www.impactjournals.com/oncotarget/ Oncotarget, 2018, Vol. 9, (No. 10), pp: 9137-9155

Research Paper

Differentially expressed microRNAs in lung adenocarcinoma invert effects of copy number aberrations of prognostic genes

Tomas Tokar[1], Chiara Pastrello[1], Varune R. Ramnarine[1,2], Chang-Qi Zhu[1], Kenneth J. Craddock[1], Larrisa A. Pikor[3], Emily A. Vucic[3], Simon Vary[1,4,5], Frances A. Shepherd[1], Ming-Sound Tsao[1,6,7], Wan L. Lam[3] and Igor Jurisica[1,6,8,9]

[1]Princess Margaret Cancer Centre, University Health Network, Toronto, Canada

[2]The Vancouver Prostate Centre, Vancouver General Hospital, Vancouver, Canada

[3]Department of Integrative Oncology, British Columbia Cancer Research Centre, Vancouver, Canada

[4]Mathematical Institute, University of Oxford, Oxford, United Kingdom

[5]Faculty of Mathematics, Physics and Informatics, Comenius University, Bratislava, Slovakia

[6]Department of Medical Biophysics, University of Toronto, Toronto, Canada

[7]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Canada

[8]Department of Computer Science, University of Toronto, Toronto, Canada

[9]Institute of Neuroimmunology, Slovak Academy of Sciences, Bratislava, Slovakia

Correspondence to: *Igor Jurisica,* ***email:*** *juris@ai.utoronto.ca*

Keywords: *lung adenocarcinoma; copy number aberrations; microRNA; gene regulatory network; prognostic signature*

Received: *August 25, 2017* ***Accepted:*** *January 02, 2018* ***Published:*** *January 08, 2018*

ABSTRACT

In many cancers, significantly down- or upregulated genes are found within chromosomal regions with DNA copy number alteration opposite to the expression changes. Generally, this paradox has been overlooked as noise, but can potentially be a consequence of interference of epigenetic regulatory mechanisms, including microRNA-mediated control of mRNA levels.

To explore potential associations between microRNAs and paradoxes in non-small-cell lung cancer (NSCLC) we curated and analyzed lung adenocarcinoma (LUAD) data, comprising gene expressions, copy number aberrations (CNAs) and microRNA expressions. We integrated data from 1,062 tumor samples and 241 normal lung samples, including newly-generated array comparative genomic hybridization (aCGH) data from 63 LUAD samples.

We identified 85 "paradoxical" genes whose differential expression consistently contrasted with aberrations of their copy numbers. Paradoxical status of 70 out of 85 genes was validated on sample-wise basis using The Cancer Genome Atlas (TCGA) LUAD data. Of these, 41 genes are prognostic and form a clinically relevant signature, which we validated on three independent datasets. By meta-analysis of results from 9 LUAD microRNA expression studies we identified 24 consistently-deregulated microRNAs. Using TCGA-LUAD data we showed that deregulation of 19 of these microRNAs explains differential expression of the paradoxical genes.

Our results show that deregulation of paradoxical genes is crucial in LUAD and their expression pattern is maintained epigenetically, defying gene copy number status.

INTRODUCTION

Integration of array comparative genomic hybridization (aCGH) with mRNA microarray data has revealed significant associations between occurrence of copy number aberrations (CNAs) and differential gene expression in diverse cancers [1–7]. However, significantly downregulated genes have been often found to reside within chromosomal regions with increased number of copies (gains) and *vice versa*, creating a paradoxical signal. For example, Phillips *et al.* reported that 14% of the genes downregulated in prostate cancer reside within regions of DNA copy number gains, and approximately 9% of upregulated ones reside in regions of DNA copy number loss [1]. Usually, this paradox is ignored as a noise, but can potentially be a consequence of interference of other regulatory mechanisms controlling mRNA transcription [8].

In recent years, the cancer research community has investigated how epigenetic regulators, known as microRNAs (miRNAs), form prognostic signatures and affect regulatory pathways that can lead to tumorigenesis. miRNAs are short non-coding RNAs that regulate the translation of mRNA by serving as guide molecules in mRNA silencing, mediated by various associated proteins [9]. Targeting most protein-coding transcripts [10], miRNAs are involved in diverse biological processes, including development and homeostasis [11, 12]. Moreover, growing evidence implicates miRNAs as factors associated with major human pathologies, including cancer [13–16].

In 2012 approximately 13% of all new cancer cases worldwide were cancers of lung (and bronchus), making lung cancer one of the most frequent cancer type (surpassed only by breast cancer in women) [17]. Despite smoking cessation, and advances in detection and treatment, lung cancer remains the main cause of cancer-related death worldwide for both men and women [18]. With nearly 160,000 deaths annually it kills more people than other common cancers combined, including colon, breast and prostate [18]. The most common type of lung cancer is lung adenocarcinoma (LUAD), comprising approximately 45% of all lung cancer cases [19, 20].

In this paper, we integratively analyzed gene expression and CNA data from 12 publicly available LUAD datasets, and new CNA data obtained from 63 LUAD samples profiled at our institution (Figure 1). By combining and analyzing data from 1,062 tumor tissue

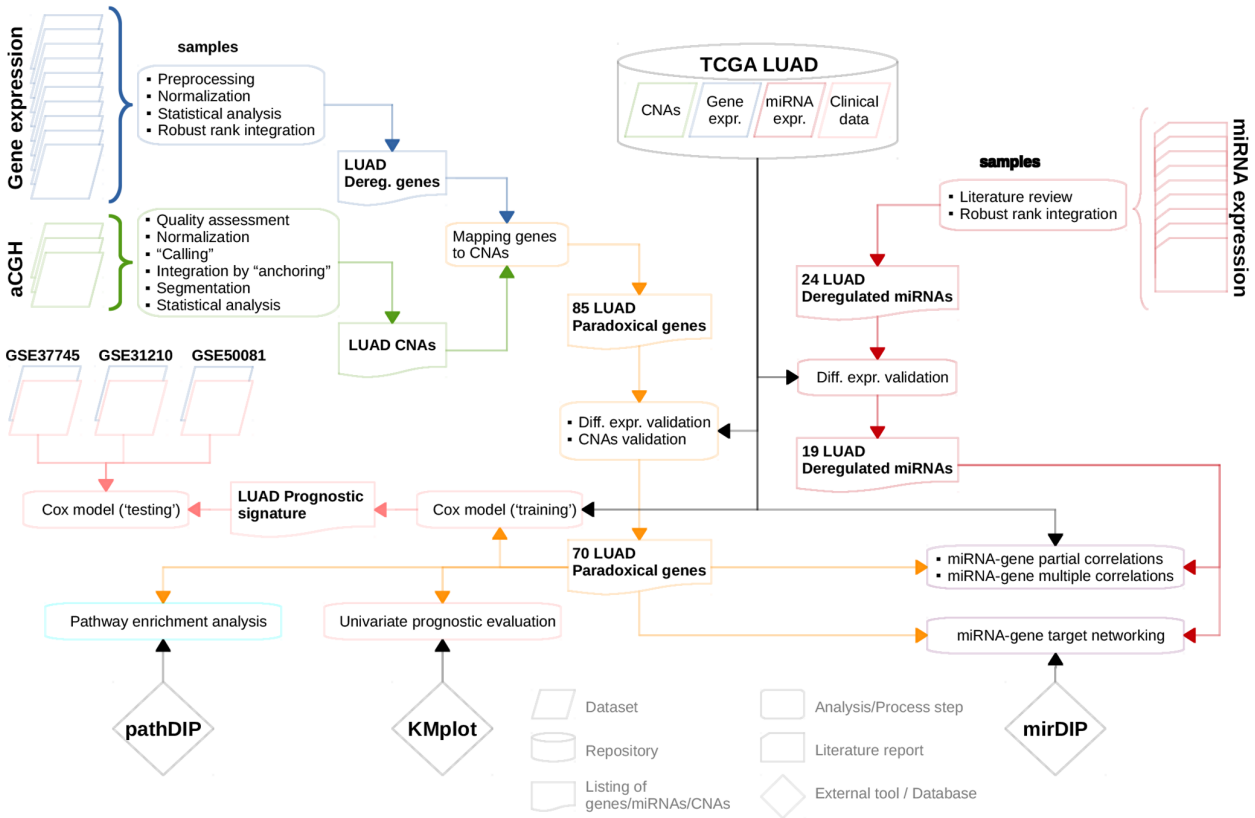

Figure 1: Flowchart depicting sequence of analyses/computational steps as performed and datasets as used. For more details see Materials and Methods section.

samples and 241 normal samples we identified genes whose differential expression consistently was in contrast to aberrations of their copy numbers. Paradoxical status of these genes was then validated on the sample level, using TCGA LUAD gene expression and CNA data. Furthermore, to assess whether the paradoxical expression patterns were caused by epigenetic disruptions in lung tumors, we compiled miRNA expression data from 406 LUAD samples and 321 normal lung samples. Using miRNA:gene associations from mirDIP [21] and the measure of co-expression, we showed that this paradox can be explained by 19 miRNAs consistently deregulated in LUAD.

RESULTS

Identification of the paradoxical genes

First, we examined the frequency and statistical significance of autosomal CNAs across 3 LUAD aCGH datasets, including our new data and two publicly available datasets (see Materials and Methods). We identified multiple chromosomal regions with significantly ($p < 0.05$, randomization test) high frequency of gains (more than two copies) or losses (less than two copies); frequencies and corresponding p-values are listed in Supplementary Data 1. The most extensive positive aberrations identified occur on the q-arm of chromosomes 1, 7 and 8 as well the p-arm of chromosomes 5 and 7 (Figure 2A). The most significant copy number losses occurred on the q-arm of chromosomes 6, 9, 13, 15, and 18, along with the p-arm of chromosomes 8 and 9.

To identify genes whose differential expression remained consistent across patient cohorts, we performed integrative analysis of 10 publicly available gene expression datasets, comprising 740 LUAD samples and 241 normal tissue samples. Among 15,323 genes that were subjected to the robust rank analysis, we identified 1,309 genes that were significantly deregulated across the datasets ($p < 0.01$, robust rank aggregation), where 701 of these were downregulated genes, and 608 are upregulated. Excluding 9 non-protein-coding genes, reduced the numbers to 600 upregulated and 700 downregulated genes (see Supplementary Data 2). Non-protein coding genes involve downregulated *C17orf91*, and eight upregulated miRNA sequences: MIR7112, MIR6847, MIR7113, MIR671, MIR4647, MIR93, MIR25, MIR4721, all of which are intragenic miRNAs residing within upregulated genes. In subsequent meta-analysis of miRNA expression in LUAD, we found none of these miRNAs to be significantly deregulated.

We identified 132 (14.6%) downregulated genes residing in regions with decreased number of copies and 102 (22%) upregulated ones residing in regions with increased number of copies ($p < 2.2E-16$, Chi-squared test). Importantly, 63 consistently downregulated and 22 consistently upregulated genes reside on chromosomal regions with opposite direction of aberration – gains and losses, respectively (Figure 2). Hereafter, we refer to these 85 genes as paradoxical genes (Supplementary Table 1).

