

RESEARCH

Open Access



Insights into long noncoding RNAs of naked mole rat (*Heterocephalus glaber*) and their potential association with cancer resistance

Jian-Jun Jiang^{1,2}, Le-Hua Cheng^{1,3}, Huan Wu^{1,2}, Yong-Han He^{1*} and Qing-Peng Kong^{1*}

Abstract

Background: Long noncoding RNAs (lncRNAs) are a class of ubiquitous noncoding RNAs and have been found to act as tumor suppressors or oncogenes, which dramatically altered our understanding of cancer. Naked mole rat (NMR, *Heterocephalus glaber*) is an exceptionally long-lived and cancer-resistant rodent; however, whether lncRNAs play roles in cancer resistance in this seductive species remains unknown.

Results: In this study, we developed a pipeline and identified a total of 4422 lncRNAs across the NMR genome based on 12 published transcriptomes. Systematic analysis revealed that NMR lncRNAs share many common characteristics with other vertebrate species, such as tissue specificity and low expression. BLASTN against with 1057 human cancer-related lncRNAs showed that only 5 NMR lncRNAs displayed homology, demonstrating the low sequence conservation between NMR lncRNAs and human cancer-related lncRNAs. Further correlation analysis of lncRNAs and protein-coding genes indicated that a total of 1295 lncRNAs were intensively coexpressed ($r \geq 0.9$ or $r \leq -0.9$, cP value ≤ 0.01) with potential tumor-suppressor genes in NMR, and 194 lncRNAs exhibited strong correlation ($r \geq 0.8$ or $r \leq -0.8$, cP value ≤ 0.01) with four high-molecular-mass hyaluronan related genes that were previously identified to play key roles in cancer resistance of NMR.

Conclusion: In this study, we provide the first comprehensive genome-wide analysis of NMR lncRNAs and their possible associations with cancer resistance. Our results suggest that lncRNAs may have important effects on anticancer mechanism in NMR.

Keywords: Naked mole rat, Long noncoding RNA, Expression profiles, Coexpression, Cancer resistance

Background

Naked mole rats (NMRs; *Heterocephalus glaber*) are small, nearly hairless, subterranean rodents that are renowned for their longevity and cancer resistance [1, 2]. In comparison with a similarly sized house mouse, the NMR exhibits unusual longevity, with a maximum lifespan exceeding 30 years, making NMR the longest-living rodent [3–5]. In addition, NMRs are able to maintain

health until almost the end of their lives and displayed exceptional resistance to multiple age-related diseases, such as cancer [6]. The high-molecular-mass hyaluronan (HMM-HA) was proved to play a key role in regulating the cancer resistance in NMR [7]. In the NMR cells, tumor resistance is mediated by signals from the HMM-HA triggering the induction of *INK4* (inhibitors of cyclin-dependent kinase 4) locus expression. The *INK4* locus encodes a novel product named pALT^{INK4a/b} which may have a crucial contribution to tumor resistance and longevity of NMR [8]. A recent study suggested that NMR-specific alternative reading frame (*ARF*) and the disruption of oncogene embryonic cell-expressed

*Correspondence: heyonghan@mail.kiz.ac.cn; kongqp@mail.kiz.ac.cn

¹ State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, the Chinese Academy of Sciences, Kunming 650223, China

Full list of author information is available at the end of the article

Ras (*ERAS*) regulate tumor resistance in NMR-induced pluripotent stem cells (NMR-iPSCs) [9]. However, the deeper mechanisms of anticancer in NMRs are not well understood, which pushes forward us to look for a better understanding of the ability of cancer resistance in NMR.

Long noncoding RNAs (lncRNAs) are operationally defined as RNA transcripts longer than 200 bp that do not appear to have coding potential [10]. Recent studies indicated that lncRNAs have become new players in cancer [11] and attracted plenty of attention in scientific community due to their altered expressions and dysregulated functions as tumor suppressors or oncogenes in various human cancer types [12, 13]. Therefore, a preferable comprehending of lncRNAs becomes important and essential for cancer research. With the rapid development of next-generation sequencing (NGS) techniques and in silico analysis, large sets of lncRNAs have been identified in many species; however, it has yet to be applied in NMR. Considering the close connections between lncRNAs and cancers, it becomes urgent and necessary to identify and feature lncRNAs in the attractive cancer-resistant rodent, NMR.

