Genomics 95 (2010) 355–362

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Identification of Long stress-induced non-coding transcripts that have altered expression in cancer ☆

Jessica M. Silva [a,b,*], Damon S. Perez [a], Jay R. Pritchett [a], Meredith L. Halling [a], Hui Tang [c], David I. Smith [a]

[a] Division of Experimental Pathology, Department of Laboratory Medicine and Pathology, Mayo Clinic and Foundation, 200 First Street, S.W., Rochester, MN 55905, USA
[b] Department of Biochemistry and Molecular Biology, Mayo Clinic and Foundation, 200 First Street, S.W., Rochester, MN 55905, USA
[c] Division of Biomedical statistics and informatics, Department of Health Sciences Research, Mayo Clinic and Foundation, 200 First Street, S.W., Rochester, MN 55905, USA

ARTICLE INFO

Article history:
Received 4 December 2009
Accepted 24 February 2010
Available online 7 March 2010

Keywords:
Long non-coding transcripts
Non-coding RNA
NNK(4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone)
Lung cancer
Breast cancer
Stress induction
Tiling arrays
Entire genome analysis

ABSTRACT

It has recently become clear that the transcriptional output of the human genome is far more abundant than previously anticipated, with the vast majority of transcripts not coding for protein. Utilizing whole-genome tiling arrays, we analyzed the transcription across the entire genome in both normal human bronchial epithelial cells (NHBE) and NHBE cells exposed to the tobacco carcinogen NNK. Our efforts focused on the characterization of non-coding transcripts that were greater than 300 nucleotides in length and whose expression was increased in response to NNK. We identified 12 Long Stress-Induced Non-coding Transcripts that we term LSINCTs. Northern blot analysis revealed that these transcripts were larger than predicted from the tiling array data. Quantitative real-time RT-PCR performed across a panel of normal cell lines indicates that these transcripts are more abundantly expressed in rapidly growing tissues or in tissues that are more prone to cellular stress. These transcripts that have increased expression after exposure to NNK also had increased expression in a number of lung cancer cell lines and also in many breast cancer cell lines. Collectively, our results identified a new class of long stress responsive non-coding transcripts, LSINCTs, which have increased expression in response to DNA damage induced by NNK. LSINCTs interestingly also have increased expression in a number of cancer-derived cell lines, indicating that the expression is increased in both, correlating cellular stress and cancer.

Introduction

The three gigabase pair human genome contains approximately 20,000 genes which themselves produce over one million proteins. However, less than 2% of the genome directly encodes proteins [1]. Recently new technologies, including whole-genome tiling arrays, have revealed that the majority of the genome is transcriptionally active and that the number of non-coding transcripts vastly exceeds the number of protein-coding transcripts [2–7].

The most abundantly expressed non-coding transcripts include the RNA components of the ribosome and the transfer RNAs, both of which are involved in translating coding transcripts into proteins. A second group of important non-coding transcripts are the microRNAs (miRNAs), which are much smaller (~20-25 nucleotides) and perform many critical functions. The miRNAs function to regulate gene expression through binding to multiple transcripts that contain sequences homologous to the "core" sequence within each miRNA [8]. Several miRNAs (e.g. those derived from the miR-15a-16-1 and miR17-92 clusters) have also been correlated with carcinogenesis, acting as either tumor suppressor miRNAs or oncogenic miRNAs [9]. Other small non-coding RNAs (ncRNAs) include the piwiRNAs and the snoRNAs. piwiRNAs regulate gene expression through mRNA degradation and translational repression [10]. snoRNAs are involved in post-transcriptional hypermodification of rRNA. snoU5 has also been found to be a candidate tumor suppressor in prostate cancer [11,12].

There are also considerably larger non-coding transcripts that play important functional roles within the cell. Perhaps the best known long ncRNA transcript is the human XIST/TSIX transcript, which is involved in chromatin remodeling events associated with X-chromosome inactivation and dosage compensation in eukaryotes. Copies of this 8 kb transcript are responsible for coating the inactive X chromosome. Two adjacent long ncRNAs that have been identified are tncRNA (NEAT1) and MALAT1 (NEAT2). Both have been demonstrated to play

☆ "Sequence data from this article have been deposited with the DDBJ/EMBL/GenBank Data Libraries under Accession Numbers: **GU228573, GU228574, GU228575, GU228576, GU228577, GU228578, GU228579, GU228580, GU228581, GU228582, GU228583, GU228584**, for LSINCT1 to LSINCT12 respectively."

* Corresponding author. Division of Experimental Pathology, Department of Laboratory Medicine and Pathology, Mayo Clinic and Foundation, 200 First Street, S.W., Rochester, MN 55905. Fax: +1 507 266 5193.
E-mail addresses: silva.jessica@mayo.edu (J.M. Silva), perez.damon@appliedbiosystems.com (D.S. Perez), jrpritchett@fas.harvard.edu (J.R. Pritchett), mhalling@gustavus.edu (M.L. Halling), h.tang@mayo.edu (H. Tang), smith.david@mayo.edu (D.I. Smith).

0888-7543/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2010.02.009

roles in cancer development. tncRNA (trophoblast derived non-coding RNA) is a 4.5 kb transcript associated with the suppression of the MHC class II antigen promoter in T cells thus, suppressing the display of fetal histocompatibility antigens to allow fetal evasion of the maternal immune system [13]. In addition, tncRNA has also been found to have increased expression in HELA cells. MALAT1 is an 8 kb ncRNA that is over-expressed in a variety of different cancers and is a marker for metastasis in early stage non-small cell lung cancer patients [14]. Another long ncRNA HOTAIR is produced within one cluster of HOX genes and is responsible for regulating another HOX gene cluster on a different chromosome [15]. Thus, the functions of the long ncRNAs are diverse and there is increasing evidence that many are important cellular regulators.

ncRNAs may modulate several cellular stress responses [16]. For instance, Adapt33 ncRNA is a stress-responsive transcript that is induced under conditions of a cytoprotective "adaptive response" and has been shown to be a stress-inducible riboregulator [17]. Another, small ncRNA, DsrA, stimulates translation of the RpoS stress response factor in *Escherichia coli* by base-pairing with the 5' leader of the RpoS mRNA and opening a stem-loop that represses translation initiation [18]. Recently, Ben Amor et al. conducted a genome-wide bioinformatic analysis and identified 76 long non-protein-coding RNAs (npcRNA) involved in *Arabidopsis* differentiation and stress responses [19]. In humans, the expression of the ncRNAs PRINS (Psoriasis susceptibility-related RNA Gene Induced by Stress) was induced by ultraviolet-B irradiation, viral infection (herpes simplex virus), and translational inhibition [20].

