Identification of differentially expressed long non-coding RNAs in bladder cancer

Stefan Peter¹, Edyta Borkowska¹, Ross M Drayton¹, Callum P. Rakhit¹, Aidan Noon¹,³, Wei Chen² and James WF Catto¹

¹. Academic Urology Unit and Unit of Molecular Oncology, University of Sheffield, UK;
². Max Delbrück Center for Molecular Medicine, Robert-Rössle-Str. 10, D-13092 Berlin, Germany;
³. Department of Urology, University of Toronto, Toronto, Canada

§ Address for correspondence;

James Catto,
Department of Oncology,
G Floor, The Medical School,
University of Sheffield,
Beech Hill Road,
SHEFFIELD,
S10 2RX, United Kingdom

Tel: +44 (0)114 226 1229
Abstract

Purpose

Loss of epigenetic gene regulation through altered long non-coding RNA (lncRNA) expression appears important in human cancer. LncRNAs have diagnostic and therapeutic potential, and offer insights into the biology disease, but little is known of their expression in urothelial cancer (UC). Here we identify differentially expressed lncRNAs with potential regulatory functions in UC.

Experimental design

The expression of 17,112 lncRNAs and 22,074 mRNAs was determined using microarrays in 83 normal and malignant urothelial (discovery) samples and selected RNAs with qPCR in 138 samples for validation. Significantly differentially expressed RNAs were identified and stratified according to tumour phenotype. siRNA knock-down, functional assays and whole genome transcriptomic profiling were used to identify potential roles of selected lncRNAs.

Results

We observed upregulation of many lncRNAs in UC that was distinct to corresponding, more balanced changes for mRNAs. In general, lncRNA expression reflected disease
phenotype. We identified 32 lncRNAs with potential roles in disease progression. Focusing upon a promising candidate, we implicate upregulation of AB074278 in apoptosis avoidance and the maintenance of a pro-proliferative state in cancer through a potential interaction with EMP1, a tumour suppressor and a negative regulator of cell proliferation.

**Conclusions**

We report differentially expression profiles for numerous lncRNA in UC. We identify phenotype-specific expression and a potential mechanistic target to explain this observation. Further studies are required to validate lncRNAs as prognostic biomarkers in this disease.
Statement of Translational Relevance

Bladder cancer is a common disease whose biology is poorly understood. Here we show that many long non-coding RNAs have altered expression in the disease and may play key roles in proliferation and disease progression. We identify one long non-coding RNA that appears a strong candidate as a prognostic biomarker and show that it affects mRNA expression, cell proliferation and death, and may have potential as a therapeutic target.
Introduction

Bladder cancer is the fourth commonest male malignancy and one of the most expensive human cancers to manage (1)(2). The majority of tumours are Urothelial Carcinoma (UC) in histological type. Clinico-pathological data suggest UC are best stratified into two distinct phenotypes characterized by low and high-grade cellular differentiation. Low-grade UC frequently have mutations of FGFR3 and STAG2, partial deletion of chromosome 9 (3), and rarely progress to muscle invasion or metastases following endoscopic resection (4). In contrast, high-grade BCs are aggressive tumours that may present before or after the onset of muscle invasion (5). These poorly differentiated UC have widespread chromosomal instability, multiple mutations and are best characterized by deficiency of p53-mediated pathways (3). High-grade tumours share many molecular alterations regardless of stage and respond poorly to chemotherapy (6).

Whilst many reports detail genetic events in UC, alterations of epigenetic gene regulation are also important in this disease (7). In general, epigenetic alterations reflect UC disease phenotypes and associated genetic events (8) (9). For example, Wolff et al. reported high-grade invasive UCs had widespread aberrant hypermethylation, whilst low-grade non-invasive tumours had regional hypomethylation (10). These
reflect chromosomal changes within the disease and can be used to predict progression (11). With regards to microRNA, low-grade UC are characterized by loss of expression of many species, including miRs-99a/100 that target \textit{FGFR3} (12, 13). In contrast, high-grade tumours have upregulation of many miRs, including miR-21 that targets the p53 pathway (14) and miR-129 (15).

Epigenetic gene regulation may occur directly or indirectly through non-coding RNA (ncRNA) species. These are classified according to size, function and gene location (16). Most reports of ncRNA in cancer have focused upon microRNA's, which modulate gene expression through targeting complementary seed sequences within mRNA 3' UTRs (17). The role of microRNAs in disease varies with cell type, biological stress, species conservation and processing (9, 18). To date, few data have reported the role of long non-coding RNA (lncRNA) in cancer and little is known of their function (16, 19). Exceptions include lncRNAs clearly implicated in carcinogenesis (such as MALAT-1 or HOTAIR, reviewed in Spizzo et al. (20)). LncRNAs are annotated according to gene location (intergenic, exonic, intronic, anti-sense etc.) and size (21). The GENCODE consortium recently annotated 9,277 long ncRNA genes, corresponding to 14,880 transcripts (22). In contrast to protein coding genes, the majority of reported lncRNAs had few gene exons, were located in the nucleus and chromatin, and many were not
conserved between species. As multiple report suggest lncRNAs are important in human cancer (19) (16, 23) and preliminary data suggest this is true for UC (24), we hypothesized that they play a role in urothelial carcinogenesis. To test this hypothesis we profiled the expression of 9,351 lncRNAs in normal and malignant urothelial samples. We identified global changes that reflect the disease phenotype and differ to patterns of mRNA alteration. We performed explorative analysis of selected species and identified those with potential prognostic roles.

