of the TS/PS genes in our list, and to verify the possibility of using them as cancer biomarkers, we proceeded to a series of "wet-bench" validations. We designed a custom-dedicated microarray with probes corresponding to our TS/PS gene list. RNAs from 18 nontumor tissues, including testis and placenta as well as somatic tissues, were probed accordingly. As expected, all the tested genes were only expressed in testis (Fig. 1B, "T") and placenta (Fig. 1B, "P"), and no measurable signal was observed when analyzing RNA from normal tissues (Fig. 1B, "Ctrl soma"). Additionally, we tested RNAs from 21 cancer samples (Fig. 1B, "Cancer") corresponding to eight different cancer types as well as three cancer cell lines (Fig. 1B, "CCL") (all samples in our dedicated array are listed in table S5). The data confirmed that TS/PS genes were ectopically activated in all tested cancer samples (Fig. 1B).

An additional validation was carried out using a quantitative reverse transcription polymerase chain reaction (qRT-PCR) assay testing the activation of 13 of our TS/PS genes, arbitrarily selected among those frequently deregulated in our transcriptomic analyses described above. In this setting, we could detect at least one positive TS/PS gene expression in 72% of the examined tumors ($n = 73$), whereas the corresponding control tissues were all negative ($n = 21$) (Fig. 1C; primers listed in table S6).

We hypothesized that the epigenetic alterations that characterize the cancer cells (3, 4) would likely drive the nonprogrammed activations of the TS/PS genes (12, 13).

This hypothesis was supported by monitoring the frequency of TS/PS gene activations in various cancers (our 1776 reference cancer samples) as a function of the status of their promoters, which demonstrated that the TS/PS genes associated with CpG-rich promoters, with high levels of methylation in somatic cells, were more frequently activated in cancers than genes associated with CpG-poor promoters (Fig. 2A).

This observation was further validated by analyzing the methylation-dependent transcriptional activity of TS/PS genes in human colon cancer cell line HCT116 bearing the inactivation of DNA methyltransferase 3b (DNMT3b) or the inactivation of both DNMT3b and DNMT1 [HCT116 DKO (double knockout)] (Fig. 2, B and C), which enabled us to identify genes activated by global DNA demethylation. Using our dedicated microarray, we showed that in HCT116 DKO, CpG-rich associated genes were also frequently activated, in contrast to genes with CpG-poor promoters, which did not respond to DNA demethylation (Fig. 2B). This was confirmed by qRT-PCR on 49 CpG-rich TS/PS genes, nearly half of which showed a clear activation in the DKO cells (Fig. 2C; primers listed in table S6).

This experiment also showed that, despite the genome-wide DNA demethylation in these cells, a significant number of CpG-rich TS/PS genes remain silent (Fig. 2C). This suggests that, although DNA demethylation may be necessary, in some cases, it is not sufficient to induce an ectopic gene expression.

The ectopic activation of 26 tissue-restricted genes in lung tumors is a strong and independent predictor of poor prognosis

Having found that the off-context expression of normally silent genes systematically occurs in cancer, we next investigated whether these genes could represent useful biomarkers by considering one cancer type, lung cancer. Lung cancer is one of the most common cancers in humans and is the most frequent cause of mortality by cancer in men and women (14). In the context of a clinical research program, we recruited a cohort of 293 lung cancer patients (recruited at Grenoble University Hospital, France), who received surgery. The group included 152 patients with early-stage cancer (T1N0) according to the TNM classification (tumor size, node positivity, and metastasis). For each of the study subjects, genome-wide transcriptomic analysis was performed on pretreatment diagnostic tumor samples, and pathological and clinical data were recorded, including overall and disease-free survival over a period of 5 to 10 years [described in table S7; full Affymetrix transcriptomic data available on the Gene Expression Omnibus (GEO) Web site under the reference GSE30219].

Applying the strategy described above, we could detect aberrant expressions of TS/PS genes in all lung tumor samples in our series (Fig. 3A), including the 152 cases of early-stage T1N0 (Fig. 3B and table S8). Moreover, a series of 15 paired tumor and corresponding nontumor lung (NL) samples confirmed that these genes are mainly activated in the tumors and only rarely in the NLs of the same patients (these rare expression events in NLs could correspond to possible presence of residual cancer cells, or the effect of the tumor on the surrounding tissues) (Fig. 3C).

To look for epigenetic deregulations potentially leading to these ectopic expressions, we analyzed the methylome of a subset of 55 patients from the cohort. An initial analysis

showed that the CpG-rich regions of the TS/PS genes, mostly hypermethylated in normal somatic tissues including lung, were globally demethylated in a large proportion of lung tumors (Fig. 3D and table S9). To explore a potential relationship between the demethylation of CpG-rich regions associated with TS/PS genes and their ectopic expression, we plotted the percentage of methylation of each CpG against the expression level of the corresponding gene (Fig. 3E and fig. S2). For most genes, ectopic activations were associated with the demethylation of at least one CpG, but, as suggested by our data from HCT116 DNMT DKO (Fig. 2C), the demethylation of any given CpG was not always associated with the activation of the corresponding gene. These observations, which agree with reported genome-wide cancer methylome studies (15, 16), suggest that, although DNA demethylation is associated with and involved in the ectopic gene activations in cancer, other factors/epigenetic aberrations also contribute.

We hypothesized that some of these aberrant expressions, by being the result of major epigenetic deregulations and/or leading to the expression of factors beneficial to cancer cells, could be associated with aggressive tumors and therefore have prognostic implications. Considering the TS/PS genes expressed in more than 1% of the tumors, we used univariate analysis to compare the global survival probabilities over a period of 5 years between the groups of patients whose cancers expressed each gene and those that did not.

This screen identified 26 TS/PS genes (fig. S1A, step 3), whose aberrant expression was individually associated with a lower survival probability in the lung cancer patients in our series [$P < 0.05$, log-rank test; hazard ratios (HRs) >1.5; table S10]. To overcome the sporadic nature of these gene expression changes and optimize the information obtained on all samples, we decided to quantify the combined activations of these 26 genes. We took a simple approach, which was as follows. The patients were first assigned into two groups: those with a tumor expressing none of the 26 genes, and those with a tumor expressing at least 1 of the 26 genes. We then further refined this latter group by distinguishing tumors expressing one or two genes from tumors expressing three genes or more. As a result of this approach, the patients were stratified into three groups—P1, P2, and P3—according to the number of 26 genes expressed in their tumors: P1 tumors expressed none, P2 tumors expressed 1 or 2, and P3 tumors expressed 3 or more of the 26 genes.

We found highly significant differences in overall survival probabilities between these three groups ($P < 0.0001$; Fig. 4A). Additionally, the prognostic power of this 26-gene classifier was independent of other parameters, including clinical stage (TNM classification) (Fig. 4B) and histological subtype (fig. S3). In particular, this 26-gene group was a very efficient predictor for overall survival of early-stage patients. A multivariate analysis confirmed that our 26-gene combination was a stronger prognostic parameter associated with overall survival than histological subtypes or TNM stages ($P < 0.0001$; table S11). The survival predictive power of our 26-gene group was also strong when considering various subsets of patients grouped by sex, age, tobacco consumption, adjuvant chemotherapy or radiotherapy, as well as their status for P53 mutation, as shown by the forest plot diagram of survival HRs between P1 and P3 patients (Fig. 4C). Finally, a comparison of the clinical outcomes between P1 and P3 patients allowed us to confirm that the tumors classified “P3” presented a particularly aggressive phenotype. Indeed, most patients with these tumors quickly relapsed

and/or developed metastases, which was generally followed by short-term fatal outcome (Fig. 4, D and E).