We were interested in identifying human non-coding transcripts that are both long and induced upon cellular stress. Our interest in stress-responsive transcripts is because of the known association between cellular stress and cancer. We performed whole-genome tiling array experiments using RNA prepared from Normal Human Bronchial Epithelial cells (NHBE cells) exposed to NNK (4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone), a potent tobacco carcinogen. NNK is one of the most carcinogenic tobacco-specific nitrosamines. Picomolar concentrations of NNK can be detected in body fluids and tissues of cancer patients and tobacco products [21,22]. NNK is an organ-selective lung carcinogen in laboratory animals [23] and is formed from nicotine during tobacco processing and smoking. NNK is metabolically activated by the cytochrome P450 enzymes (predominately CYP2A6 and CYP2A13) to intermediates that methylate and pyridyloxobutylate DNA, resulting in DNA adduct formation, single strand breaks, and increased levels of 8-oxodeoxyguanosine in the DNA [24]. The cellular stress NNK elicits has been shown *in vivo* and *in vitro* dysregulating various genes affecting viability, cell movement, cell cycle, and cell proliferation [25–28]. Lung cancers induced by NNK are primarily adenocarcinomas [22,29]. In a recent paper, NNK has been shown to also promote tumor growth and metastasis in mouse models of lung cancer [30]. It has also been shown to induce several other cancers including nasal, oral cavity, liver, pancreas, and cervix[31,32]. Interestingly, NNK has been utilized to induce cancer in human breast epithelial cells [32] and has been shown to significantly increase mammary tumor incidence and/or multiplicity in human HRAS proto-oncogene transgenic (Hras128) rats [33].

We hypothesize that long stress-induced non-coding transcripts are encoded in the genome and that they are not only involved in the response to cellular stress, but also have altered expression in cancer derived cells. Our tiling array data has identified a unique group of long stress-induced non-coding transcripts, which we term LSINCTs. These LSINCTs accumulate to higher levels in NNK-induced cells, and are more abundantly expressed in normal human tissues that are more proliferative. In addition, LSINCTs have increased expression in a number of lung cancer-derived cell lines and are also over-expressed in many breast cancers. These results support the concept that the expression of these non-coding transcripts is correlated with cellular stress and carcinogenesis.

Materials and Methods

Cell Lines, Cell Culture and Chemical Treatment

All cell lines were grown in 5% CO_2 at 37 °C incubator. HMEC, HCC1500, and HCC1569 cells were purchased from Lonza (Basel, Switzerland). Cryopreserved NHBE cells were purchased from Cambrex Bio Science (Walkerville, MD) and grown in defined BEBM medium (Cambrex) with supplements at manufacturer-suggested concentrations. All lung cancer cell lines were purchased from ATCC (Manassas, Virginia) and grown in ATCC RPMI-1640 medium containing 10% fetal bovine serum (FBS). T47D, BT474, HCC1500, HCC1569, and UACC893 were grown in ATCC RPMI-1640 medium containing 10% fetal bovine serum (FBS). MCF7, MCF10A, MDA157, and MDA435 were grown in Gibco DMEM with L-glutamine and 10% FBS. The MDA468 cell line was cultured in L15 medium with L-glutamine and 10% FBS.

NNK-induced cells were prepared by treating NHBE and MCF7 cells with 200 μM NNK (4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone) (Toronto Research Chemicals) prepared in dimethyl sulfoxide for 24 hours. We measured a strong induction of three DNA damage genes; GADD45, CYP2A6, and CYP2A13 as an indication of DNA damage response to determine optimal concentration of NNK [32,34].

RNA and RNA isolation

Human normal total RNA was purchased from Applied Biosystems Inc. (Austin, TX) and served as a control when analyzing LSINCTs expression across a panel of cancer specimens.

Total RNA was isolated from cells or tissues using the Qiagen RNAeasy Plus Mini Kit (Valencia, CA) and Gentra Systems Versagene Total RNA Cell Kit (Minneapolis, MN) according to the manufacturer's protocol.

Whole-genome tiling array design

The GeneChip® Human 35 bp Tiling Array 1.0R Set (Affymetrix) design is based on the human genome version 34 according to NCBI versioning system, as downloaded from the UCSC website (http:/www.genome.ucsc.edu/) and repeat masked. Probes are roughly at a 35 bp resolution (center-to-center spacing of consecutive 25-mers), subject only to requirements of synthesis and probe quality. Arrays were subdivided according to genome location into 14 separate microarrays. Each probe pair is formed by a perfectly-matched 25-mer (PM), identical to the genome sequence at the selected position, and a mismatch (MM) probe that differs from PM at the central base. This tiling array design contained 41 804 probe pairs representing 1 364 427 919 (91%) nucleotides of the repeat masked (http://www.repeatmasker.org/) human genome sequence that could be grouped into windows containing at least 5 probes.

Tiling array experimental procedure

Analysis was performed at the Microarray Core Facility at Mayo Clinic (Rochester, MN). Briefly, total RNA was isolated from NHBE cultured in T-150 flasks in triplicate using Gentra Systems Versagene Total RNA Cell Kit (Minneapolis, MN), including DNase treatment. To focus on longer transcripts, total RNA was isolated using a column with low affinity for RNAs less than 80 nt. RNA quality was assessed by ultraviolet spectrophotometry. Agilent traces were then obtained for each sample.

The GeneChip® WT Double-Stranded DNA Terminal Labeling Kit and GeneChip® WT Amplified Double-Stranded cDNA Synthesis Kit from Affymetrix (Santa Clara, CA) was used. This protocol entails first and second strand cDNA synthesis using random hexamers, RNA removal and cDNA purification, a quality control cDNA step, a cDNA fragmentation step, TdT labeling, prehybridization of the chips, and

the final hybridization of the labeled cDNA onto the GeneChip® Human 35 bp Tiling Array 1.0R Set. Labeled samples were re-hybridized to up to a total of four chips, as suggested by the protocol. Each tiling array chip was completed in triplicate (i.e., three complete 14 chip sets were hybridized and scanned for both control NHBE and NHBE exposed to NNK).

Each array was scanned using the Affymetrix GeneChip® 300 G7 scanner. The GeneChip® Operating Software (GCOS) automatically generated four files required for data analysis, including the CEL file. The raw CEL files were used in our data analysis. Using Integrated Genome Browser (IGB) from Affymetrix and the UCSC Human Genome Browser, we were able to visualize the raw and normalized hybridization signals from different data sets, together with annotation information from public and private data banks.