Materials and Methods

Patient samples and cell lines.

We analysed two patient cohorts in this work (Table 1). For RNA microarray profiling we examined urothelial samples from 57 patients with UC and 26 disease free controls (urothelium from radical prostatectomy cases). For validation, we examined a second cohort of n=138 samples from patients with UC. Tumors were obtained following trans-urethral resection of the first cancer within a patient, naïve to BCG, chemotherapy or other pre-treatments, and classified using the 2004 WHO/ISUP criteria. Tumors were selected to reflect the UC spectrum: low-grade non-muscle invasive (NMI), high-grade NMI and invasive UC. All tissues were fresh frozen in liquid nitrogen. Histologic
confirmation was obtained before use. We also analysed the human telomerase reverse transcriptase (hTERT) immortalised normal human urothelial (NHU) cells maintained in keratinocyte serum-free medium containing bovine pituitary extract, epidermal growth factor (Invitrogen) and cholera toxin. The hTERT NHU cells were obtained as a gift directly from Prof. M.A. Knowles, University of Leeds, and tested by this lab using CGH and mutational analysis of CDKN2A and TP53 as described (25). The cells were not authenticated in Sheffield prior to use.

**Long non-coding RNA expression profiling.**

From each sample, 10 x 10µm thick sections were microdissected to extract normal and malignant urothelial cells (>90% content). Total RNA was extracted using the mirVana™ kit (Ambion, TX) according to manufacturer's protocol (26), and measured using a 2100 Bioanalyzer (Agilent, Cheshire, UK). The expression of long ncRNAs and protein coding mRNAs was determined using microarrays (NCode™ Human Non-coding RNA Microarrays, Invitrogen, Paisley, UK). Each sample was prepared according to Agilent's one-color microarray protocol. Briefly, 200ng RNA samples were mixed with spike in control RNA (Agilent), labelled with cyanine 3-CTP (Low Input Quick Amp Labeling Kit, one-color, Agilent) and hybridised to the microarray (Gene Expression Hybridization Kit, Agilent) according to the manufacturer's protocol. After washing
(Gene Expression Wash Buffer 1 and 2, Agilent), the microarray slides were scanned using the Agilent Microarray Scanner platform (High Resolution Microarray Scanner C) and raw probe fluorescence extracted. The microarray data is deposited with Gene Expression Omnibus (GSE55433). The NCode™ microarray contains duplicate probes to 17,112 ncRNAs and 22,074 mRNAs. The IncRNAs were identified by various strategies without annotation (27). To annotate the array we converted all probes to hg 19.0 loci (using LiftOver, UCSC genome browser), matched to target gene, structure and the Gencode v7.0 annotation (22), and the nearest CpG island/differentially methylated region (DMR) (28). For microarray validation, individual IncRNA expression was measured in triplicate using real-time qPCR. Total RNA was transcribed using random hexamer primers (Applied Biosystems, Cheshire, UK) and the High Capacity Reverse Transcription Kit (Applied Biosystems) before diluting 10-fold in nuclease-free water. cDNA was mixed with SyberGreen MasterMix (Applied Biosystems) and PCR primers, and analysed using the ABI 7900HT real-time PCR system. cDNA expression was calculated using ΔCt values normalized by subtraction of the mean of TEGT and HSP90AB1 expression (as reference genes) and fold change (FC = 2^{-ΔΔCt}) calculated (29).

Protein coding potential score of ncRNAs
We investigated the protein coding potential of selected lncRNAs using geneID (30) and CPC against UniProt Reference Clusters (31). For both, negative scores suggest low protein coding potential.

**siRNA knock-down of lncRNA expression**

We selected representative lncRNAs for further analysis and modulated expression using custom siRNA. All experiments were performed in triplicate using hTERT immortalised NHU cells at 70% confluence transfected with siRNAs (synthesized with LifeTechnologies BLOCK-iT™ RNAi designer) specific to lncRNAs and controls (scrambled RNA sequence). Cells were seeded into twelve-well dishes and incubated for 3 hrs before transfection with 80 nmol/L RNAi using 2 µl Lipofectamine RNAiMAX (Invitrogen) per well. Transfection efficiency was determined 72hrs later by real-time qPCR (Applied Biosystems).

**Growth analysis of siRNA modulated cells**

The growth characteristics of siRNA transfected cells were analysed for cell cycle regulation (propidium iodide flow cytometry), apoptosis (Caspase-3 activation) and proliferation (MTT assay). For cell cycle analysis, cells were harvested after 24 hrs, 48 hrs and 72 hrs of knock-down, centrifuged (3 min at 1,400 g) and washed in Phosphate...
buffered saline (PBS). Cells \((1 \times 10^6)\) were re-suspended in 1 ml PBS, mixed with 3 ml ice cold absolute ethanol, fixed over night at 4°C before washing twice in PBS. Cells were then mixed with 5µl RNase solution (2 mg/ml) and 300 µl propidium iodide solution (50 µg/ml), and incubated over night at 4°C. The next day, cells were analysed on a FACSCalibur flow cytometry analyzer (Becton Dickinson, Oxford, UK). For apoptosis we determined Caspase-3 activity in 3 x 10^5 cells, 24 hrs after transfection using the CaspGLOW™ Fluorescein Active Caspase-3 Staining Kit (MBL, Woburn, MA). Briefly, we resuspended the cells in 300 µl PBS, added 1 µl FITC-DEVD-FMK and incubated them for 1 h at 37°C. After pelleting the cells and washing them in washing buffer, they were re-suspended in 300 µl wash buffer and analysed on a FACSCalibur flow cytometry analyzer (Becton Dickinson). For proliferation, cells were seeded into 96 well plates following siRNA mediated gene knock-down. The culture medium was replaced with 50 µl MTT-PBS solution (3 mg MTT / 1 ml PBS) 24 hrs, 48 hrs and 72 hrs after knock-down, respectively. Cells were incubated at 37°C for 3 h before 200 µl MTT dissolvent (DMSO) was added to each well. MTT absorbance was measured at 570 nm using a microplate reader.