Validation of the CNAs and differential expression of the paradoxical genes

While we identified paradoxical genes using data from diverse patient cohorts, we sought to validate our findings in an independent, homogeneous datasets. We used data from TCGA comprising: CNA, mRNA-seq and miRNA-seq LUAD data from 514 LUAD and 57 normal samples. This dataset was selected for validation solely on the basis of the CNA, mRNA and miRNA expression data availability, without considering clinicopathological characteristics of the data. Five of the 85 paradoxical genes could not be evaluated due to the missing CNA or expression data. From the remaining 80, we successfully validated 70 genes ($p < 1E-4$, randomization test), whose copy aberration status and differential expression confirmed results from the integrative analysis (Supplementary Figure 1, Supplementary Table 1). Further analysis only considers these 70 validated paradoxical genes.

To test whether paradoxical deregulation occurs in the individual samples, we measured frequencies of paradoxical co-occurrence of up-/downregulation (expression z-score >/< +/-1.647) and losses/gains (log2 CNA >/< +/-0.2) of the paradoxical genes across individual TCGA LUAD samples (Figure 3). We found that frequency of paradoxical deregulation ranges from 9% (NFKBIA) to 74% (NPR1) of samples, median frequency equal to 46% and mean 44%. For all the 70 genes frequency of paradoxical deregulation exceeds the frequency of regular (non-paradoxical) deregulation.

Co-occurrence between deregulated miRNAs and paradoxical genes

We hypothesize that differential expression of paradoxical genes can be to a large extent explained by deregulation of the miRNAs that target these genes, either directly or through regulatory mediators, such as transcription factors. We thus performed meta-analysis of 9 papers reporting differentially expressed miRNAs in LUAD to identify consistently deregulated miRNAs. We found 24 such miRNAs ($p < 0.05$, robust rank analysis, see Methods section), 13 of which were upregulated (hsa-mir-21, hsa-mir-182, hsa-mir-210, hsa-mir-9, hsa-mir-183, hsa-mir-135b, hsa-mir-130b, hsa-mir-200b, hsa-mir-191, hsa-mir-31, hsa-mir-196b, hsa-mir-196a, hsa-mir-200a, ordered by significance) and 11 downregulated (hsa-mir-126, hsa-mir-145, hsa-mir-30a, hsa-mir-218, hsa-mir-139, hsa-mir-195, hsa-mir-486, hsa-

mir-143, hsa-mir-144, hsa-mir-34c, hsa-mir-16) (Figure 4, Supplementary Data 3).

We then tested co-occurrence between the miRNAs deregulation and emergence of the paradoxical genes within the same cohort of patients. We first assessed differential expression of the 24 consistently deregulated miRNAs using TCGA-LUAD miRNA-seq data. Nineteen significantly deregulated miRNAs (out of 24, p = 1.18E-8, hypergeometric test) validated the results of the meta-analysis (hsa-mir-130b, hsa-mir-135b, hsa-mir-139, hsa-mir-143, hsa-mir-144, hsa-mir-182, hsa-mir-183, hsa-mir-195, hsa-mir-196a, hsa-mir-196b, hsa-mir-200a, hsa-mir-200b, hsa-mir-21, hsa-mir-210, hsa-mir-218, hsa-mir-30a, hsa-mir-31, hsa-mir-486, hsa-mir-9). While five miRNAs (hsa-mir-191, -34c, -126, -145, -16) did not validate; hsa-mir-191, -34c failed to pass the validation criteria only due to insufficient expression fold change, although their expression was altered significantly (Supplementary Table 2). Therefore, only 19 validated miRNAs are used for further analysis.

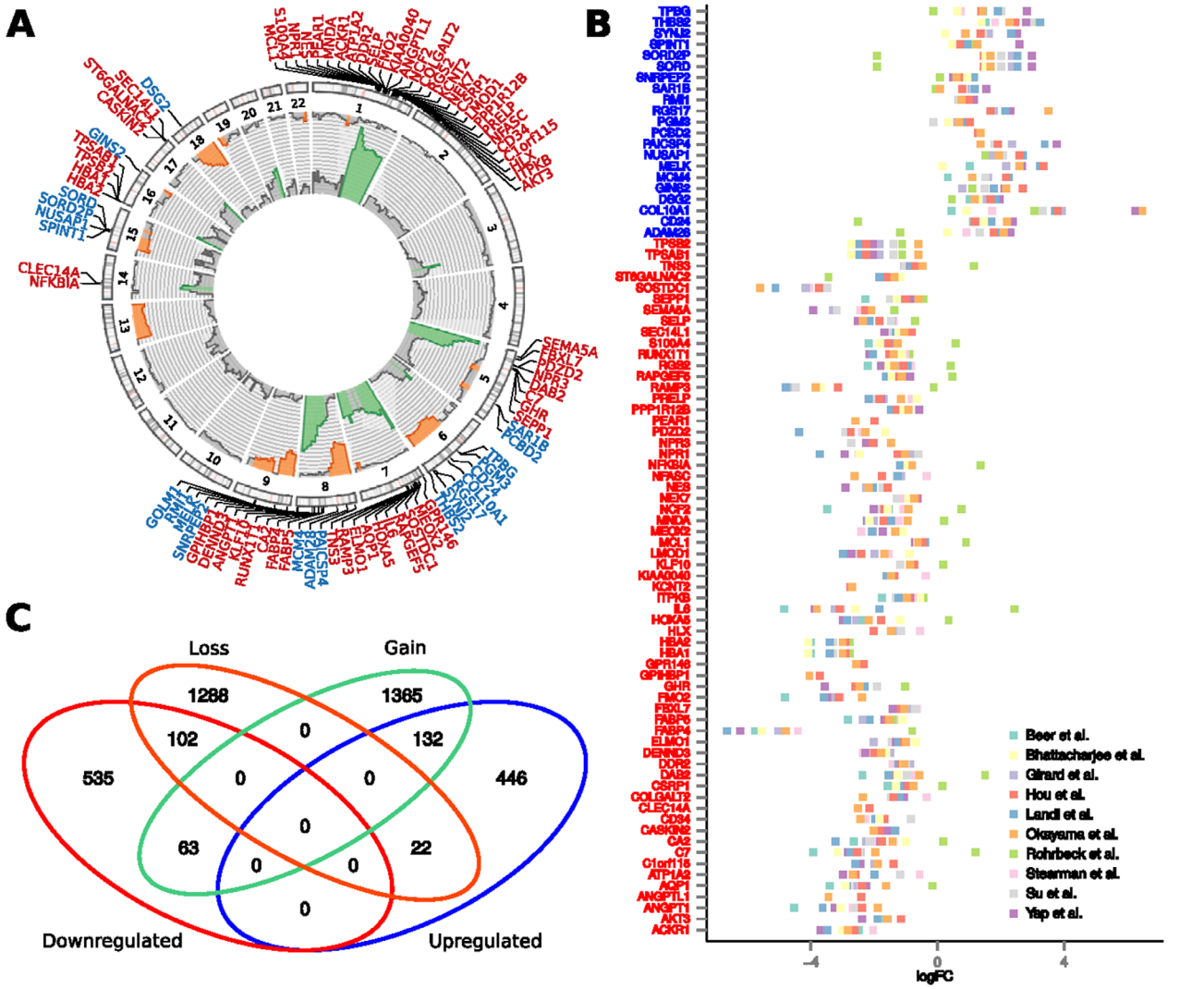

Figure 2: Association between chromosomal copy number aberrations and differential expression of genes. (A) Frequencies of gains (pointing outbound) and losses (pointing inbound) of the given chromosomal region as obtained from integrative analysis of three aCGH datasets. The aberration frequencies are depicted in range from 0-50% and regions with significant frequency of aberrations are highlighted by color (orange – losses, green – gains). Precise chromosomal locations of these paradoxical genes are depicted in the circular plot. **(B)** Tumor-vs-normal expression fold change of the paradoxical genes, obtained across 10 publicly available datasets. In Figures A and B, symbols of the downregulated genes are labeled red, while the symbols of the upregulated genes are labeled blue. **(C)** Venn diagram showing overlaps between up-/downregulated genes and genes residing within the regions of chromosomal copy number gain, or loss.

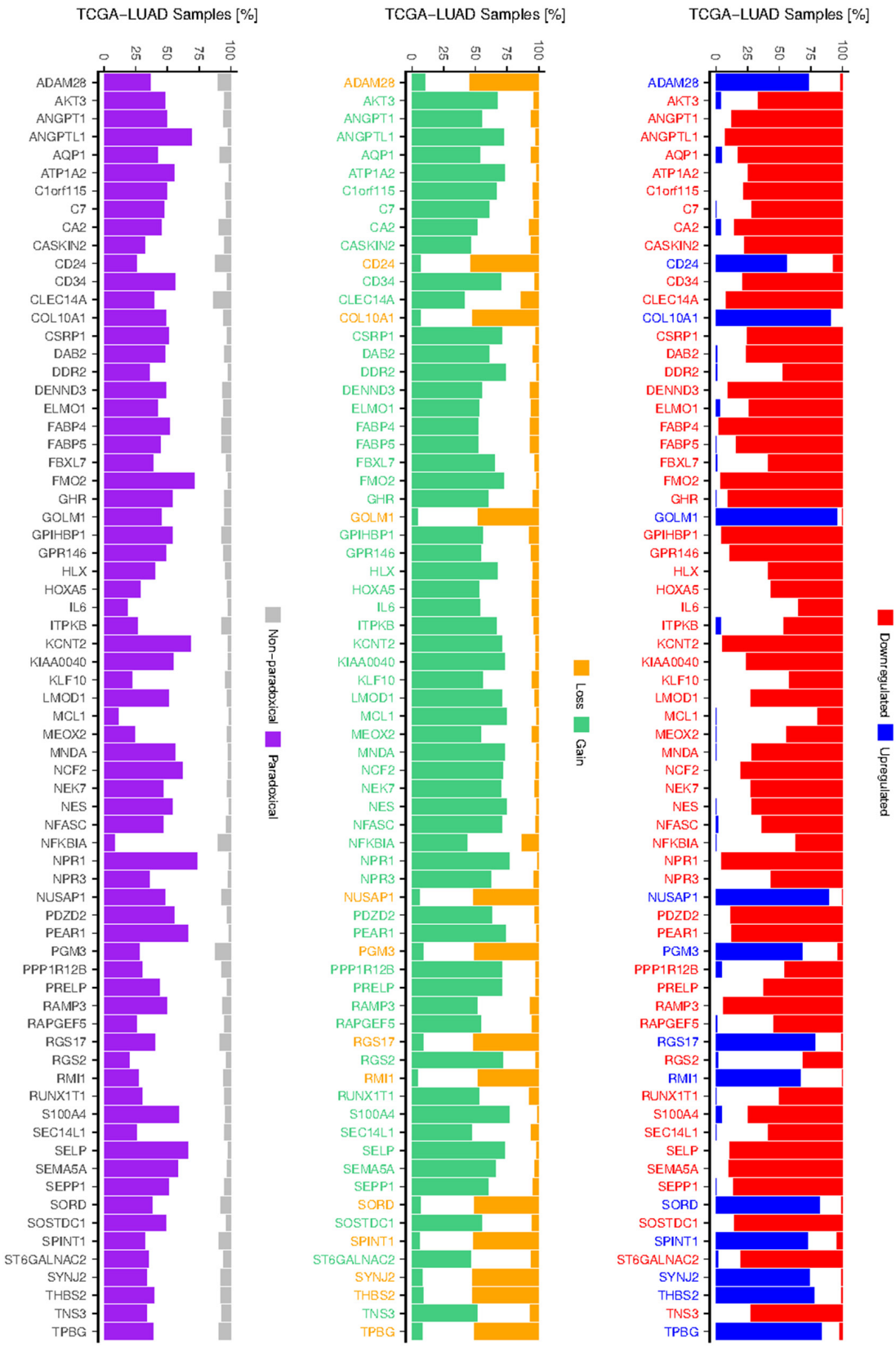

Figure 3: Frequency of deregulation and CNAs of paradoxical genes. Barplot at the left shows frequencies of paradoxical and non-paradoxical co-occurrence of deregulated expression and CNAs. Barplots depicting frequencies of up- and downregulation (middle), gain and losses (right) of 70 validated paradoxical genes, as occur across TCGA LUAD samples. Colors of the gene labels indicate their deregulation/CNA status as obtained from the integrative analysis.