Northern Blot analysis

Probes were created as PCR products (300 bp) based on LSINCT-specific primers, designed using the Primer3 program (http://biotools.umassmed.edu/bioapps/primer3_www.cgi). Sequences are available in Supplemental data. PCR products were extracted from 2% low-melting point agarose gels and purified using the QIAquick Gel Extraction Kit (Qiagen). Probes were then labeled with [±P32]-labeled deoxynucleotide triphosphates using the Megaprime DNA Labeling Systems kit from Amersham Biosciences (Piscataway, NJ). Formaldehyde 1% agarose gels were loaded with 5-20 µg total RNA, which was transferred onto Nitrocellulose Membrane (Bio-Rad) overnight and then fixed to the membrane by UV-crosslinking (Stratalinker, Stratagene). Membranes were hybridized at 42° with labeled probes overnight using ULTRAhyb Ultrasensitive hybridization buffer (Ambion) in a rotating incubator. Membranes were washed in buffers and were exposed to CL-X Posure Film (Thermo Scientific) for 1 to 3 days.

Quantitative Real-Time RT-PCR

Primers for LSINCTs were designed using Primer3 (Supp Table 7b). cDNA was synthesized using the Thermoscript RT-PCR System from Invitrogen using 2 µg of total RNA and random primers. cDNA quantitation was then performed with specific primers using the SYBR green method (ABI 7900HT Fast Real-time PCR system). Primers were optimized for qPCR with β-actin as a control gene and then with the transcript region of interest. When the optimal primer concentration produced a linear response to input cDNA concentration, RNA samples were analyzed in triplicate for each tested transcript. To normalize the expression levels (ΔC_T), the threshold cycle (C_T) for each transcript was subtracted from the C_T of the more abundantly-expressed control gene (β-actin). Transcripts displaying consistently altered expression levels consistently in each of the triplicate experiments were included for statistical analysis comparing normal versus cancer samples from breast and cells induced with NNK treatment. Stressed or cancer samples that displayed a 2-fold difference in expression compared to appropriate controls underwent further analysis to determine statistical significance.

Statistical analysis

Affymetrix provided two software programs for initial data analysis, Tiling Analysis Software (TAS) and Integrated Genome Browser (IGB). Tiling Analysis Software v1.1 User Guide (Affymetrix) was consulted. Briefly, the analyses provided within the TAS application included analyzing feature intensity data stored in CEL files to produce signal and p-values for each genomic position, computation of genomic intervals based on those computed signal and p-values, and computation of summary statistics and visualizations for assessing the quality of the array data. The results of this analysis were imported into applications such as IGB or the UCSC Human Genome Browser. With IGB, annotation variations were compared in different data sets. IGB combined in one viewer its own experimental or computational results, common reference information, and access to public and private data banks.

LSINCT Selection using Model-Based Analysis of Tiling Arrays (MAT)

The peak-finding algorithm, MAT, was used to analyze tiling array data. Raw CEL files were first normalized across all replicates using quantile normalization [35], then the differential expression between NNK and NHBE cells was studied using MAT [36]. Using a stringent threshold of FDR rate equal to zero, the process identified nearly 1000 significantly differentially expressed genomic intervals. The length of these intervals ranges from 307 nt to 6680 nt (average 1017 nt).

University of California Santa Cruz Genome Bioinformatics Browser (UCSC)

UCSC *Homo sapiens* genome browser gateway March 2006 (NCBI Build 36.1) was used to analyze the LSINCT regions. BED files created with IGB were used to locate potential LSINCTs throughout the genome. Sequence coordinates from the BED files were imported into the UCSC browser where these regions underwent additional investigation in order to confirm that these sequences are indeed non-coding. This analysis included several of the UCSC browser fields. Transcript analysis examined LSINCTs that contain no open reading frames, no homology to any RefSeq Genes, and no known mRNA homology.

Results

Whole genome tiling array of RNA from NHBE and NHBE-NNK induced cells

A tiling array experiment was performed to obtain an unbiased view of transcription changes across the entire genome in NHBE cells exposed to NNK compared to unexposed NHBE cells. Initial experiments determined optimal NNK concentrations for induction of a DNA damage response (data not shown). NHBE cells were exposed to different NNK concentrations for a 24 hour period and the expression of three genes: GADD45, CYP2A6, and CYP2A13 were measured. These are genes known to be induced in response to DNA damage [22]. We found that 200 µM NNK produced strong induction of these. It should be noted that this inducing NNK concentration is extremely high relative to the picomolar concentrations of NNK that are detected in the blood of humans who smoke cigarettes however, ***in vitro*** studies in bronchial epithelial cells confirm this concentration to be optimal for metabolizing NNK [32,34].

To identify NHBE NNK-induced Transcriptionally Active Regions (NNK- TARs), we utilized the Tiling Analysis Software (TAS) and Integrated Genome Browser (IGB). We analyzed NHBE NNK-TARs as compared with transcripts found from non-treated NHBE cells based on a p-value of <0.01 using TAS. We analyzed the NNK-TARs by TAS/IGB and Model-Based Analysis of Tiling Arrays (MAT) to identify candidate NNK-TARs for further investigation. A number of the NNK-TARs were identified with both methods, but there were many NNK-TARs that were only identified with one of the two methods. In total, over 231, 305 NNK-TARs were identified with their lengths ranging from 100 nt to greater than 3000 nt. This analysis was further characterized by identifying NNK-induced TARs (119 305) or NNK-decreased TARs (112 000). In this study we focused our efforts on the identification of TARs which were induced by NNK. In order to identify the longer non-coding transcripts and to intentionally exclude short non-coding RNAs including some pre-miRNAs; we excluded transcripts smaller than 300 nts, thereby identifying 1 305 long NNK-induced TARs (Fig. 1).

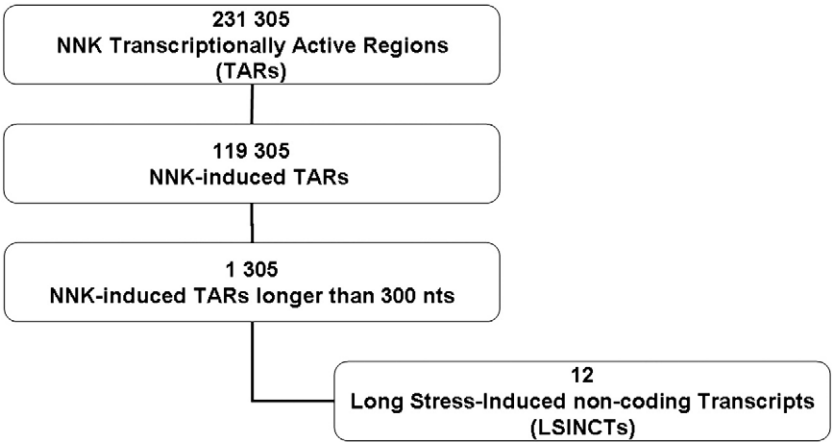

Fig. 1. Whole genome tiling array was conducted using RNA isolated from NHBE cells treated with NNK. 231,305 NNK-TARs, were identified with lengths ranging from 100 nt to greater than 3000 nt. Further analysis identified 119 305 NNK-induced TARs. 1 305 NNK-induced TARs were then identified by excluding only TARs longer than 300 nts. Utilizing criteria for non-coding RNAs identified 12 Long Stress-Induced Non-Coding Transcripts, LSINCTs.

