Published online 16 October 2008 *Nucleic Acids Research, 2008, Vol. 36, No. 20* ***6535–6547***
doi:10.1093/nar/gkn697

Whole genome exon arrays identify differential expression of alternatively spliced, cancer-related genes in lung cancer

Liqiang Xi[1], Andrew Feber[1], Vanita Gupta[1], Maoxin Wu[1], Andrew D. Bergemann[1], Rodney J. Landreneau[3], Virginia R. Litle[2], Arjun Pennathur[3], James D. Luketich[3] and Tony E. Godfrey[1,2,*]

[1]Department of Pathology, [2]Department of Cardiothoracic Surgery, Mount Sinai School of Medicine, New York, NY 10029 and [3]Heart, Lung and Esophageal Surgery Institute, and Pittsburgh Cancer Institute, University of Pittsburgh, Pittsburgh, PA 15260, USA

Received August 25, 2008; Revised September 24, 2008; Accepted September 25, 2008

ABSTRACT

Alternative processing of pre-mRNA transcripts is a major source of protein diversity in eukaryotes and has been implicated in several disease processes including cancer. In this study we have performed a genome wide analysis of alternative splicing events in lung adenocarcinoma. We found that 2369 of the 17 800 core Refseq genes appear to have alternative transcripts that are differentially expressed in lung adenocarcinoma versus normal. According to their known functions the largest subset of these genes (30.8%) is believed to be cancer related. Detailed analysis was performed for several genes using PCR, quantitative RT-PCR and DNA sequencing. We found overexpression of *ERG* variant 2 but not variant 1 in lung tumors and overexpression of *CEACAM1* variant 1 but not variant 2 in lung tumors but not in breast or colon tumors. We also identified a novel, overexpressed variant of *CDH3* and verified the existence and overexpression of a novel variant of P16 transcribed from the *CDKN2A* locus. These findings demonstrate how analysis of alternative pre-mRNA processing can shed additional light on differences between tumors and normal tissues as well as between different tumor types. Such studies may lead to the development of additional tools for tumor diagnosis, prognosis and therapy.

INTRODUCTION

Non-small cell lung cancer (NSCLC) is the most common cause of cancer-related death in the USA (1). While early detection may improve outcomes, overall 5-year survival rates for NSCLC are currently only 16% (2). New molecular diagnostic tests and novel therapeutic strategies are needed for this terrible disease. NSCLC is one of the most studied tumor types in the scientific literature demonstrated by the number of excellent studies on global gene expression (3–6) and genome-wide DNA copy number changes (7,8) that have been conducted in NSCLC. These studies have enhanced our knowledge of lung cancer biology, led to proposals for multicenter trials of primary tumor gene expression for prognosis and treatment and may identify avenues for novel therapeutic development. A promising area that remains relatively unexplored, however, is alternative splicing (AS) of mRNA to produce functionally different proteins. Such studies may lead to improved diagnostic and prognostic tools and may identify additional therapeutic targets for NSCLC.

Alternative splicing of pre-mRNA is an important process in normal metazoan development (9,10). Furthermore, recent bioinformatics analysis suggests that 65% of human genes are alternatively spliced (11–14); a large increase over prior estimates as low as 5% (15). AS is not only involved in normal development, but is also associated with human diseases including cancer (16–27). For some genes, alternative transcripts are differentially expressed between tumor and normal tissue and in a few cases, the expression of AS variants has been associated with tumor progression (28–32). However, most studies of AS in human disease have used a targeted approach and focused on individual genes. There is a great deal of potential for novel discovery from genome-wide studies of alternative splicing. Until recently such large scale studies have been a considerable technical and bioinformatic challenge but the introduction of new technology and powerful data analysis software now makes them more feasible.

*To whom correspondence should be addressed. Tel: 585 273 3112; Fax: 585 276 2576; Email: tony_godfrey@urmc.rochester.edu

© 2008 The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this study we have used the GeneChip Human Exon 1.0 ST Array from Affymetrix to explore genome-wide AS events in the most predominant histologic type of NSCLC, lung adenocarcinoma. The study was designed to identify cancer-associated alternative splicing events, verify splice variants and to validate differential expression of selected splice variants in independent tissue sets. Our results demonstrate that a large number of known genes, including well known oncogenes and tumor suppressors, are alternatively spliced and differentially expressed between normal lung and lung adenocarcinoma. These findings may provide a new resource for diagnosis and treatment of NSCLC.

MATERIALS AND METHODS

Specimens and RNA isolation

Snap frozen lung tissue specimens were obtained from tissue banks at the Heart, Lung and Esophageal Surgery Institute, University of Pittsburgh Medical Center. This study involving human tissue was approved by the Institutional Review Boards from both the University of Pittsburgh and Mount Sinai School of Medicine. In total, 36 pairs of lung adenocarcinoma and adjacent normal lung tissue plus 43 additional adenocarcinoma and squamous cell carcinoma specimens were analyzed (see clinical information for all patients in Supplementary Table S1). All tumor specimens were determined to comprise >70% tumor and adjacent normal specimens contained no histologically evident tumor or contaminating tissues. Forty, 5-micron sections from each tissue block were cut and placed immediately in Qiagen RNA lysis buffer. RNA was isolated using Qiagen kits with on-column DNAse treatment to remove genomic DNA followed by precipitation. Purified RNA was then quantified using a NanoDrop spectrophotometer and RNA integrity was determined by running aliquots on an Agilent Bioanalyzer. RNA integrity numbers were >6 in all cases.

RNA labeling, hybridization, data processing and quality assessment

A total of 2 µg of RNA from each of 20 tumor/normal paired specimens ($n = 40$) was labeled with reagents from Affymetrix according to the manufacturers instructions. Hybridization cocktails containing 5–5.5 µg of fragmented, end-labeled single-stranded cDNA were prepared and hybridized to GeneChip Human Exon 1.0 ST arrays. These arrays survey both gene expression and alternative splicing patterns on a whole-genome scale on a single array. The array contains ~5.4 million, 5-µm features (probes) grouped into 1.4 million probesets, interrogating over 1 million exon clusters (33). Processed arrays were scanned using the GeneChip Scanner 3000 7G. Affymetrix Expression Console Software™ (version 1.0) was used to perform quality assessment.

Data analysis

All exon array data was analyzed using tools in Partek Genomic Suite 6.4 software (Partek Inc., St. Louis, MO). The Robust Multi-array Analysis (RMA) (34) algorithm was used for probeset (exon-level) intensity analysis. Exon-level data was then filtered to include only those probesets that are in the 'core' meta-probe list, which represents 17 800 RefSeq genes and full-length GenBank mRNAs. Within this gene set, the Analysis of Variance (ANOVA) and multi test correction for *P*-values in Partek Genomic Suite were used to identify alternative splicing events. Tissue type (tumor versus normal in this case) was chosen as the candidate variable in the ANOVA model to obtain tumor-related splicing events. ANOVA *P*-values were corrected using Bonferroni method. A list of genes with significant alternative spliced events was generated by using a 0.05 FDR criterion as a significant cutoff. Then the genes were sorted based on gene function using Ingenuity Pathway Analysis software (Ingenuity® Systems, www.ingenuity.com). Subsequent verification and validation of splicing events was restricted to those genes with functions identified as being associated with cancer such as invasion, cell movement, apoptosis, cell death, tumorigenesis, and differentiation, etc. The flow chart for data analyses was as follows:

(0) QC with Affymetrix expression console.
(1) Probe-level analysis with GC-RMA in Partek GS.
(2) Alternative splice analysis of exon data with ANOVA in Partek GS.
(3) Function analysis of genes with alternative splicing in Ingenuity Pathway Analysis and generation of a cancer-related gene list.
(4) Manual review of Partek gene view plots to identify alternate splicing forms and determine the frequency of changes observed in the patient set.
(5) Detailed manual analysis focusing on genes with simple forms of alt splicing and high frequency of changes. Reviewed Affymetrix probeset sequences, RefSeq database, Blatted probe sequences in UCSC Genomic Browser.
(6) Identification of genes with alternative splicing at high frequency in this patient set (>50% patients with same change).
(7) Verification and validation.

Exon array data verification, splice event validation, and variant transcript quantification

Reverse transcription of 2 µg RNA was performed in 100 µl reaction volumes with random hexamer priming and MMLV reverse transcriptase (Epicentre, Madison, WI) (35). Diluted cDNA was used in all following PCR or qPCR reactions. Standard PCR was carried out for 35 cycles using Titanium™ Taq DNA Polymerase (ClonTech, Mountain View, CA) starting with 25 ng of cDNA as template in a 50 µl reaction. PCR products were separated on 10% Criterion™ Precast TBE Gels (Bio-Rad, Hercules, CA) for visualization or 1% agarose gels for DNA extractions using the MinElute™ Gel Extraction Kit (Qiagen, Inc., Valencia, CA). Quantitative real-time PCR was performed using the Brilliant SYBR Green QPCR kit (Stratagene, La Jolla, CA) with 10 ng of cDNA as template in a 5 µl reaction

on an ABI PRISM® 7900HT instrument (Applied Biosystems, Foster City, CA). PCR was run in triplicate for each sample. Relative expression was calculated using the delta-C_T methods previously described (36) and with *B2M* as the endogenous control gene. *B2M* was chosen as it showed very little variability in the exon array data among all samples. Bidirectional DNA sequencing of novel transcripts was performed by the DNA Core Facility at the Mount Sinai School of Medicine using forward and reverse PCR primers.

Detailed methods for analysis of alternative splicing and primer design for PCR are included in the Supplementary Methods.