Association between deregulated miRNAs and paradoxical genes

According to mirDIP [21], 46 of the 70 paradoxical genes (65.7%) are targeted by subsets of the 19 deregulated miRNAs (p = 4.8E-3, hypergeometric test; Figure 5). Moreover, for 35 of these 46 targeted paradoxical genes, we found at least one deregulated miRNA that targets the given gene with its expression status contrasting the expression status of the given gene, implying that paradoxical expression of this gene could be explained by the miRNA deregulation.

To further asses the association between expression of individual miRNAs and paradoxical genes we calculated partial correlation between them [22], using copy number status of the paradoxical genes as a controlling variable. We found 369 significantly co-expressed miRNA:gene pairs (27% of all miRNA:gene combinations), 362 (98.1%) of which are explanatory, i.e., there is a positive correlation between miRNAs and genes deregulated in the same direction, or negative correlation between inversely deregulated ones (Figure 6A). We found 64 paradoxical genes (91.4%) whose correlation with at least one of the 19 validated miRNAs is among the top 5% of the correlations measured between 10E+4 random pairs of genes and miRNAs, and whose paradoxical expression can be explained by deregulation of the upstream miRNA.

As downstream effects of the deregulation of individual miRNAs may be combined, we aimed to evaluate potential associations between expression of individual paradoxical genes and *en bloc* expression of the deregulated miRNAs. We calculated coefficients of multiple (multivariate) correlation (CMC) between the paradoxical genes and deregulated miRNAs. The value of CMC can be interpreted as the correlation between dependent variable (gene expression) and its best prediction that can be computed linearly from the set of independent variables (expression of miRNAs). We found 23 paradoxical genes (32.9%) with CMCs in the top 5% of the values of the same measure as calculated across 17,745 genes covered by TCGA-LUAD RNA-seq data (p = 1.7E-14, hypergeometric test; Figure 6B).

Clinical significance of the paradoxical genes

Using KMplot [23], we assessed the association of the 70 paradoxical genes with patient disease-free survival. We found 41 (58.6%) of these genes as significantly associated with survival (FDR < 0.05). We assume that down-regulation of significantly positive genes (HR < 1, FDR < 0.05) as well as the up-regulation of significantly negative ones (HR > 1, FDR < 0.05) worsens the survival prognosis. Under this assumption, with the exception of three genes (ELMO1, DENND3,

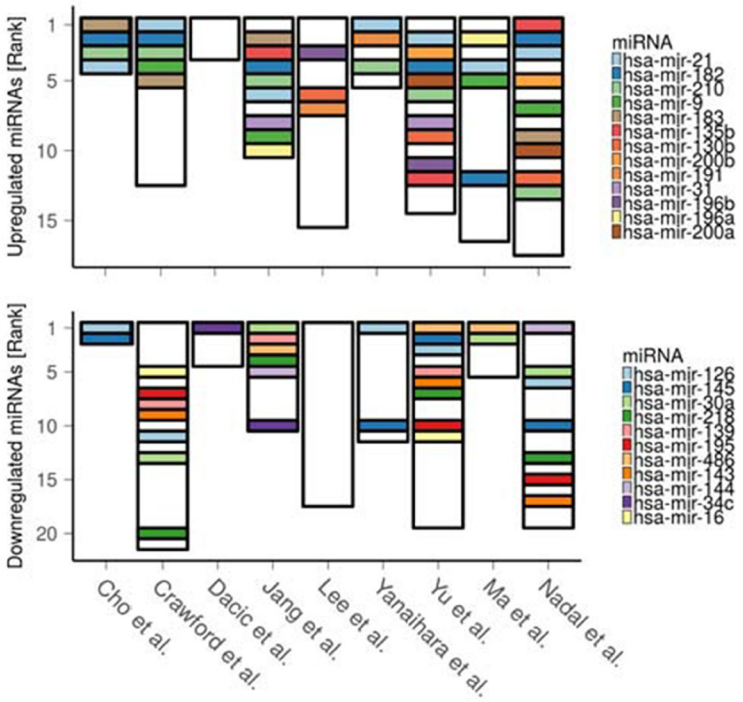

Figure 4: Ranking of the differentially expressed miRNAs as reported across 9 LUAD miRNA studies. The lower the rank the greater the reported significance (and/or expression fold change) of the corresponding miRNA. Height of the bars denotes total number of reported miRNAs in each study.

SPINT1), the actual deregulation of paradoxical genes is associated with a negative impact on patient survival (Figure 7A; Supplementary Table 3). This implies that the gene expression paradoxes we identified here mostly worsen the prognosis of LUAD patients.

Using TCGA-LUAD RNA-seq and matching clinical data, we constructed a multivariate Cox prognostic model, where expressions of the paradoxical genes served as prognostic variables. The model was validated using three independent publicly available gene expression datasets and associated clinical data: Botling *et al.* [24], Okayama *et al.* [25] and Der *et al.* [26] (Figures 7B-7D). The resulting concordance index, area under ROC curve, hazard ratio between risk groups and associated P-value, demonstrated robust prognostic potential of paradoxical genes signature.

Pathway enrichment analysis of the paradoxical genes

To elucidate biological functions of the paradoxical genes we performed a comprehensive pathway enrichment analysis. Using Pathway Data Integration Portal (pathDIP) [27] we identified 22 pathways significantly enriched by the 70 paradoxical genes (FDR < 0.05, hypergeometric test). A list of all pathways and respective gene memberships is provided in Supplementary Data 4. Interestingly, several of the enriched pathways are related to lipid metabolism and signaling (adipogenesis, LPA receptor mediated events, regulation of lypolysis) are key players in carcinogenesis [28]. Moreover, several enriched guidance molecule pathways (ephrin signaling, semaphorin interactions, integrin, DCC-mediated attractive signaling)

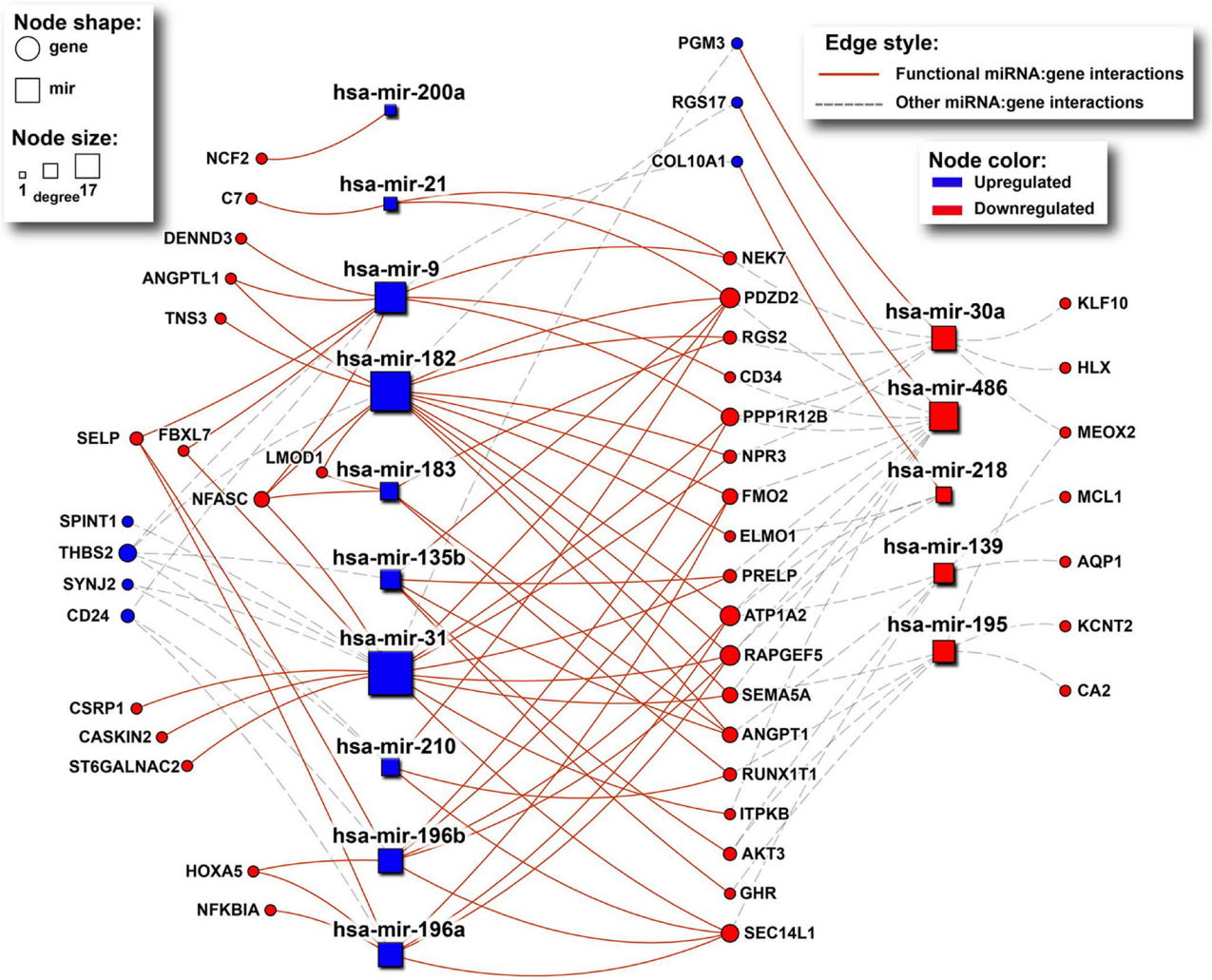

Figure 5: The network of interactions between deregulated miRNAs and their paradoxical gene targets as obtained from mirDIP. Rectangles and circles represent miRNAs and genes, respectively. Red color denotes downregulated transcripts, while blue denotes upregulated ones. Size of nodes corresponds to number of interactions (degree). Solid red lines indicate miRNA:gene interactions between inversely deregulated transcripts, indicating potential causal associations.

are noted as cancer-drug targets [29]. PTEN-dependent cell cycle arrest, apoptosis pathway as well as hemostasis are known to play a role in cancer [30–32].