In this study, we characterized for the first time the genome-wide lncRNA profiles in NMR and identified the association of lncRNAs with tumor-associated genes. Our results provide new insights into the longevity and cancer resistance in NMR, which is of essential significance in enhancing our understanding of cancer and especially broaden our knowledge on mechanisms of cancer resistance.

Results

Genome-wide identification and characterization of NMR lncRNAs

In order to comprehensively identify NMR lncRNAs, 12 transcriptomes of three tissues (kidney, liver, and brain) from newborn, 4-year-old, 4-year-old with low-oxygen-treated and 20-year-old NMRs were collected [3] (Additional file 1: Table S1). Raw datasets were first subjected to SolexaQA (-h 20, -l 30) [14] to remove low-quality and short reads, and 67.8 million reads were retained for further analysis. Using a stringent filtering pipeline (Fig. 1), 4422 potential lncRNAs yielded, consisting of 3684 long intergenic noncoding RNAs (lincRNAs), 733 antisense lncRNAs (anti-lncRNA) and 5 lncRNAs transcribed from intronic regions (in-lncRNA) (Table 1). The average length of NMR lncRNAs is 16,625 bp, of which in-lncRNAs have a significant shorter length (3922 bp) than anti-lncRNA (19,668 bp) and lincRNA (16,037 bp). Compared with protein-coding genes (PCGs), lncRNAs exhibited a shorter length (Fig. 2a), less average number of exons per lncRNA transcript (2.51 vs. 8.05, Fig. 2b), but higher GC content (Fig. 2c). As expected, lncRNAs showed a

significant lower expression level than PCGs by comparing the fragments per kilobase of exon per million fragments mapped (FPKM) values (Fig. 2d), which has also been observed in other species such as rainbow trout [15] and pacific oyster [16]. In conclusion, lncRNAs in NMR displayed higher GC content but shorter length, less exon numbers and lower expression level in comparison with PCGs (Fig. 2).

Expression profiles of NMR lncRNAs

RNA-seq datasets from newborn, 4 year-old and 20-year-old NMRs were obtained to characterize the expression pattern of the lncRNAs. More than 80% of lncRNAs are expressed at all ages of NMR. There are 45 lncRNAs, 29 lncRNAs and 31 lncRNAs specifically expressed in newborn, 4-year-old and 20-year-old NMRs, respectively. The age-specific expression of lncRNAs indicates that some lncRNAs may play roles in the growth and development of NMR. In addition, we found 67 lncRNAs expressed particularly in 4-year-old NMR with low-oxygen treatment, suggesting them to be likely involved in low-oxygen metabolism (Fig. 3a, Table Additional file 2: S2).

Tissue-specific expression of lncRNAs was investigated using RNA-seq datasets from NMR livers, kidneys and brains. Consistent with other vertebrate species [17, 18], NMR lncRNAs displayed a significant tissue-specific expression pattern. 3667 lncRNAs are expressed at least in one tissue type, 190 are expressed merely in brains, 226 specifically in kidneys, and 205 in livers (Fig. 3b, Additional file 2: Table S2).

Differential expression of lncRNAs across developmental tissues

We next utilized these 12 RNA-seq datasets to explore the expression dynamics of lncRNAs in the NMR genome. A total of 40 lncRNAs were found differentially expressed across the developmental tissues with a fold change >2 and false discovery rate (FDR) < 0.05 (Fig. 4, Additional file 3: Figure S1, Additional file 4: Table S3). Recent studies indicated that transcription of ncRNAs including some lncRNAs may act to regulate the capacities of their chromosomal neighboring PCGs, both negatively and positively [19, 20]. Therefore, we selected 13 lncRNAs whose expressions significantly varied in livers or brains and acquired their nearest neighboring PCGs for further analysis (Fig. 4). The 13 lncRNAs are likely to be highly expressed in either liver or brain of the newborn (NB) NMR or in the liver of the older individuals, and we found that their nearest PCGs also exhibited metabolism-related or brain-related functions, indicating that some lncRNAs may have a functional connection with their neighboring PCGs.

