Published in final edited form as:

Lung Cancer. 2014 July ; 85(1): 31–39. doi:10.1016/j.lungcan.2014.03.020.

Identification of mRNAs and lincRNAs associated with lung cancer progression using next-generation RNA sequencing from laser micro-dissected archival FFPE tissue specimens

Matthew L. Morton^{a,1}, Xiaodong Bai^{b,1}, Callie R. Merry^{c,d}, Philip A. Linden^g, Ahmad M. Khalil^{c,f}, Rom S. Leidner^{#e,f}, and Cheryl L. Thompson^{#a,f,*}

^aDepartment of Family Medicine and Community Health, Case Western Reserve University, Cleveland, OH, USA

^bCenter for RNA Molecular Biology, Case Western Reserve University, Cleveland, OH, USA

^cDepartment of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH, USA

^dDepartment of Biochemistry, Case Western Reserve University, Cleveland, OH, USA

^eDepartment of Medicine, Case Western Reserve University, Cleveland, OH, USA

^fCase Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH, USA

^gDepartment of Surgery, University Hospitals Case Medical Center, Cleveland, OH, USA

[#] These authors contributed equally to this work.

Abstract

Objectives—Adenocarcinoma in situ (AIS) is an intermediate step in the progression of normal lung tissue to invasive adenocarcinoma. However, molecular mechanisms underlying this progression remain to be fully elucidated due to challenges in obtaining fresh clinical samples for downstream analyses. Formalin fixation and paraffin embedding (FFPE) is a tissue preservation system widely used for long-term storage. Until recently, challenges in working with FFPE precluded using new RNA sequencing technologies (RNA-seq), which would help clarify key pathways in cancer progression. Also, isolation techniques including laser-capture micro-dissection provide the ability to select histopathologically distinct tissues, allowing researchers to study transcriptional variations between tightly juxtaposed cell and tissue types.

Materials and methods—Utilizing these technologies and new alignment tools we examined differential expression of long intergenic non-coding RNAs (lincRNAs) and mRNAs across normal, AIS and invasive adenocarcinoma samples from six patients to identify possible markers of lung cancer progression.

© 2014 Elsevier Ireland Ltd. All rights reserved.

*Corresponding author at: Department of Family Medicine and Community Health, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA. Tel.: +1 216 368 3956; fax: +1 216 369 4348. Cheryl.L.Thompson@case.edu, cheryl@case.edu.

¹The authors wish it to be known that, in their opinion, the first 2 authors should be regarded as joint first authors.

Conflicts of interest statement The authors declare that they have no competing interests.

Results—RNA extracted and sequenced from these 18 samples generated an average of 198 million reads per sample. After alignment and filtering, uniquely aligned reads represented an average 35% of the total reads. We detected differential expression of a number of lincRNAs and mRNAs when comparing normal to AIS, or AIS to invasive adenocarcinoma. Of these, 5 lincRNAs and 31 mRNAs were consistently up- or down-regulated from normal to AIS and more so to invasive carcinoma. We validated the up-regulation of two mRNAs and one lincRNA by RT-qPCR as proof of principle.

Conclusion—Our findings indicate a potential role of not only mRNAs, but also lincRNAs in the progression to invasive adenocarcinoma. We anticipate that these findings will lay the groundwork for future experimental studies of candidate RNAs from FFPE to identify their functional roles in lung cancer.

Keywords

RNA-Seq; Adenocarcinoma in situ of the lung; Invasive lung adenocarcinoma; Formalin-fixed paraffin embedded (FFPE); lincRNA; Gene expression; Laser capture microdissection (LCM)

1. Introduction

Adenocarcinoma in situ (AIS) of the lung is a non-invasive, adenomatous form of non-small cell lung cancer (NSCLC) [1,2]. The epidemiology of AIS mimics that of other forms of lung cancer with several primary exceptions. Compared to patients with other forms of cancer, patients with AIS are more likely to be nonsmokers or have a minimal smoking history. Additionally, the proportion of AIS patients who are female is significantly higher, closer to 50%, than for other forms of lung cancer. Current treatment for AIS includes complete resection (either lobar or sublobar) for early-stage disease. Several studies report 5-year survival rates close to 100% in AIS patients with complete resection [1]. These rates are consistent with studies published on patients with tumors comprised of pure AIS without invasion, AIS with a small (i.e. < 5 mm) focus of invasion, and predominantly AIS with invasion localized to the periphery of a central scar. However, for patients with AIS with obvious invasion, the 5-year survival rate is much lower [2], highlighting the need for understanding of factors influencing the progression of AIS to invasive adenocarcinoma.