To validate the potential of our genes to be used as a prognosis-classifying tool, we tested a subset of our patients using qRT-PCR detection of our genes. A clear-cut specific amplification was obtained for 25 of our classifying genes (Fig. 5A, left panel). A total of 61 patients in our series were tested using qRT-PCR and were assigned into P1, P2, and P3 groups, as defined previously. The comparison of survival probabilities between these three groups demonstrated our ability to use qRT-PCR to predict prognosis (Fig. 5A, right panel). A comparison between qRT-PCR and the transcriptomic approach for the detection of ectopic expressions of our prognostic genes showed that 44 of the 61 patients (72%) were assigned into similar groups by the two approaches and that all of the patients in the poor-prognosis group P3 were correctly assigned using the qRT-PCR approach. A total of 17 patients were differently classified depending on the approach. Overall, these differences were due to the higher sensitivity of qPCR compared to the transcriptomic analysis. Indeed, five patients with short survival (<12 months) but assigned to the P1 group using the transcriptomic approach were better classified into the P2 to P3 groups by qRT-PCR, and three patients with intermediate survival (>12 and <36 months) were all assigned to the P1 group by the transcriptomic approach and to the P2 group by qRT-PCR detection. However, nine patients with survival of >36 months were more appropriately assigned to the P1 group using the transcriptomic approach than by qRT-PCR, which allocated them into the P2 group. An explanation for this observation is the “leaky” silence of our marker genes, which could be detected by qRT-PCR, because of the high sensitivity of this approach, leading to the misclassification of P1 tumors into the P2 group. Although these misclassifying expressions picked by qRT-PCR are not frequent, to remain “safe” in our prognosis classification, we decided to consider P2 tumors as an intermediate population where prognosis could not be accurately predicted.

This association between activations within our 26-gene group and a reduced global survival time was validated by applying the same 26-gene classification system to two external lung cancer populations, for whom survival follow-up times and transcriptomic data were obtained with the same Affymetrix technology as ours. These two studies included 82 patients (ADC, SQC, and large cell carcinomas) from (17) (GEO reference GSE19188) and 138 patients (ADC and SQC) from (18) (GEO reference GSE8894). According to the expression of our 26 classifying markers, these patients were assigned into P1, P2, and P3, and their survival times were compared. As mentioned above, because the P2 prognostic classification is not highly reliable for predicting prognosis, the corresponding survival curves are indicated as dotted lines. However, despite the very different origins of the two populations and the differences in the proportions of genders and histological subtypes, our 26-gene classifying system enabled us to reliably identify the very aggressive P3 type tumors in both studies (Fig. 5B), confirming the validity of our 26-gene classification in these independent external studies. In addition, the predictive power of a subset of these 26 genes was tested in a study of 443 patients with lung ADC (19), although, not using the full capacity of our classifier because only 11 of the 26 genes were represented on the Affymetrix platform used in this study, we could also confirm the discriminative power of our approach in this large and independent group (fig. S4).

Metastasis-prone aggressive lung tumors identified by the 26 genes have a characteristic molecular profile

The fact that our classifying genes could distinguish a group of very aggressive lung tumors (P3) independently of other parameters prompted us to perform a supervised analysis of the transcriptome of these aggressive P3 tumors compared to the P1 group (fig. S1A, step 4) to identify, beyond our 26 classifying genes, gene expression patterns associated with this aggressive phenotype. This analysis revealed a gene expression profile composed of 1447 up-regulated and 1825 down-regulated genes (heat maps shown in Fig. 6A; genes listed in table S12), characterizing very aggressive tumors. A close inspection of the P3 versus P1 gene expression profile showed that a subset of patients that were ranked P1 (good prognostic) according to our 26 classifying genes presented a "P3-like" aggressive expression profile (Fig. 6A, blue rectangle). The survival of P1 patients with this P3-like profile was not significantly different from that of the other P1 patients, and differed from that of the P3 patients (Fig. 6B), further demonstrating that the expression of the 26 classifying genes is a better prognostic predictor/censor of lung cancer survival than the P3 gene expression profile alone. This analysis also revealed another tumor group showing an intermediate gene expression profile: ones that did not overexpress P3 up-regulated genes but did present the P3 down-regulated gene profile (Fig. 6A, gray rectangle). All of these "intermediate" tumors corresponded to carcinoid tumors, which are associated with good prognosis.

On the basis of the expression profile of the highly aggressive P3 tumors, we attempted a biological characterization of these lung cancer "killer cells." A GSEA enabled an in-depth investigation of the molecular profile of these aggressive tumors. The aggressive signature was highly enriched in genes normally predominantly expressed in embryonic stem or germline cells (Fig. 6C, upper panels). The analysis of the most enriched Gene Ontology (GO) terms showed that the up-regulated factors were mostly nuclear and related to the cell cycle and proliferation, whereas the depleted functions were related to the immune response and cell interactions and signaling (Fig. 6C, middle and lower panels). Accordingly, the analysis of the expression patterns of the down-regulated genes showed that 20% were genes that are normally predominantly expressed in immune organs (data obtained with normal samples of lymphocytes, spleen, or tonsil; listed in table S1), suggesting a reduced number of infiltrating immune cells within the tumors and/or in their environment. These data paint a molecular portrait of very aggressive and metastasis-prone cancer cells as those pushing their proliferative and self-renewal capacities while escaping immune surveillance systems and/or depleted in immune cells.

The aggressive lung tumor profile overlaps with signatures of other metastasis-prone tumors and could orient therapeutic strategies

The GSEA analysis also confirmed the shared characteristics between the aggressive signature of our series of lung cancer and the profile of the lung tumors associated with poor prognosis identified in the lung ADC study from (19) (Fig. 6D, left panels). The general transcriptomic profile corresponding to the aggressive tumors of this study largely overlapped that of the P3 tumor group we identified with the 26-gene signature, which

further supported the relevance of our approach for the classification of the metastasis-prone lung tumors.

Other overlapping data sets confirmed our characterization of the P3 subset of lung tumors as an aggressive clinical phenotype, because we have also found very large overlaps with signatures from metastatic tumors of different origins (Fig. 6D, right panels, and table S13).

The expression profile of our aggressive lung tumors also suggested that they could be resistant to doxorubicin and to gefitinib (Fig. 6E). Other significantly enriched gene sets suggested that the genes that are highly up-regulated in these aggressive tumors could be down-regulated by specific therapeutic approaches, including treatment with the Ras inhibitor salirasib or Aplidin, a marine-derived compound with potential anticancer properties (20). Indeed, transcriptomic data of cells treated with these compounds revealed that the P3 molecular profile could be shifted to a more P1-like situation after administering these drugs (Fig. 6F and table S13).

DISCUSSION

Here, we demonstrate systematic ectopic activation of hundreds of tissue-specific genes in many cancers, a phenomenon that had been previously reported to occur only sporadically. A key to this finding was the previous definition of tissue-restricted gene expression, which identified a specific category of genes that are normally expressed in the germline and placenta, and present a characteristic “locked” epigenetic configuration in all non-germline cell types, including embryonic stem cells.