RESULTS

Data quality assessment identified no outlier arrays using Expression Console Software. Spearman Rank correlation of the hybridization control signal values between any two chips was high ($r^2 \geq 0.92$). Thus, data from all arrays was included in the alternative splicing ANOVA (Alt-splice ANOVA). Analyzing the 17 800 genes represented in 'Core' Probeset list, we identified a total of 2369 genes that appear to have differential expression of alternate transcripts between normal and tumor tissue (FDR correction to establish a cutoff *P*-value of 3.94e-6, corresponding to the 0.05 FDR level). Gene function analysis indicated that the largest subset (729/2369, 30.8%) of these genes were cancer related (a full list of genes is included in Supplementary Table S2) followed by other functional categories such as tissue development, cellular growth and proliferation, tissue morphology, and immune response. Of the 729 cancer-related genes, 47 showed the same alternate splicing event in 50% or more (≥10 of 20) of patients. In addition, one gene (*CDH3*) showed alternative splicing in only 8 (40%) patients, but all of these were female (8 of 14; 57%) indicating the possibility of gender-specific, cancer related alternative splicing. Of the 48 genes, 20 have reported alternate splice variants in the Entrez Gene database and the remaining 28 genes have only one known transcript. The 48 genes can also be further divided into categories as follows and listed in Table 1: (i) six genes with known splice variants where relevance of splice variants has been identified in cancer. These include *ADAM12*, *CEACAM1*, and *FGFR4* which have been associated specifically with lung cancer; (ii) 14 genes with known splice variants but where relevance of the splice variants to cancer has not been determined; (iii) 28 potentially novel splice variants differentially expressed between normal lung and lung adenocarcinoma.

Verification of exon array data

The first step in verification of alternative splicing was to verify the exon array expression data for regions of genes (exons) that appeared to have differential expression in tumor and normal samples. Eleven genes (*ARMET*, *CDKN1A*, *CDKN2A*, *CDKN2B*, *CEACAM1*, *ERG*, *FOXP1*, *KLF2*, *RASIP1*, *VEGFC* and *CDH3*) were selected for qRT-PCR verification based on the frequency of alternative splicing and interest level from a review of the literature. Expression of two exons from differentially expressed portions of each gene was quantified using qRT-PCR on the same samples used in the array experiments (Supplementary Figures). Array data was considered verified if qRT-PCR demonstrated the same, directional difference in expression between the different regions of the gene, or if novel alternative transcripts were specifically identified based on PCR product size. For 4 of the 11 genes (*CEACAM1*, *ERG*, *RASIP1* and *VEGFC*), the qPCR data was in clear agreement with the array data. One gene (*CDKN2A*) demonstrated differential expression that was statistically borderline (Supplementary Figures). In addition, while *CDH3* exon-specific qPCR did not appear to validate differential expression, a novel alternative transcript was identified that matched the array data (described in detail below). Thus, 6 of the 11 selected genes were considered validated while the remaining five genes (*ARMET*, *CDKN1A*, *FOXP1*, *KLF2* and *CDKN2B*) were not verifiable by PCR. In three of five cases, we noted that the exon array probes that did not show differential expression were located in very G-C rich regions, raising the possibility of cross hybridization as the reason for erroneous exon array data. Of the six validated genes, four (*ERG*, *CDH3*, *CDKN2A*, and *CEACAM1*) were explored in more detail.

Differential expression of *ERG* splice variants in lung adenocarcinoma

Exon array analysis of *ERG* demonstrated differential expression of splice variants in 20 of 20 tumor/normal tissue pairs. Furthermore, the expression pattern observed was consistent with differential expression of the two known transcripts of *ERG* curated in the RefSeq database (Figure 1a and b). Variant 1 (NM_182918.2) encodes a short form with a unique 5′ first exon and an additional exon towards the 3′ end while variant 2 (NM_004449.3) encodes a long form with three unique exons on the 5′ end. In order to validate the array data, we first designed two sets of qRT-PCR primers to quantify the differential expression of unique versus shared exons. The results demonstrate that the qRT-PCR data correlated with the array data in the same individuals (Supplementary Figures). Secondly, to prove that two transcripts exist in this set of lung tumor and matched normal samples, we designed a common primer set that spanned exon 7-8 of variant 2 of *ERG* and exon 5-6-7 of variant 1. These data demonstrate the existence of both *ERG* variants (Figure 1c) in both tumor and normal tissue. Two additional transcript variants, identified in the UCSC genome browser (uc002ywz.1 and uc002yxc.1), were also evaluated but no expression was found in either tumor or normal tissues. Finally, to specifically demonstrate differential expression of variants 1 and 2 we designed primer sets unique to the two transcripts and quantified expression of each. This was performed in 35 tumor/normal pairs (which included the 20 pairs used in the original array analysis) (Figure 1c). We observed that *ERG* variant 2 was significantly overexpressed (mean tumor/normal ratio 4.73; paired *t*-test *P*-value = 0.0005) in tumor

Table 1. Alternative spliced genes between tumor and normal in high frequency from Exon Array analysis

Symbol	Entrez ID	Gene name	Frequency (%)
Genes with known splice variants where relevance of splice variants has been identified in cancer			
ADAM12	8038	*ADAM metallopeptidase domain 12 (meltrin alpha)*	70
BCL6	604	*B-cell CLL/lymphoma 6 (zinc finger protein 51)*	55
CDKN2A	1029	*Cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)*	65
CEACAM1	634	*Carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein)*	70
DSCR1	1827	*Down syndrome critical region gene 1*	85
FGFR4	2264	*Fibroblast growth factor receptor 4*	90
Genes with known variants but where relevance of splice variants to cancer has not been identified			
ADRA1A	148	*Adrenergic, alpha-1A-, receptor*	90
AKAP12	9590	*A kinase (PRKA) anchor protein (gravin) 12*	90
CCNE1	898	*Cyclin E1*	75
CDCA1	83 540	*Cell division cycle associated 1*	90
DKK3	27 122	*Dickkopf homolog 3 (Xenopus laevis)*	75
ERG	2078	*v-ets erythroblastosis virus E26 oncogene like (avian)*	100
FEZ1	9638	*Fasciculation and elongation protein zeta 1 (zygin I)*	90
FPRL1	2358	*Formyl peptide receptor-like 1*	70
GSN	2934	*Gelsolin (amyloidosis, Finnish type)*	75
KL	9365	*Klotho*	80
RASGRF1	5923	*Ras protein-specific guanine nucleotide-releasing factor 1*	95
TBX3	6926	*T-box 3 (ulnar mammary syndrome)*	50
TNFSF11	8600	*Tumor necrosis factor (ligand) superfamily, member 11*	55
ZBTB16	7704	*Zinc finger and BTB domain containing 16*	80
Genes with novel splice variants			
ANKRD1	27 063	*Ankyrin repeat domain 1 (cardiac muscle)*	75
APCDD1	147 495	*Adenomatosis polyposis coli down-regulated 1*	65
ARHGEF3	50 650	*Rho guanine nucleotide exchange factor (GEF) 3*	85
ARMET	7873	*Arginine-rich, mutated in early stage tumors*	85
CDH1	999	*Cadherin 1, type 1, E-cadherin (epithelial)*	60
CDH3	1001	*Cadherin 3, type 1, P-cadherin (placental)*	40
CXCL5	6374	*Chemokine (C-X-C motif) ligand 5*	60
EMP2	2013	*Epithelial membrane protein 2*	65
FSTL3	10 272	*Follistatin-like 3 (secreted glycoprotein)*	70
GATA2	2624	*GATA binding protein 2*	95
GATA6	2627	*GATA binding protein 6*	90
HCK	3055	*Hemopoietic cell kinase*	80
IGFBP6	3489	*Insulin-like growth factor binding protein 6*	75
ITGA5	3678	*Integrin, alpha 5 (fibronectin receptor, alpha polypeptide)*	55
KLF2	10 365	*Kruppel-like factor 2 (lung)*	85
NFIL3	4783	*Nuclear factor, interleukin 3 regulated*	90
PGR	5241	*Progesterone receptor*	70
PIAS1	8554	*Protein inhibitor of activated STAT, 1*	60
PTHR1	5745	*Parathyroid hormone receptor 1*	100
RAMP2	10 266	*Receptor (calcitonin) activity modifying protein 2*	100
RASIP1	54 922	*Ras interacting protein 1*	100
SMAD6	4091	*SMAD, mothers against DPP homolog 6 (Drosophila)*	90
SPARC	6678	*Secreted protein, acidic, cysteine-rich (osteonectin)*	80
SRPX	8406	*Sushi-repeat-containing protein, X-linked*	100
SSBP2	23 635	*Single-stranded DNA binding protein 2*	80
TBX2	6909	*T-box 2*	95
TGFBR3	7049	*Transforming growth factor, beta receptor III (betaglycan, 300kDa)*	55
VEGFC	7424	*Vascular endothelial growth factor C*	85

Known or Novel splice variants were based on Entrez Gene database and relevance to cancer was based on PubMed search. Both performed on 31 December 2007.

samples compared to paired normal while *ERG* variant 1 was slightly under-expressed in tumor samples compared to normal (ratio = 0.74; $P = 0.0186$).