**Genetic consequences of IncRNA modulation**
mRNA expression was determined in transfected cells using HG-U133 Plus 2.0 microarrays (Affymetrix, Cal.) (32). Briefly, RNA was prepared with the Affymetrix protocol and annealed to an oligo-d(T) primer with a T7 polymerase binding site. cDNA was generated using superscript II and E. coli DNA ligase and polymerase I, before the reaction was completed with T4 DNA polymerase and EDTA. Amplified cDNA was cleaned, biotin-labeled and fragmented, before hybridizing to the microarray for 16 hours (45°C in a rotating oven at 60rpm). After washing and staining, the arrays were scanned (GC3000 scanner) and data processed using Gene Chip Operating System software. mRNA expression was determined using Microarray Analysis Suite 5 (Affymetrix) and defined as expressed (perfect match probeset intensity greater than mismatch intensity) or absent (mismatch probeset intensity greater or equal to perfect match intensity). ANOVA analysis was then performed using Partek Genomic Suite 6.5 beta and differentially expressed transcripts were defined as ≥2 relative fold change using a cut-off of p < 0.05. Real-time qPCR was used to confirm the expression of individual mRNAs of interest in the second patient cohort.

Statistical methods

Raw intensity values for each microarray probe were used to calculate lncRNA expression. The NCode™ microarray includes multiple probes for each target RNA. We
excluded RNAs whose targeting probes were non-concordant (defined as signal ratio less than 0.5 or greater than 2.0) in all the arrays. Normalisation was achieved according to the 1-color default normalisation of Agilent’s Gene Spring software (version 7), by dividing each raw intensity value by the median of the chip and the median expression of that RNA in all samples. Changes in lncRNA expression and statistical significance were calculated and illustrated using Volcano plots. Significant differences in expression between malignant versus normal urothelium, or between UC phenotypes was defined using the Significance Analysis of Microarray software (33) as a T Test p-value of <0.05, a false discovery rate (FDR) of <0.01 and an expression FC of +/- 2. Hierarchical clustering was performed using Cluster 3.0 and visualized in Tree view (Eisen Lab 9). Correlation between variables was assessed using Pearson's correlation coefficient within SPSS (version 14.0, SPSS, Inc.). Area proportional Venn diagrams were produced using BioVenn (34). lncRNA expression and tumour outcome were investigated using the log-rank test and plotted by the Kaplan-Meier method within SPSS. Tumour progression was defined as the presence of pathologic, radiological, or clinical evidence of an increase in tumour stage and measured from the time of surgery to the time of proven event.
Results

IncRNA expression in bladder cancer

We investigated the expression of 17,112 IncRNAs and 22,074 mRNAs. The data are accessible through GEO Series accession number GSE55433. We filtered to 9,351 IncRNAs and 7,922 mRNAs, for which the microarray probe signals were concordant (Figure 1a, Supplementary Table 1). Comparison between UC and normal urothelium revealed 2,075 differentially expressed IncRNAs. In general, there was an increased expression of long ncRNAs in UC (1,788 (86%) were upregulated (FC > 2) and 287 (14%) were downregulated (FC < 0.5)) when compared to normal urothelium. Fewer protein coding mRNAs (n=1,410) were differentially expressed between the malignant and normal tissues, and their distribution was more balanced (836 (59%) upregulated and 574 (41%) downregulated in UC) than for IncRNAs (T Test p<0.001). We filtered the microarray transcripts to those in the Gencode v7 catalogue (ref (22); n=3,885, Supplementary Table 2), of which 355 were IncRNAs differentially expressed in UC. Whilst the majority were long intergenic ncRNA (lincRNA, n=225), we identified 130 IncRNAs located within protein coding genes (termed genic). These included RNAs coded in the sense (intronic (n=20) and overlapping (n=3)) and antisense (exonic (n=50), intronic (n=51) and overlapping (n=6)) direction with respect to the mRNA gene (Supplementary Figure 1). The proportion of aberrantly expressed IncRNAs in UC,
when compared to normal urothelium, did not vary with gene location (range from 15% for antisense exonic to 22% for overlapping antisense lncRNAs, Figure 1b).