The large-scale activation of this specific category of silenced genes in all cancers primarily reflects a general loss of cell identity, most likely due to a profound transformation of their overall epigenetic landscape. However, loss of methylation, although affecting most of the TS/PS genes promoters, cannot fully explain their ectopic activation. Changes in other epigenetic marks, such as histone modifications, should be considered to thoroughly evaluate the impact of the cancer epigenetic alterations in these aberrant gene activations.

Of particular interest is a group of 26 of these genes whose combined activation in lung tumors is associated with a particularly aggressive cancer phenotype.

The analysis of the literature on these 26 classifying genes showed that most of them had not previously been associated with cancers and that they have completely unknown functions.

This study therefore leaves us with important questions regarding the oncogenic potentials of these 26 genes. In particular, it is unclear how the ectopic activation of each of these genes or their combinations could account for the characteristic aggressive gene expression profile evidenced here in metastatic-prone aggressive tumors. It would also be important to understand the relationship between the expression of these genes and well-known oncogenic drivers.

An analysis of the literature shows that some of the 26 genes have already been identified as cancer-associated genes, mainly as so-called C/T antigens, and others are paralogs or pseudogenes corresponding to important cancer-related genes.

One of them, EBI3 (Epstein-Barr virus–induced 3), has already been discovered not only as an independent predictor of poor prognosis in lung cancer but also as a driver (21). PIWIL1, also linked with poor prognosis in various cancer types including lung cancer, has been found associated with stem cell renewal (22–25), in agreement with the embryonic stem cell–like gene expression pattern observed in our aggressive P3 tumors. In addition, the identity of some of our genes suggests yet unknown oncogenic mechanisms. We also found two PTEN-related sequences among our 26 genes. TPTE encodes a testis-specific protein with high sequence similarity to the tumor suppressor PTEN (26), and TPTE2P2 encodes a testis-specific TPTE pseudogene. It is therefore possible that the ectopic activation of TPTE or a truncated protein produced by TPTE2P2 could interfere with the tumor suppressor activity of PTEN and act as a dominant negative factor with oncogenic activity. A similar mechanism could also explain the role of the NBPF4 neuroblastoma breakpoint family. Indeed, an amplified member of the family, NBPF23, has been recently shown to be specifically associated with pediatric neuroblastoma and could drive the malignant transformation through an unknown mechanism (27). The ectopic activation of NBPF4 could increase the NBPF gene dosage and elicit a similar oncogenic effect.

Our findings leave room for speculations and new investigations, but the very typical and homogeneous clinical, biological, and molecular portrait of aggressive tumors identified with this 26-gene signature implies that they might be associated with fundamental oncogenic mechanisms, which cooperate to generate such aggressive tumors.

Our approach also indicated several drugs that might efficiently target P3 lung tumors. Indeed, GSEA identified major overlaps between the sets of genes down-regulated by these drugs and the genes overexpressed in our P3 lung tumors. Some of these drugs, including salirasib (28) and Aplidin (29, 30), have been unsuccessfully tested against non–small cell lung cancer in phase 2 clinical trials. Our observation suggests that they might only be effective in a restricted number of tumors with a characteristic P3-like gene expression signature. The actual sensitivity of these tumor cells remains to be demonstrated in vitro as well as in vivo, including in patients, but the present data suggest potential therapeutic strategies to explore for these deadly tumors.

MATERIALS AND METHODS

Identification of genes with tissue-restricted expression using EST and transcriptomic data

Combining large-scale analysis of ESTs (31) (http://www.ncbi.nlm.nih.gov/sites/entrez?db=unigene&cmd=search&term=9606[taxid] AND tissue[restricted]) and transcriptomic data on normal human tissues available online, we systematically looked for genes with a tissue-restricted pattern of expression. Briefly, only genes predominantly expressed in germline or placenta cells with no ESTs in other tissues were retained as restricted. The genes on this list underwent another consideration, which was the assessment of their

promoter/epigenetic status. This led to the establishment of a list of 506 genes considered to have a restricted pattern of expression in germline and placenta.

Analysis of the epigenetic characteristics of the promoter regions of TS/PS genes in somatic cells

Publicly available pangenomic data were used to determine the epigenetic status of the promoter regions of the TS/PS genes. The list of selected studies and data processing is detailed in the Supplementary Materials and Methods.

Transcriptomic data analysis and statistics for normal tissues and cancer samples

Transcriptomic data from normal tissues and cancer samples were obtained with the Affymetrix technology "Affymetrix.GeneChip. HG-U133_Plus_2." We used data available either on the GEO Web site (http://www.ncbi.nlm.nih.gov/geo/) or from our own lung tumor series (GSE30219). The data were analyzed as described in the Supplementary Materials and Methods.

Thresholds of expression for germline- and placenta-restricted genes were established as described in the Supplementary Materials and Methods.

Analysis of the expression of TS/PS genes on a dedicated microarray and by qRT-PCR

RNA of each of the various control, cancer, and cell line samples was extracted or purchased from Cytomix (see table S5 for information on the origin of the RNA samples) and then processed for hybridization or analyzed by qRT-PCR as described in the Supplementary Materials and Methods.

Analysis of the effect of DNA demethylation in inducing the expression of TS/PS genes in somatic cells using our dedicated microarray and qRT-PCR experiments

The objective was to compare the expression levels of TS/PS genes between HCT116 cell line with a DKO for DNMT3b and DNMT1 and its wild-type counterpart by analysis on our dedicated microarray (described above; the values were expressed as fold changes over the threshold of signal obtained in normal somatic tissues; genes were considered activated when the normalized signal value was above the threshold value of 1.2) and qRT-PCR analysis (described above).

TS/PS gene DNA methylation analysis

Whole-genome DNA methylation was analyzed in 55 patients from our cohort with the Illumina Infinium HumanMethylation450 assay (32), as described in the Supplementary Materials and Methods.

Assessment of prognostic value of the off-context expression of TS/PS genes in lung cancer patients

The prognostic value of the expression of TS/PS genes in lung cancer was assessed in univariate and multivariate analyses as described in the Supplementary Materials and Methods.

Identification and characterization of the molecular profile of aggressive lung tumors

The supervised transcriptomic analysis carried out to identify the genes differentially expressed between patients with best prognosis (P1) and those with poorest prognosis (P3) and the GSEA of the gene expression profile characterizing aggressive tumors are described in the Supplementary Materials and Methods.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank J. Kim and his collaborators from Sungkyunkwan University (Seoul, South Korea) for sharing the raw data of their published work. We acknowledge the efficient contribution of the Centre de Ressources Biologiques of Grenoble Hospital in collecting and managing the cancer samples.

Funding: The work in S.K./S.R. and E.B.'s laboratories is supported by INCa-DHOS, ANR "EpiSperm," and "ARC Subvention libre" programs. A.D. has been fully supported by INCa-DHOS, AGIRDOM, and ANR grants. D.G.B. was supported by grant R01 CA154365. The clinical research on lung cancer was funded by PNES POUMON INCA 2005 and BIOMARKSCAN PHRC 2003. The transcriptomic analyses were performed thanks to the program "Carte d'Identité des Tumeurs" supported by the Ligue Nationale Contre Le Cancer. S.R. and S.K. are recipients of a "contrat d'interface" from INSERM-Aviesan.