DISCUSSION

Integrating three LUAD aCGH datasets, we identified several chromosomal regions with extensive copy number aberrations. Weir and colleagues [33] reported similar profile of copy number aberrations in 371 LUAD samples, confirming gains on 1q, 5p, 7p, 7q and 8q as well as deletions on 6q, 8p, 9p, 9q, 13q, 18q (Table 1). Lee *et al.* [34] obtained similar results using Molecular Inversion Probe assays on 12 LUAD samples, confirming gains on 1q, 5p, 7p, 7q and 8q and losses on 6q, 8p, 18q (but not losses on 9p, 9q, 13q and 15q). While the individual aberrations vary greatly among individuals, as even the most frequent aberrations appear only in less than 50% of samples, the overall CNA profile of LUAD is conserved across the patient cohorts.

By integrative analysis of multiple gene expression and copy number datasets, we found significant association between CNA status and differential expression of genes. Similar associations were previously reported in other cancer types [2–5, 7]. However, we also discovered 85 paradoxical genes whose expression was in opposite direction to their CNA. Seventy of these genes were subsequently validated across a homogeneous

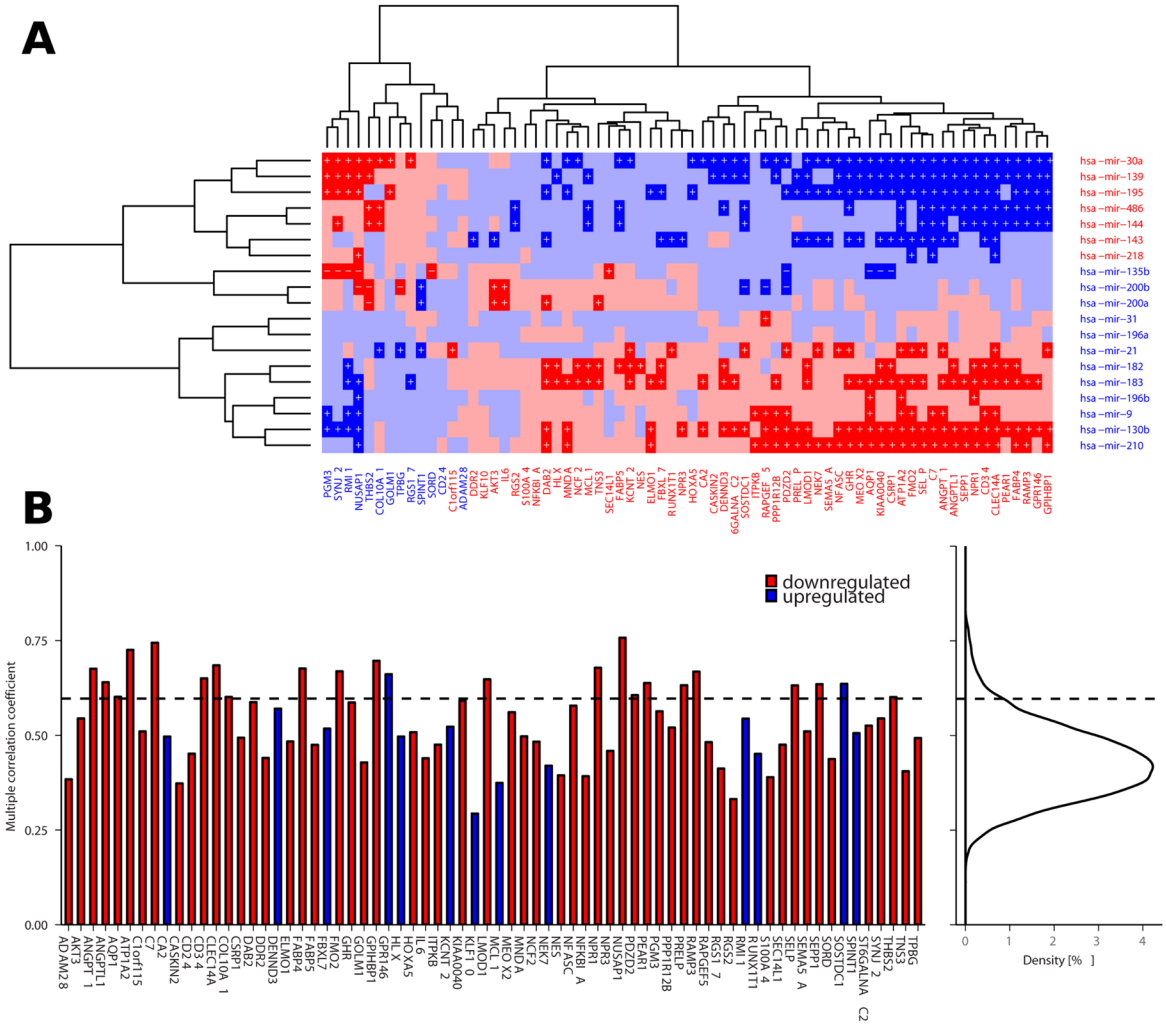

Figure 6: Correlation between deregulated miRNAs and paradoxical genes. (A) Partial correlations between deregulated miRNAs and paradoxical genes as measured across TCGA LUAD data (red denotes negative correlation, blue denotes positive correlation, darker shade indicates significant correlations, $p < 0.05$). Plus signs denote partial correlation with causal explanation of gene deregulation, minus signs denote correlations that are significant but non-explanatory. **(B)** Barplot showing multiple correlations between paradoxical genes and *en block* deregulated miRNAs as calculated across TCGA LUAD data. Curve on the right depicts distribution of the same measure across all the genes in the TCGA LUAD data. Dashed line denotes 95th percentile of the distribution; there are 23 (32.9%) paradoxical genes whose multiple correlation coefficient falls among the top 5% of the highest values.

LUAD data cohort from TCGA. Paradoxical expression of these genes was validated even on the individual samples, proving that such paradoxical gene expression is a well preserved feature of the molecular profile of LUAD.

Expression of paradoxical genes is associated with miRNAs consistently deregulated across multiple LUAD patient cohorts. Deregulation of these miRNAs inverts the effects of the genomic aberrations occurring in tumors. This is demonstrated by significant overlap between paradoxical genes and targets of these miRNAs, as well as by the two methods we applied here to measure correlation between the paradoxical genes and given miRNAs. Although correlation does not imply causality, the results of our analysis strongly suggest that differential expression of paradoxical genes is caused by deregulation of miRNAs.

We tested prognostic relevance of the paradoxical genes and found 41 (58.6%) paradoxical genes significantly associated with patient survival. Paradoxical expression of 39 genes has negative impact on prognosis. We also developed paradoxical gene signature that was validated on a three independent validation datasets [24–26].

While the majority of the paradoxical genes are novel and their association with LUAD prognosis has not been investigated thoroughly, there are 17 validated paradoxical genes that have been previously associated with prognosis of other cancers: *ADAM28* [35, 36], *ANGPT1* [37], *CA2* [38], *CA24* [39], *DAB2* [40], *HOXA5* [41, 42], *IL6* [43], *KLF10* [44], *MCL1* [45], *NES* [46], *NUSAP1* [47], *PDZD2* [48], *RGS17* [49, 50], *RUNX1T1* [51, 52], *SELC14L1* [53], *SEMA5A* [54] and *SEPP1* [55] (Table 2).

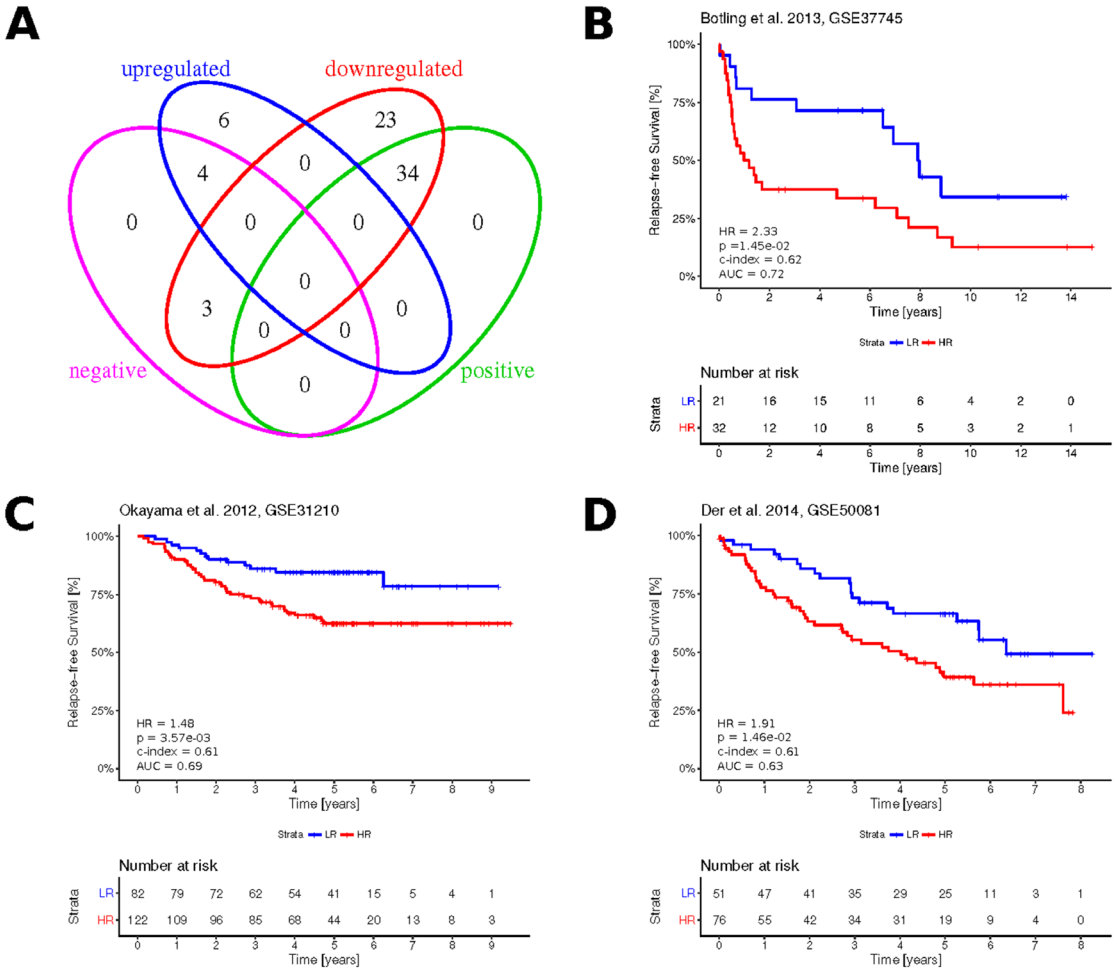

Figure 7: Clinical significance of the paradoxical genes. (A) Venn diagram showing overlapping subsets of paradoxical genes with up-/ downregulated expression and subsets with positive (HR < 1, FDR < 0.05) and negative (HR > 1, FDR < 0.05) association with LUAD prognosis, as obtained from KMplot. **(B–D)** Kaplan-Meier plots showing survival curves in the three independent validation cohorts, as stratified based on the Cox proportional hazards calculated from paradoxical genes expression. Numbers in the bottom left, indicate resulting hazard ratio (HR), associated statistical significance of patient stratification (p), concordance index (c-index), area under ROC curve (AUC) calculated at five years.