Expression of lncRNAs with respect to tumour phenotype

We compared RNA expression across UC phenotypes (low grade NMI, high grade NMI and invasive) with normal urothelium. Unsupervised hierarchical clustering revealed three different expression profiles fitting the disease-free, the low-grade, and the high-grade/invasive phenotypes (Figure 1c). For all tumour phenotypes, we observed more up than downregulated lncRNAs, although the extent varied between groups (91% in invasive tumours, 61% in low-grade and 53% in high-grade, Figure 2a and Supplementary Figure 2a). Accordingly, the magnitude of differential lncRNAs expression varied between the invasive cohort (average fold change 2.57), and the low-grade and high-grade cohort (average fold change 1.75 and 1.73 respectively, Supplementary Figure 2b). There was less difference in the magnitude of change in expression for mRNAs than lncRNAs (average fold change of 2.11 in invasive, 1.59 in low-grade and 1.45 in high-grade, T test p<0.01). The majority of aberrantly expressed lncRNAs and mRNAs belonged to the invasive subset [1,800/2,034 lncRNAs (88%) and 1050/1410 mRNAs (74%)], followed by the low-grade cohort [n=560 lncRNAs (28%) and n=757 mRNAs (54%)] and the high-grade cohort [n=416 lncRNAs (20%) and
n=661 mRNAs (47%) [(Figure 2b). We defined a specific signature for each tumour phenotype by identifying RNAs significantly differentially expressed between each tumor type. Of these, 75% [1,356/1,800] of lncRNAs in the invasive phenotype, compared to 19% [109/560] of the low-grade cohort and only 3% [15/416] of the high-grade cohort were phenotype specific. Similarly for mRNAs, 45% in the invasive cohort, 17% and 4% of the low-grade and high-grade cohort, respectively, were phenotype specific. We identified n=188 (9%) lncRNAs and n=374 (26%) mRNAs that were altered in UC regardless of tumour phenotype.

**LncRNA expression and tumour progression**

To identify lncRNAs with roles in urothelial carcinogenesis we searched for those related to disease progression. We performed univariate log rank analysis and identified 32/2,075 (1.5%) lncRNAs associated with tumour progression (Bonferonni adjusted log rank p-value<0.05, Table 2, Figure 3a). More of these tumour progression related lncRNAs were phenotype specific (n=16) than shared between all UC (n=6). Eight were annotated within GENCODE. LncRNAs associated with tumour progression were balanced between up (n=17) and downregulation (n=15). Protein coding potential was calculated using two algorithms (Table 2) (24, 25). The resultant scores were
closely correlated (Pearson’s r=0.81) and suggested only one member was likely to be a misclassified mRNA (CR611332).

A role for AB074278 expression in cell proliferation and apoptosis

For functional validation, we investigated an aberrantly expressed lncRNA using siRNA in normal urothelial (hTERT-NHU) cells. We selected AB074278, as it was i). associated with disease progression, ii). upregulated in all UC phenotypes (we were interested in lncRNAs generic to UC not sub-type specific), iii). had low predicted protein-coding scores (thus likely to be a ncRNA), iv). worse outcomes with high expression (thus a potential oncogenic role), iv). also upregulated in UC, and v). appeared of particular interest as it was intronic (sense direction) to an protein coding host gene (sense to TANC2; as were most validated ncRNAs in GENCODE) also upregulated in UC (thus potentially regulated by the lncRNA (35)). siRNA Transfection in hTERT-NHU cells reduced expression by 80% after 24 hours and significantly increased cell apoptosis/death (Figure 3b). Knock-down also significantly reduced proliferation (Figure 3c) when compared to scrambled RNA controls. For comparison, we also knocked-down two aberrantly expressed lncRNAs not associated with tumour progression (namely: G36639 and U50531). Proliferation was reduced for both these
IncRNAs, but significantly less dramatically than for AB074278 (Supplementary Figure 3).

**mRNA interactions for AB074278: Epithelial membrane protein 1**

To explore interactions for AB074278, we compared changes in mRNA expression between siRNA knock-down and control NHU cells using HG-U133 Plus 2.0 microarrays in triplicate. We identified 471 mRNAs that were significantly differentially expressed (359 upregulated and 112 down regulated, Figure 4a). As the cellular phenotype of cells with AB074278-siRNA is increased apoptosis and reduced proliferation, we selected the 87/471 mRNAs with roles in proliferation, cell death and apoptosis (gene functions identified using DAVID Bioinformatics Resources, Vsn. 6.7 (36), bold font in supplementary Table 3). We identified 3/87 mRNAs (EMP1, CKS2 and PTGS2) correlated to AB074278 expression in the discovery NCode™ microarray dataset (Pearson's r<-0.35 or >0.2, p<0.001, Supplementary Table 4). We measured the expression of these 3 mRNAs and TANC2 (the host protein-coding gene for AB074278), in the validation cohort of 138 urothelial samples (Supplementary Figure 4) to look for associations with disease phenotype. We identified low expression of Epithelial membrane protein 1 (*EMP1*) in UC was associated with increased risk of progression and BC-specific mortality (Figure 4b, 20% vs. 38% progression rate for tumors with
high and low expression respectively, Log rank p<0.03) and EMP1 was down regulated
in UC (Figure 4c, ANOVA p=0.005), in contrast to the changes seen for AB07428 (worse
outcomes with high expression (Log rank p<0.05) and upregulation in UC (ANOVA
p=0.02)). To support a direct regulation, we also saw increased EMP1 expression (3.8 x
fold) following AB074278 knock-down in NHU-TERT cells (Figure 4d, p<0.01). No
significant changes were seen with respect to phenotype for TANC2, PTGS2 and CKS2.