Identification of long stress-induced non-coding transcripts (LSINCTs)

Long Stress-Induced Non-Coding Transcripts (LSINCTs) were selected from the 1 305 long NNK-induced TARs using IGB, the University of California Santa Cruz Genome Browser (UCSC) March 2006 version, as well as the Ensembl Browser (refer to Methods section). Utilizing these browsers, the long NNK-induced TARs underwent additional investigation in order to confirm that these sequences are indeed long and contained non-coding potential. This analysis included size, verifying a lack of significant open reading frames, the absence of repeats, and no homology to any RefSeq Genes or any potential mRNAs.

Utilizing the UCSC Genome browser (Supp Fig. 2 lower panel showing LSINCT5), we are able to identify a group of 12 LSINCTs that satisfied these criteria and these have been termed LSINCT 1-12. Two of the identified LSINCTs, LSINCT2 and LSINCT5, are shown in Fig. 2.

The 12 LSINCTs selected for analysis have predicted lengths ranging from 300-800 nt based on the hybridization pattern to adjacent oligonucleotide tiling array probes. One LSINCT (LSINCT10) was located in intron 3 of the SDC5 gene, while the remainder of LSINCTs are located within intergenic regions of the genome (Supp Table 1). In addition to these 12 LSINCTs, our selection criteria also recovered two previously characterized long non-coding transcripts, tncRNA and MALAT1 (Supp. Fig. 3).

LSINCTs are longer than expected from the tiling array data

We analyzed the full size of the LSINCTs by utilizing Northern Blots. This revealed that each of the LSINCTs was larger than expected from the tiling array data (Supp Table 1). This result is unexpected however not surprising as the threshold value for IGB identified only transcripts which contain a p-value of less than one. As seen in LSINCT10 (Supp

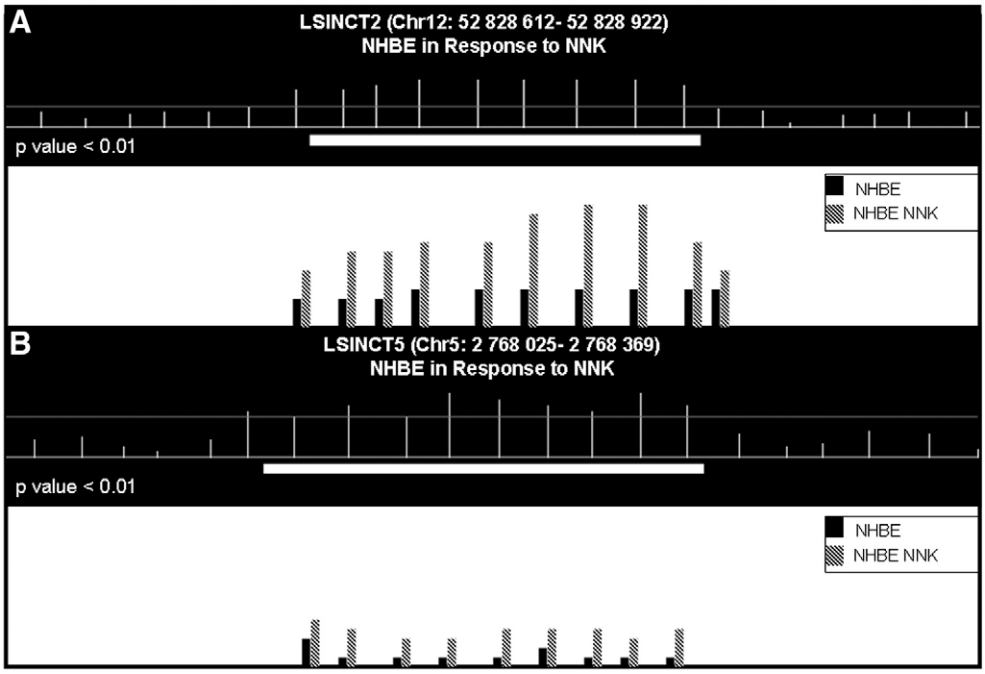

Fig. 2. Tiling array data analyzed with the Integrated Genome Browser for LSINCT2 (Panel A) and LSINCT5 (Panel B) showing location. Top panels exhibiting difference of LSINCT signal (vertical bars) with significant p value (white horizontal bar) in NHBE NNK-induced cells. Lower panels exhibit LSINCTs with a decreased signal in untreated NHBE cells (black bars) compared to an increased signal in NHBE cells exposed to NNK (striped bars).

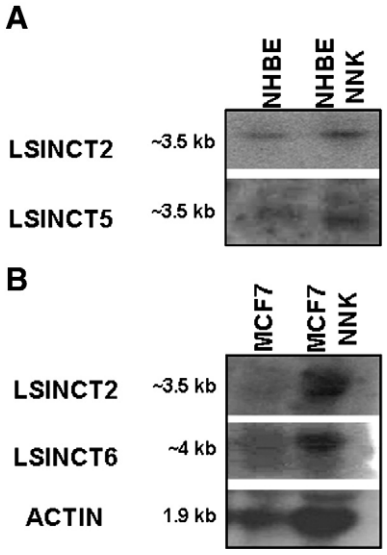

Fig. 3. Northern Blots for (A) LSINCT2 and LSINCT5 expression in NHBE and NHBE NNK-induced cells showing length ~3.5 kb and increased expression in NNK-induced cells. (B) LSINCT2 (~3.5 kb) and LSINCT6 (~4 kb) expression in MCF7 and MCF7 NNK-induced cells. LSINCTs show higher expression in MCF7 cells compared to MCF7-NNK-induced cells analyzed to actin.

Fig. 1) if we lowered this threshold this would give us a longer transcript instead of multiple smaller ones, which may be seen with some of the other LSINCTs as well. For example, LSINCT2 has a RNA transcript of ~3.5 kb as compared to an expected size of 311 nt based on tiling array data. (Fig. 3)

In addition, we also used the Northern Blots to validate that each of the LSINCTs did indeed have increased expression after exposure of cells to NNK as seen in LSINCT2 and LSINCT5 (Fig. 3) which was also validated by qPCR (Fig. 5A). This further validated the tiling array data. Similar results were seen for the other 11 LSINCTs.