Identification of a novel C*DH3* transcript variant in lung adenocarcinoma

CDH3 is encoded by 16 exons and has only one known transcript in the RefSeq database. Exon array data suggested alternative splicing of exon 2 since all other exons were expressed considerably higher in tumor than in normal tissues. Interestingly, this was observed only in female patients (8 of 14 cases) and not in any of six male patients (Supplementary Figures). Given the established gender differences in lung adenocarcinoma incidence and survival (37), this finding was evaluated further. QRT-PCR, with primers in exon 2 and in exon 3, verified that exon 3 was expressed at much higher levels (mean 24-fold tumor/normal) in tumor compared with normal in these eight patients while exon 2 was expressed only 3-fold higher in tumor versus normal. However, exon 2 and exon 3 were expressed similarly

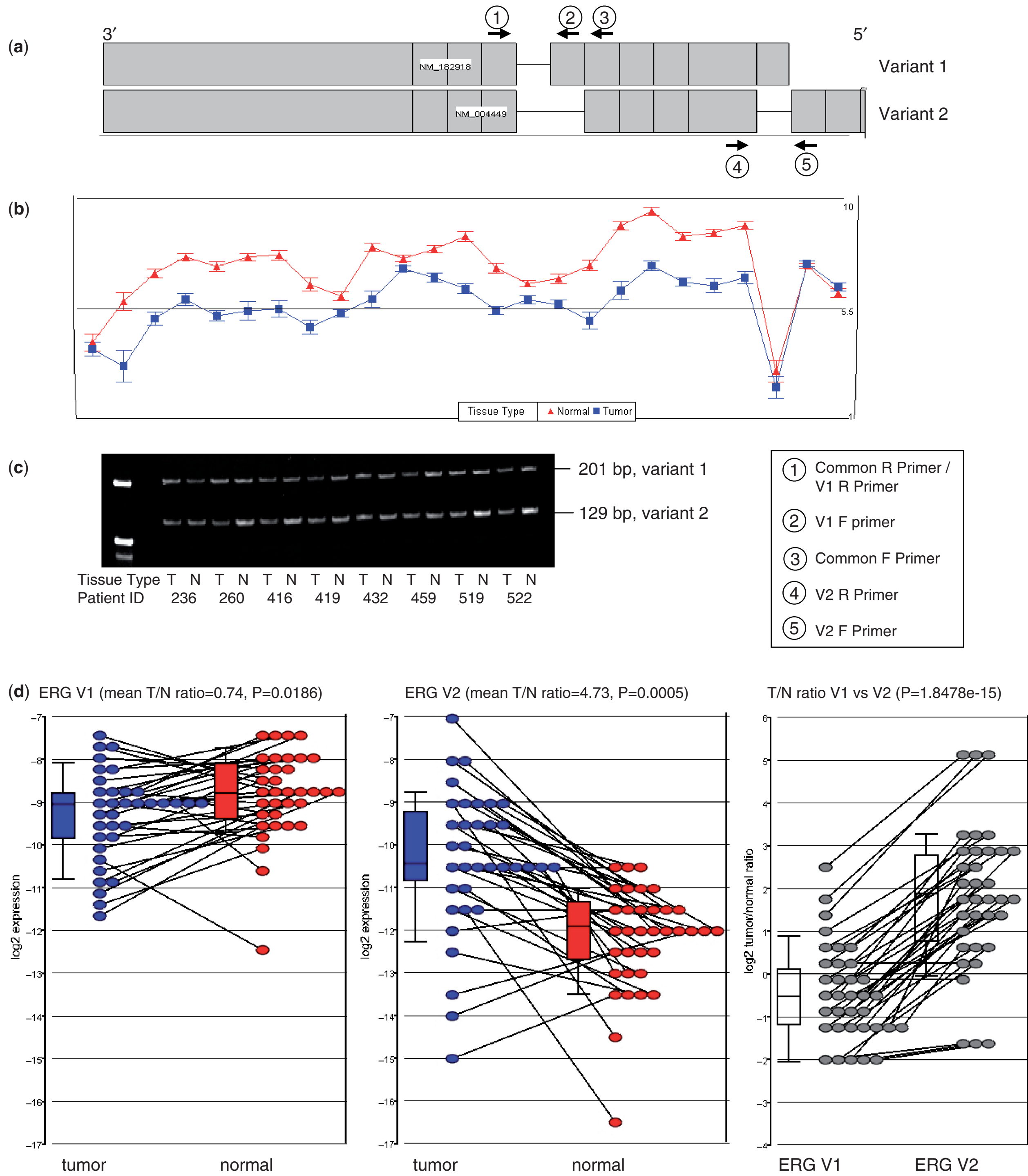
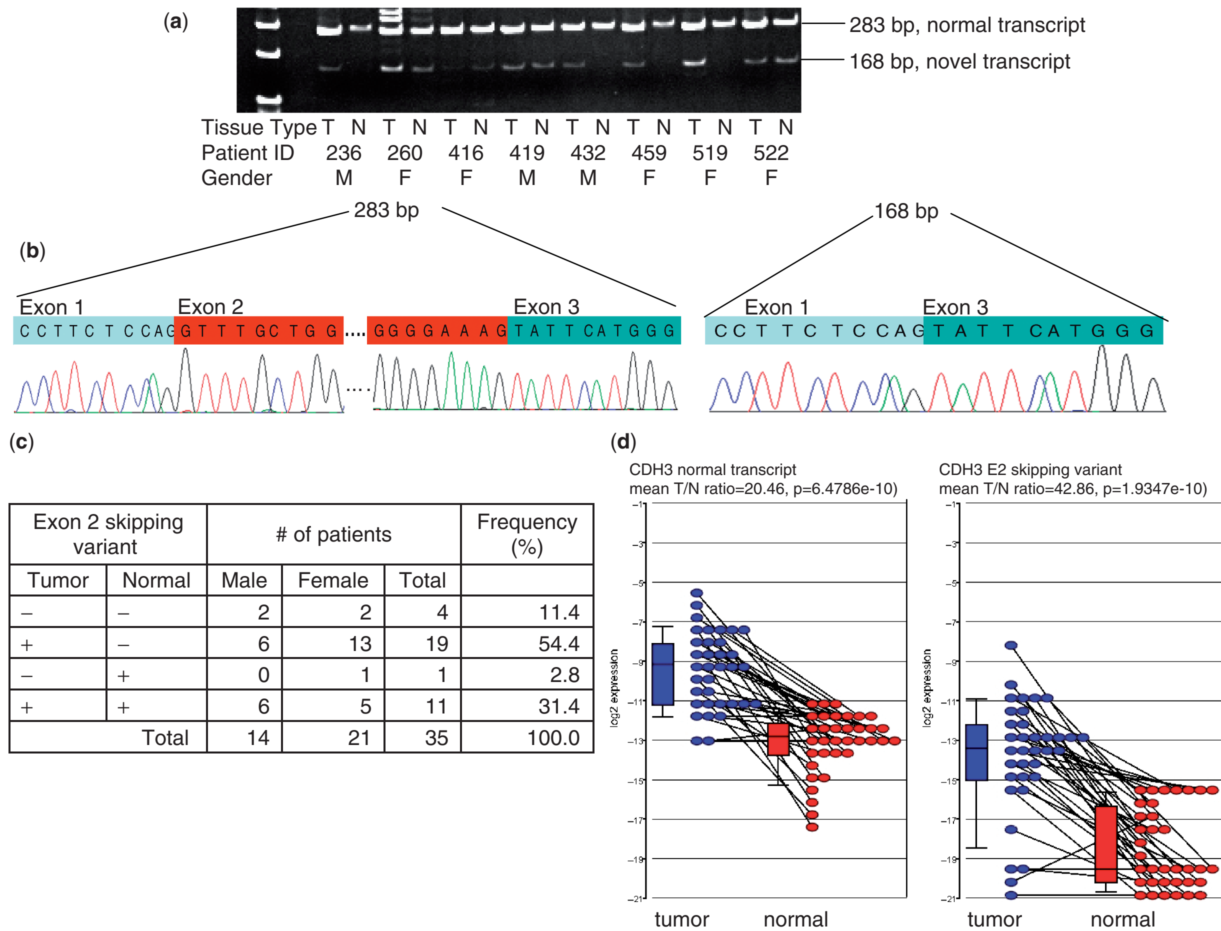


Figure 1. Exon array analysis and PCR analysis for alternative transcript variants of *ERG*. (**a**) Exon structure for ERG variant 1 (NM_182918.2) and variant 2 (NM_004449.3) showing the location of PCR primers used for expression analyses. (**b**) Partek GS alternative splice analysis of exon expression data in 20 patients. The graph shows mean expression value and standard error for each probe set in tumour (blue) and normal (normal) groups. (**c**) verification of PCR product size difference for two variants using primers 1 and 3. (**d**) Quantification of the two variants using primer set 1 and 2 for variant 1 and primers 4 and 5 for variant 2 in tumour/normal paired samples (tissue pairs are joined by solid lines) from 35 patients.

with a 2- to 3-fold difference in tumor versus normal in the other six female patients and in all male patients, thus validating the array data. Next, in order to further characterize the area of exon 1-3, we designed a forward primer in exon 1 and reverse primer across the junction of exons 3 and 4. RT-PCR products, visualized on a gel, showed the expected 283 bp band for the normal transcript and a shorter (168 bp) band (Figure 2a). Sequencing of these two bands (Figure 2b) confirmed the the presence of the normal transcript (283 bp band)



Exon 2 skipping variant		# of patients			Frequency (%)
Tumor	Normal	Male	Female	Total	
–	–	2	2	4	11.4
+	–	6	13	19	54.4
–	+	0	1	1	2.8
+	+	6	5	11	31.4
	Total	14	21	35	100.0

Figure 2. Identification, validation, and quantification of novel *CDH3* transcript variants. (**a**) Identification of an alternative *CDH3* transcript using PCR primers located in exon 1 and across the exon 3-4 boundary. (**b**) DNA sequencing results for the two PCR products demonstrating skipping of exon 2. (**c**) Frequency of the E2 skipping transcript expressed in 35 patients. (**d**) Specific quantification of the two variants in tumour/normal paired samples (tissue pairs are joined by solid lines) from 35 patients.

and the presence of a shorter transcript long and missing exon 2 (168 bp). To evaluate expression of this novel alterative transcript (referred to as *CDH3* E2 skipping transcript, or variant transcript, from this point forward), qRT-PCR was performed with primers unique to each *CDH3* transcript. This was performed in paired tumor/normal tissue from 35 patients and the results are shown in Figure 2. Based on a threshold of 40 cycles, four patients were negative for the E2 skipping transcript in both tumor and normal tissue, one patient was negative in tumor and positive in normal, 19 patients showed expression in tumor but not in normal, and 11 patients showed expression in both tissue types (Figure 2c). In all 11 patients, expression of the E2 skipping transcript was higher in tumor than in normal tissues. Therefore the E2 skipping transcript was overexpressed in 86% (30 of 35) of tumors (Figure 2d). However in this expanded patient cohort there was no statistically significant gender difference in the frequency or expression level of the E2 skipping transcript when compared with gender. Finally, the normal *CDH3* transcript was also found to be significantly overexpressed in tumor compared with normal (Figure 2d).