The use of archival, formalin-fixed and paraffin-embedded (FFPE) tissue samples for RNA expression analyses has recently received interest as a result of improved techniques for recovering RNA from FFPE samples, as well as the vast number of these samples often available to researchers. Many of these samples not only carry extensive clinical data, but also have well-preserved tissue architecture that is amenable to dissection using techniques such as laser-capture microdissection (LCM) [3,4], allowing researchers to accurately extract regions enriched with cells of interest. The use of FFPE samples for RNA expression analyses requires overcoming several challenges. Formalin fixation results in RNA degradation and fragmentation, as well as modification to nucleotides, all of which decrease the yield of RNA extracted from such samples [5,6]. As a result of numerous technological advances in addressing these challenges, several recent studies utilizing real-time quantitative PCR (RT-qPCR) comparing fresh frozen samples to FFPE samples as old as 40

years have demonstrated the utility of RNA extracted from FFPE samples for mRNA and miRNA expression [7–11].

While the majority of work on RNA extracted from FFPE samples has been completed using RT-qPCR and microarray approaches [5], there has been recent interest in the use of next generation RNA Sequencing (RNA-Seq) technologies for examining gene expression from FFPE samples [12], as high-throughput RNA-Seq has become more accessible and provides comprehensive and quantitative measurements of gene expression. The combination of LCM, which allows for isolation of well-defined regions of distinct tissues, and RNA-Seq technology, which permits agnostic transcriptome-wide analysis, represents significant opportunities for discovery-phase research.

The AIS progression model allows for identification of genomic alterations occurring in the transition from normal epithelium to AIS, which may be important for early detection and prevention, as well as from AIS to invasive adenocarcinoma, which may represent important targets for treatment. The ability to isolate regions of normal tissue, AIS, and invasive cancers using LCM from FFPE surgical resection specimens and subsequently perform whole transcriptome analysis by next generation RNA-Seq provides an opportunity to probe large reservoirs of samples for identification of early and late markers of cancer progression. In this study, we tested the feasibility of combining LCM with FFPE RNA-Seq to identify novel alterations in the normal-AIS-invasive adenocarcinoma progression model of lung cancer.

2. Materials and methods

2.1. FFPE tissue microdissection and RNA isolation

Archival FFPE blocks were reviewed by a pulmonary pathologist to select and de-identify lung resection samples from the three preceding years, each of which included large components of pure AIS, separate invasive adenocarcinoma, and normal lung tissue, under an IRB approved research protocol. From each tissue block, six sections of 4- μ m thickness on PEN membrane slides were prepared under RNAase free conditions. Laser Capture Microdissection (LCM) of distinct areas of normal lung, pure AIS and invasive adenocarcinoma was performed using the Applied Biosystems Arcturus Veritas Microdissection System (Life Technologies, Grand Island, NY). Additionally, a companion 12- μ m section from a block of normal lung tissue from each case was prepared under identical conditions. This larger slice size allowed for capturing of sufficient normal tissue from a non-microdissected specimen within a single cut. Total RNA was isolated from microdissected AIS, invasive adenocarcinoma, and companion normal lung sections using the Ambion RecoverAll kit (Life Technologies, Grand Island, NY) according to manufacturer's protocol. RNA concentration was determined using the Invitrogen Qubit fluorometer (Life Technologies, Grand Island, NY) and RNA Integrity Number (RIN) by the 2100 Bioanalyzer (Agilent Technologies, Inc., Santa Clara, CA). To assess feasibility across a range of histologies, we selected 18 initial samples for RNA-Seq: normal, AIS and invasive components from six individuals. Total RNA yield averaged 928 ng per sample (283–3150 ng), at mean concentration of 20.5 ng/ μ L (5.7–41.6 ng/ μ L); RIN values ranged from 2.1 to 2.4 (mean 2.2) which are higher than expected for FFPE samples.

2.2. Transcriptome sequencing and alignment

RNA sequencing libraries were prepared using the Ovation RNA-Seq FFPE (7150-08) and Encore NGS Library System 1 kits (300-08) according to manufacturer's protocol (NuGEN Technologies, Inc., San Carlos, CA). Non-multiplexed sequencing of 100 bp paired-end reads was performed using the Illumina GAIIX (Illumina, Inc., San Diego, CA) according to manufacturer's protocols through Case Western Reserve University Sequencing Core facility. Alignment of paired-end reads (FASTQ) to human genome build 19 (hg19) was performed using Genentech's Genomic Short-Read Nucleotide Alignment Program (GSNAP version 2012-06-25) [13–15], allowing for a maximum of 5 mismatches, including penalties for insertions and deletions and local or distant splicing, and configured to recognize novel splicing and detect known splice-sites from RefGene provided through the UCSC Genome Browser. While other alignment software packages require a higher mismatch threshold, (e.g. 10 maximum), in order to accommodate alignments overlapping intron-exon boundaries, the current implementation of GSNAP can identify terminal alignments and integrates the Genomic Mapping and Alignment Program (GMAP), reducing the threshold of allowed mismatches required. Single nucleotide polymorphism (SNP) tolerant alignment was not configured. Other GSNAP options utilized in order to increase accuracy included turning trimming off, setting clipping of ambiguous known splicing at ends of reads to extend into introns instead of clipping at splice sites, and setting unknown (N) characters in a query as mismatches, in order to increase stringency. The Sanger quality protocol was selected for compatibility with the Illumina Pipeline 1.8, which uses the NCBI/Sanger Phred scale for quality scores. Reads were additionally aligned to known human lincRNA databases [16].