Table 1: List of the large scale copy number aberrations in LUAD

CNA	Arm	Frequency [%]	Weir et al., 2007 [33]	Lee et al., 2012 [34]
Gain:	1q	30-45	x	x
	5p	22-45	x	x
	7p	19-34	x	x
	7q	16-20	x	x
	8q	15-35	x	x
Loss:	6q	7-13	x	x
	8p	7-26	x	x
	9p	13-20	x	
	9q	7-12	x	
	13q	7-12	x	
	15q	7-8	x	
	18q	8-21	x	x

Frequency indicates range of relative number of aberrations occurring across the span of the given region. Checkmarks indicate aberrations confirmed by the above studies.

Table 2: List of validated paradoxical genes whose association with cancer prognosis has previously been reported

Gene	Association to cancer prognosis	Ref.
ADAM28	Overexpression correlates with cell proliferation and lymph node metastasis	[35]
	Serological and histochemical marker for NSCLC	[36]
ANGPT1	Role in the prognosis of patients with oral squamous cell cancer	[37]
CA2	Downregulated in gastric cancer and proposed as an independent prognostic factor for patient survival	[38]
CD24	Expression at early stages of breast cancer indicates a highly invasive tumor	[39]
DAB2	An important tumor suppressor, frequently downregulated in various tumors	[40]
HOXA5	Downregulation is associated with poor prognosis in NSCLC	[41]
	Shown to prevent tumor progression and metastasis in colon cancer	[42]
IL6	SNP associated with risk of multiple cancers	[43]
KLF10	Potential clinical predictor for progression of pancreatic cancer	[44]
MCL1	Key molecule for acquiring epithelial-to-mesenchymal transition-associated chemo-resistance in NSCLC	[45]
NES	Marker of cancer stem cells	[46]
NUSAP1	Encodes protein that is proposed biomarker for prostate cancer recurrence	[47]
PDZD2	Shown to induce senescence or quiescence of prostate, breast and liver cancer cells via transcriptional activation of p53	[48]
RGS17	Induces tumor cell proliferation lung and prostate cancers	[49]
	Regulator of cell survival and chemoresistance in ovarian cancer	[50]
RUNX1T1	Predictor of liver metastasis in pancreatic endocrine tumours	[51]
	Associated with proliferation and senescence inhibition in t(8;21)-positive leukaemic cells	[52]
SELC14L1	Proposed marker for predicting prognosis and progression of prostate cancer	[53]
SEMA5A	Over-expressed pancreatic cancer cells, regulates tumorigenesis, proliferation, invasion and metastasis, and serve as a target for diagnosis and treatment of pancreatic cancer	[54]
SEPP1	Inversely related to pancreatic cancer risk	[55]

Table 3: Summary of the public datasets used in this study

	Author & Year	No. of samples (total/normal)	Source & Notes
CNAs:	Chitale et al., 2009 [93]	199	http://cbio.mskcc.org/public/
	Job et al., 2010 [97]	60	E-TABM-926, ArrayExpress
Gene expression:	Bhattacharjee et al., 2001 [98]	207/17	http://www.broadinstitute.org/MPR/lung/
	Beer et al., 2002 [99]	96/10	GSE68571
	Stearman et al., 2005 [100]	39/19	GSE2514
	Yap et al., 2005 [101]	58/9	E-MEXP-231, ArrayExpress
	Su et al., 2007 [102]	54/27	GSE7670
	Landi et al., 2008 [103]	107/50	GSE10072
	Rohrbeck et al., 2008 [85]	15/5	GSE6044
	Hou et al., 2010 [86]	109/64	GSE19188
	Girard et al., 2011	50/20	GSE31547, Unpublished
	Okayama et al., 2012 [25]	246/20	GSE31210
miRNA expression:	Yanaihara et al., 2006 [87]	208/104	
	Cho et al., 2009 [88]	20/10	
	Crawford et al., 2009 [89]	20/8	
	Dacic et al., 2010 [90]	12/6	
	Yu et al., 2010 [91]	40/20	
	Lee et al., 2011 [92]	12/6	
	Jang et al., 2012 [94]	206/103	
	Ma et al., 2014 [95]	108/54	
	Nadal et al., 2014 [96]	101/10	

From the panel of consistently deregulated miRNAs, several were recently associated with cancer progression and prognosis. Most notably, miRNAs hsa-mir-21 and -196b are well established oncomirs [56], and hsa-mir-196a is known to be associated with NSCLC [57]. The role of hsa-mir-30a varies across human cancer types [58]. Hsa-mir-135b reverses chemoresistance of NSCLC [59]. Hsa-mir-139 is associated with aggressive tumor behavior and disease progression in breast cancer [60], and is believed to inhibit bladder cancer proliferation and self-renewal [61]. Hsa-mir-143 inhibits tumor growth of breast cancer [62]. Hsa-mir-144 has been shown to induce cell cycle arrest and apoptosis in pancreatic cancer cells [63]. Hsa-mir-182 promotes prostate cancer progression [64]. Hsa-mir-195 inhibits the proliferation and invasion of pancreatic cancer cells [65]. Hsa-mir-218 downregulation contributes to epithelial-mesenchymal transition and tumor metastasis in lung cancer [66]. Hsa-mir-130b is a part of the new prognostic marker for patient risk assessment and as an indicator of therapy resistance in prostate cancer [67]. Similarly, hsa-mir-183 in combination with hsa-mir-19b were recently proposed as biomarkers of lung cancer [68], and miRNAs hsa-mir-145 and -9 as biomarkers for early-stage cervical cancer [69].

Genomic instability is one of the cancer hallmarks [70] that results in CNAs and differential expression of various genes. However, paradoxical genes with expression patterns opposite to gene dosage status often are dismissed as noise and overlooked in genome-wide cancer gene discovery efforts [8]. Prognostic significance of the paradoxical genes suggests that their deregulation in LUAD is crucial for cancer progression and is maintained by the cancer cells despite the CNAs affecting expression of these genes in an inverse manner. We found that deregulation of the paradoxical genes is maintained at the epigenetic level by a group of deregulated miRNAs. These

findings highlight the importance of integrative analysis, which combines information across diverse types of high-throughput data.

MATERIALS AND METHODS

Copy number aberrations: sample collection, preparation and data processing

The samples used in this study were from the banked resected tumors collected in the BR.10 adjuvant chemotherapy trial [71]. The study has received approval by the Institutional Research Ethics Board. A total of 142 formalin-fixed paraffin embedded (FFPE) and 16 snap-frozen samples were included. Haematoxylin and eosin stained slides from FFPE blocks first were reviewed by a lung pathologist to locate tumor rich areas (tumor cellularity > 60%), and then the block was cored at this area. Cored specimens were de-paraffinized by incubation in xylene overnight, and then washed with ethanol and air dried. Qiagen ATL buffer (QIAamp© DNA extraction kit cat. 51306, Germantown, MD) were added. Specimens were digested by proteinase K at 55°C overnight at 450rpm (Eppendorf© Thermomixer R, Fisher Scientific). DNA isolation followed the manufacturer's protocol (Qiagen, Cat. 51306, Germantown, MD). Samples of isolated genomic DNA were quantified by Nanodrop 1000 (Thermo Scientific, Wilmington, DE) and electrophoresed in 0.8% agarose gel to visualize DNA size distribution. Severely degraded samples (80% of DNA fragments with size < 20bp) were excluded. Eight out of 142 FFPE and none of the snap-frozen samples were excluded. The final cohort contains 134 FFPE and 16 snap-frozen samples.

Test and reference DNA were labeled using Cy3 and Cy5 dCTPs respectively; 200 ng of genomic DNA was labeled using the BioPrime DNA labeling system (Invitrogen). Prior to hybridization, test and reference labeled DNA were combined and purified using a ProbeQuant Sephadex G-50 Column (Amersham, GE Healthcare Life Sciences, Chicago, IL) to remove unincorporated nucleotides. Then 100 µg of Human Cot-1 DNA (Invitrogen) was added to the labeled sample prior to precipitation with 0.1 volume 3M sodium acetate and 2.5 volumes of ethanol. The DNA pellet was resuspended in 20 µl DIG Easy hybridization solution (Roche, Indianapolis, IN), 2.5 µl (20 µg/µl) sheared herring sperm DNA and 2.5 µl (100µg/µl) yeast tRNA (Calbiochem, San Diego, CA). DNA was denatured at 85°C for 10 minutes and repetitive sequences were blocked at 37°C for one hour prior to hybridization.

Prehybridization was carried out using 20 µl DIG Easy hybridization buffer (Roche), 2.5 µl 10% BSA and 2.5 µl (20 µg/µl) sheared herring sperm DNA, at 45°C for 1 hour. Hybridization was carried out at 45°C for 24-48 hours. Arrays were washed for 5 x 5 min., in 0.1 x SSC, 0.1% SDS at room temperature in the dark with agitation. Each array was then rinsed 5 times in a clean slide box containing 0.1 x SSC with agitation. Slides were then dried with (oil free) nitrogen air stream and stored in the dark until imaging.

Array image capture and data normalization were performed as previously described [72]. Briefly, post-hybridization arrays were scanned using a CCD-based imaging system (Virtek ChipReader), and quantitated using Soft-Worx Tracker spot analysis software (Applied Precision, Issaquah, WA).

Data were log2 transformed, and replicate clones having standard deviations > 0.075 or signal-to-noise ratios in each dye channel of < 3 were filtered out. A multi-step normalization was then performed to control for biases caused by the array (e.g., spatial biases or differences in background signal), the dyes used for labeling, or the DNA sample quality [73, 74]. The amount of "copycat'" correction required for each sample was plotted in a histogram of all samples; those that required too much correction and did not lie within a normal distribution were deemed to be poor quality DNA, and were eliminated from analysis. By these criteria, 35 samples were eliminated, leaving 115 samples (including 63 LUAD samples used here) from 113 patients for further analysis. Data from all 115 samples are publicly accessible through: http://ophid.utoronto.ca/aCGH/.

Analysis of the copy number aberrations data

In addition to our newly-produced CNA data, we analyzed two publicly available aCGH datasets acquired from LUAD tumor samples (see Table 3). Each of the public datasets was first normalized, segmented and additionally underwent post-segmentation normalization using methods provided by Bioconductor package CGHcall (v2.22.0) [75]. All three datasets then underwent a "calling" process using the CGHcall method from the same package, converting the continuous log-ratios on each probe, to one of the three discrete values (calls) corresponding to: (i) decreased number of copies (loss), (ii) normal copy and (iii) increased number of copies (gain) [75].