REFERENCES AND NOTES

1. Mellor J, Dudek P, Clynes D. A glimpse into the epigenetic landscape of gene regulation. Curr Opin Genet Dev. 2008; 18:116–122. [PubMed: 18295475]
2. Ong CT, Corces VG. Enhancer function: New insights into the regulation of tissue-specific gene expression. Nat Rev Genet. 2011; 12:283–293. [PubMed: 21358745]
3. Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. Nat Rev Genet. 2007; 8:286–298. [PubMed: 17339880]
4. Jones PA, Baylin SB. The epigenomics of cancer. Cell. 2007; 128:683–692. [PubMed: 17320506]
5. Berdasco M, Esteller M. Aberrant epigenetic landscape in cancer: How cellular identity goes awry. Dev Cell. 2010; 19:698–711. [PubMed: 21074720]
6. Bert SA, Robinson MD, Strbenac D, Statham AL, Song JZ, Hulf T, Sutherland RL, Coolen MW, Stirzaker C, Clark SJ. Regional activation of the cancer genome by long-range epigenetic remodeling. Cancer Cell. 2013; 23:9–22. [PubMed: 23245995]
7. Radu A, Pichon C, Camparo P, Antoine M, Allory Y, Couvelard A, Fromont G, Hai MT, Ghinea N. Expression of follicle-stimulating hormone receptor in tumor blood vessels. N Engl J Med. 2010; 363:1621–1630. [PubMed: 20961245]
8. Simpson AJ, Caballero OL, Jungbluth A, Chen YT, Old LJ. Cancer/testis antigens, gametogenesis and cancer. Nat Rev Cancer. 2005; 5:615–625. [PubMed: 16034368]
9. Weber M, Hellmann I, Stadler MB, Ramos L, Pääbo S, Rebhan M, Schübeler D. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet. 2007; 39:457–466. [PubMed: 17334365]
10. Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. A chromatin landmark and transcription initiation at most promoters in human cells. Cell. 2007; 130:77–88. [PubMed: 17632057]
11. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. High-resolution profiling of histone methylations in the human genome. Cell. 2007; 129:823–837. [PubMed: 17512414]
12. Rousseaux S, Khochbin S. New hypotheses for large-scale epigenome alterations in somatic cancer cells: A role for male germ-cell-specific regulators. Epigenomics. 2009; 1:153–161. [PubMed: 22122641]

13. Akers SN, Odunsi K, Karpf AR. Regulation of cancer germline antigen gene expression: Implications for cancer immunotherapy. Future Oncol. 2010; 6:717–732. [PubMed: 20465387]
14. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. CA Cancer J Clin. 2013; 63:11–30. [PubMed: 23335087]
15. Fang F, Turcan S, Rimner A, Kaufman A, Giri D, Morris LG, Shen R, Seshan V, Mo Q, Heguy A, Baylin SB, Ahuja N, Viale A, Massague J, Norton L, Vahdat LT, Moynahan ME, Chan TA. Breast cancer methylomes establish an epigenomic foundation for metastasis. Sci Transl Med. 2011; 3:75ra25.
16. Selamat SA, Chung BS, Girard L, Zhang W, Zhang Y, Campan M, Siegmund KD, Koss MN, Hagen JA, Lam WL, Lam S, Gazdar AF, Laird-Offringa IA. Genome-scale analysis of DNA methylation in lung adenocarcinoma and integration with mRNA expression. Genome Res. 2012; 22:1197–1211. [PubMed: 22613842]
17. 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, Grosveld F, Philipsen S. Gene expression-based classification of non-small cell lung carcinomas and survival prediction. PLoS One. 2010; 5:e10312. [PubMed: 20421987]
18. Lee ES, Son DS, Kim SH, Lee J, Jo J, Han J, Kim H, Lee HJ, Choi HY, Jung Y, Park M, Lim YS, Kim K, Shim Y, Kim BC, Lee K, Huh N, Ko C, Park K, Lee JW, Choi YS, Kim J. Prediction of recurrence-free survival in postoperative non–small cell lung cancer patients by using an integrated model of clinical information and gene expression. Clin Cancer Res. 2008; 14:7397–7404. [PubMed: 19010856]
19. Shedden K, Taylor JM, Enkemann SA, Tsao MS, Yeatman TJ, Gerald WL, Eschrich S, Jurisica I, Giordano TJ, Misek DE, Chang AC, Zhu CQ, Strumpf D, Hanash S, Shepherd FA, Ding K, Seymour L, Naoki K, Pennell N, Weir B, Verhaak R, Ladd-Acosta C, Golub T, Gruidl M, Sharma A, Szoke J, Zakowski M, Rusch V, Kris M, Viale A, Motoi N, Travis W, Conley B, Seshan VE, Meyerson M, Kuick R, Dobbin KK, Lively T, Jacobson JW, Beer DG. Director's Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma. Gene expression-based survival prediction in lung adenocarcinoma: A multi-site, blinded validation study. Nat Med. 2008; 14:822–827. [PubMed: 18641660]
20. Depenbrock H, Peter R, Faircloth GT, Manzanares I, Jimeno J, Hanauske AR. In vitro activity of aplidine, a new marine-derived anti-cancer compound, on freshly explanted clonogenic human tumour cells and haematopoietic precursor cells. Br J Cancer. 1998; 78:739–744. [PubMed: 9743292]
21. Nishino R, Takano A, Oshita H, Ishikawa N, Akiyama H, Ito H, Nakayama H, Miyagi Y, Tsuchiya E, Kohno N, Nakamura Y, Daigo Y. Identification of Epstein-Barr virus–induced gene 3 as a novel serum and tissue biomarker and a therapeutic target for lung cancer. Clin Cancer Res. 2011; 17:6272–6286. [PubMed: 21849417]
22. Grochola LF, Greither T, Taubert H, Möller P, Knippschild U, Udelnow A, Henne-Bruns D, Würl P. The stem cell-associated Hiwi gene in human adenocarcinoma of the pancreas: Expression and risk of tumour-related death. Br J Cancer. 2008; 99:1083–1088. [PubMed: 18781170]
23. Liu X, Sun Y, Guo J, Ma H, Li J, Dong B, Jin G, Zhang J, Wu J, Meng L, Shou C. Expression of hiwi gene in human gastric cancer was associated with proliferation of cancer cells. Int J Cancer. 2006; 118:1922–1929. [PubMed: 16287078]
24. Taubert H, Greither T, Kaushal D, Würl P, Bache M, Bartel F, Kehlen A, Lautenschläger C, Harris L, Kraemer K, Meye A, Kappler M, Schmidt H, Holzhausen HJ, Hauptmann S. Expression of the stem cell self-renewal gene Hiwi and risk of tumour-related death in patients with soft-tissue sarcoma. Oncogene. 2007; 26:1098–1100. [PubMed: 16953229]
25. Taubert H, Würl P, Greither T, Kappler M, Bache M, Bartel F, Kehlen A, Lautenschläger C, Harris LC, Kaushal D, Füssel S, Meye A, Böhnke A, Schmidt H, Holzhausen HJ, Hauptmann S. Stem cell-associated genes are extremely poor prognostic factors for soft-tissue sarcoma patients. Oncogene. 2007; 26:7170–7174. [PubMed: 17525744]
26. Tapparel C, Reymond A, Girardet C, Guillou L, Lyle R, Lamon C, Hutter P, Antonarakis SE. The TPTE gene family: Cellular expression, subcellular localization and alternative splicing. Gene. 2003; 323:189–199. [PubMed: 14659893]