The RNA-SeQC [17] software package for quality control analysis of RNA-Seq data was used to generate quality control metrics including mean coverage, mate pair information, the number of genes and transcripts detected, and strand specificity. RefSeq annotations on hg19 were used for RNA-SeQC analysis.

2.3. Differential expression analysis

Uniquely aligned reads were further analyzed using HTseq-count to obtain raw read counts of genes in RefSeq annotation and lincRNAs in Cabili et al. [16]. The statistical significance of differences in expression between the six AIS and six normal samples was assessed using normalized read counts in DESeq package in R. This was repeated comparing the six AIS and six invasive samples. Expression values of genes and lincRNAs were calculated as FPKM (fragment per kilobase of exon per million of mapped fragments) and used for heatmap generation after taking base-2 logarithmic values.

Gene ontology enrichment analysis of statistically significantly altered mRNAs was done using GOEAST [18] for mRNAs showing a consistent upward or downward expression trend from normal to AIS to invasive.

2.4. qPCR validation of candidate lincRNA and mRNAs

To validate our findings, we selected one lincRNA and three mRNAs for validation. Of these, *linc-CBR1-2*, and two mRNAs, *RAB25* and *HMBG3* were up-regulated in AIS

compared to normal tissue. The last mRNA, *SCAMP3*, was chosen as one that was identified from sequencing as not significantly altered between normal, AIS or invasive (unadjusted $p > 0.05$) (negative control). Sequence for *linc-CBR1-2* was obtained from the UCSC genome browser and used to design qPCR primers with Primer3 Input (version 0.4.0) [19]. *Linc-CBR1-2* primers are within exon 1 (forward: 5'-TTT GGA GAG CTG TGA ACG TG-3', reverse: 5'-GAG CGA GCT TGA TAA GCT GAA-3'). Primer sequences for mRNA validation were obtained from Primer-Bank [20]: *RAB25* (forward: 5'-TCGTGGGTAACAAAAGTGACC-3', reverse: 5'-AGCTCAACATTGGTAGAGTCCA-3'), *SCAMP3* (forward: 5'-CTGTGCTAGGAATTGTCATGCT-3', reverse: 5'-CCGCAGGGTTGGAGAAGAC-3'), and *HMGB3* (forward: 5'-CCCAGAGGTCCCTGTCAATTT-3', reverse: 5'-CGATCATAGCGCACTTTATCTGC-3'). cDNA was synthesized using RNA to cDNA EcoDry Premix Random Hexamer (Clontech, Mountain View, CA). qPCR was performed using Maxima® SYBR Green/ROX qPCR Master Mix (Fermentas Life Sciences, Glen Burnie, MD) on the three patients for which sufficient quantity of RNA remained for their normal, AIS and invasive components. Samples were normalized using ACTB RNA as an endogenous control. Fold enrichment for each patient was calculated in reference to the expression of the paired normal tissue sample.

3. Results

3.1. Transcriptome sequencing and alignment

Quality metrics produced by RNA-SeQC [17] are provided for each sample in Supplementary Table I. A mean read count of 198 million was observed for the 18 samples with an average unique mapping rate of 35% after filtering. On average, 28,572 transcripts were detected. After removal of unaligned reads and reads with multiple matching alignments, an average of 69 million uniquely aligned reads per sample were retained for further analysis. Mean coverage as generated by RNA-SeQC [17] is presented for low-, medium-, and high-expression tertiles of 1000 representative transcripts in Fig. 1A–C, respectively.

Supplementary Table I related to this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.lungcan.2014.03.020>.

3.2. Gene expression analyses

Table 1 lists top differentially expressed transcripts after filtering to select those showing consistent upward or downward trends of Fold Change (FC) values across tissues. A full list of those altered from normal to AIS and AIS to invasive are found in Supplementary Tables II and III, respectively. A total of five lincRNAs and 31 mRNAs with $p < 0.05$ were differentially expressed with a trend (20 were classified as new genes not already existing in the UCSC annotations). Interestingly, all of these RNAs with statistically significant associations and this trend were all down-regulated in the progression. Several additional RNAs were either down or up-regulated in one comparison, but not both. Our most statistically significantly altered lincRNA with a consistent trend was *linc-COX4NB-9*, which showed a significant reduction of more than 2 fold from normal to AIS and a further

reduction of more than 2 fold from AIS to invasive carcinoma. Several mRNAs showed similar large drops in expression in the progression from normal to invasive cancer (Table 1).