LSINCT location relative to genes

The LSINCTs analyzed here are located throughout the intergenic and intragenic regions of the genome. We analyzed the upstream or downstream genes nearest to each LSINCT. As indicated in Supplemental Table 2, there are several significant genes adjacent to LSINCTs, including Epidermal Growth Factor Receptor (EGFR), HOXC4, and Insulin Growth Factor Binding Protein 4 (IGFBP4) adjacent to LSINCT1, LSINCT2, and LSINCT6, respectively. LSINCT RNA transcripts range from 3.5 kb to 4 kb in length, suggesting that several of these non-coding transcripts could be part of the transcriptional units of the closest associated genes. The close proximity of several of these LSINCTs to the closest adjacent genes suggested that these transcripts could be part of the transcriptional output from that gene. In order to rule this out we created PCR primers within each of these LSINCTs and multiple primers within different regions of the closest adjacent genes but could not amplify products with these primer pairs utilizing cDNA made from NHBE-NNK induced RNA. This suggests that these LSINCTs are not part of the major transcript arising from these adjacent genes (data not shown).

LSINCTs are induced by stress

In order to further validate that the expression of the LSINCTs is induced by stress, random primed cDNA made from NHBE cells and NHBE cells exposed to NNK was analyzed for the expression of each LSINCT using quantitative real time RT-PCR (qPCR). LSINCTs 1-12 displayed an average of 2 to 4-fold greater expression in NHBE cells after NNK treatment (Fig. 5A). We also studied the expression of LSINCTs in NNK-induced MCF7 cells (Fig. 5B). The expression of LSINCTs 2, 4, 5, 6, and 9 was increased by NNK, expression of LSINCTs 1, 8, and 12 was decreased, and there was no change in the expression of LSINCTs 10 or 11. These data indicate that there are differences in LSINCT inducibility in various tissue-derived cell lines.

LSINCT Expression in a panel of normal human tissues

In addition to examining the expression of LSINCTs in NNK-induced cells, we monitored LSINCT expression in ten normal tissues (Supp Table 3). LSINCT expression was generally higher in pancreas, colon, and spleen. We found the LSINCTs to have lower expression in other normal tissues including liver, brain, ovary, kidney, gall bladder, cervix, and prostate (Fig. 4). These results demonstrate that LSINCTs are more highly expressed in proliferative stress-responsive normal tissues whereas non-proliferative tissues such as brain showed lower LSINCT expression.

LSINCT expression in a panel of lung and breast cancer cell lines

Increased expression of LSINCTs in NNK-induced cell lines, as well as in proliferative stress-responsive normal tissue, led us to examine the correlation of LSINCTs in carcinogenesis. As NNK is one of the most carcinogenic tobacco-specific nitrosamines and has been shown to induce lung cancer [37], we first chose to examine LSINCT expression in a lung cancer cell line panel (Supp Table 4). We found that all LSINCTs contained at least a 2-fold overexpression in at least one of the lung cancer cell lines (Fig. 6A). The cell line that showed the greatest increases in expression of the largest number of LSINCTs was NCI-1437.

NNK treatment has been previously shown to induce breast carcinogenesis and has also been used to identify O6-Methylguanine DNA adducts in maternal rats mammary glands [32,38]. In order to identify if our newly discovered NNK-induced LSINCTs also have an

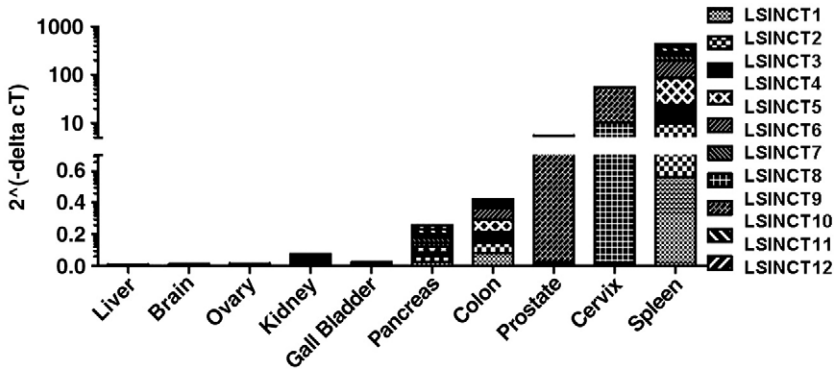

Fig. 4. LSINCT expression in a panel of normal tissues. LSINCTs contain higher expression in normal tissues that are localized in stress environments. Experiments were repeated three times in triplicate.

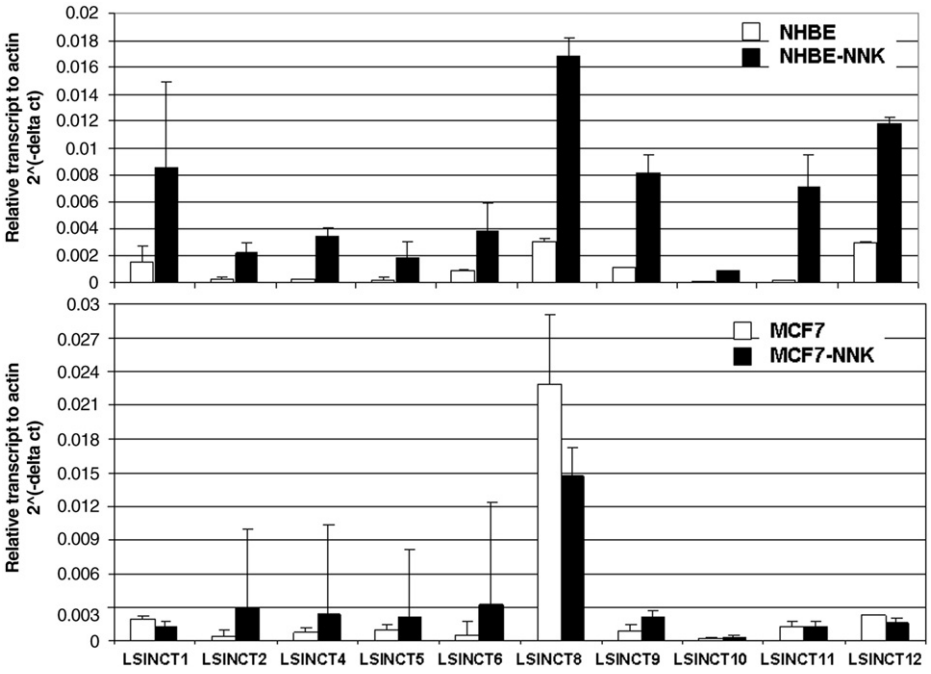

Fig. 5. (A) LSINCT1-LSINCT12 expression in NHBE cells (white bars), and NHBE cells treated with NNK (black bars). (B) LSINCT1-LSINCT12 expression in MCF7 cells (white bars) and MCF7 cells induced with NNK (black bars). Experiments were repeated three times and done in triplicates.

association with breast cancer, we then determined LSINCT expression in a breast cancer cell line panel derived from both primary and metastatic tumors. (Supp Table 5). We found that all LSINCTs displayed increased expression in at least several of the breast cancer cell lines examined. We also found that all of the LSINCTs were overexpressed in two of the breast cancer cell lines: HCC1500 and HCC1569 (Fig. 6B). In addition to qPCR analysis we also compared LSINCT expression in MCF7 with or without NNK treatment by Northern Blot (Fig. 3B). LSINCT2 and LSINCT6 were expressed higher in the breast cancer cell line MCF7 when exposed to NNK demonstrating that here is indeed higher expression of LSINCTs under stressful conditions.