Discussion
Technological improvements have revealed the importance of ncRNA in cellular
function (20). Whilst the recent GENCODE annotated catalogue details an abundance
and the distribution of many IncRNAs, little is known about their function. Here we have
performed the first large comprehensive screen of IncRNAs in UC to identify those likely
to play roles in the biology of this cancer. Putative IncRNAs were identified using an
algorithm scoring characteristics of protein-coding genes, including open reading frame
length, synonymous/non-synonymous base substitution rates and similarity to known
proteins (27). We identified differential expression of many IncRNAs, without selection
for genetic location, and found that the overall pattern of altered expression (mostly
upregulation) was more extensive to that seen for protein coding mRNAs. Recently
Wang et al. reported balanced changes in lncRNA expression in 12 UC and preliminary data to suggest lncRNAs have malignant roles in mTOR and p53 signaling, and other cancer pathways (24). Whilst our observations do not suggest such a balanced alteration in expression, data in breast and neurological cancers (23) and in human primary keratinocytes (35) support our findings. Given a lack of knowledge regarding lncRNAs and that our panel was designed computationally, we focused upon the minority (17%) of lncRNAs catalogued within GENCODE.(22) This percentage compares to the 12% overlap for lincRNAs reported by Kapranov et al. (37), and suggests much work is needed to fully map the lncRNA transcriptome. To date, various profiling strategies have been used to identify differential lncRNA expression in cancer (e.g. analysis of candidate ncRNA (38, 39), microarrays (40, 41) and transcriptome sequencing (23, 42)). These reports identify lncRNA expression is often tissue or disease specific (23). Our findings support these data and also identify the majority of lncRNAs have tumour phenotype specific expression (71% of all aberrant expressed lncRNAs were phenotype specific vs. 9% were common to all UC). This reflects (but varies in extent) changes in mRNA expression (44% phenotype-specific expression vs. 26% commonly expressed), suggesting co-ordinated anatomical dysregulation (e.g. through regional chromosomal instability or regional epigenetic silencing) or direct interaction (43). To explore this, we mapped adjacent RNAs in our dataset. We identified many (n=3,048) neighbouring
pairs where the expression was correlated (Pearson’s correlation \( r \geq \pm 0.5 \) and \( p < 0.05 \)). This relationship was strongest for genes within 3,000 base pairs (Supplementary Figure 5a). There were 803 IncRNA/mRNA pairs (26%), where both members were significantly differentially expressed in UC. Expression was directly correlated for the vast majority of these pairs (82%, Supplementary Figure 5b), suggesting common transcriptional control or the epigenetic mediation of mRNA expression by nearby IncRNAs (22, 35), antisense RNAs (44) or promoter-associated IncRNAs (45).

The most altered transcripts were found in invasive UC. This was especially true for upregulated IncRNAs (~64%), making the invasive phenotype distinguishable from the low- and high-grade tumors. This observation is in contrast to microRNA profiles within UC, which often share differences between high-grade and invasive tumours (12), and suggests an exciting role to aid pathological disease staging of high grade tumors. Whilst our data require validation, we noted fewer progression events in our high-grade NMI tumours than typical (usually around 25% (5)) suggesting a chance enrichment for more indolent disease may have affected our comparisons. We focused upon IncRNAs that play potential roles in urothelial carcinogenesis, through selecting those associated with tumour progression. Most of those identified were phenotype specific, reflecting global trends within UC, and have not been reported previously in cancer. We used loss-
of-function studies for AB074278 in normal human urothelial cells to explore carcinogenic roles. This cell line has more intact cellular processes (e.g. epigenetic regulation) and fewer genetic events than most malignant cell lines, and so is better to model the subtle impact of epigenetic changes on gene expression (changes from genetic events, such as chromosomal loss/amplification, may dominate subtle modifications arising from epigenetic alterations). AB074278 was chosen because it was significantly upregulated in all tumour phenotypes (it’s overexpression in UC samples could be confirmed by qPCR, data not shown), strongly associated with tumour progression, and has a low protein coding potential score. Furthermore, lncRNA AB074278 is intronic to TANC2, which was also upregulated in UC, suggesting the potential for direct regulation. Whilst the expression of lncRNA AB074278 and TANC2 were not correlated in UC cases, TANC2 expression did drop significantly following AB0724278 knockdown (Supplementary Figure 6), allowing for the speculation that lncRNA AB074278 may regulate (by enhancing) the expression of TANC2 in cis (35). Little is known about the function of TANC2, although it is believed to play a role in embryonic development (46). Whilst these data are promising, we were keen to select mRNAs whose expression was abnormal in UC and correlated to AB074278. Using two microarray screens we identified a dynamic correlation between AB074278 and EMP1, and selected this gene due to its potential involvement in disease progression and its
role in proliferation consistent with our observations. Our expression data of EMP1 matched its functional description in the literature (over-expression of EMP1 was found to inhibit the proliferation of EC9706 cells (47)) and could explain the observed decrease in proliferation in transfected NHU-TERT cells. EMP1 is a putative tumour suppressor gene whose decreased expression is associated with advanced clinical stage and metastasis in oral squamous cell carcinoma (48) and reported to be directly involved in the inhibition of proliferation (47). To explore the link between AB074278 and EMP1 we compared the sequence of both RNAs and EMP1 associated genes (identified through Ingenuity Pathway Analysis, Ingenuity Inc. and EpiTect ChIP qPCR Primers by Transcription Factor search algorithm by SABiosciences) using BLAT (UCSC, GRCh37/hg19) to investigate potential direct interactions. We found a 21nt stretch of identical sequence for AB074278 and SND1 (Staphylococcal nuclease domain-containing protein 1) (Supplementary Table 5). SND1 is a transcription cofactor that regulates EMP1 through its interaction with EMP1’s transcription factor STAT5A (Supplementary Figure 7, Supplementary Material S1-S3). These data suggest but do not confirm the exact relationship between AB074278 and EMP1, which now requires further investigation.
In summary, we have identified many IncRNAs significantly altered in UC and associated with disease progression and tumour subtypes. We specifically implicate AB074278 in apoptosis avoidance and cell proliferation, potentially through regulating the expression of EMP1.