As the individual datasets come with different probesets, obtained copy numbers calls correspond to different chromosomal segments and cannot be compared directly. We then integrated results acquired from individual datasets by assigning a set of chromosomal positioning "anchors" that comprised the starts and ends of chromosomal locations of all the probes in the four datasets, as described by Guo et al. [76]. Then for each anchor, if the anchor was within the chromosomal location of the probe from any of the datasets, acquired vector of states corresponding to the probe were assigned to this anchor. Conversely, if an anchor was outside of any of the probes of the given dataset, a vector of missing values was created and assigned to the anchor. Anchors with missing

values from more than one datasets were removed. Number of losses and gains were calculated across the anchors and their statistical significance was evaluated by p-values, calculated by comparing actual gains/losses counts to those obtained from 10^6 permutations.

Analysis of gene expression datasets

We analyzed 10 publicly available gene expression datasets (Table 1) which satisfied the following criteria: (i) were originally from studies on tissue samples from surgically resected human LUAD tumors, (ii) contained at least one sample of noncancerous normal tissue for comparison, and (iii) were produced by using Affymetrix platforms to enable uniform processing and analysis of all the datasets. We first normalized and summarized each dataset by Gene Chip Robust Multiarray Averaging (gcrma, v2.38.0) [77]. For each individual dataset, we then evaluated differential expression of the genes using Bioconductor package limma (v3.32.7) [78]. Based on the expression fold change, genes were classified as either up- or downregulated, and then ranked according to statistical significance, which was evaluated by FDR-adjusted p-value. Analyzing 10 datasets resulted in 10 rankings for upregulated genes and 10 for downregulated ones. To identify consistently deregulated genes, obtained rankings were subjected to robust rank aggregation analysis implemented in R package RobustRankAggreg (v1.1) [79]. This analysis detects genes that are ranked consistently better than expected under the null hypothesis of uncorrelated inputs, and assigns a p-value as a significance score for each gene. The stability of the resulting significance scores was assessed by the leave-one-out correction, in which the same analysis was repeated 10 times, each time excluding one of the rankings. Acquired p-values from each round were averaged into a corrected p-value. Genes whose significance score was smaller than chosen threshold (corrected $p < 0.01$) were further considered as consistently significantly deregulated genes.

Meta-analysis of the miRNA expression

Compared to gene expression studies, fewer miRNA expression profiles from LUAD are available, and various platforms are used, often including custom arrays. Therefore, to provide analysis of miRNA expression in LUAD, instead of acquiring and processing expression data, we summarized reported results of 9 published miRNA expression studies (Table 1). Full text and (if applicable) supplementary data of each of the studies were carefully examined, and miRNAs with significantly altered expression were selected for further analysis. miRNA names were standardized according to miRBase (release 21) [80]. All miRNAs were classified as either up- or downregulated and ranked according to their reported statistical significance. If this was not reported, expression fold change was used instead. Examining 9 studies we obtained 9 rankings for upregulated miRNAs and 9 for downregulated ones. Analogously to gene expression analysis described in the previous section, obtained miRNA rankings subsequently were subjected to the robust rank aggregation analysis and leave-one-out correction of the obtained p-values. miRNAs whose significance score was smaller than a chosen threshold (corrected $p < 0.05$), comprised the resulting list of consistently significantly deregulated miRNAs.

Acquiring chromosomal locations and copy number status of the deregulated genes

Using Bioconductor package biomaRt (v2.22) [81] we determined the chromosomal locations of the deregulated genes from the Ensembl (v75, Feb. 2014) database and compared these locations with chromosomal coordinates of the aberrant regions. Deregulated genes whose chromosomal locations overlapped with aberrant regions were counted and statistical significance of the association between the aberrations and differential gene expression was then evaluated using Chi-square test.

Identification of the miRNA-target pairs

We used microRNA Data Integration Portal, v2.3.2.0 (mirDIP; http://ophid.utoronto.ca/mirDIP) [21] to acquire data on human miRNAs and their respective targets. mirDIP integrates data from 14 miRNA resources and supports a search for miRNA-target pairs under user-defined filters, including a number of independent confirmations of given pairs, confidence criteria, etc. We restricted our search to only miRNA-target pairs that fell among the top third of the most confident predictions from at least two different sources. miRNA names were standardized as described above, and symbols of their gene targets were standardized by HGNC symbol checker (http://www.genenames.org, version from September 2015). To assess the significance of overlap between targets of deregulated miRNAs and paradoxical genes, we performed hypergeometric testing, using 15,323 genes that were subjected to robust rank analysis as a population and 7,836 of these genes that, according to mirDIP are targeted by at least one of the deregulated miRNAs as a number of "successes" in the population.

Calculation of the miRNA:target partial correlation and multiple correlation coefficients

Partial correlation coefficients between gene and miRNA expressions were calculated using R package ppcor v1.1, using copy number of the given gene as a third – controlling variable. Statistical significance of the obtained values was calculated by two-sided comparison

with distribution of the same measure obtained across 10^4 random miRNA:gene pairs. To distinguish whether correlations between miRNA and given genes may also imply causal association, we compared the sign of the correlation and copy number status of the miRNA and gene. If the miRNA and gene were deregulated in the same direction, mutual correlation must be positive to indicate a causal association. If the miRNA and genes were inversely deregulated, mutual correlation must be negative to indicate causal association.

Multiple correlation coefficients between gene and *en bloc* miRNA expression C was calculated as follows:

$$C^2 = c^T R^{-1} c$$

where c denotes a vector of Pearson coefficients of correlation between a given genes and miRNA expressions, R denotes a matrix of Pearson coefficients of correlation between miRNA expressions.

Evaluation of the prognostic significance of the paradoxical genes

Prognostic properties of the individual genes were evaluated by KMplot (http://kmplot.com/analysis/) [23], version 2015, using only LUAD patient data and corresponding disease-free survival censored at 10 years. If multiple probe sets were mapped to the same gene, we used only JetSet probes mapping to a given gene. Obtained hazard ratios (HR) and associated p-values were then summarized and multiple testing adjustment of the p-values was subsequently computed using false discovery rate (FDR) method.

To evaluate the multivariate prognostic potential of the paradoxical genes we developed a Cox proportional hazards model, where expressions of 70 validated paradoxical genes served as covariates. The model was derived using R package glmnet [82] (v2.0.2), applying LASSO (L1) regularization to prevent over-fitting. TCGA-LUAD RNA-seq data were standardized by converting to z-scores and along with the corresponding clinical data were used as "training data". The resulting model was then validated on three independent datasets, and its predictive performance was first evaluated by a concordance index (function survConcordance from R package survival [83], v2.38.3), and an area under receiver operating characteristics curve (AUC), measured at the fifth year after initial time point (function AUC.cd from the R survAUC package, v 1.0.5). Patients were then separated into two groups based on the predicted risk score, using its 40^{th} percentile as a threshold. This threshold was selected based on ROC analysis of the model using training data. Validated HR between these two groups, as well as associated statistical significance (log-rank test) were calculated (function survdiff from the survival package) and Kaplan-Meier survival curves of both groups were plotted (for more details see [84]).

Pathway enrichment analysis

Using Pathway Data Integration Portal v2.5.1.2 (http://ophid.utoronto.ca/pathDIP), we performed comprehensive pathway enrichment analysis across 20 major pathway databases [27]. We considered literature curated gene:pathway memberships as well as those predicted according to experimentally detected protein-protein interactions (including interactions experimentally detected between orthologues plus FpClass interactions with minimum confidence level for predicted associations equal 0.95; for more details see pathDIP documentation).

Abbreviations

AUC: area under receiver operating characteristics; aCGH: array comparative genomic hybridization; CMC: coefficient of multiple correlation; CNAs: copy number aberrations; FDR: false discovery rate; FFPE: formalin-fixed paraffin-embedded; GCRMA: gene chip robust multiarray averaging; HR: hazard ratio; LUAD: lung adenocarcinoma; TCGA: the cancer genome atlas.

Author contributions

TT – conception and design of the work, data collection, data analysis and interpretation, data visualization, drafting and critical revision of the article, final approval of the version to be published; CP - conception and design of the work, critical revision of the article; VR - conception and design of the work; CZ – data collection, data analysis and interpretation, critical revision of the article; KC – data collection, critical revision of the article; LP – data collection, critical revision of the article; EV - data collection, critical revision of the article; SV - data analysis and interpretation, critical revision of the article; FS - project management; critical revision of the article; MT – project management; critical revision of the article; WL - project management; critical revision of the article; IJ - conception and design of the work, data visualization, drafting and critical revision of the article, project management, final approval of the version to be published.

ACKNOWLEDGMENTS

SV would like to acknowledge support from travel bursary covering visit to Jurisica Lab from Faculty of Mathematics, Physics and Informatics, Comenius University, Bratislava. The authors would like to thank Richard Lu, Wing Xie, Adrian M. Teisanu and Dan Strumpf for the original development of mirDIP and CDIP portals.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

FUNDING

This work was supported in part by Ontario Research Fund (ORF #GL2-01-030), Canada Research Chair Program (CRC #203373 and #225404), Canada Foundation for Innovation (CFI #12301, #29272, #30865), and IBM. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