27. Diskin SJ, Hou C, Glessner JT, Attiyeh EF, Laudenslager M, Bosse K, Cole K, Mossé YP, Wood A, Lynch JE, Pecor K, Diamond M, Winter C, Wang K, Kim C, Geiger EA, McGrady PW, Blakemore AI, London WB, Shaikh TH, Bradfield J, Grant SF, Li H, Devoto M, Rappaport ER, Hakonarson H, Maris JM. Copy number variation at 1q21.1 associated with neuroblastoma. Nature. 2009; 459:987–991. [PubMed: 19536264]

28. Riely GJ, Johnson ML, Medina C, Rizvi NA, Miller VA, Kris MG, Pietanza MC, Azzoli CG, Krug LM, Pao W, Ginsberg MS. A phase II trial of salirasib in patients with lung adenocarcinomas with KRAS mutations. J Thorac Oncol. 2011; 6:1435–1437. [PubMed: 21847063]

29. Eisen T, Thatcher N, Leyvraz S, Miller WH Jr, Couture F, Lorigan P, Lüthi F, Small D, Tanovic A, O'Brien M. Phase II study of weekly plitidepsin as second-line therapy for small cell lung cancer. Lung Cancer. 2009; 64:60–65. [PubMed: 18692272]

30. Peschel C, Hartmann JT, Schmittel A, Bokemeyer C, Schneller F, Keilholz U, Buchheidt D, Millan S, Izquierdo MA, Hofheinz RD. Phase II study of plitidepsin in pretreated patients with locally advanced or metastatic non-small cell lung cancer. Lung Cancer. 2008; 60:374–380. [PubMed: 18054408]

31. Boguski MS, Lowe TM, Tolstoshev CM. dbEST—Database for "expressed sequence tags". Nat Genet. 1993; 4:332–333. [PubMed: 8401577]

32. Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, Esteller M. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics. 2011; 6:692–702. [PubMed: 21593595]

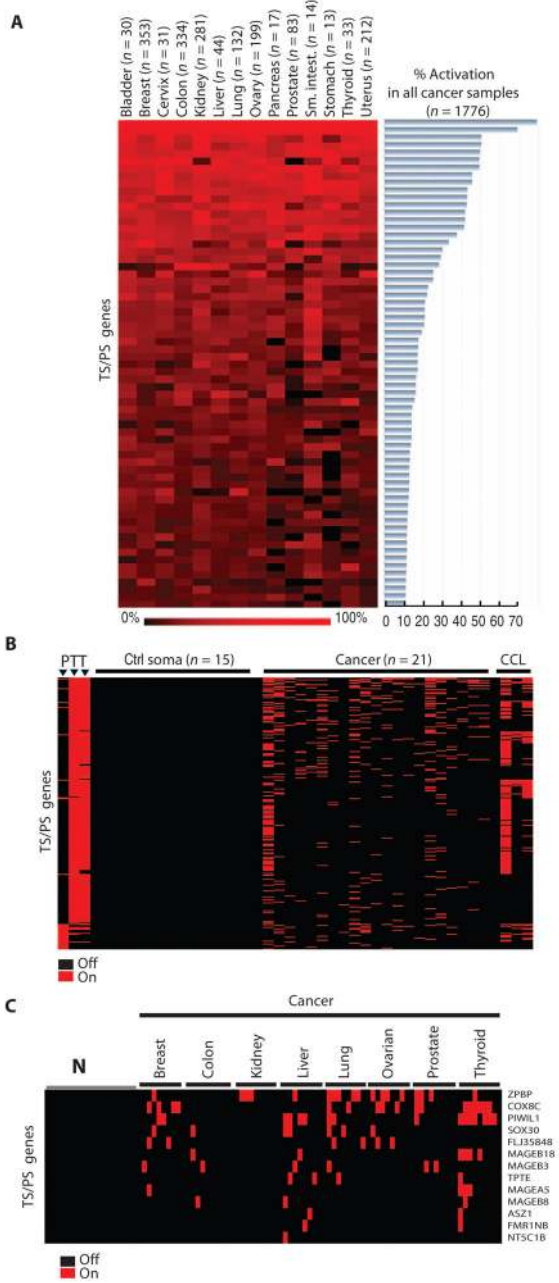

Fig. 1. Ectopic expressions of TS/PS genes detected in multiple cancers

(**A**) Frequencies of activation of TS/PS genes in 14 different types of solid tumors from our analysis of transcriptomic data available online obtained from 1776 solid tumor samples (GSE2109, described in table S3) shown here for the 65 most frequently activated genes on a black (0%) to red (100%) scale (heat map). Overall frequencies of gene activation in all tumor samples are presented in the histogram on the right. (**B**) Transcriptomic profiles of TS/PS genes on a dedicated microarray in normal human tissues and in tumor samples and cell lines: "P," "T," "Ctrl soma," "Cancer," and "CCL" respectively indicate placenta, testis ($n = 2$), adult somatic tissues, cancer samples of various origins, and cancer cell lines ($n =$

3). All samples are listed in table S5. Color code: black, no expression ("Off"); red, gene activation ("On"). (**C**) Expression of a selection of 13 frequently expressed TS/PS genes detected by qRT-PCR in 73 tumor samples of eight different origins (including breast, colon, kidney, liver, lung, ovarian, prostate, and thyroid). The corresponding nontumor samples ("N") are shown on the left part of the heat map in the same order.