Acknowledgment

We thank Ian Sudbery, University of Oxford and Magnus Rattray, Manchester University for help with microarray annotation, and Dr Paul Heath University of Sheffield for help with microarray studies.

Financial Support. JWFC was supported by a GSK Clinician Scientist fellowship and project/program grants from Yorkshire Cancer Research (Grant number S305PA), Astellas Educational Foundation and the European Union (European Community's Seventh Framework Programme. Grant Numbers: FP7/2007-2013, HEALTH-F2-2007-201438). EB was supported by the Ministry of Science and Higher Education, Poland (Grant No 617/MOB/2011/0).
Disclosure of potential conflicts of interest. The authors declare no competing conflicts of interest.
References


### Table 1. Details of patients and tissue samples analysed in this study.

<table>
<thead>
<tr>
<th></th>
<th>a). Microarray cohort</th>
<th>b). Validation cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UCC</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>Low grade NMI</td>
<td>High grade NMI</td>
</tr>
<tr>
<td>Total Gender</td>
<td>24</td>
<td>13</td>
</tr>
<tr>
<td>Male</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>Female</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>76.3</td>
<td>76.4</td>
</tr>
<tr>
<td>Range</td>
<td>48-82</td>
<td>54-84</td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pTa</td>
<td>24</td>
<td>2</td>
</tr>
<tr>
<td>pTis</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>pT1</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>pT2-4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Progression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>No</td>
<td>22</td>
<td>11</td>
</tr>
<tr>
<td>Follow-up (mo)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>31.8</td>
<td>31.6</td>
</tr>
<tr>
<td>Range</td>
<td>13-98</td>
<td>15-107</td>
</tr>
</tbody>
</table>
Table 2. Detailed information on all lncRNA associated with tumour progression.