Supplementary material related to this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.lungcan.2014.03.020>.

The top lincRNAs and mRNAs are presented in heatmaps by sample in Figs. 2 and 3, respectively. These RNAs represent a signature that changes from normal to AIS to invasive carcinoma.

The gene ontology enrichment analysis of mRNAs in Table 1 identified neuropeptide signaling as the top enriched biological process ($p = 6.5 \times 10^{-21}$). A number of processes associated with invasiveness were also in our top enriched pathways, including filopodium assembly ($p = 4.0 \times 10^{-17}$), blood vessel maturation ($p = 1.5 \times 10^{-11}$) and angiogenesis ($p = 3.6 \times 10^{-3}$). For molecular functions, receptor activity was the most enriched ($p = 1.9 \times 10^{-14}$), as was Transforming Growth Factor β (TGF β) receptor activity ($p = 5.2 \times 10^{-5}$).

3.3. Validation of differentially expressed lincRNAs and mRNAs by RT-qPCR

Using RT-qPCR for *RAB25*, *SCAMP3* and *HMGB3*, we validated what we observed by RNA-seq for these mRNAs using RT-qPCR—*RAB25* and *HMGB3* up-regulated in AIS compared with normal, as well as the lack of significant change in *SCAMP3* (Figs. 4). We also confirmed a stepwise up-regulation of *linc-CBR1-2* from normal to AIS to invasive by RT-qPCR (Fig. 5). Overall, this data suggests excellent correlation between RT-qPCR and RNA-Seq findings.

4. Discussion

We successfully micro-dissected and sequenced RNA from 18 samples representing normal tissue, AIS, and invasive carcinoma from six patients. We confirmed the expression patterns of several RNAs via RT-qPCR. However, because of the small sample size, confirmation in additional samples is warranted prior to functional studies. An important limitation of this study, which should be noted, is that we did not also perform these experiments using paired fresh frozen tissues from the same patients. As is generally the case clinically, co-collection is not routine practice. Although fresh tissue is widely considered the “gold standard”, FFPE samples are not only more accessible and more available, but also more amenable to dissection via LCM, allowing for progression studies such as this or other studies of small regions of a sample. Earlier studies have demonstrated concordance between FFPE and fresh frozen samples in RT-qPCR studies of mRNAs [7–11], and therefore we feel that this methodology will be useful in future studies.

Our findings, although limited by the relatively small sample size, show interesting and promising findings. Some of the differentially expressed genes we identified in this study are known to play important roles in cancer. For example, CDH5, an established tumor suppressor gene associated with angiogenesis, was shown to be up-regulated in lung cancers in an earlier candidate gene study [21]. ACVRL1 (also known as ALK1) has well known

importance in lung cancer, and ALK-rearrangement is a target in the treatment of lung cancers with this fusion. FABP4 interacts with PTEN, a well established tumor suppressor, and has been associated with a number of cancers [22]. Our GO analysis showed that our top mRNAs fell into a number of related molecular functions known to be associated with cancer. The identification of a number of angiogenesis related genes led to these pathways being among our top hits, as well as pathways related to TGF β , a well known tumor suppressor gene.

To the best of our knowledge there have been no studies characterizing the expression profile of the whole transcriptome of normal and AIS tissues using RNA-Seq. Furthermore, our study is the first to compare the expression of lincRNAs in normal, AIS and invasive lung adenocarcinoma. However, there have been microarray studies that examined gene expression in AIS. Comparing our findings to two published studies that utilized microarray experiments suggests significant overlap in findings. In the first genome-wide expression studies in AIS, Goodwin et al. [23] used a microarray to identify differentially expressed mRNAs in AIS tumors and benign tissues from three patients. We were able to replicate a number of their findings in our AIS vs. normal comparison. Specifically, we noted up-regulation of *CRABP2*, *CD24*, *XBP1* and *AGR2* as well as down-regulation of *FCN2* and *IL7R*. Zhang et al. [24] recently reported differential gene expression from AIS vs. adenocarcinoma with AIS features. They report that *CLDN18* was up-regulated and *ATM* and *ATR* were down-regulated. We saw similar patterns, although the changes were only minor in our samples, and only *CLDN18* was statistically significantly up-regulated. There is no current literature on the role of lincRNAs in this type of lung cancer. Thus, not unexpectedly, this study is the first to identify lincRNAs that are differentially expressed in AIS and invasive adenocarcinoma.