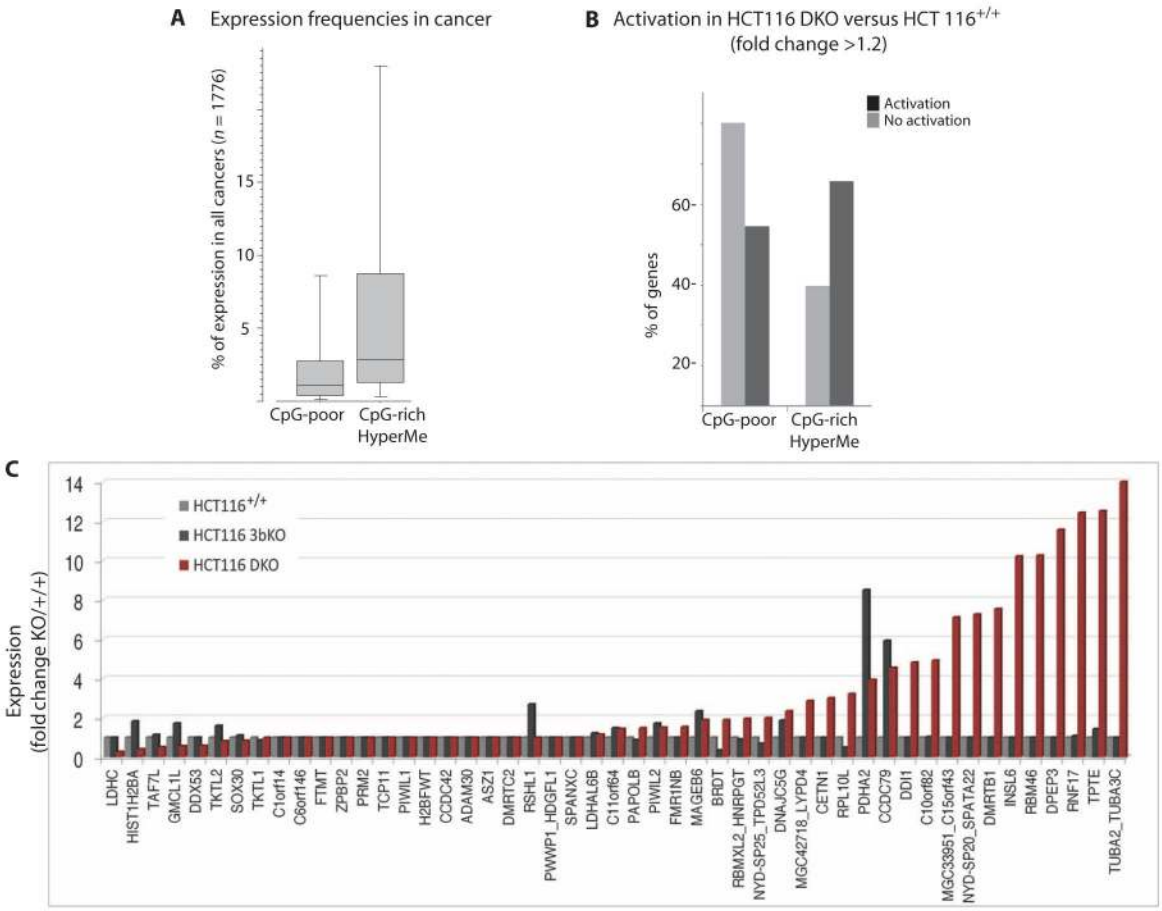

Fig. 2. TS/PS genes with hypermethylated promoters more susceptible to deregulation in cancer

(**A**) Box plots showing the respective distributions of activation frequencies in cancer (considering all 1776 cases of solid tumors analyzed in the data set GSE2109) of the two groups of genes, associated either with CpG-poor promoters (left) or with hypermethylated CpG-rich promoters (right). (**B**) Activation of TS/PS genes in response to DNA demethylation in HCT116 DNMT DKO (HCT116 cell line with double inactivation of DNMT1 and DNMT3b) compared to HCT116$^{+/+}$ (wild type) using our dedicated microarray; the histogram shows the numbers of genes activated (black) or not (gray) according to their promoter category, either CpG-poor or CpG-rich hypermethylated ("CpG-rich HyperMe"). (**C**) qRT-PCR detecting the expression of 49 TS/PS genes among those associated with a CpG-rich hypermethylated promoter (listed on the *x* axis of the histogram) in HCT116 cell line wild type (+/+, light gray bars), KO for DNMT3b (dark gray bars), and DKO for DNMT1 and DNMT3b (red bars); values are fold changes in reference to the normalized values obtained in HCT116$^{+/+}$.

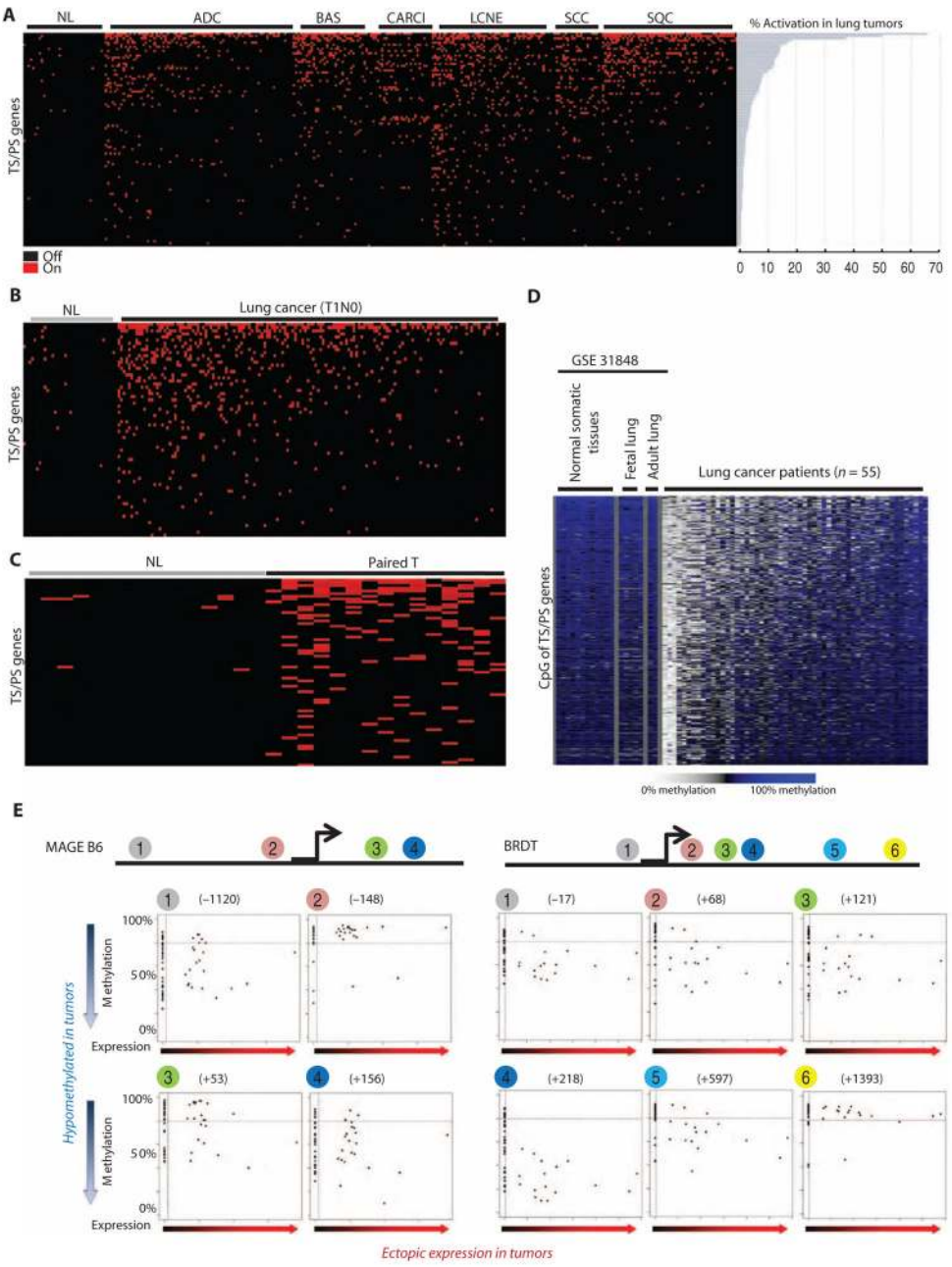

Fig. 3. Ectopic expression of TS/PS genes in the series of 293 lung cancer patients at all stages of the disease