Fig. 7 shows that LSINCT2 and LSINCT5 were overexpressed in most breast cancer cell lines (6 of 7) with LSINCT5 showing significant overexpression in all breast cancer cell lines. This is not seen in the lung cancer panel as LSINCT2 has overexpression in half of the cell lines whereas LSINCT5 does not have significant overexpression in any lung cancer cell lines. These data indicate altered cancer expression of the LSINCTs and show a correlation of LSINCTs with carcinogenesis.

Discussion

The discovery that the majority of the human genome is transcriptionally active has resulted in a number of studies aimed at characterizing the large number of non-coding transcripts produced across the genome. While most of the current work has focused on the identification of novel smaller ncRNAs such as the miRNAs, a number of considerably larger non-coding transcripts have been identified. A number of these have been characterized and many of them have been found to be important cellular regulators. Several of these have been found to be stress responsive and associated with cancer.

We therefore utilized a unique genome-wide search for long stress-induced non-coding transcripts. NHBE cells were exposed to high concentrations of NNK, a carcinogen produced in cigarette smoke, and transcripts were compared to untreated cells by hybridization to an Affymetrix whole-genome tiling array. Utilizing a stringent selection process we have now identified 12 Long Stress-Induced Non-Coding Transcripts. In addition, our screen identified two previously known adjacent long non-coding transcripts, tncRNA and MALAT1. These long ncRNAs are recognized as important regulators in various cancers. Our analysis also identifies these two transcripts as stress-responsive non-coding transcripts.

Currently, there is a growing number of long regulatory ncRNAs associated with stress and cancer. These ncRNAs include well-known XIST and HOTAIR (Hox transcript antisense RNA) [15,39]. There is evidence that XIST, HOTAIR, and others interact with chromatin modulators to regulate gene expression patterns through epigenetic effects. It is unknown whether the LSINCTs we identified are included in this class of long regulatory ncRNAs. Other regulatory ncRNAs associated with stress and cancer include the 2 kb transcribed antisense to beta-secretase-1 (BACE1-AS) gene. BACE1-AS has been found to regulate BACE1 in Alzheimer disease and has also been shown to regulate the BACE1 protein when induced by various cellular stressors.[40,41] Another important long ncRNA associated with prostate cancer is the 1.6 kb long prostate tissue-specific and prostate cancer-associated ncRNA PCGEM1 (Prostate Cancer Gene Expression Marker 1). Fu et. al has recently identified overexpression of PCGEM1 in a LNCAP model that led to inhibition of apoptosis by the induction of doxorubicin[42]. This list does not entirely cover all long ncRNAs [43], however our data raises the possibility that LSINCT expression also plays a role in cellular stress and carcinogenesis.

We used both qPCR and Northern Blot analysis to show that the 12 LSINCTs were indeed induced by NNK exposure. Northern blotting also showed the 12 LSINCTs to be considerably larger in size than expected from the tiling array results with each transcript being between 2 and 4 kb in length. This could be caused by two reasons. The first is that there was an arbitrary cut-off for significant signals and if several oligos had lower hybridization than that cut off it would arbitrary cut a larger transcript into multiple smaller ones. In addition, since there is repeat masking in the tiling array design there are instances where that would result in the false impression that one larger transcript was actually several smaller transcripts. In addition to examining LSINCT expression in normal lung cells following exposure to NNK, we also

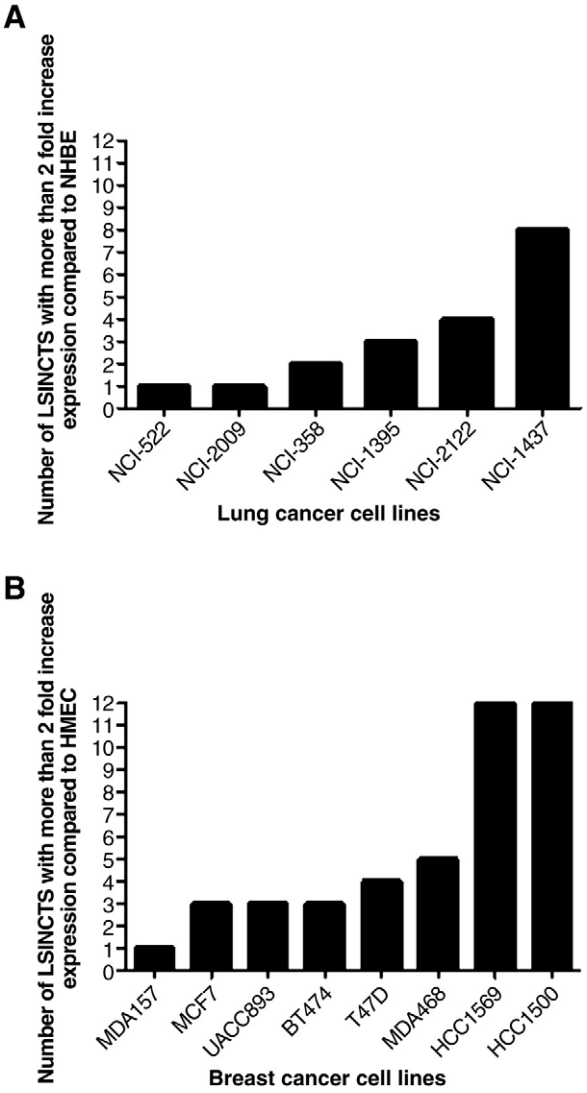

Fig. 6. Panel of lung cancer cell lines (A) and breast cancer cell lines (B) showing the number of LSINCTs with more than 2 fold increase expression as compared to NHBE or HMEC cells. Experiments were repeated two times and done in triplicate.