1. Phillips JL, Hayward SW, Wang Y, Vasselli J, Pavlovich C, Padilla-Nash H, Pezullo JR, Ghadimi BM, Grossfeld GD, Rivera A, Linehan WM, Cunha GR, Ried T. The consequences of chromosomal aneuploidy on gene expression profiles in a cell line model for prostate carcinogenesis. Cancer Res. 2001; 61:8143–9.
2. Hyman E, Kauraniemi P, Hautaniemi S, Wolf M, Mousses S, Rozenblum E, Ringnér M, Sauter G, Monni O, Elkahloun A, Kallioniemi OP, Kallioniemi A. Impact of DNA amplification on gene expression patterns in breast cancer. Cancer Res. 2002; 62:6240–5.
3. Pollack JR, Sørlie T, Perou CM, Rees CA, Jeffrey SS, Lonning PE, Tibshirani R, Botstein D, Børresen-Dale AL, Brown PO. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. Proc Natl Acad Sci. 2002; 99:12963–8.
4. Järvinen AK, Autio R, Haapa-Paananen S, Wolf M, Saarela M, Grenman R, Leivo I, Kallioniemi O, Mäkitie AA, Monni O. Identification of target genes in laryngeal squamous cell carcinoma by high-resolution copy number and gene expression microarray analyses. Oncogene. 2006; 25:6997–7008.
5. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C, Tyler-Smith C, Carter N, Scherer SW, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science. 2007; 315:848–53.
6. Chari R, Thu KL, Wilson IM, Lockwood WW, Lonergan KM, Coe BP, Malloff CA, Gazdar AF, Lam S, Garnis C, MacAulay CE, Alvarez CE, Lam WL. Integrating the multiple dimensions of genomic and epigenomic landscapes of cancer. Cancer Metastasis Rev. 2010; 29:73–93.
7. Peifer M, Fernández-Cuesta L, Sos ML, George J, Seidel D, Kasper LH, Plenker D, Leenders F, Sun R, Zander T, Menon R, Koker M, Dahmen I, et al. Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. Nat Genet. 2012; 44:1104.
8. Huang N, Shah PK, Li C. Lessons from a decade of integrating cancer copy number alterations with gene expression profiles. Brief Bioinform. 2012; 13:305–16.
9. Huntzinger E, Izaurralde E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat Rev Genet. 2011; 12:99–110.
10. Ha M, Kim VN. Regulation of microRNA biogenesis. Nat Rev Mol Cell Biol. 2014; 15:509–24.
11. Carrington JC, Ambros V. Role of microRNAs in plant and animal development. Science. 2003; 301:336–8.
12. Becker-Santos DD, Thu KL, English JC, Pikor LA, Martinez VD, Zhang M, Vucic EA, Luk MT, Carraro A, Korbelik J, Piga D, Lhomme NM, Tsay MJ, et al. Developmental transcription factor NFIB is a putative target of oncofetal miRNAs and is associated with tumour aggressiveness in lung adenocarcinoma. J Pathol. 2016; 240:161–72.
13. Lin PY, Yu SL, Yang PC. MicroRNA in lung cancer. Br J Cancer. 2010; 103:1144–8.
14. Thu KL, Chari R, Lockwood WW, Lam S, Lam WL. miR-101 DNA copy loss is a prominent subtype specific event in lung cancer. J Thorac Oncol. 2011; 6:1594–8.
15. Enfield KSS, Pikor LA, Martinez VD, Lam WL. Mechanistic roles of noncoding RNAs in lung cancer biology and their clinical implications. Genet Res Int. 2012; 2012.
16. Di Leva G, Garofalo M, Croce CM. MicroRNAs in cancer. Annu Rev Pathol. 2014; 9:287–314.
17. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015; 136:E359--E386.
18. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. CA Cancer J Clin. 2015; 65:5–29.
19. Goldstraw P, Ball D, Jett JR, Le Chevalier T, Lim E, Nicholson AG, Shepherd FA. Non-small-cell lung cancer. Lancet. 2011; 378:1727–1740.
20. Pikor LA, Ramnarine VR, Lam S, Lam WL. Genetic alterations defining NSCLC subtypes and their therapeutic implications. Lung Cancer. 2013; 82:179–89.
21. Shirdel EA, Xie W, Mak TW, Jurisica I. NAViGaTing the micronome-using multiple microRNA prediction databases to identify signalling pathway-associated microRNAs. PLoS One. 2011; 6:e17429.
22. De La Fuente A, Bing N, Hoeschele I, Mendes P. Discovery of meaningful associations in genomic data using partial correlation coefficients. Bioinformatics. 2004; 20:3565–74.
23. Gyorffy B, Surowiak P, Budczies J, Lánczky A. Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. PLoS One. 2013; 8:e82241.
24. Botling J, Edlund K, Lohr M, Hellwig B, Holmberg L, Lambe M, Berglund A, Ekman S, Bergqvist M, Pontén F, König A, Fernandes O, Karlsson M, et al. Biomarker discovery in Non--Small cell lung cancer: Integrating gene

expression profiling, meta-analysis, and tissue microarray validation. Clin Cancer Res. 2013; 19:194–204.

25. Okayama H, Kohno T, Ishii Y, Shimada Y, Shiraishi K, Iwakawa R, Furuta K, Tsuta K, Shibata T, Yamamoto S, Watanabe S, Sakamoto H, Kumamoto K, et al. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. Cancer Res. 2012; 72:100–11.
26. Der SD, Sykes J, Pintilie M, Zhu CQ, Strumpf D, Liu N, Jurisica I, Shepherd FA, Tsao MS. Validation of a histology-independent prognostic gene signature for early-stage, non-small-cell lung cancer including stage IA patients. J Thorac Oncol. 2014; 9:59–64.
27. Rahmati S, Abovsky M, Pastrello C, Jurisica I. pathDIP: an annotated resource for known and predicted human gene-pathway associations and pathway enrichment analysis. Nucleic Acids Res. 2017; 45:D419–D426.
28. Beloribi-Djefaflia S, Vasseur S, Guillaumond F. Lipid metabolic reprogramming in cancer cells. Oncogenesis. 2016; 5:e189.
29. Nasarre P, Potiron V, Drabkin H, Roche J. Guidance molecules in lung cancer. Cell Adhes Migr. 2010; 4:130–45.
30. Di Cristofano A, Pandolfi PP. The multiple roles of PTEN in tumor suppression. Cell. 2000; 100:387–90.
31. Cavallaro U, Christofori G. Cell adhesion and signalling by cadherins and Ig-CAMs in cancer. Nat Rev Cancer. 2004; 4:118–32.
32. Garnier D, Magnus N, D'Asti E, Hashemi M, Meehan B, Milsom C, Rak J. PL-05 genetic pathways linking hemostasis and cancer. Thromb Res. 2012; 129:S22–9.
33. Weir BA, Woo MS, Getz G, Perner S, Ding L, Beroukhim R, Lin WM, Province MA, Kraja A, Johnson LA, Shah K, Sato M, Thomas RK, et al. Characterizing the cancer genome in lung adenocarcinoma. Nature. 2007; 450:893–8.
34. Lee HW, Seol HJ, Choi YL, Ju HJ, Joo KM, Ko YH, Lee JI, Nam DH. Genomic copy number alterations associated with the early brain metastasis of non-small cell lung cancer. Int J Oncol. 2012; 41:2013–20.
35. Ohtsuka T, Shiomi T, Shimoda M, Kodama T, Amour A, Murphy G, Ohuchi E, Kobayashi K, Okada Y. ADAM28 is overexpressed in human non-small cell lung carcinomas and correlates with cell proliferation and lymph node metastasis. Int J Cancer. 2006; 118:263–73.
36. Kuroda H, Mochizuki S, Shimoda M, Chijiiwa M, Kamiya K, Izumi Y, Watanabe M, Horinouchi H, Kawamura M, Kobayashi K, Okada Y. ADAM28 is a serological and histochemical marker for non-small-cell lung cancers. Int J Cancer. 2010; 127:1844–56.
37. Jung S, Sielker S, Purcz N, Sproll C, Acil Y, Kleinheinz J. Analysis of angiogenic markers in oral squamous cell carcinoma-gene and protein expression. Head Face Med. 2015; 11:19.
38. Hu X, Huang Z, Liao Z, He C, Fang X. Low CA II expression is associated with tumor aggressiveness and poor prognosis in gastric cancer patients. Int J Clin Exp Pathol. 2014; 7:6716.
39. Rostoker R, Abelson S, Genkin I, Ben-Shmuel S, Sachidanandam R, Scheinman EJ, Bitton-Worms K, Orr ZS, Caspi A, Tzukerman M, LeRoith D. CD24+ cells fuel rapid tumor growth and display high metastatic capacity. Breast Cancer Res. 2015; 17:1–14.
40. Tong JH, Ng DC, Chau SL, So KK, Leung PP, Lee TL, Lung RW, Chan MW, Chan AW, Lo KW, To KF. Putative tumour-suppressor gene DAB2 is frequently down regulated by promoter hypermethylation in nasopharyngeal carcinoma. BMC Cancer. 2010; 10:253.
41. Zhang M, Nie F, Sun M, Xia R, Xie M, Lu KH, Li W. HOXA5 indicates poor prognosis and suppresses cell proliferation by regulating p21 expression in non small cell lung cancer. Tumor Biol. 2015; 36:3521–31.
42. Ordóñez-Morán P, Dafflon C, Imajo M, Nishida E, Huelsken J. HOXA5 counteracts stem cell traits by inhibiting Wnt signaling in colorectal cancer. Cancer Cell. 2015; 28:815–29.
43. Du Y, Gao L, Zhang K, Wang J. Association of the IL6 polymorphism rs1800796 with cancer risk: a meta-analysis. Genet Mol Res GMR. 2014; 14:13236–46.
44. Chang VH, Chu PY, Peng SL, Mao TL, Shan YS, Hsu CF, Lin CY, Tsai KK, Yu WC, Ch'ang HJ. Krüppel-like factor 10 expression as a prognostic indicator for pancreatic adenocarcinoma. Am J Pathol. 2012; 181:423–30.
45. Toge M, Yokoyama S, Kato S, Sakurai H, Senda K, Doki Y, Hayakawa Y, Yoshimura N, Saiki I. Critical contribution of MCL-1 in EMT-associated chemo-resistance in A549 non-small cell lung cancer. Int J Oncol. 2015; 46:1844–8.
46. Neradil J, Veselska R. Nestin as a marker of cancer stem cells. Cancer Sci. 2015; 106:803–11.
47. Gulzar ZG, McKenney JK, Brooks JD. Increased expression of NuSAP in recurrent prostate cancer is mediated by E2F1. Oncogene. 2013; 32:70–7.
48. Tam CW, Liu VW, Leung WY, Yao KM, Shiu SY. The autocrine human secreted PDZ domain-containing protein 2 (sPDZD2) induces senescence or quiescence of prostate, breast and liver cancer cells via transcriptional activation of p53. Cancer Lett. 2008; 271:64–80.
49. James MA, Lu Y, Liu Y, Vikis HG, You M. RGS17, an overexpressed gene in human lung and prostate cancer, induces tumor cell proliferation through the cyclic AMP-PKA-CREB pathway. Cancer Res. 2009; 69:2108–16.
50. Hooks SB, Callihan P, Altman MK, Hurst JH, Ali MW, Murph MM. Regulators of G-Protein signaling RGS10 and RGS17 regulate chemoresistance in ovarian cancer cells. Mol Cancer. 2010; 9:1.
51. Nasir A, Helm J, Turner L, Chen DT, Strosberg J, Hafez N, Henderson-Jackson EB, Hodul P, Bui MM, Nasir NA, Hakam A, Malafa MP, Yeatman T, et al. RUNX1T1: a novel predictor of liver metastasis in primary pancreatic endocrine neoplasms. Pancreas. 2011; 40:627.

52. Martinez N, Drescher B, Riehle H, Cullmann C, Vornlocher HP, Ganser A, Heil G, Nordheim A, Krauter J, Heidenreich O. The oncogenic fusion protein RUNX1-CBFA2T1 supports proliferation and inhibits senescence in t(8;21)-positive leukaemic cells. BMC Cancer. 2004; 4:1.

53. Agell L, Hernández S, Nonell L, Lorenzo M, Puigdecanet E, de Muga S, Juanpere N, Bermudo R, Fernández PL, Lorente JA, Serrano S, Lloreta J. A 12-gene expression signature is associated with aggressive histological in prostate cancer: SEC14L1 and TCEB1 genes are potential markers of progression. Am J Pathol. 2012; 181:1585–94.

54. Sadanandam A, Varney ML, Singh S, Ashour AE, Moniaux N, Deb S, Lele SM, Batra SK, Singh RK. High gene expression of semaphorin 5A in pancreatic cancer is associated with tumor growth, invasion and metastasis. Int J Cancer. 2010; 127:1373–83.

55. Geybels MS, van den Brandt PA, Schouten LJ, van Schooten FJ, van Breda SG, Rayman MP, Green FR, Verhage BA. Selenoprotein gene variants, toenail selenium levels, and risk for advanced prostate cancer. J Natl Cancer Inst. 2014; 106:dju003.

56. Esquela-Kerscher A, Slack FJ. Oncomirs — microRNAs with a role in cancer. Nat Rev Cancer. 2006; 6:259–69.

57. Li Q, Yang Z, Chen M, Liu Y. Downregulation of microRNA-196a enhances the sensitivity of non-small cell lung cancer cells to cisplatin treatment. Int J Mol Med. 2016; 37:1067–74.