Identification of *CDKN2A* transcript variant 2 in lung adenocarcinoma

RefSeq databases indicate several transcript *CDKN2A* variants which differ in their first exons. At least three alternatively spliced variants, each encoding distinct proteins have been reported with variants 1, 3, and 4 encoding P16, Isoform 3, and ARF respectively (Figure 3a). The data from the exon arrays (Figure 3b) and qRT-PCR verification (Supplementary Figures) suggested the presence of a further splice variant of this gene. Analysis of the Affymetrix probe design and NCBI GenBank databases identifed a cloned sequence originating from testis tissue (Accession #: BG717152, GenBank ID: 13996339) and a mRNA sequence, *CDKN2A* transcript variant 2 (Accession #: NM_058196, GenBank ID: 17738295). However, the record was temporarily removed by NCBI staff since the variant has not been confirmed. Four unique PCR primer sets were designed to assess all four possible transcript variants in this set of lung tumor/normal tissue. Variant 3 was not expressed in lung tissues but the three other variants were all expressed (Figure 3c). Variant 2 was also expressed in lung cancer cell lines

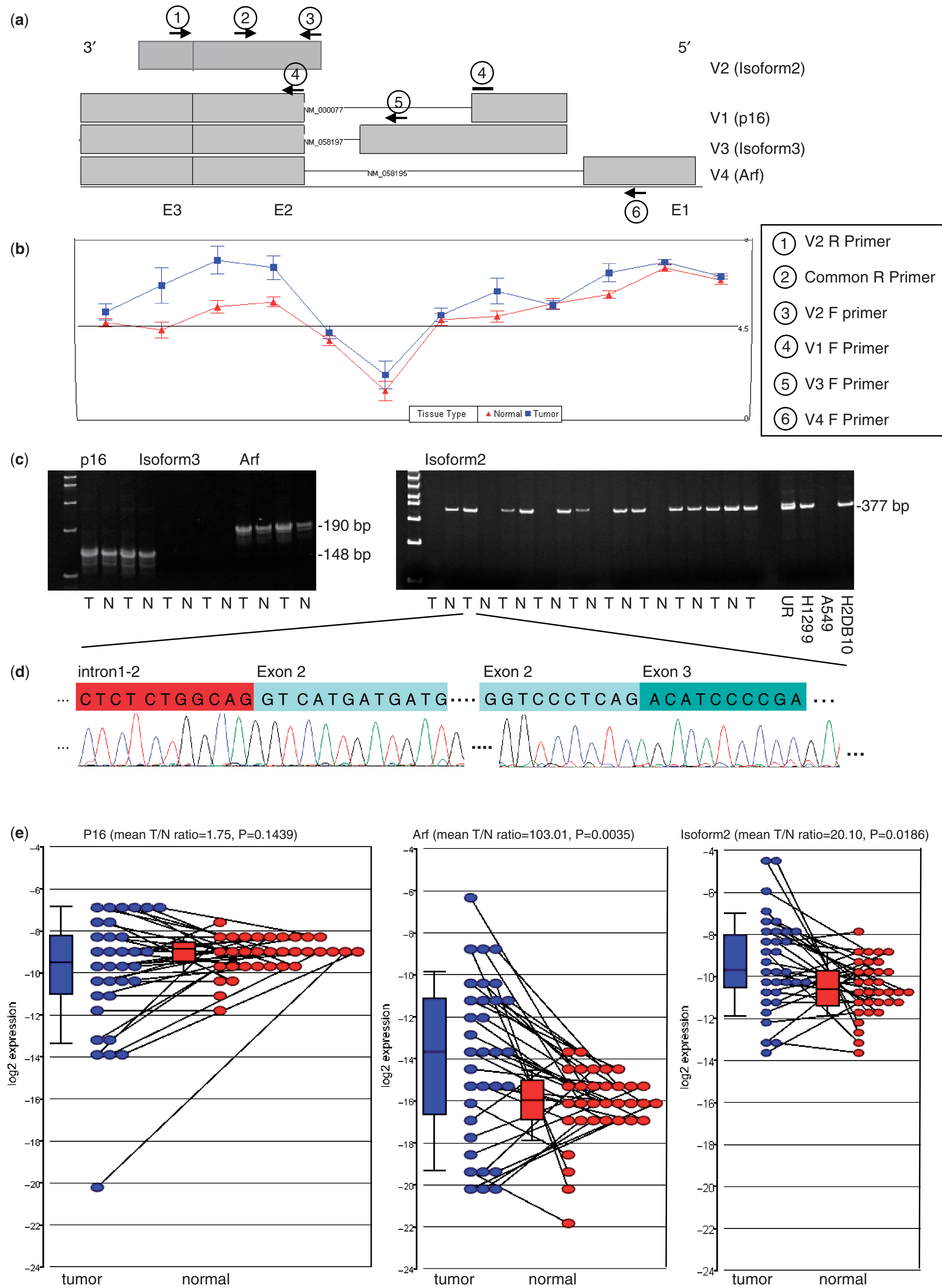


Figure 3. Exon array analysis and PCR analysis for alternative transcript variants of *CDKN2A*. (**a**) Exon structure for *CDKN2A* transcript variants 1 (NM_000077), 2 (NM_058196), 3 (NM_058197) and 4 (NM_058195) indicating the location of PCR primers used for expression analyses. (**b**) Partek GS® alternative splice analysis of exon expression data in 20 patients. The graph shows mean expression value and standard error for each probe set in tumour (blue) and normal (normal) groups. (**c**) verification of PCR product size difference for three variants using primers 1 and 3 (variant 2/isoform2), 2 and 4 (variant 1/p16) and 2 and 6 (variant 4/Arf). Variant 3 was not detected using primers 2 and 5. (**d**) DNA sequencing results of isoform 2. Sequencing shows intronic sequence upstream of exon 2 but also reads directly into exon 3 eliminating the possibility of genomic DNA as the source of the PCR product. **(e)** Quantification of variant P16 (using primers 2 and 4), Arf (2 and 6), and Isoform 2 (2 and 3) in tumour/normal paired samples (tissue pairs are joined by solid lines) from 33 patients.