One recent study used micro and macro dissection to isolate areas of normal tissue and invasive tumors [25]. Using data available from Gene Expression Omnibus (GSE31552) [25] on available paired normal and tumor samples from adenocarcinomas of the lung, we compared our top mRNA hits (listed in Table 1) to their data using GEO2R. *FABP4* was the highest deregulated tumor marker in their sample (adjusted $p = 7.8 \times 10^{-19}$). In addition, both *AGER* (adjusted $p = 0.039$) and *SLC6A4* (adjusted $p = 0.0017$) were also down-regulated in their tumors. We also noted that, although they did not hold up to multiple testing, of the other 28 mRNAs, 11 were nominally significantly altered in the same direction in their samples.

The small sample size allowed feasibility testing of our approach and represents an important proof-of-principle in support of RNASeq of microdissected FFPE tissue. Many of our RNAs were found at low levels, particularly the lincRNAs. However, lincRNAs are, on average, expressed at much lower levels than mRNAs [16,26,27]. These molecules are regulatory in nature and therefore can be expressed at one copy per cell in some cases. A total of five lincRNAs and 31 mRNAs showed a consistent upward or downward trend of differential expression across the progression sequence in our RNA-Seq analysis [28]. Using RT-qPCR, we validated up-regulation of *linc-CBR1-2* from normal to AIS and from normal to invasive as a proof of principle, though as with most lincRNAs, the structure and function of *linc-CBR1-2* has not yet been investigated. We realize that the small sample size limited

our ability to detect lincRNAs and mRNAs that may only be differentially expressed in a minority of patients, or whose levels of expression are only moderately differentially expressed. Another limitation to our study, which is true of any study using LCM, is that components of the tumor that were not the target tissue, such as adjacent normal cells, could have been included. To minimize this limitation, we performed all LCM under the guidance of a pathologist.

5. Conclusions

This study is the first, to our knowledge, that comprehensively identifies both mRNAs and lincRNAs that are differentially expressed in the AIS progression model. We also describe a method for accurately isolating and sequencing RNA from distinct regions of FFPE specimens. This work, while built on a small number of samples, will help lay the groundwork for a broad range of future studies incorporating FFPE tissue specimens into downstream RNA-Seq analysis pipelines by providing important data that can be used in designing future experiments. Future work will be required to extend these findings to larger sample sizes and confirm in independent populations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work used the High Performance Computing Resource at Case Western Reserve University. The data discussed in this publication has been deposited in NCBI's Gene Expression Omnibus [29] (GSE52248) (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE52248>).

Funding This work was supported by the National Cancer Institute (K07 CA136758); the Case Comprehensive Cancer Center (P30 CA043703); the Center for RNA Molecular Biology at Case Western Reserve University; an American Society of Clinical Oncology Career Development Award; and an NCI Clinical Oncology Research Career Development Award (K12CA076917).

References

- [1]. Arenberg D. Bronchioloalveolar carcinoma. *Semin Respir Crit Care Med*. 2011; 32:52–61. [PubMed: 21500124]
- [2]. Schmidt L, Myers J. Bronchioloalveolar carcinoma and the significance of invasion: predicting biologic behavior. *Arch Pathol Lab Med*. 2010; 134:1450–4. [PubMed: 20923299]
- [3]. Burgemeister R. Nucleic acids extraction from laser microdissected FFPE tissue sections. *Methods Mol Biol*. 2011; 724:117–29. [PubMed: 21370010]
- [4]. Joseph A, Gnanaprasagam VJ. Laser-capture microdissection and transcriptional profiling in archival FFPE tissue in prostate cancer. *Methods Mol Biol*. 2011; 755:291–300. [PubMed: 21761313]
- [5]. Farragher SM, Tanney A, Kennedy RD, Paul Harkin D. RNA expression analysis from formalin fixed paraffin embedded tissues. *Histochem Cell Biol*. 2008; 130:435–45. [PubMed: 18679706]
- [6]. Gnanaprasagam VJ. Unlocking the molecular archive: the emerging use of formalin-fixed paraffin-embedded tissue for biomarker research in urological cancer. *BJU Int*. 2010; 105:274–8. [PubMed: 19519763]
- [7]. Liu A, Xu X. MicroRNA isolation from formalin-fixed, paraffin-embedded tissues. *Methods Mol Biol*. 2011; 724:259–67. [PubMed: 21370018]