<table>
<thead>
<tr>
<th>Genebank ID</th>
<th>lncRNA Size [nt]</th>
<th>Exons</th>
<th>Protein coding score GeneID score</th>
<th>CPC score</th>
<th>Genecode subcategory</th>
<th>UCC expression*</th>
<th>Expression (n)$</th>
<th>Progression rate$</th>
<th>Log rank (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AK122774</td>
<td>2,196</td>
<td>1</td>
<td>-2.42 -1.04</td>
<td>exonic antisense / divergent</td>
<td>All</td>
<td>0.2</td>
<td>27</td>
<td>29</td>
<td>7%</td>
</tr>
<tr>
<td>AK055767</td>
<td>2,282</td>
<td>1</td>
<td>-0.26 -1.04</td>
<td>intergenic opposite strand</td>
<td>Inv</td>
<td>2.2</td>
<td>28</td>
<td>28</td>
<td>46%</td>
</tr>
<tr>
<td>uc004dpc</td>
<td>2,001</td>
<td>1</td>
<td>-0.28 -1.14</td>
<td>intergenic sense</td>
<td>Inv</td>
<td>0.5</td>
<td>28</td>
<td>28</td>
<td>39%</td>
</tr>
<tr>
<td>AF009302</td>
<td>206</td>
<td>1</td>
<td>-2.64 -1.38</td>
<td>intronic sense</td>
<td>Inv</td>
<td>2.2</td>
<td>27</td>
<td>29</td>
<td>44%</td>
</tr>
<tr>
<td>AK026718</td>
<td>2,193</td>
<td>1</td>
<td>-2.52 -1.30</td>
<td>intergenic same strand</td>
<td>UCC only</td>
<td>0.4</td>
<td>27</td>
<td>29</td>
<td>11%</td>
</tr>
<tr>
<td>YA174161</td>
<td>239</td>
<td>1</td>
<td>-4.22 -1.08</td>
<td>intergenic strand</td>
<td>Inv</td>
<td>0.5</td>
<td>27</td>
<td>29</td>
<td>11%</td>
</tr>
<tr>
<td>CR611332</td>
<td>5,304</td>
<td>1</td>
<td>61.16</td>
<td>exonic sense</td>
<td>Inv</td>
<td>0.5</td>
<td>27</td>
<td>29</td>
<td>44%</td>
</tr>
<tr>
<td>AK096965</td>
<td>2,193</td>
<td>1</td>
<td>-2.52 -1.30</td>
<td>intergenic same strand</td>
<td>UCC only</td>
<td>0.4</td>
<td>28</td>
<td>28</td>
<td>39%</td>
</tr>
<tr>
<td>AB074278</td>
<td>1,714</td>
<td>1</td>
<td>-2.46 -1.37</td>
<td>intronic sense</td>
<td>All</td>
<td>3.2</td>
<td>28</td>
<td>28</td>
<td>39%</td>
</tr>
<tr>
<td>AF043897</td>
<td>3,949</td>
<td>1</td>
<td>-2.60 -1.04</td>
<td>exonic antisense</td>
<td>Inv</td>
<td>0.4</td>
<td>27</td>
<td>29</td>
<td>11%</td>
</tr>
<tr>
<td>AK022361</td>
<td>1,774</td>
<td>1</td>
<td>-1.11 -0.79</td>
<td>intronic sense</td>
<td>Inv</td>
<td>0.5</td>
<td>27</td>
<td>29</td>
<td>11%</td>
</tr>
<tr>
<td>AF075063</td>
<td>3,949</td>
<td>1</td>
<td>-1.15 -1.25</td>
<td>intergenic sense</td>
<td>Inv</td>
<td>0.5</td>
<td>28</td>
<td>28</td>
<td>39%</td>
</tr>
<tr>
<td>G36954</td>
<td>457</td>
<td>1</td>
<td>-1.82 -1.10</td>
<td>intergenic same strand</td>
<td>Inv</td>
<td>2.3</td>
<td>27</td>
<td>29</td>
<td>11%</td>
</tr>
<tr>
<td>AK127730</td>
<td>5,015</td>
<td>1</td>
<td>-2.00 -0.75</td>
<td>intergenic opposite strand</td>
<td>All</td>
<td>4.0</td>
<td>27</td>
<td>29</td>
<td>39%</td>
</tr>
<tr>
<td>uc004cem</td>
<td>2,906</td>
<td>1</td>
<td>17.86 -0.24</td>
<td>exonic antisense</td>
<td>LG</td>
<td>0.5</td>
<td>28</td>
<td>28</td>
<td>14%</td>
</tr>
<tr>
<td>uc003unx</td>
<td>4,533</td>
<td>6</td>
<td>17.52 -0.69</td>
<td>exonic antisense</td>
<td>LG, Inv</td>
<td>2.2</td>
<td>27</td>
<td>29</td>
<td>41%</td>
</tr>
<tr>
<td>AK130230</td>
<td>1,538</td>
<td>1</td>
<td>1.58 -1.07</td>
<td>exonic antisense</td>
<td>All</td>
<td>3.1</td>
<td>27</td>
<td>29</td>
<td>41%</td>
</tr>
<tr>
<td>uA837665</td>
<td>995</td>
<td>3</td>
<td>-1.28 -0.64</td>
<td>intergenic same strand</td>
<td>Inv</td>
<td>2.2</td>
<td>28</td>
<td>28</td>
<td>39%</td>
</tr>
<tr>
<td>G36954</td>
<td>457</td>
<td>1</td>
<td>-1.82 -1.10</td>
<td>intergenic same strand</td>
<td>Inv</td>
<td>2.3</td>
<td>27</td>
<td>29</td>
<td>11%</td>
</tr>
<tr>
<td>AK127730</td>
<td>5,015</td>
<td>1</td>
<td>-2.00 -0.75</td>
<td>intergenic opposite strand</td>
<td>All</td>
<td>4.0</td>
<td>28</td>
<td>28</td>
<td>39%</td>
</tr>
<tr>
<td>uc003ey</td>
<td>253</td>
<td>1</td>
<td>-1.67 -1.18</td>
<td>exonic sense</td>
<td>Inv</td>
<td>2.1</td>
<td>27</td>
<td>29</td>
<td>41%</td>
</tr>
<tr>
<td>AK128855</td>
<td>4,453</td>
<td>1</td>
<td>-0.88 -1.02</td>
<td>intronic sense</td>
<td>LG, Inv</td>
<td>0.4</td>
<td>27</td>
<td>29</td>
<td>15%</td>
</tr>
<tr>
<td>uc002hx</td>
<td>2,282</td>
<td>1</td>
<td>0.00 -0.94</td>
<td>intergenic same strand</td>
<td>Inv</td>
<td>2.1</td>
<td>28</td>
<td>28</td>
<td>39%</td>
</tr>
<tr>
<td>BC015007</td>
<td>1,523</td>
<td>1</td>
<td>2.99 -0.88</td>
<td>exonic antisense</td>
<td>LG, Inv</td>
<td>0.4</td>
<td>27</td>
<td>29</td>
<td>15%</td>
</tr>
<tr>
<td>uc001qgv</td>
<td>7,375</td>
<td>2</td>
<td>-0.57 -1.12</td>
<td>intergenic same strand</td>
<td>Inv</td>
<td>2.2</td>
<td>28</td>
<td>28</td>
<td>39%</td>
</tr>
<tr>
<td>AK129609</td>
<td>1,830</td>
<td>1</td>
<td>-1.13 -1.25</td>
<td>intronic sense</td>
<td>LG, HG</td>
<td>2.1</td>
<td>27</td>
<td>29</td>
<td>41%</td>
</tr>
<tr>
<td>BC033908</td>
<td>3,868</td>
<td>3</td>
<td>0.33 -1.14</td>
<td>intergenic same strand</td>
<td>Inv</td>
<td>2.3</td>
<td>27</td>
<td>29</td>
<td>41%</td>
</tr>
<tr>
<td>AF075063</td>
<td>557</td>
<td>1</td>
<td>-1.87 -1.35</td>
<td>intergenic same strand</td>
<td>All</td>
<td>0.2</td>
<td>27</td>
<td>29</td>
<td>15%</td>
</tr>
<tr>
<td>AF086032</td>
<td>303</td>
<td>1</td>
<td>-2.50 -1.27</td>
<td>exonic sense</td>
<td>HG, Inv</td>
<td>0.4</td>
<td>27</td>
<td>29</td>
<td>15%</td>
</tr>
<tr>
<td>AK124944</td>
<td>1,867</td>
<td>1</td>
<td>-0.71 -0.71</td>
<td>exonic antisense</td>
<td>Inv</td>
<td>2.1</td>
<td>28</td>
<td>28</td>
<td>39%</td>
</tr>
<tr>
<td>AF086164</td>
<td>561</td>
<td>1</td>
<td>-2.63 -1.39</td>
<td>intergenic same strand</td>
<td>LG, Inv</td>
<td>0.4</td>
<td>28</td>
<td>28</td>
<td>14%</td>
</tr>
<tr>
<td>BX648713</td>
<td>3,726</td>
<td>1</td>
<td>7.90 -0.90</td>
<td>intronic sense</td>
<td>All</td>
<td>2.3</td>
<td>27</td>
<td>29</td>
<td>41%</td>
</tr>
<tr>
<td>AK022801</td>
<td>1,880</td>
<td>1</td>
<td>-2.51 -1.22</td>
<td>intronic sense</td>
<td>Inv</td>
<td>2.1</td>
<td>27</td>
<td>29</td>
<td>37%</td>
</tr>
</tbody>
</table>