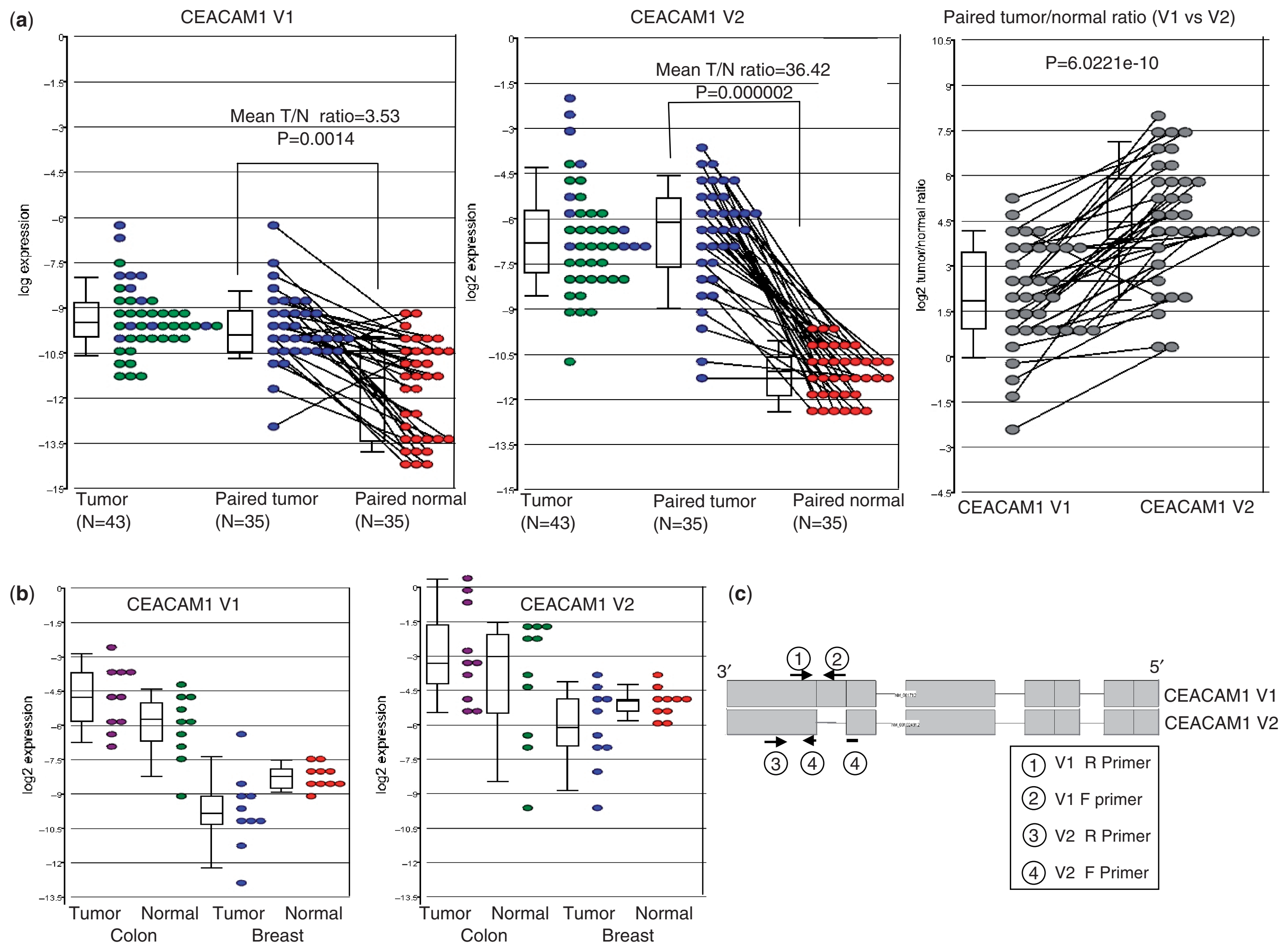


Figure 4. Quantification of *CEACAM1* variants in NSCLC, colon and breast cancer patients. (**a**) Expression of *CEACAM1* variant 1 and variant 2 in 35 tumour/normal paired samples from lung adenocarcinoma patients (tissue pairs are joined by solid lines) and in 43 lung adenocarcinoma ($n = 11$) and squamous cell carcinoma ($n = 32$) samples only (without matched normals). The third graph shows the tumour/normal expression ratio for variant 1 and 2 in the 35 matched tissue pairs. (**b**) Expression of variant 1 and variant 2 in non-paired tumour and normal samples from colon and breast cancer patients (10 patients each). (**c**) Exon location and PCR primers locations for qPCR of variant 1 (NM_001712; primers 1 and 2), variant 2 (NM_001024912; primers 3 and 4).

H1299 and H2DB10, and Universal Reference (UR) RNA which contains 10 combined cancer cell line RNAs, but not in the lung cancer cell line A549. Furthermore, DNA sequencing of the 377 bp variant 2 PCR product confirmed that the transcript encodes a portion of intron 1, consistent with the reported variant 2 mRNA (Figure 3d). Finally, quantification of three variants in 33 paired tumor and normal tissues showed that variant 1 (p16) was not expressed significantly differently between tumor and normal (mean tumor/normal ratio = 1.75; $P = 0.1439$). Both variant 2 (Isoform 2, mean tumor/normal ratio = 20.1; $P = 0.0186$) and variant 4 (ARF, mean tumor/normal ratio = 103.01; $P = 0.0035$) were significantly overexpressed in tumor compared to normal (Figure 3e).

Quantification of *CEACAM1* transcript variants in NSCLC

CEACAM1 is encoded by nine exons with two known variants in the RefSeq database. Variant 1 uses all nine exons while variant 2 is missing exon 7 (Figure 4c). Our array data demonstrated higher expression of exons 1 through 6 as well as 8 and 9 in tumors versus normal while expression of exon 7 was essentially equal (Supplementary Figures). This was observed in 14 of 20 (70%) tissue pairs and suggested differential expression of the two known *CEACAM1* variants in lung tumor and normal tissues as reported previously (38). Furthermore, the exon array data was verified by qRT-PCR using PCR primer sets designed to amplify exon 7 and exon 8 specifically in the same samples (Supplementary Figures).

Expression of *CEACAM1* was reduced in malignant tissues as compared with corresponding normal tissues deriving from breast (39), prostate (40), colon (41), and endometrium (42). These findings indicated that *CEACAM1* might suppress carcinogenesis. However, in contrast, high expression of *CEACAM1* protein was seen in lung tumor tissues and also correlated with poor survival in lung cancer (43–45). Since these studies did not examine expression of the two *CEACAM1* variants

specifically, we designed qRT-PCR assays unique to variants 1 and 2 of *CEACAM1* and analyzed expression of each in lung, breast and colon tumors plus normal tissues from each organ site. In 35 lung adenocarcinoma patients we found that *CEACAM1* variant 2 was highly and significantly overexpressed in paired tumor versus normal (tumor/normal ratio = 36.42; $P = 0.000002$) while variant 1 was only slightly overexpressed in tumor (tumor/normal ratio = 3.53; $P = 0.0014$). Furthermore, analysis of an additional 43 lung tumors, including squamous cell cancers, revealed expression levels indistinguishable from the original 35 adenocarcinoma samples (Figure 4a). In the breast and colon tissues however, we found no significant differences in expression of either variant 1 or variant 2 in tumor versus normal (unmatched) samples (Figure 4b). Given this data we postulated that overexpression of *CEACAM1* variant 2 in lung cancer may be responsible for the survival differences observed between tumor types. However, an analysis of disease-free survival in our cohort of 78 lung cancer patients including 48 stage I and 30 higher stages (median follow up 24 months) showed no association of *CEACAM1* variant 1 or 2 expression with patient survival (Cox regression *P*-values 0.715 and 0.536, respectively).

DISCUSSION

In this study we have performed an extensive identification and verification of alternative splice variant gene expression in NSCLC. To our knowledge, this study is the first such genome wide analysis of alternative splicing events in NSCLC or any other tumor type. Our results indicate that approximately 13% of the 17 800 core RefSeq genes appear to have alternative transcripts that are differentially expressed between lung adenocarcinoma and adjacent normal lung tissue. Furthermore, the largest subsets of these alternatively spliced genes appear to be cancer related and/or involved in cellular processes such as growth and proliferation. For some genes, alternative transcripts have already been identified but we now demonstrate their differential expression in cancer. In many cases however, our microarray data indicates the presence of differentially expressed alternative transcripts that are currently unidentified. Thus it appears that differential expression of alternative transcripts is frequent in NSCLC and that this may be a valuable resource for the development of novel diagnostic, prognostic and therapeutic tools.

While the ability to analyze alternative transcript expression on a genome wide scale is very powerful, verification and validation of this data is labor intensive. For this reason, we chose to focus on genes that have previously been associated with cancer, and where differential expression occurred in >50% of tumor/normal tissue pairs. In total, 11 genes were examined and we were able to validate the array data for six of them. Thus, verification and validation of data from the exon arrays is clearly required. Alternative transcript expression for four of these genes was studied in more detail.

The *ERG* (V-*ETS* avian erythroblastosis virus E26 oncogene homolog) protein shares significant homology with both 5′and 3′ regions of the viral *ETS1* oncogene, *ETS1*, suggesting that it belongs to the *ETS* oncogene family. *ERG* is located at chromosome band 21q22 and has been identified as the target of genomic rearrangement events in acute myeloid leukemia (46), Ewing's sarcoma (47) and prostate cancer (48–50). In acute myeloid leukemia (AML) and Ewing's sarcoma *ERG* has been associated with several translocation fusion partners including *ELF4*, *FUS1* and *EWSR1* (46,51,52). Furthermore, high expression of *ERG* in the absence of karyotypic rearrangement or amplification was demonstrated to be an adverse prognostic factor in patients with AML (53). In prostate cancer, *ERG* is frequently fused to a nearby gene, *TMPRSS2* resulting in androgen regulation of *ERG* and several reports now indicate that the presence of this fusion is a poor prognostic indicator in prostate cancer (54,55). Interestingly, none of the reports cited above discuss the existence of two *ERG* variants and how these are related to the fusion product. Our analysis of variant-specific expression clearly showed that variant 2 has much higher expression in tumor compared to the paired normal lung tissue while variant 1 has similar or lower expression in tumors. Thus it seems that the oncogenic effect of *ERG* may be exerted through functions encoded by variant 2 and expression of fusion gene products should also be closely evaluated for which *ERG* variant is present.Further investigations are required to identify the functional differences between the two variants and could lead to more targeted drug discovery.

The cadherins are a family of transmembrane proteins that mediate calcium-dependent cell-cell adhesion at adherens junctions. The cytoplasmic domain of cadherins binds to A and G catenins and is linked to the actin cytoskeleton via A catenin (56). These interactions are vital for stable cell-cell interactions and maintenance of normal cell physiology. In cancer, disruption of the adherens junctions, for example by downregulation or inactivating mutation of cadherins, can result in epithelial-to-mesenchymal transition, increased proliferation, invasion and metastasis (56,57). In part, this may be mediated by the release and accumulation of B catenin which, when translocated to the nucleus induces transcription of genes such as cyclin D1 and c-myc.

While the role of the prototypic cadherin, E-cadherin (*CDH1*) as a classic tumor suppressor gene in cancer is well established the role of P-cadherin remains unclear as it behaves differently depending on the tumor type being studied. For example, in melanoma, the loss of P-cadherin (and E-Cadherin) allows invasion and migration of cells and thus P-cadherin appears to be acting as a pro-adhesion tumor suppressor (58,59). In breast cancer however, high expression of P-cadherin strongly correlated with high histologic grade, increased proliferation and poor patient survival (60,61). Furthermore, in pancreatic cancer cell lines, overexpression of P-cadherin resulted in increased cell motility, cytoplasmic accumulation of catenins and activation of the Rho GTPases, Rac1 and Cdc42 (62).