- [8]. Lu X, van der Straaten T, Tiller M, Li X. Evidence for qualified quantitative mRNA analysis in formalin-fixed and paraffin-embedded colorectal carcinoma cells and tissue. *J Clin Lab Anal.* 2011; 25:166–73. [PubMed: 21567463]
- [9]. Ludyga N, Grunwald B, Azimzadeh O, Englert S, Hofler H, Tapio S, Aubele M. Nucleic acids from long-term preserved FFPE tissues are suitable for downstream analyses. *Virchows Arch.* 2012; 460:131–40. [PubMed: 22270699]
- [10]. Stewart GD, Baird J, Rae F, Nanda J, Riddick AC, Harrison DJ. Utilizing mRNA extracted from small, archival formalin-fixed paraffin-embedded prostate samples for translational research: assessment of the effect of increasing sample age and storage temperature. *Int Urol Nephrol.* 2011; 43:961–7. [PubMed: 21448682]
- [11]. Waldron L, Simpson P, Parmigiani G, Huttenhower C. Report on emerging technologies for translational bioinformatics: a symposium on gene expression profiling for archival tissues. *BMC Cancer.* 2012; 12:124. [PubMed: 22458912]
- [12]. Sinicropi D, Qu K, Collin F, Crager M, Liu ML, Pelham RJ, Pho M, Rossi AD, Jeong J, Scott A, Ambannavar R, Zheng C, Mena R, Esteban J, Stephans J, Morlan J, Baker J. Whole transcriptome RNA-Seq analysis of breast cancer recurrence risk using formalin-fixed paraffin-embedded tumor tissue. *PLoS ONE.* 2012; 7:e40092. [PubMed: 22808097]
- [13]. Grant GR, Farkas MH, Pizarro AD, Lahens NF, Schug J, Brunk BP, Stoeckert CJ, Hogenesch JB, Pierce EA. Comparative analysis of RNA-Seq alignment algorithms and the RNA-Seq unified mapper (RUM). *Bioinformatics.* 2011; 27:2518–28. [PubMed: 21775302]
- [14]. Wu TD, Watanabe CK. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics.* 2005; 21:1859–75. [PubMed: 15728110]
- [15]. Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics.* 2010; 26:873–81. [PubMed: 20147302]
- [16]. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 2011; 25:1915–27. [PubMed: 21890647]
- [17]. DeLuca DS, Levin JZ, Sivachenko A, Fennell T, Nazaire MD, Williams C, Reich M, Winckler W, Getz G. RNA-SeqQC. RNA-seq metrics for quality control and process optimization. *Bioinformatics.* 2012; 28:1530–2. [PubMed: 22539670]
- [18]. Zheng Q, Wang XJ. GOEAST: a web-based software toolkit for gene ontology enrichment analysis. *Nucleic Acids Res.* 2008; 36:W358–63. [PubMed: 18487275]
- [19]. Rozen, S.; Skaletsky, HJ. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz, SA.; Misener, S., editors. *Bioinformatics methods and protocols: methods in molecular biology.* Humana Press; Totowa, NJ: 2000. p. 365-86.
- [20]. Wang X, Spandidos A, Wang H, Seed B. PrimerBank: a PCR primer database for quantitative gene expression analysis, 2012 update. *Nucleic Acids Res.* 2012; 40:D1144–9. [PubMed: 22086960]
- [21]. Metodieva SN, Nikolova DN, Cherneva RV, Dimova II, Petrov DB, Toncheva DI. Expression analysis of angiogenesis-related genes in Bulgarian patients with early-stage non-small cell lung cancer. *Tumori.* 2011; 97:86–94. [PubMed: 21528670]
- [22]. Gorbenko O, Panayotou G, Zhyvoloup A, Volkova D, Gout I, Filonenko V. Identification of novel PTEN-binding partners: PTEN interaction with fatty acid binding protein FABP4. *Mol Cell Biochem.* 2010; 337:299–305. [PubMed: 19911253]
- [23]. Goodwin LO, Mason JM, Hajdu SI. Gene expression patterns of paired bronchioloalveolar carcinoma and benign lung tissue. *Ann Clin Lab Sci.* 2001; 31:369–75. [PubMed: 11688848]
- [24]. Zhang Z, Wang A, Sun B, Zhan Z, Chen K, Wang C. Expression of CLDN1 and CLDN10 in lung adenocarcinoma in situ and invasive lepidic predominant adenocarcinoma. *J Cardiothorac Surg.* 2013; 8:95. [PubMed: 23591077]
- [25]. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva ANCB. GEO: archive for functional genomics data sets – update. *Nucleic Acids Res.* 2013; 41:D991–5. [PubMed: 23193258]

- [26]. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*. 2009; 458:223–7. [PubMed: 19182780]
- [27]. Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A, Regev A, Lander ES, Rinn JL. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci U S A*. 2009; 106:11667–72. [PubMed: 19571010]
- [28]. Zhang WC, Zhang ZF, You J, Wang CL. Global gene expression differentiating pure bronchioloalveolar carcinoma from adenocarcinoma with bronchi-oloalveolar carcinoma features. *Eur J Cardiothorac Surg*. 2012
- [29]. Edgar R, Domrachev M, Lash AE. Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 2002; 30:207–10. [PubMed: 11752295]

Fig. 1.