*UC Expression (Phenotype: In which phenotype was abnormal expression seen (LG = Low-grade NMI, HG = High-grade NMI, Inv = Invasive UC). FC: Fold change in UC vs. normal urothelium).

$ Expression (n): the number of cases with high or low expression and Progression: the progression rates in these tumours, respectively.

‡ Protein coding score calculated using GeneID and CPC (Coding Potential Calculator). Negative values suggest a low protein coding risk.
Figure Legends

Figure 1. (a) Percentage of significantly differentially expressed RNAs in UC when compared to normal urothelium. Analyzed lncRNAs were more frequently aberrantly expressed than mRNAs, and less balanced (most were upregulated). (b) Significantly aberrant expressed lncRNAs in UC when annotated by the Gencode v7 catalogue and mapped to genomic structure (expressed as a percentage of the total number of probes present in the database for that genomic locus). (c) LncRNA expression in normal and malignant urothelium. Centroid linkage hierarchical clustering was performed after selecting lncRNAs with >80% expression frequency and median centering data using correlation similarity in Cluster 3.0. First dendrogram row, followed by phenotype description.

Figure 2. (a) LncRNA expression in low-, high-grade and invasive bladder cancer compared with normal urothelium from non-UC cases. Changes in expression (fold change values plotted as log 2) and statistical significance (SAM p-values plotted as −log 10) reveal both upregulation and downregulation of numerous lncRNAs. For all three phenotypes, there is a general upregulation in lncRNA expression. (b) Differentially expressed lncRNAs according to each UC phenotype. The majority of the altered lncRNAs was phenotype specific, with most transcripts found in the invasive phenotype.
**Figure 3.** Tumour progression with respect to IncRNA expression. (a) Tumour progression was analysed with respect to low (dashed line) and high (solid line) IncRNA expression (dichotomized around median). Examples of aberrant up (AK127730, AK130203 and AB074278) and down (AF075063, BC015007 and AK122774) IncRNA regulation in UC (plotted by the Kaplan Meier method, Log rank p values shown). (b) siRNA mediated gene knock-down of the IncRNA AB074278 led to 84% reduction of its expression in hTERT NHU cells and caused cell death / apoptosis. Cell death was determined by propidium iodide flow cytometry and apoptosis by Caspase-3 activation and flow cytometry. (c) Cell proliferation, measured by MTT assay, was also reduced following knock-down of AB074278.

**Figure 4.** Changes in gene expression in hTERT-NHU cells after knock-down of IncRNA AB074278. (a). The altered expression of 17,601 mRNAs was recorded using HG-U133 Plus 2.0 microarrays. Complete linkage hierarchical clustering was performed after selecting mRNAs with >80% expression frequency (n = 536) and mean centering data using correlation similarity in Cluster 3.0. Arrow identifies EMP1. (b) Tumour progression and UC specific mortality with respect to EMP1 and AB074278 expression (dichotomized around the median), was analysed and plotted by the Kaplan Meier
method, Log rank p values shown. (c). Expression of EMP1 and AB074278 in the validation cohort of urothelial samples (box plots show median and 95%CI DCt values (normalized to TERT expression), ANOVA p values shown). (d). Expression of lncRNA AB074278 after siRNA mediated knock-down and transfection with control scrambled siRNA (Scr RNA) and corresponding changes in expression of mRNA EMP1 after knock-down of AB074278 as measured by quantitative real-time PCR.
Figure 1
Figure 2
Figure 3
Figure 4
Clinical Cancer Research

Identification of differentially expressed long non-coding RNAs in bladder cancer

Stefan Peter, Edyta Borkowska, Ross M Drayton, et al.

Clin Cancer Res  Published OnlineFirst August 27, 2014.

Updated version

Access the most recent version of this article at: doi:10.1158/1078-0432.CCR-14-0706

Supplementary Material

Access the most recent supplemental material at: http://clincancerres.aacrjournals.org/content/suppl/2014/08/27/1078-0432.CCR-14-0706.DC1

Author Manuscript

Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts

Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions

To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions

To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.