In our study we found overexpression of P-cadherin in lung tumors compared to normal lung but also identified overexpression of an alternative splice variant in which exon 2 is missing. Analysis of the resulting mRNA indicates that the normal ATG initiation codon is placed out of frame and would result in a truncated protein after only 27 amino acids. This would clearly result in an inactive protein and would fit with a tumor suppressor function for P-cadherin in lung cancer if it were not for the fact that full length P-cadherin mRNA is actually overexpressed in our tumors. However, upon further analysis we identified several alternative in frame ATG codons downstream of the known translation start site. Furthermore, at least two of these putative alternative start sites have kozak sequences that are believed to be active in other genes. Protein translation initiated at either of these sites would result in a P-cadherin protein lacking the signal peptide and most of the extracellular domain, while retaining the transmembrane domain, juxtamembrane domain and the catenin binding domain. If such a protein were to be overexpressed in tumors one can easily envision disruption of the adherens junctions in a dominant manner leading to catenin accumulation and tumorigenesis.

Its is well known that multiple transcripts are transcribed from the *CDKN2A* (cyclin-dependent kinase inhibitor 2A) locus. *CDKN2A* is an extensively studied tumor suppressor locus that is frequently mutated or deleted in a wide variety of tumor types. Exploration of the RefSeq database identified four transcript variants potentially transcribed from this locus of which three are considered verified. Variant 1 gives rise to the p16 protein and variant 4 gives rise to the alternative reading frame p14/ARF protein. Variant 4 also gives rise to a shorter protein product (p19smARF) which results from an alternative translation start site (63). Variant 3 gives rise to a longer protein that shares the same reading frame as p16 and appears to be specifically expressed in the pancreas. In addition, another transcript variant (p16γ) was recently identified (64) but has yet to be curated in the RefSeq databases. Finally, variant 2 lacks exons 1α and 1β, and exon 2 is slightly longer due to inclusion of an additional 100 bases of intronic sequence. Variant 2 may also have a shorter 3′ UTR than p16 or Arf. Variant 2 was originally cloned from testis tissue but has been temporarily removed by RefSeq staff for further evaluation.

In cancer, inactivation of the p16INK4a/ARF tumor suppressor genes is frequently mediated through genomic deletion, promoter methylation or inactivating mutation leading to loss of p53 and Rb dependent cell cycle regulation (65). In NSCLC, loss of heterozygosity and/or homozygous deletion of the *CDKN2A* locus on chromosome 9p21 has been reported at frequencies up to 40% (8). In our study, ~30% of tumors demonstrated reduced expression of all three measured transcripts (p16, ARF and variant 2) and this is likely a result of genomic deletions. However, in the remaining tumors expression of ARF and variant 2 (but not p16) were significantly higher than in paired normal tissue. Overexpression of ARF in cancer has now been reported several times (66–68) and has been associated with poor differentiation status in hepatocellular carcinoma (67) and worse outcome in B-cell lymphomas (66). Similarly, overexpression of p16 has also been observed and has been associated with progression and poor survival in ovarian cancer (69), prostate cancer (70) and breast cancer (71). While overexpression of p16 and ARF appears to contradict their known cellular functions as tumor suppressors, mechanisms have been proposed whereby this event may be explained through activating mutations in Rb or induction of *myc* and *ras* (68,72,73). However, our data suggests an alternative: that the variant 2 transcript may account for the previously observed overexpression. Variant 2 was originally believed to give rise to a new isoform (Isoform 2) of P16 with the first amino acid encoded by an in frame ATG that is present in the original exon 2. However, we also identified an alternative ATG codon that is in the extended exon 2 and is in frame with ARF. This alternative ATG has a reasonably good Kozak sequence (CCGTCATGC) and, being upstream of the putative p16 isoform 2 start site, would presumably dominate translation initiation. This putative ARF isoform would lack the amino terminal portion of ARF and would therefore be unable to bind TBP-1, E2F, Myc, FoxM1, CTBP1 or mdm2 (74) and may be unable to block cell cycle progression. However others have shown that the carboxy terminus of artificially truncated ARF still accumulates in the nucleolus (75–77) and thus this putative ARF isoform could theoretically act as a dominant negative, thus explaining how overexpression of ARF may be protumorigenic.

CEACAM1 [carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein)] is a cell–cell adhesion molecule that also plays a role in signal transduction. Two common variants are known for *CEACAM1*; one with a long cytoplasmic domain (L form or variant 1) and one with a short cytoplasmic domain (S form or variant 2). The expression of *CEACAM1* in cancer has been extensively studied but early reports appeared to be contradictory. For example, reduced expression of *CEACAM1* was reported in breast, colon, prostate and endometrial cancer (39,41,78) and *CEACAM1* was therefore considered to be a negative regulator of tumor cell growth. However, in melanoma (79) and lung cancer (43,44), several reports indicated that *CEACAM1* was overexpressed in tumors and that this was associated with disease progression and poor outcome. In 1997 Turbide *et al.* (80) found that the L form of *CEACAM1* exhibited a tumor suppressive phenotype and that this was dominant over expression of the S form. Furthermore, using semi-quantitative RT-PCR Wang *et al.* (38) found that the L form of *CEACAM1* predominated in normal lung while the S form appeared more abundant in tumors. Thus they proposed that isoform switching rather than *CEACAM1* downregulation occurs in NSCLC as opposed to other tumor types. Our quantitative analysis clearly demonstrates a switch in abundance from the L form (variant 1) to the S form (variant 2) in NSCLC and we also demonstrate that no such switch appears to occur in breast cancer or colon cancer. Furthermore, we also analyzed a publicly available GeneChip Human Exon 1.0 ST array data set from colon (33) and found no significant

differential expression of *CEACAM1* variants in those 10 pairs of colon tumor/normal samples (data not shown). Thus our findings support the hypothesis that the tumor suppressive or oncogenic effects of *CEACAM1* are splice variant dependent and that expression of the two variants is differentially regulated in different tissue types.

In conclusion, our data demonstrates that differential expression of alternative splice variants is a common event in NSCLC. It also shows that in addition to identification of novel, cancer-related splice variants, additional information can be gained even with regard to extensively studied, cancer-related genes. Splice variant expression should be considered in future genome-wide expression studies and may lead to novel diagnostic, prognostic or therapeutic strategies in the fight against cancer.

(GeneChip Human Exon 1.0 ST array cell files along with GC-RMA data from core gene probsets and patient information have been submitted to GEO databases and GEO Accession # is GSE12236).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

Lung cancer cell lines were kindly provided by Dr Stuart Aaronson at the Mount Sinai School of Medicine.

FUNDING

Funding for open access charge: The work was funded in part by NIH/NCI grant R01 CA94059.

Conflict of interest statement. None declared.