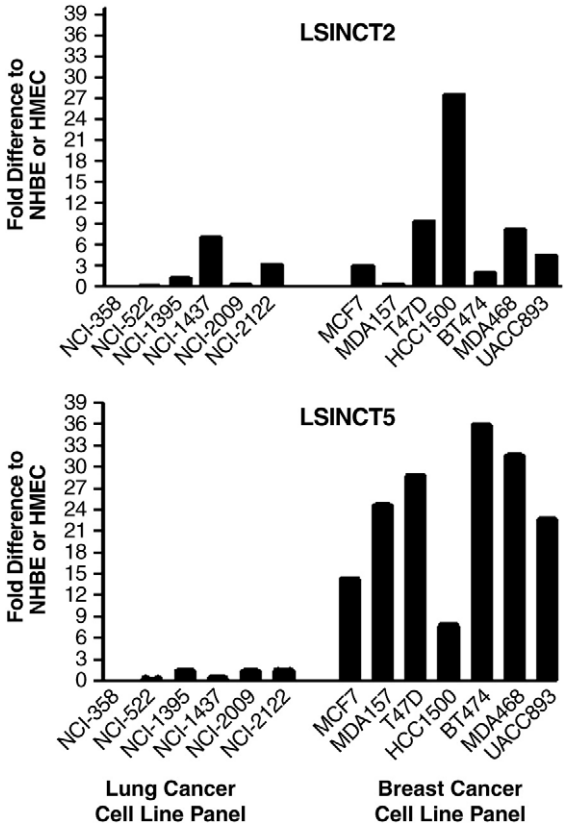

Fig. 7. LSINCT2 (panel A) and LSINCT5 (panel B) fold expression difference in a panel of lung cancer cell lines and a panel of breast cancer cell lines.

demonstrated increased expression of several LSINCTs in breast cancer cell lines following a similar exposure to NNK, as NNK has been shown to be associated with breast carcinogenesis.

We measured LSINCT expression in a panel of normal human tissues. These data revealed that, although LSINCTs expression could be detected in each of the tissues tested, expression was higher in proliferative or stress-responsive tissues (spleen and colon). In contrast there was very low expression in non-proliferative tissues such as brain. We also measured the expression of the LSINCTs in a panel of lung cancer cell lines as compared to the normal cell line NHBE. This analysis revealed that the LSINCTs are not just induced by cellular stress, but that they were frequently over-expressed in lung and other cancers. The lung cancer cell lines we utilized were non-small cell lung cancer adenocarcinomas with the exception of NCI-358, which is a bronchioalveolar carcinoma cell line. Although NNK has been shown to primarily induce lung adenocarcinomas, we are the first to show the expression of our NNK-induced transcripts in a bronchioalveolar cell line, indicating that NNK may also promote these types of cancers.

NNK's ability to transform immortalized breast cells and promote tumorgenesis in breast tumor models led us to examine LSINCT expression in breast cancer cell lines. This revealed that each of the LSINCTs was overexpressed in several of the breast cancer cell lines as compared to normal breast epithelial cell line (HMEC) and immortalized MCF10 cells. Further analysis indicates some LSINCT overexpression in the breast cell lines that are HER2- and TP53+ (Supplemental Table 6).

Further analysis of the LSINCTs which were over-expressed in the breast cancer panel has allowed us to fully characterize several LSINCTs in detail. For instance, we have validated the larger transcript of LSINCT5 by RNA Amplification of cDNA Ends (RACE) to corroborate the Northern Blots, showing it is indeed a 2.5 kb transcript. In this process, we have also found LSINCT5 to be polyadenylated and transcribed in trans. It also has increased expression in several other cancers including cancers of the cervix and ovary.

Our preliminary studies have therefore identified a group of long stress-responsive non-coding transcripts. It should be realized that there are probably many more stress responsive transcripts throughout the genome as we have utilized a very stringent screen which may have ruled out a number of viable non-coding transcripts and examined only those induced by NNK. We also identified a similar number of non-coding transcripts that had decreased expression after exposure to NNK. We are currently characterizing several of the LSINCTs in more detail to determine the precise role that they play in cellular stress and in cancer development.

Authors' contributions

JS carried out the analysis of identifying non-coding transcripts and analysis of all data, conducted the northern blots and qPCR experiments, and drafted manuscript. DP participated in the design and analysis of the tiling array and helped draft manuscript. JP conducted qPCR experiments. MH conducted qPCR experiments. HT participated in design and analysis of tiling array. DS conceived of the study, participated in design of tiling array, supervised all individuals, and helped draft manuscript. All authors read and approved the final manuscript.

Conflict of interest statement
The authors declare no conflict of interest.

Acknowledgments

We acknowledge the Mayo Clinic Microarray Core Facility for assistance with tiling array experiments. We would like to thank Louis Maher III, Ph.D. for revisions and comments on this manuscript. This work was funded by the Mayo Foundation and the Department of Defense Breast Cancer Program for funding this as a Concept Award (W81XWH-07-1-0545).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2010.02.009.