(**A**) The heat map (left panel) shows the detection of TS/PS gene expression in all 293 patients (*x* axis) including adenocarcinoma (ADC), basaloid (BAS), carcinoid tumors (CARCI), large cell neuroendocrine tumors (LCNE), small cell carcinoma (SCC), and squamous cell carcinoma (SQC) histological subtypes, as well as in NL samples. Heat map color code: black, no activation; red, activation. The histogram (right panel) shows the frequency of lung cancer tumors (*x* axis, in %) aberrantly expressing each of the same TS/PS genes (*y* axis). (**B**) Heat map showing the expression of TS/PS genes focusing on the early lung cancer (T1N0) cases ($n = 152$) of the series, as described above; all patients are

included in (A), but here they are sorted by decreasing number of ectopically expressed genes (not by histological subtypes). (**C**) Heat map focusing on the expression of TS/PS genes in a subset of 15 paired tumor samples ("Paired T") and their corresponding NL [these samples are also shown in the previous two heat maps in (A) and (B), but in different order because they are of different TNM stages and histology]; this figure shows only the genes activated in this subset of patients. (**D**) Heat map showing the methylation levels (β values from 0 to 100% on a light gray to blue color scale) of 347 CpGs associated with the transcription start site (TSS) and 5′ untranslated regions of 88 TS/PS genes in normal somatic tissue samples (from left to right: mean methylation value in adipose tissue, adrenal gland, bladder, blood, brain, heart, lymph node, pancreas, skeletal muscle, spleen, stomach, ureter, and five fetal and two adult lung samples; data available on the GEO Web site under the reference GSE31848), as well as in 55 lung tumor samples of our series. β Values are shown in table S9. (**E**) Scatter plots corresponding to the individual CpGs localized near the TSS (−1500 to +1500 base pairs) of two genes, BRDT and MAGEB6, showing that the methylation levels (β values on the *y* axis) correlated with their respective expression levels ($\log_2$ ratios) in the 55 lung cancer patients. The positions of the CpGs relative to the TSS are indicated between brackets.

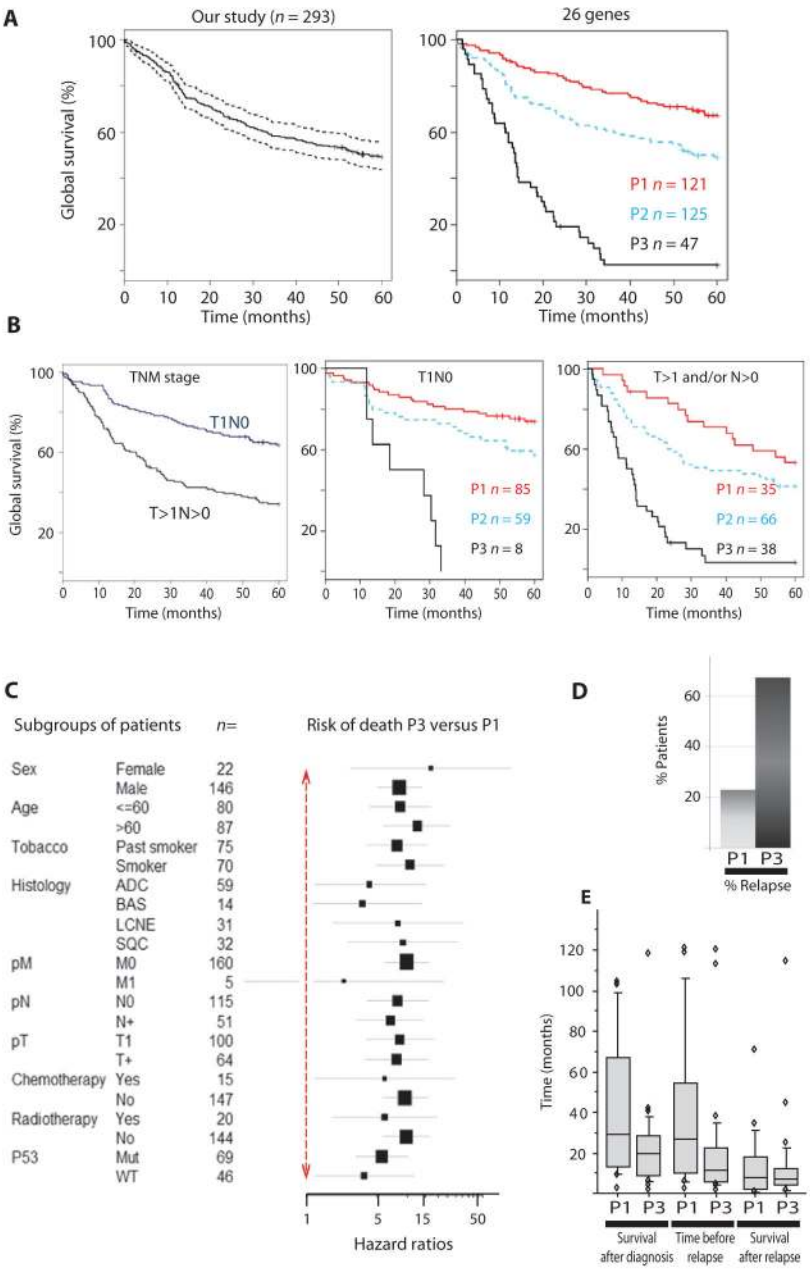

Fig. 4. Off-contextactivationsof26TS/PSgenesindependentlyassociatedwithpoorprognosis in lung cancer

(**A** and **B**) Cumulative global Kaplan-Meier survival estimates of the 293 patients in our series either grouped together (A, left panel) or divided into three groups according to the number of ectopic expressions found within the subset of 26 genes. The groups were defined as follows: P1 (no expression, red curve, $n = 121$), P2 (one or two expressed genes, blue curve, $n = 125$), and P3 (three or more ectopically expressed genes, black curve, $n = 47$) (A, right panel). (B) The left panel shows the survival probabilities of patients according to the TNM stage (as indicated). The middle panel shows the survival probabilities of the three groups (P1, P2, and P3) defined by our classifying genes, considering only the T1N0

patients. The same approach was used to classify the T>1/N>0 patients (right panel). (**C**) Forest plot of HRs (P3 versus P1; on a log scale) for overallriskofdeath(more than5 years). A univariateCoxproportional hazard model estimated HRs for the overall risk of death. The horizontal lines provide the 95% confidence interval for the ratios. The vertical red dotted double-arrow line corresponds to an HR of 1. (**D**) Histograms showing the frequencies of relapse (local recurrence and/or metastasis) observed in P1 and P3 patients. (**E**) Box plots showing the distribution of times (in months) for P1 and P3 patients, corresponding to overall survival (left), survival before relapse (middle), and survival after the diagnosis of relapse (right).