REFERENCES

1. Jemal,A., Murray,T., Ward,E., Samuels,A., Tiwari,R.C., Ghafoor,A., Feuer,E.J. and Thun,M.J. (2005) Cancer statistics. *CA Cancer J. Clin.*, **55**, 10–30.
2. Ries,L.A.G., Melbert,D., Krapcho,M., Mariotto,A., Miller,B.A., Feuer,E.J., Clegg,L., Horner,M.J., Howlader,N., Eisner,M.P. *et al.* (2006) *SEER Cancer Statistics Review, 1975–2004*. National Cancer Institute, Bethesda, MD.
3. Bhattacharjee,A., Richards,W.G., Staunton,J., Li,C., Monti,S., Vasa,P., Ladd,C., Beheshti,J., Bueno,R., Gillette,M. *et al.* (2001) Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc. Natl Acad. Sci. USA*, **98**, 13790–13795.
4. Beer,D.G., Kardia,S.L., Huang,C.C., Giordano,T.J., Levin,A.M., Misek,D.E., Lin,L., Chen,G., Gharib,T.G., Thomas,D.G. *et al.* (2002) Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat. Med.*, **8**, 816–824.
5. Potti,A., Mukherjee,S., Petersen,R., Dressman,H.K., Bild,A., Koontz,J., Kratzke,R., Watson,M.A., Kelley,M., Ginsburg,G.S. *et al.* (2006) A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer. *N. Engl. J. Med.*, **355**, 570–580.
6. Chen,H.Y., Yu,S.L., Chen,C.H., Chang,G.C., Chen,C.Y., Yuan,A., Cheng,C.L., Wang,C.H., Terng,H.J., Kao,S.F. *et al.* (2007) A five-gene signature and clinical outcome in non-small-cell lung cancer. *N. Engl. J. Med.*, **356**, 11–20.
7. Tonon,G., Wong,K.K., Maulik,G., Brennan,C., Feng,B., Zhang,Y., Khatry,D.B., Protopopov,A., You,M.J., Aguirre,A.J. *et al.* (2005) High-resolution genomic profiles of human lung cancer. *Proc. Natl Acad. Sci. USA*, **102**, 9625–9630.
8. Weir,B.A., Woo,M.S., Getz,G., Perner,S., Ding,L., Beroukhim,R., Lin,W.M., Province,M.A., Kraja,A., Johnson,L.A. *et al.* (2007) Characterizing the cancer genome in lung adenocarcinoma. *Nature*, **450**, 893–898.
9. Maniatis,T. and Tasic,B. (2002) Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature*, **418**, 236–243.
10. Black,D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.*, **72**, 291–336.
11. Mironov,A.A., Fickett,J.W. and Gelfand,M.S. (1999) Frequent alternative splicing of human genes. *Genome Res.*, **9**, 1288–1293.
12. Brett,D., Hanke,J., Lehmann,G., Haase,S., Delbruck,S., Krueger,S., Reich,J. and Bork,P. (2000) EST comparison indicates 38% of human mRNAs contain possible alternative splice forms. *FEBS Lett.*, **474**, 83–86.
13. Kan,Z., States,D. and Gish,W. (2002) Selecting for functional alternative splices in ESTs. *Genome Res.*, **12**, 1837–1845.
14. Modrek,B., Resch,A., Grasso,C. and Lee,C. (2001) Genome-wide detection of alternative splicing in expressed sequences of human genes. *Nucleic Acids Res.*, **29**, 2850–2859.
15. Sharp,P.A. (1994) Split genes and RNA splicing. *Cell*, **77**, 805–815.
16. Collesi,C., Santoro,M.M., Gaudino,G. and Comoglio,P.M. (1996) A splicing variant of the RON transcript induces constitutive tyrosine kinase activity and an invasive phenotype. *Mol. Cell Biol.*, **16**, 5518–5526.
17. Gayther,S.A., Barski,P., Batley,S.J., Li,L., de Foy,K.A., Cohen,S.N., Ponder,B.A. and Caldas,C. (1997) Aberrant splicing of the TSG101 and FHIT genes occurs frequently in multiple malignancies and in normal tissues and mimics alterations previously described in tumours. *Oncogene*, **15**, 2119–2126.
18. Scotet,E. and Houssaint,E. (1998) Exon III splicing switch of fibroblast growth factor (FGF) receptor-2 and -3 can be induced by FGF-1 or FGF-2. *Oncogene*, **17**, 67–76.
19. Ge,K., DuHadaway,J., Du,W., Herlyn,M., Rodeck,U. and Prendergast,G.C. (1999) Mechanism for elimination of a tumor suppressor: aberrant splicing of a brain-specific exon causes loss of function of Bin1 in melanoma. *Proc. Natl Acad. Sci. USA*, **96**, 9689–9694.
20. Baudry,D., Hamelin,M., Cabanis,M.O., Fournet,J.C., Tournade,M.F., Sarnacki,S., Junien,C. and Jeanpierre,C. (2000) WT1 splicing alterations in Wilms' tumors. *Clin. Cancer Res.*, **6**, 3957–3965.
21. Slawin,K.M., Shariat,S.F., Nguyen,C., Leventis,A.K., Song,W., Kattan,M.W., Young,C.Y., Tindall,D.J. and Wheeler,T.M. (2000) Detection of metastatic prostate cancer using a splice variant-specific reverse transcriptase-polymerase chain reaction assay for human glandular kallikrein. *Cancer Res.*, **60**, 7142–7148.
22. Liu,H.X., Cartegni,L., Zhang,M.Q. and Krainer,A.R. (2001) A mechanism for exon skipping caused by nonsense or missense mutations in BRCA1 and other genes. *Nat. Genet.*, **27**, 55–58.
23. Lukas,J., Gao,D.Q., Keshmeshian,M., Wen,W.H., Tsao-Wei,D., Rosenberg,S. and Press,M.F. (2001) Alternative and aberrant messenger RNA splicing of the mdm2 oncogene in invasive breast cancer. *Cancer Res.*, **61**, 3212–3219.
24. Kwabi-Addo,B., Ropiquet,F., Giri,D. and Ittmann,M. (2001) Alternative splicing of fibroblast growth factor receptors in human prostate cancer. *Prostate*, **46**, 163–172.
25. Bartel,F., Taubert,H. and Harris,L.C. (2002) Alternative and aberrant splicing of MDM2 mRNA in human cancer. *Cancer Cell*, **2**, 9–15.
26. Barbour,A.P., Reeder,J.A., Walsh,M.D., Fawcett,J., Antalis,T.M. and Gotley,D.C. (2003) Expression of the CD44v2-10 isoform confers a metastatic phenotype: importance of the heparan sulfate attachment site CD44v3. *Cancer Res.*, **63**, 887–892.
27. Steinman,H.A., Burstein,E., Lengner,C., Gosselin,J., Pihan,G., Duckett,C.S. and Jones,S.N. (2004) An alternative splice form of Mdm2 induces p53-independent cell growth and tumorigenesis. *J. Biol. Chem.*, **279**, 4877–4886.
28. Brinkman,B.M. (2004) Splice variants as cancer biomarkers. *Clin. Biochem.*, **37**, 584–594.
29. Venables,J.P. (2004) Aberrant and alternative splicing in cancer. *Cancer Res.*, **64**, 7647–7654.
30. Kalnina,Z., Zayakin,P., Silina,K. and Line,A. (2005) Alterations of pre-mRNA splicing in cancer. *Genes Chromosomes. Cancer*, **42**, 342–357.