References

[1] S.A. Barciszewski, Beyond the proteome: non-coding regulatory RNAs, Gen Biol. 3 (2002).
[2] T.E.P. Consortium, Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project, Nature 447 (7146) (2007) 799–816.
[3] S.R. Eddy, Non-coding RNA genes and the modern RNA world, Nat. Rev. Genet. 2 (12) (2001) 919–929.
[4] Fabrý cio F. Costa*, Non-coding RNAs: New players in eukaryotic biology, Gene 357 (2) (2005) 83–94. Review.
[5] M.B. Gerstein, *What is a gene, post-ENCODE?* History and updated definition, Genome Res. 17 (2007) 669–681.
[6] J.S. Mattick, I.V. Makunin, Non-coding RNA, Hum. Mol. Genet. 15 (Spec No 1) (2006) R17–R29.
[7] V. Moulton*, Tracking down noncoding RNAs, PNAS 102 (7) (February 15 2005) 2269–2270.
[8] W. Filipowicz, Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nature 9 (2008) 102–114.
[9] E. Barbarotto, MicroRNAs and cancer: Profile, Profile, profile, Int. J. Cancer 122 (2007) 969–977.
[10] D. XF, A novel class of testis-specific small RNAs: piRNAs, Zhonghua Nan Ke Xue (2008) 163–165.
[11] P. Das, Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the Caenorhabditis elegans germline, Mol. Cell 1 (2008) 79–90.
[12] X. Dong, SnoRNA U50 is a candidate tumor-suppressor gene at 6q14.3 with a mutation associated with clinically significant prostate cancer, Hum. Mol. Genet. 7 (2008) 1031–1042.
[13] A. Geirsson, Inhibition of alloresponse by a human trophoblast non-coding RNA suppressing class II transactivator promoter III and major histocompatibility class II expression in murine B-lymphocytes, J. Heart Lung Transplant. 9 (2004) 1077–1081.
[14] P. Ji, et al., MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer, Oncogene 22 (39) (2003) 8031–8041.
[15] J.L. Rinn, Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs, Cell (2007) 129.
[16] F.F. Costa, Non-coding RNAs: new players in eukaryotic biology, Gene 357 (2) (2005) 83–94.
[17] Y. Wang, Characterization of adapt33, a stress-inducible riboregulator, Gene Expr. 11 (2003) 85–94.
[18] T. Soper, the rpoS mrNA leader recruits Hfq to facilitate annealing with DsrA sRNA, RNA 14 (2008) 1907–1917.
[19] M. Crespi, Novel long non-protein coding RNAs involved in Arabidopsis differentiation and stress responses, Genome Res. (2008).
[20] E. Sonkoly, *Identification and characterization of a novel, psoriasis susceptibility-related noncoding RNA gene,* PRINS, J. Biol. Chem. 25 (2005) 21459–21467.
[21] S. Hecht, Quantitation of 4-oxo-4-(3-pyridyl) butanoic acid and enantiomers of 4-hydroxy-4-(3pyridyl) butanoic acid in human urine: a substantial pathway of nicotine metabolism, Chem. Res. Toxicol. 59 (1999) 172–179.
[22] H.D. Hecht SS, 4-(Methylnitrosamino)-1-(3-pyridyl)-1-butanone, a nicotine-derived tobacco-specific nitrosamine, and cancer of the lung and pancreas in Humans, Origins of Human Cancer: A Comprehensive Review, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1991, pp. 745–755.
[23] H.D. Hecht SS, Tobacco-specific nitrosamines, an important group of carcinogens in tobacco and tobacco smoke, Carcinogenesis 9 (1988) 875–884.
[24] S.S. Hecht, DNA adduct formation from tobacco-specific N-nitrosamines, Fundam. Mol. Mutat. Res. 424 (1–2) (1999) 127–142. Review.
[25] D.K. Lonardo F., S.J. Freemantle, Y. Ma, N. Memoli, D. Sekula, E.A. Knauth, J.S. Beebe, E. Dmitrovsky, Evidence for the epidermal growth factor receptor as a target for lung cancer prevention, Clin. Cancer Res. 8 (1) (2002) 54–60.
[26] H.M. Chuang C.H., Synergistic DNA damage and lipid peroxidation in cultured human white blood cells exposed to 4-(methyl-nitrosamino)-1-(3-pyridyl)-1-butanone and ultraviolet A, Environ. Mol. Mutagen. 47 (2) (2006) 73–81.
[27] C.A. Jorquera R, H.M. Schuller, DNA single-strand breaks and toxicity induced by 4-(methyl-nitrosamino)-1-(3- pyridyl)-1-butanone or N-nitrosodimethylamine in hamster and rat liver, Carcinogenesis 15 (2) (1994) 389–394.
[28] T.I. Abdel-Aziz HO, Y. Tabuchi, K. Nomoto, Y. Murai, K. Tsuneyama, Y. Takano, High-density oligonucleotide microarrays and functional network analysis reveal extended lung carcinogenesis pathway maps and multiple interacting genes in NNK [4-(methylnitrosamino)-1-(3-pyridyle)-1-butanone] induced CD1 mouse lung tumor, J. Cancer Res. Clinic Oncol. 133 (2) (2007) 107–115.
[29] A.S. Carmella SG, S.S. Hecht, Metabolites of the tobacco specific nitrosamine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone in smokers' urine, Cancer Res. 53 (1993) 721–724.
[30] R.W. Davis R., S. Banerjee, M. Kovacs, E. Haura, D. Coppola, S. Chellappan, Nicotine promotes tumor growth and metastasis in mouse models of lung cancer, PLoS ONE 10 (2009).
[31] N. Gerald, A. Wogan, S. Stephen, B. Hecht, S. James, C. Felton, L.A.L. Allan H., D. Conney, Environmental and chemical carcinogenesis, Semin. Cancer Biol. 14 (2004) 473–486.
[32] N. Siriwardhana, Precancerous model of human breast epithelial cells induced by NNK for prevention, Breast Cancer Treat. 25 (2007).
[33] F.K. Ohnishit, F.K. Ohnishit, Y. Ohshima, X. Jiegou, S. Ueda, M. Iigo, N. Takasuka, A. Naito, K. Fujita, Y. Matsuoka, K. Izumi, H. Tsuda, Possible application of human c-Ha-ras proto-oncogene transgenic rats in a medium-term bioassay model for carcinogens, Toxicol. Pathol. 35 (3) (2007) 436–443.
[34] L.I. Proulx, 4-(Methylnitrosamino)-1-(3-pyridyl)-1-butanone, a component of tobacco smoke, modulates mediator release from human bronchial and alveolar epithelial cells, Clin. Exp. Immunol. 10 (1111) (2005) 46–53.
[35] B.M. Bolstad, R.A. Irizarry, M. Astrand, T.P. Speed, A Comparison of Normalization Methods for High Density Oligonucleotide Array Data Based on Bias and Variance, Bioinformatics 19 (2) (2003) 185–193.
[36] W.E.*. Johnson, L.W., C.A.*. Meyer, R. Gottardo, J.S. Carroll, M. Brown, X.S. Liu, Model-based analysis of tiling-arrays for ChIP-chi, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 12457–12462.
[37] R.A. Hoffmannd, S.S. Hecht, The biological significance of tobacco-specific N-nitrosamines: smoking and adenocarcinoma of the lung, Crit. Rev. Toxicol. 26 (1996) 199–211.
[38] Saranjit K. Chhabra, L.M.A., *,1, Christine Perella*†, Dimant Desai ‡, Shantu Amin ‡, a.V.L.S. Soterios A. Kyrtopoulos, Coexposure to Ethanol with N-Nitrosodimethylamine of 4-(Methylnitrosamino)-1-(3-pyridyl)-1-butanone during Lactation of Rats: Marked Increase in O6-Methylguanine–DNA Adducts in Maternal Mammary Gland and in Suckling Lung and Kidney, Toxicol. Appl. Pharmacol. 169 (2000) 191–200.
[39] C.-L. Tsai, Higher order chromatin structure at the X-inactivation center via looping DNA, Dev. Biol. 319 (2) (2008) 416–425.
[40] G. St. Laurent III, F.M., C. Wahlestedt, Non-coding RNA transcripts: sensors of neuronal stress, modulators of synaptic plasticity, and agents of change in the onset of Alzheimer's disease, Neurosci. Lett. 466 (2) (2009) 81–88.
[41] Mohammad Ali Faghihi1, Farzaneh Modarresi1, Ahmad M. Khalil1, Douglas E. Wood3, Barbara G. Sahagan3, C.E.F. Todd E. Morgan4, Georges St. Laurent III5,6, Paul J. Kenny7, Claes Wahlestedt1, Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of b-secretase, Nat. Med. (2008).
[42] L.R. Xiaoqin Fu, Nicholas Tran, a.S.S. Gyorgy Petrovics, Regulation of Apoptosis by a Prostate-Specific and Prostate Cancer-Associated Noncoding Gene, PCGEM1, DNA Cell Biol. 25 (3) (2006) 135–141.
[43] J. Mattick, The Genetic Signatures of Noncoding RNAs, PLoS Genet. 5 (4) (2009).