Mean coverage (the number of reads that cover a given genomic position in units of reads per base) for bottom, middle, and top 1000 transcripts with low, medium, and high expression plotted against distance (in base pairs) from the 3' end. Mean coverage across our samples for (A) the bottom 1000 transcripts with low expression, (B) the middle 1000 transcripts with medium expression, and (C) the top 1000 transcripts with high expression was generated and plotted using RNA-SeQC. Three individual cases are designated as A, B, C, and corresponding tissue samples are designated as “ais” – adenocarcinoma in situ, “inv” – invasive adenocarcinoma, and “nor” normal lung.

Fig. 2.

Heatmaps of differentially expressed lincRNAs. Log2 FPKM values of differentially expressed lincRNAs for each sample were plotted in normal (NOR) vs. adenocarcinoma in situ (AIS) (A) and AIS vs. invasive adenocarcinoma (INV) (B). The heat map color legend is provided in the figure.

Fig. 3. Heat maps of differentially expressed mRNAs. Log2 FPKM values of differentially expressed lincRNAs for each sample were plotted in normal (NOR) vs. adenocarcinoma in situ (AIS) (A) and AIS vs. invasive adenocarcinoma (INV) (B). The heat map color legend is provided in the figure.

Fig. 4. qPCR validation of differentially expressed mRNAs from RNA-seq data. Total RNA was isolated from normal, AIS, and invasive tissues from three distinct patients A, B, and C. cDNA synthesis was performed prior to carrying out RT-qPCR analysis on three mRNAs that showed differential expression by RNA-seq.

Fig. 5. RT-qPCR validation of *linc-CBR1-2*. RT-qPCR validation was performed on samples from patients A, B, and C. Results of the validation for each patient are presented as a histogram. The average of replicates across conditions is also presented.

Table 1

Differentially expressed lincRNAs and mRNAs with consistent trends. All statistically significant (unadjusted) lincRNAs and mRNAs that demonstrated a trend toward up or down regulation in invasive lung cancer. Chr – chromosome, FC – fold change, NOR – normal lung, AIS – adenocarcinoma in situ, INV – invasive adenocarcinoma.

Gene symbol	Accession #	FPKM			Location		Statistics						
		NOR	AIS	INV	Chr	Start	End	FC AIS v. NOR	p-val AIS v. NOR	p-adj AIS v. NOR	FC INV v. AIS	p-val INV v. AIS	p-adj INV v. AIS
lincRNAs <i>linc-ADCY1-4</i>	Accession #'s not available for lincRNA	101.36	33.03	8.87	7	45,228,588	45,232,084	-1.62	0.0201	0.76	-1.9	0.0347	1
<i>linc-APPBP2-2</i>		69.55	25.94	6.65	17	59,470,733	59,477,096	-1.42	0.0165	0.71	-1.96	0.0387	1
<i>linc-C3orf30-6</i>		104.79	39.9	8.55	3	116,958,058	116,963,238	-1.39	0.0347	0.91	-2.22	0.0204	1
<i>linc-COX4NB-9</i>		1775.78	410.84	85.34	16	86,508,051	86,524,485	-2.11	9.43E-05	0.07	-2.27	0.0001	0.76
<i>linc-GRIN2A-1</i>		233.49	103.23	33.94	16	10,608,824	10,619,948	-1.18	0.0357	0.92	-1.6	0.0241	1
mRNAs <i>ACVRL1</i>	NM 001077401, NM 000020	1354.15	351.09	133.14	12	52,301,202	52,317,145	-1.95	0.0003	0.11	-1.4	0.0195	1
<i>ADAMTS8</i>	NM 007037	433.73	131.94	56.33	11	130,274,818	130,298,539	-1.72	0.0039	0.4	-1.23	0.0176	1
<i>AGER</i>	NM 172,197, NR 038190, NM 001136, NM 001206932, NM 001206934, NM 001206936, NM 001206929, NM 001206940, NM 001206966, NM 001206954	14,949.32	2906.79	644.24	6	32,148,920	32,152,099	-2.36	0.0002	0.1	-2.17	0.0006	1
<i>BTNL8</i>	NM 001159710, NM 001159708, NM 001159707, NM 001159709, NM 001040462, NM 024850	64.22	8.39	0	5	180,326,077	180,377,906	-2.94	0.0065	0.48	-Inf	0.0024	1
<i>BTNL9</i>	NM 152547	3882.18	974.82	224.52	5	180,467,225	180,488,523	-1.99	0.0012	0.24	-2.12	0.0044	1
<i>C19orf69</i>	NM 001130514	46.18	6.34	2.17	19	41,949,063	41,950,670	-2.86	0.0022	0.32	-1.55	0.0214	1
<i>CDH5</i>	NM 001795	890.56	194.26	83.93	16	66,400,525	66,438,689	-2.2	0.0001	0.08	-1.21	0.0167	1
<i>CLDN18</i>	NM 001002026, NM 016369	2139.31	890.63	343.63	3	137717658	137752494	-1.26	0.028	0.85	-1.37	0.0021	1
<i>FABP4</i>	NM 001442	190.16	34.84	6.08	8	82390732	82395473	-2.45	0.0003	0.11	-2.52	0.0018	1