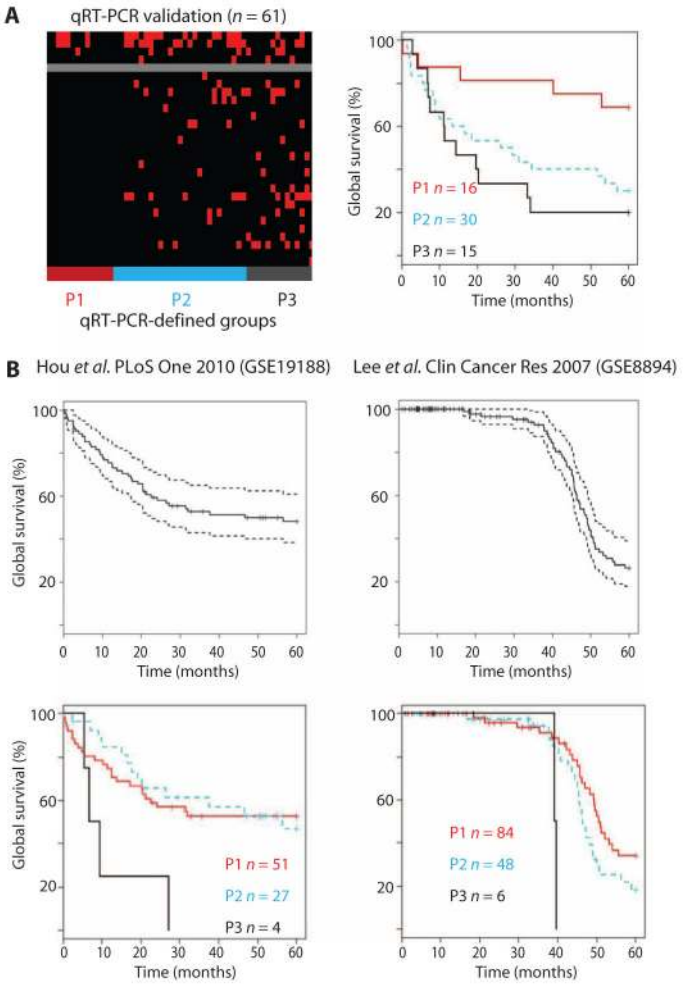

Fig. 5. Validation of the 26-gene prognosis-classifying strategy for lung cancer patients

(**A**) qRT-PCR detection of the expression of our classifying genes. The heat map (left) shows the detection of four frequently expressed TS/PS genes (upper panel) and of the prognosis-classifying genes (lower panel) in a subset of 61 patients from our lung cancer patient series. The survival curves (right) compare the survival probabilities between patients assigned to the P1, P2, and P3 groups by qRT-PCR. (**B**) Cumulative global Kaplan-Meier survival estimates of the patients from two external lung cancer populations either combined (top panels: the continuous lines show the mean survival probability, and the dotted lines correspond to the 95th percentile) or divided into the three groups P1, P2, and P3 defined as in Fig. 4A.

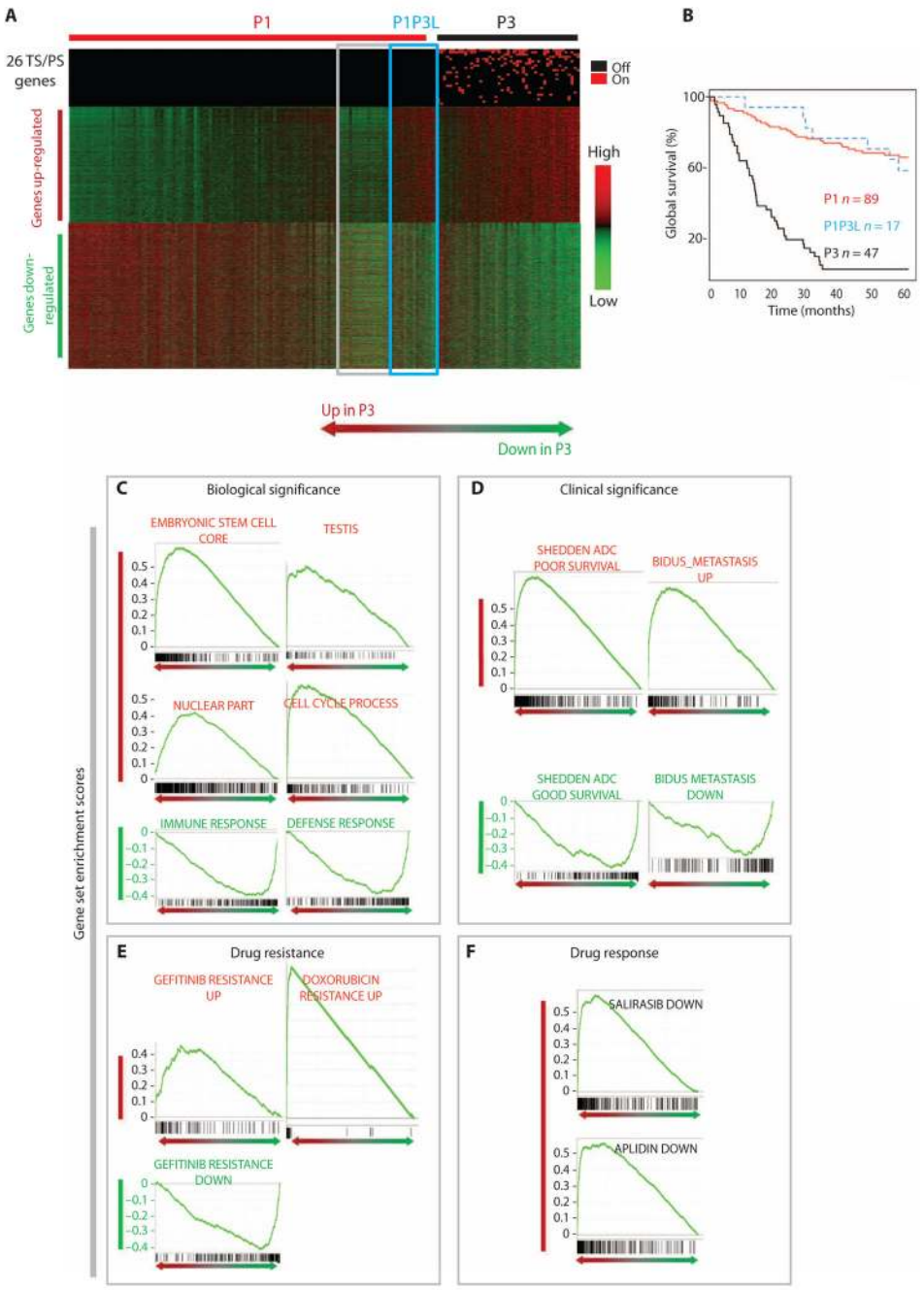

Fig. 6. Biological and clinical characteristics of aggressive lung tumors revealed by differential expression profiling

(**A**) Heat maps showing the expression of the 26 TS/PS genes (upper panel), as well as genes up-regulated (middle part) and down-regulated (lower part) in aggressive tumors (P3) compared with the "good prognostic" group of tumors (P1). The robust multiarray average normalized values of expression of these genes are represented in the indicated color scales (green, low expression; red, high expression); the patients, represented on the *x* axis, were classified by prognostic groups and ranked by increasing value of the differential expression between up-and down-regulated genes; genes were ranked by decreasing difference of

expression values between P1 and P3 samples. Blue frame: subset of the patients classified as group P1 presenting a P3-like molecular profile (P1P3L); gray frame: carcinoid tumors displaying an atypical molecular profile. (**B**) The Kaplan-Meier curves represent the respective survival probabilities of the subgroups of P1 patients with the P3-like expression profile (P1P3L, blue curve), other P1 patients (red curve), and P3 patients (black curve). (**C** to **F**) Enrichment plots displaying the normalized enriched scores of some of the highly significant overlapping gene sets identified with gene set enrichment analysis (GSEA). The green curve shows the running enrichment score (*y* axis) for the gene set as the analysis walks down the ranked list of genes (*x* axis). The black bars along the *x* axis represent the genes of the gene set, ranked according to their fold change of expression in the "P3 versus P1" transcriptomic analysis (from left to right: up- to down-regulated genes).