58. Yang X, Chen Y, Chen L. The Versatile Role of microRNA-30a in Human Cancer. Cell Physiol Biochem. 2017; 41:1616–32.

59. Su W, Mo Y, Wu F, Guo K, Li J, Luo Y, Ye H, Guo H, Li D, Yang Z. miR-135b reverses chemoresistance of non-small cell lung cancer cells by downregulation of FZD1. Biomed Pharmacother. 2016; 84:123–9.

60. Dai H, Gallagher D, Schmitt S, Pessetto ZY, Fan F, Godwin AK, Tawfik O. Role of miR-139 as a surrogate marker for tumor aggression in breast cancer. Hum Pathol. 2017; 61:68–77.

61. Luo H, Yang R, Li C, Tong Y, Fan L, Liu X, Xu C. MicroRNA-139-5p inhibits bladder cancer proliferation and self-renewal by targeting the Bmi1 oncogene. Tumor Biol. 2017; 39:101042831771841.

62. Zhai L, Ma C, Li W, Yang S, Liu Z. miR-143 suppresses epithelial-mesenchymal transition and inhibits tumor growth of breast cancer through down-regulation of ERK5. Mol Carcinog. 2016; 55:1990–2000.

63. Li J, Sun P, Yue Z, Zhang D, You K, Wang J. miR-144-3p induces cell cycle arrest and apoptosis in pancreatic cancer cells by targeting proline-rich protein 11 expression via the mitogen-activated protein kinase signaling pathway. DNA Cell Biol. 2017; 36:619–26.

64. Yao J, Xu C, Fang Z, Li Y, Liu H, Wang Y, Xu C, Sun Y. Androgen receptor regulated microRNA miR-182-5p promotes prostate cancer progression by targeting the ARRDC3/ITGB4 pathway. Biochem Biophys Res Commun. 2016; 474:213–9.

65. Xu Z, Li C, Qu H, Li H, Gu Q, Xu J. MicroRNA-195 inhibits the proliferation and invasion of pancreatic cancer cells by targeting the fatty acid synthase/Wnt signaling pathway. Tumor Biol. 2017; 39:101042831771132.

66. Shi ZM, Wang L, Shen H, Jiang CF, Ge X, Li DM, Wen YY, Sun HR, Pan MH, Li W, Shu YQ, Liu LZ, Peiper SC, et al. Downregulation of miR-218 contributes to epithelial-mesenchymal transition and tumor metastasis in lung cancer by targeting Slug/ZEB2 signaling. Oncogene. 2017; 36:2577–88.

67. Cannistraci A, Federici G, Addario A, Di Pace AL, Grassi L, Muto G, Collura D, Signore M, De Salvo L, Sentinelli S, Simone G, Costantini M, Nanni S, et al. C-Met/miR-130b axis as novel mechanism and biomarker for castration resistance state acquisition. Oncogene. 2017; 36:3718–28.

68. Zaporozhchenko IA, Morozkin ES, Skvortsova TE, Ponomaryova AA, Rykova EY, Cherdyntseva NV, Polovnikov ES, Pashkovskaya OA, Pokushalov EA, Vlassov VV, Laktionov PP. Plasma miR-19b and miR-183 as potential biomarkers of lung cancer. PLoS One. 2016; 11:e0165261.

69. Azizmohammadi S, Safari A, Azizmohammadi S, Kaghazian M, Sadrkhanlo M, Yahaghi E, Farshgar R, Seifoleslami M. Molecular identification of miR-145 and miR-9 expression level as prognostic biomarkers for early-stage cervical cancer detection. QJM. 2017; 110:11–5.

70. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011; 144:646–74.

71. Zhu CQ, Ding K, Strumpf D, Weir BA, Meyerson M, Pennell N, Thomas RK, Naoki K, Ladd-Acosta C, Liu N, Pintilie M, Der S, Seymour L, et al. Prognostic and predictive gene signature for adjuvant chemotherapy in resected non-small-cell lung cancer. J Clin Oncol. 2010; 28:4417–24.

72. Watson SK, deLeeuw RJ, Horsman DE, Squire JA, Lam WL. Cytogenetically balanced translocations are associated with focal copy number alterations. Hum Genet. 2007; 120:795–805.

73. Khojasteh M, Lam WL, Ward RK, MacAulay C. A stepwise framework for the normalization of array CGH data. BMC Bioinformatics. 2005; 6:274.

74. Chi B, Coe BP, Ng RT, MacAulay C, Lam WL, deLeeuw RJ. MD-SeeGH: a platform for integrative analysis of multi-dimensional genomic data. BMC Bioinformatics. 2008; 9:243.

75. de Wiel MA, Kim KI, Vosse SJ, Van Wieringen WN, Wilting SM, Ylstra B. CGHcall: calling aberrations for array CGH tumor profiles. Bioinformatics. 2007; 23:892–4.

76. Guo X, Ma X, An J, Shang Y, Huang Q, Yang H, Chen Z, Xing J. A meta-analysis of array-CGH studies implicates antiviral immunity pathways in the development of hepatocellular carcinoma. PLoS One. 2011; 6:e28404.

77. Wu Z, Irizarry RA. Preprocessing of oligonucleotide array data. Nat Biotechnol. 2004; 22:656–8.

78. Smyth GK. Limma: linear models for microarray data. In: Bioinformatics and computational biology solutions using R and Bioconductor. Springer; 2005. p. 397–420.

79. Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. Bioinformatics. 2012; 28:573–80.

80. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 2014; 42:D68–73.

81. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 2009; 4:1184–91.

82. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010; 33:1.

83. Terry M. Therneau, Patricia M. Grambsch. Modeling survival data: extending the cox model. New York: Springer; 2000.

84. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. BMC Med Res Methodol. 2013; 13:1.

85. Rohrbeck A, Neukirchen J, Rosskopf M, Pardillos GG, Geddert H, Schwalen A, Gabbert HE, von Haeseler A, Pitschke G, Schott M, Kronenwett R, Haas R, Rohr UP. Gene expression profiling for molecular distinction and characterization of laser captured primary lung cancers. J Transl Med. 2008; 6:69.

86. 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.

87. Yanaihara N, Caplen N, Bowman E, Seike M, Kumamoto K, Yi M, Stephens RM, Okamoto A, Yokota J, Tanaka T, Calin GA, Liu CG, Croce CM, et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. Cancer Cell. 2006; 9:189–98.

88. Cho W, Chow AS, Au JS. Restoration of tumour suppressor hsa-miR-145 inhibits cancer cell growth in lung adenocarcinoma patients with epidermal growth factor receptor mutation. Eur J Cancer. 2009; 45:2197–206.

89. Crawford M, Batte K, Yu L, Wu X, Nuovo GJ, Marsh CB, Otterson GA, Nana-Sinkam SP. MicroRNA 133B targets pro-survival molecules MCL-1 and BCL2L2 in lung cancer. Biochem Biophys Res Commun. 2009; 388:483–9.

90. Dacic S, Kelly L, Shuai Y, Nikiforova MN. miRNA expression profiling of lung adenocarcinomas: correlation with mutational status. Mod Pathol. 2010; 23:1577–82.

91. Yu L, Todd NW, Xing L, Xie Y, Zhang H, Liu Z, Fang H, Zhang J, Katz RL, Jiang F. Early detection of lung adenocarcinoma in sputum by a panel of microRNA markers. Int J Cancer. 2010; 127:2870–8.

92. Lee YM, Cho HJ, Lee SY, Yun SC, Kim JH, Lee SY, Kwon SJ, Choi E, Na MJ, Kang JK, Sonf JW. MicroRNA-23a: a novel serum based diagnostic biomarker for lung adenocarcinoma. Tuberc Respir Dis (Seoul). 2011; 71:8–14.

93. Chitale D, Gong Y, Taylor BS, Broderick S, Brennan C, Somwar R, Golas B, Wang L, Motoi N, Szoke J, Reinersman JM, Major J, Sander C, et al. An integrated genomic analysis of lung cancer reveals loss of DUSP4 in EGFR-mutant tumors. Oncogene. 2009; 28:2773–83.

94. Jang JS, Jeon HS, Sun Z, Aubry MC, Tang H, Park CH, Rakhshan F, Schultz DA, Kolbert CP, Lupu R, Park JY, Harris CC, Yang P, et al. Increased miR-708 expression in NSCLC and its association with poor survival in lung adenocarcinoma from never smokers. Clin Cancer Res. 2012; 18:3658–67.

95. Ma J, Mannoor K, Gao L, Tan A, Guarnera MA, Zhan M, Shetty A, Stass SA, Xing L, Jiang F. Characterization of microRNA transcriptome in lung cancer by next-generation deep sequencing. Mol Oncol. 2014; 8:1208–19.

96. Nadal E, Zhong J, Lin J, Reddy RM, Ramnath N, Orringer MB, Chang AC, Beer DG, Chen G. A MicroRNA cluster at 14q32 drives aggressive lung adenocarcinoma. Clin Cancer Res. 2014; 20:3107–17.

97. Job B, Bernheim A, Beau-Faller M, Camilleri-Broët S, Girard P, Hofman P, Mazières J, Toujani S, Lacroix L, Laffaire J, Dessen P, Fouret P; LG Investigators. Genomic aberrations in lung adenocarcinoma in never smokers. PLoS One. 2010; 5:e15145.

98. Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, Ladd C, Beheshti J, Bueno R, Gillette M, Loda M, Weber G, Mark EJ, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc Natl Acad Sci. 2001; 98:13790–5.

99. Beer DG, Kardia SL, Huang CC, Giordano TJ, Levin AM, Misek DE, Lin L, Chen G, Gharib TG, Thomas DG, Lizyness ML, Kuick R, Hayasaka S, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat Med. 2002; 8:816–24.

100. Stearman RS, Dwyer-Nield L, Zerbe L, Blaine SA, Chan Z, Bunn Jr PA, Johnson GL, Hirsch FR, Merrick DT, Franklin WA, Baron AE, Keith RL, Nemenoff RA, et al. Analysis of orthologous gene expression between human pulmonary adenocarcinoma and a carcinogen-induced murine model. Am J Pathol. 2005; 167:1763–75.

101. Yap YL, Lam DC, Luc G, Zhang XW, Hernandez D, Gras R, Wang E, Chiu SW, Chung LP, Lam WK, Smith DK, Minna JD, Danchin A, et al. Conserved transcription factor binding sites of cancer markers derived from primary lung adenocarcinoma microarrays. Nucleic Acids Res. 2005; 33:409–21.

102. Su LJ, Chang CW, Wu YC, Chen KC, Lin CJ, Liang SC, Lin CH, Whang-Peng J, Hsu SL, Chen CH, Huang CY. Selection of DDX5 as a novel internal control for Q-RT-PCR from microarray data using a block bootstrap re-sampling scheme. BMC Genomics. 2007; 8:140.

103. Landi MT, Dracheva T, Rotunno M, Figueroa JD, Liu H, Dasgupta A, Mann FE, Fukuoka J, Hames M, Bergen AW, Murphy SE, Yang P, Pesatori A, et al. Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. PLoS One. 2008; 3:e1651.