31. Venables,J.P. (2006) Unbalanced alternative splicing and its significance in cancer. *Bioessays*, **28**, 378–386.
32. Skotheim,R.I. and Nees,M. (2007) Alternative splicing in cancer: noise, functional, or systematic? *Int. J. Biochem. Cell Biol.*, **39**, 1432–1449.
33. Gardina,P.J., Clark,T.A., Shimada,B., Staples,M.K., Yang,Q., Veitch,J., Schweitzer,A., Awad,T., Sugnet,C., Dee,S. *et al.* (2006) Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. *BMC Genomics*, **7**, 325.
34. Irizarry,R.A., Hobbs,B., Collin,F., Beazer-Barclay,Y.D., Antonellis,K.J., Scherf,U. and Speed,T.P. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, **4**, 249–264.
35. Godfrey,T.E., Kim,S.-H., Chavira,M., Ruff,D.W., Warren,R.S., Gray,J.W. and Jensen,R.H. (2000) Quantitative mRNA expression analysis from formalin-fixed, paraffin-embedded tissues using 5′ nuclease quantitative RT-PCR. *J. Mol. Diagn.*, **2**, 84–91.
36. Tassone,F., Hagerman,R.J., Taylor,A.K., Gane,L.W., Godfrey,T.E. and Hagerman,P.J. (2000) Elevated levels of FMR1 mRNA in carrier males: a new mechanism of involvement in the fragile-X syndrome. *Am. J. Hum. Genet.*, **66**, 6–15.
37. Wisnivesky,J.P. and Halm,E.A. (2007) Sex differences in lung cancer survival: do tumors behave differently in elderly women? *J. Clin. Oncol.*, **25**, 1705–1712.
38. Wang,L., Lin,S.H., Wu,W.G., Kemp,B.L., Walsh,G.L., Hong,W.K. and Mao,L. (2000) C-CAM1, a candidate tumor suppressor gene, is abnormally expressed in primary lung cancers. *Clin. Cancer Res.*, **6**, 2988–2993.
39. Riethdorf,L., Lisboa,B.W., Henkel,U., Naumann,M., Wagener,C. and Loning,T. (1997) Differential expression of CD66a (BGP), a cell adhesion molecule of the carcinoembryonic antigen family, in benign, premalignant, and malignant lesions of the human mammary gland. *J. Histochem. Cytochem.*, **45**, 957–963.
40. Luo,W., Tapolsky,M., Earley,K., Wood,C.G., Wilson,D.R., Logothetis,C.J. and Lin,S.H. (1999) Tumor-suppressive activity of CD66a in prostate cancer. *Cancer Gene Ther.*, **6**, 313–321.
41. Neumaier,M., Paululat,S., Chan,A., Matthaes,P. and Wagener,C. (1993) Biliary glycoprotein, a potential human cell adhesion molecule, is down-regulated in colorectal carcinomas. *Proc. Natl Acad. Sci. USA*, **90**, 10744–10748.
42. Bamberger,A.M., Riethdorf,L., Nollau,P., Naumann,M., Erdmann,I., Gotze,J., Brummer,J., Schulte,H.M., Wagener,C. and Loning,T. (1998) Dysregulated expression of CD66a (BGP, C-CAM), an adhesion molecule of the CEA family, in endometrial cancer. *Am.. J. Pathol.*, **152**, 1401–1406.
43. Laack,E., Nikbakht,H., Peters,A., Kugler,C., Jasiewicz,Y., Edler,L., Brummer,J., Schumacher,U. and Hossfeld,D.K. (2002) Expression of CEACAM1 in adenocarcinoma of the lung: a factor of independent prognostic significance. *J. Clin. Oncol.*, **20**, 4279–4284.
44. Sienel,W., Dango,S., Woelfle,U., Morresi-Hauf,A., Wagener,C., Brummer,J., Mutschler,W., Passlick,B. and Pantel,K. (2003) Elevated expression of carcinoembryonic antigen-related cell adhesion molecule 1 promotes progression of non-small cell lung cancer. *Clin. Cancer Res.*, **9**, 2260–2266.
45. Dango,S., Sienel,W., Schreiber,M., Stremmel,C., Kirschbaum,A., Pantel,K. and Passlick,B. (2008) Elevated expression of carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM-1) is associated with increased angiogenic potential in non-small-cell lung cancer. *Lung Cancer*, **60**, 426–433.
46. Moore,S.D., Offor,O., Ferry,J.A., Amrein,P.C., Morton,C.C. and Dal,C.P. (2006) ELF4 is fused to ERG in a case of acute myeloid leukemia with a t(X;21)(q25-26;q22). *Leuk. Res.*, **30**, 1037–1042.
47. Sorensen,P.H., Lessnick,S.L., Lopez-Terrada,D., Liu,X.F., Triche,T.J. and Denny,C.T. (1994) A second Ewing's sarcoma translocation, t(21;22), fuses the EWS gene to another ETS-family transcription factor, ERG. *Nat. Genet.*, **6**, 146–151.
48. Tomlins,S.A., Rhodes,D.R., Perner,S., Dhanasekaran,S.M., Mehra,R., Sun,X.W., Varambally,S., Cao,X., Tchinda,J., Kuefer,R. *et al.* (2005) Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science*, **310**, 644–648.
49. Lapointe,J., Kim,Y.H., Miller,M.A., Li,C., Kaygusuz,G., van de,R.M., Huntsman,D.G., Brooks,J.D. and Pollack,J.R. (2007) A variant TMPRSS2 isoform and ERG fusion product in prostate cancer with implications for molecular diagnosis. *Mod. Pathol.*, **20**, 467–473.
50. Liu,W., Ewing,C.M., Chang,B.L., Li,T., Sun,J., Turner,A.R., Dimitrov,L., Zhu,Y., Sun,J., Kim,J.W. *et al.* (2007) Multiple genomic alterations on 21q22 predict various TMPRSS2/ERG fusion transcripts in human prostate cancers. *Genes Chromosomes. Cancer*, **46**, 972–980.
51. Panagopoulos,I., Mandahl,N., Mitelman,F. and Aman,P. (1995) Two distinct FUS breakpoint clusters in myxoid liposarcoma and acute myeloid leukemia with the translocations t(12;16) and t(16;21). *Oncogene*, **11**, 1133–1137.
52. Desmaze,C., Brizard,F., Turc-Carel,C., Melot,T., Delattre,O., Thomas,G. and Aurias,A. (1997) Multiple chromosomal mechanisms generate an EWS/FLI1 or an EWS/ERG fusion gene in Ewing tumors. *Cancer Genet. Cytogenet.*, **97**, 12–19.
53. Marcucci,G., Baldus,C.D., Ruppert,A.S., Radmacher,M.D., Mrozek,K., Whitman,S.P., Kolitz,J.E., Edwards,C.G., Vardiman,J.W., Powell,B.L. *et al.* (2005) Overexpression of the ETS-related gene, ERG, predicts a worse outcome in acute myeloid leukemia with normal karyotype: a Cancer and Leukemia Group B study. *J. Clin. Oncol.*, **23**, 9234–9242.
54. Nam,R.K., Sugar,L., Yang,W., Srivastava,S., Klotz,L.H., Yang,L.Y., Stanimirovic,A., Encioiu,E., Neill,M., Loblaw,D.A. *et al.* (2007) Expression of the TMPRSS2:ERG fusion gene predicts cancer recurrence after surgery for localised prostate cancer. *Br. J. Cancer*, **97**, 1690–1695.
55. Demichelis,F., Fall,K., Perner,S., Andren,O., Schmidt,F., Setlur,S.R., Hoshida,Y., Mosquera,J.M., Pawitan,Y., Lee,C. *et al.* (2007) TMPRSS2:ERG gene fusion associated with lethal prostate cancer in a watchful waiting cohort. *Oncogene*, **26**, 4596–4599.
56. Conacci-Sorrell,M., Zhurinsky,J. and Ben-Ze'ev,A. (2002) The cadherin-catenin adhesion system in signaling and cancer. *J. Clin. Invest.*, **109**, 987–991.
57. Paredes,J., Correia,A.L., Ribeiro,A.S., Albergaria,A., Milanezi,F. and Schmitt,F.C. (2007) P-cadherin expression in breast cancer: a review. *Breast Cancer Res.*, **9**, 214.
58. Sanders,D.S., Blessing,K., Hassan,G.A., Bruton,R., Marsden,J.R. and Jankowski,J. (1999) Alterations in cadherin and catenin expression during the biological progression of melanocytic tumours. *Mol. Pathol.*, **52**, 151–157.
59. Hsu,M.Y., Wheelock,M.J., Johnson,K.R. and Herlyn,M. (1996) Shifts in cadherin profiles between human normal melanocytes and melanomas. *J. Investig. Dermatol. Symp. Proc.*, **1**, 188–194.
60. Paredes,J., Albergaria,A., Oliveira,J.T., Jeronimo,C., Milanezi,F. and Schmitt,F.C. (2005) P-cadherin overexpression is an indicator of clinical outcome in invasive breast carcinomas and is associated with CDH3 promoter hypomethylation. *Clin. Cancer Res.*, **11**, 5869–5877.
61. Peralta,S.A., Knudsen,K.A., Salazar,H., Han,A.C. and Keshgegian,A.A. (1999) P-cadherin expression in breast carcinoma indicates poor survival. *Cancer*, **86**, 1263–1272.
62. Taniuchi,K., Nakagawa,H., Hosokawa,M., Nakamura,T., Eguchi,H., Ohigashi,H., Ishikawa,O., Katagiri,T. and Nakamura,Y. (2005) Overexpressed P-cadherin/CDH3 promotes motility of pancreatic cancer cells by interacting with p120ctn and activating rho-family GTPases. *Cancer Res.*, **65**, 3092–3099.
63. Reef,S. and Kimchi,A. (2006) A smARF way to die: a novel short isoform of p19ARF is linked to autophagic cell death. *Autophagy*, **2**, 328–330.
64. Lin,Y.C., Diccianni,M.B., Kim,Y., Lin,H.H., Lee,C.H., Lin,R.J., Joo,S.H., Li,J., Chuang,T.J., Yang,A.S. *et al.* (2007) Human p16gamma, a novel transcriptional variant of p16(INK4A), coexpresses with p16(INK4A) in cancer cells and inhibits cell-cycle progression. *Oncogene*, **26**, 7017–7027.
65. Gil,J. and Peters,G. (2006) Regulation of the INK4b-ARF-INK4a tumour suppressor locus: all for one or one for all. *Nat. Rev. Mol. Cell Biol.*, **7**, 667–677.
66. Sanchez-Aguilera,A., Sanchez-Beato,M., Garcia,J.F., Prieto,I., Pollan,M. and Piris,M.A. (2002) p14(ARF) nuclear overexpression in aggressive B-cell lymphomas is a sensor of malfunction of the common tumor suppressor pathways. *Blood*, **99**, 1411–1418.
67. Ito,T., Nishida,N., Fukuda,Y., Nishimura,T., Komeda,T. and Nakao,K. (2004) Alteration of the p14(ARF) gene and p53 status in human hepatocellular carcinomas. *J. Gastroenterol.*, **39**, 355–361.

68. Ferru,A., Fromont,G., Gibelin,H., Guilhot,J., Savagner,F., Tourani,J.M., Kraimps,J.L., Larsen,C.J. and Karayan-Tapon,L. (2006) The status of CDKN2A alpha (p16INK4A) and beta (p14ARF) transcripts in thyroid tumour progression. *Br. J. Cancer*, **95**, 1670–1677.
69. Dong,Y., Walsh,M.D., McGuckin,M.A., Gabrielli,B.G., Cummings,M.C., Wright,R.G., Hurst,T., Khoo,S.K. and Parsons,P.G. (1997) Increased expression of cyclin-dependent kinase inhibitor 2 (CDKN2A) gene product P16INK4A in ovarian cancer is associated with progression and unfavourable prognosis. *Int. J. Cancer*, **74**, 57–63.
70. Henshall,S.M., Quinn,D.I., Lee,C.S., Head,D.R., Golovsky,D., Brenner,P.C., Delprado,W., Stricker,P.D., Grygiel,J.J. and Sutherland,R.L. (2001) Overexpression of the cell cycle inhibitor p16INK4A in high-grade prostatic intraepithelial neoplasia predicts early relapse in prostate cancer patients. *Clin. Cancer Res.*, **7**, 544–550.
71. Hui,R., Macmillan,R.D., Kenny,F.S., Musgrove,E.A., Blamey,R.W., Nicholson,R.I., Robertson,J.F. and Sutherland,R.L. (2000) INK4a gene expression and methylation in primary breast cancer: overexpression of p16INK4a messenger RNA is a marker of poor prognosis. *Clin. Cancer Res.*, **6**, 2777–2787.
72. Palmero,I., Pantoja,C. and Serrano,M. (1998) p19ARF links the tumour suppressor p53 to Ras. *Nature*, **395**, 125–126.
73. Bates,S., Phillips,A.C., Clark,P.A., Stott,F., Peters,G., Ludwig,R.L. and Vousden,K.H. (1998) p14ARF links the tumour suppressors RB and p53. *Nature*, **395**, 124–125.
74. Chen,Y.W., Paliwal,S., Draheim,K., Grossman,S.R. and Lewis,B.C. (2008) p19Arf inhibits the invasion of hepatocellular carcinoma cells by binding to C-terminal binding protein. *Cancer Res.*, **68**, 476–482.
75. Rizos,H., Darmanian,A.P., Mann,G.J. and Kefford,R.F. (2000) Two arginine rich domains in the p14ARF tumour suppressor mediate nucleolar localization. *Oncogene*, **19**, 2978–2985.
76. Ayrault,O., Karayan,L., Riou,J.F., Larsen,C.J. and Seite,P. (2003) Delineation of the domains required for physical and functional interaction of p14ARF with human topoisomerase I. *Oncogene*, **22**, 1945–1954.
77. Moulin,S., Llanos,S., Kim,S.H. and Peters,G. (2008) Binding to nucleophosmin determines the localization of human and chicken ARF but not its impact on p53. *Oncogene*, **27**, 2382–2389.
78. Prall,F., Nollau,P., Neumaier,M., Haubeck,H.D., Drzeniek,Z., Helmchen,U., Loning,T. and Wagener,C. (1996) CD66a (BGP), an adhesion molecule of the carcinoembryonic antigen family, is expressed in epithelium, endothelium, and myeloid cells in a wide range of normal human tissues. *J. Histochem. Cytochem.*, **44**, 35–41.
79. Thies,A., Moll,I., Berger,J., Wagener,C., Brummer,J., Schulze,H.J., Brunner,G. and Schumacher,U. (2002) CEACAM1 expression in cutaneous malignant melanoma predicts the development of metastatic disease. *J. Clin. Oncol.*, **20**, 2530–2536.
80. Turbide,C., Kunath,T., Daniels,E. and Beauchemin,N. (1997) Optimal ratios of biliary glycoprotein isoforms required for inhibition of colonic tumor cell growth. *Cancer Res.*, **57**, 2781–2788.