Gene symbol	Accession #	FPKM			Location			Statistics					
		NOR	AIS	INV	Chr	Start	End	FC AIS v. NOR	p-val AIS v. NOR	p-adj AIS v. NOR	FC INV v. AIS	p-val INV v. AIS	p-adj INV v. AIS
<i>FCV</i> ₃	NM 173452, NM 003665	290.41	55.98	8.49	1	27695601	27701315	-2.38	0.0011	0.23	-2.72	0.0006	1
<i>FOXF</i> ₁	NM 001451	226.91	73.05	31.14	16	86544133	86548070	-1.64	0.0058	0.46	-1.23	0.0355	1
<i>FOXF</i> _{1-AS1}	Not available	1280.35	283.88	62	16	86508131	86542466	-2.17	3.15E-05	0.04	-2.2	6.68E-05	0.76
<i>GPM</i> _{6 A}	NM 201592, NM 201591, NM 005277, NR 048571, NM 001261448, NM 001261447	131.92	27.16	9.3	4	176554088	176923842	-2.28	0.0087	0.54	-1.55	0.035	1
<i>GRIA</i> ₁	NM 001258022, NM 001258023, NM 001258019, NM 00114183, NM 001258020, NM 000827, NR 047578, NM 001258021	540.87	218.55	76.1	5	152870084	153193429	-1.31	0.0181	0.74	-1.52	0.0305	1
<i>HBB</i>	NM 000518	587.27	119.46	30.1	11	5246696	5248301	-2.3	0.0003	0.11	-1.99	0.0028	1
<i>KCNK</i> ₃	NM002246	219.35	97.76	27.96	2	26915581	26954066	-1.17	0.0492	1	-1.81	0.0055	1
<i>LOC</i> ₇₂₈₆₄₃	Not available	388.05	130.51	29.16	10	47133295	47133836	-1.57	0.0068	0.49	-2.16	0.0097	1
<i>LPN</i> ₃	NM 015236	238.31	52.13	8.76	4	62362839	62938168	-2.19	9.51E-05	0.07	-2.57	0.0005	1
<i>NOTCH</i> ₄	NM 004557	427.04	89.06	34.5	6	32162620	32191844	-2.26	8.67E-05	0.07	-1.37	0.0488	1
<i>OVCH</i> ₂	NM 198185	76.84	20.02	3.78	11	7710669	7727941	-1.94	0.0251	0.82	-2.4	0.0063	1
<i>PCDH</i> ₁₀	NM 032961, NM 020815	67.73	18.69	5.29	4	134070470	134112732	-1.86	0.0033	0.37	-1.82	0.0258	1
<i>PROK</i> ₂	NM 001126128, NM 021935	47.3	11.67	2.31	3	71820806	71834357	-2.02	0.0226	0.79	-2.34	0.0466	1
<i>PRX</i>	NM 181882, NM 020956	124.27	27.27	8.94	19	40899671	40919271	-2.19	0.0002	0.1	-1.61	0.0148	1
<i>RAMP</i> ₃	NM 005856	166.84	54.22	12.67	7	45197367	45223849	-1.62	0.0167	0.71	-2.1	0.0084	1
<i>ROBO</i> ₄	NM 019055	166.53	59.27	14.14	11	124754114	124767831	-1.49	0.018	0.74	-2.07	0.0188	1
<i>SEMA</i> _{3 G}	NM 020163	649.97	77.42	17.1	3	52467268	52479043	-3.07	0.001	0.22	-2.18	0.0446	1
<i>SLC</i> _{5 A 1}	NM 000343, NM 001256314	131.5	14.81	2.14	22	32439019	32509016	-3.15	0.001	0.22	-2.79	0.0103	1
<i>SLC6A</i> ₄	NM 001045	1772.74	174.72	15.25	17	28523376	28562954	-3.34	4.89E-05	0.05	-3.52	0.0001	0.76

Gene symbol	Accession #	FPKM			Location			Statistics					
		NOR	AIS	INV	Chr	Start	End	FC AIS v. NOR	p-val AIS v. NOR	p-adj AIS v. NOR	FC INV v. AIS	p-val INV v. AIS	p-adj INV v. AIS
<i>STXBP₆</i>	NM 014178	177.61	38.55	7.28	14	25281293	25519095	-2.2	0.0007	0.19	-2.4	0.003	1
<i>TCEAL₂</i>	NM 080390	456.47	89.21	23.14	X	101380660	101382684	-2.36	0.0002	0.1	-1.95	0.0018	1
<i>TIE₁</i>	NM 001253357, NM 005424	289.92	73.51	44.76	1	43766566	43788781	-1.98	0.0009	0.21	-0.72	0.0443	1