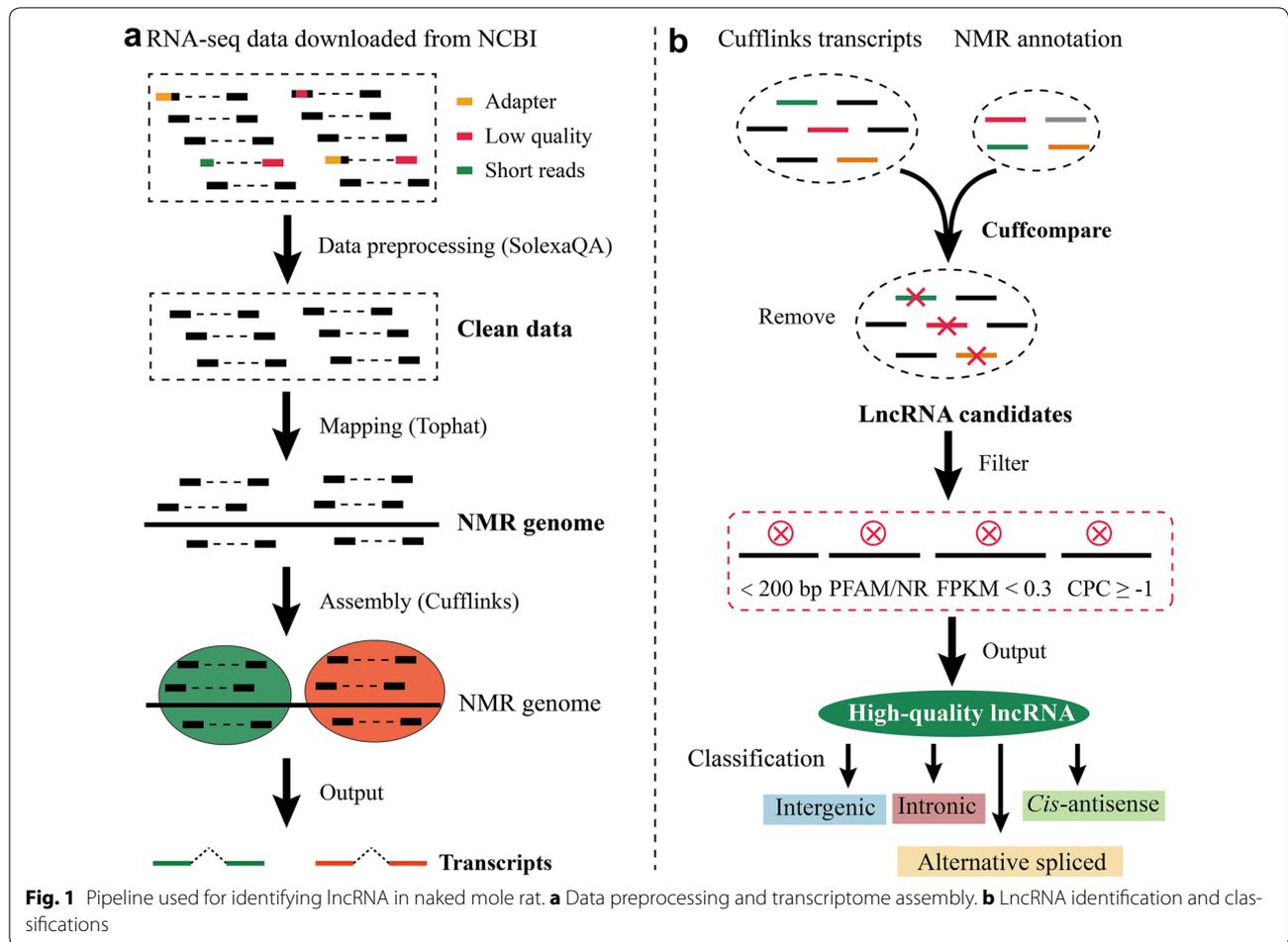


Table 1 High-quality lncRNAs identified in naked mole rat

	Number	Avg length (bp)	Exon number	Avg ExpLev ^a
Intergenic	3684	16,037	2.51	33.74
Intronic	5	3922	1.00	9.31
Antisense	733	19,668	2.55	7.50
Total	4422	16,625	2.51	29.33

^a FPKM value was estimated by TopHat and Cufflinks with default parameters; Avg Average

Conservation analysis between NMR, human and mouse lncRNAs

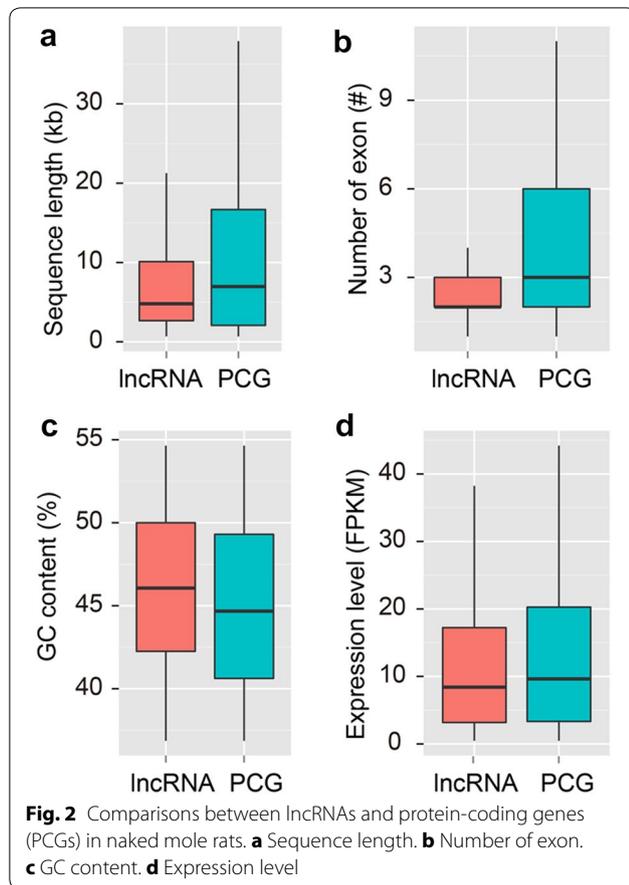
To identify homologs of the NMR lncRNAs in humans and mice, lncRNA sequences of two species were downloaded from GENCODE database. In all 27, 817 lncRNAs of humans (version 23) and 12,169 of mice (version M7) were requested. Orthologous analysis of lncRNAs of three species was performed using OrthoMCL with BLASTN program [21], which identified more than 4359 (98%) NMR lncRNAs with no detectable homologs in

both humans and mice lncRNA (Fig. 5a). Finally, merely 11 orthologous groups are retained across three species, revealing the deficiency of sequence conservation in lncRNAs (Additional file 5: Table S4).

lncRNAs display higher conservation in genomic position than sequence in diverse organisms [18, 22, 23]. Therefore, we conducted the analysis of lncRNA positional conservation among three species. As a consequence, 35 and 41 NMR lncRNAs show conservation with human and mice, respectively (Fig. 5b).

Homologous analysis of NMR lncRNA and human cancer-related lncRNAs

NMR, a strictly subterranean and extraordinarily long-lived eusocial mammal, was found resistant to both spontaneous cancer and experimentally induced tumorigenesis [3, 24]. To investigate the sequence conservation between NMR lncRNAs and cancer-related lncRNAs, homologous analysis was performed. In this study, a total of 1057 experimentally supported lncRNAs associated with 86 human cancers were requested from Lnc2Cancer



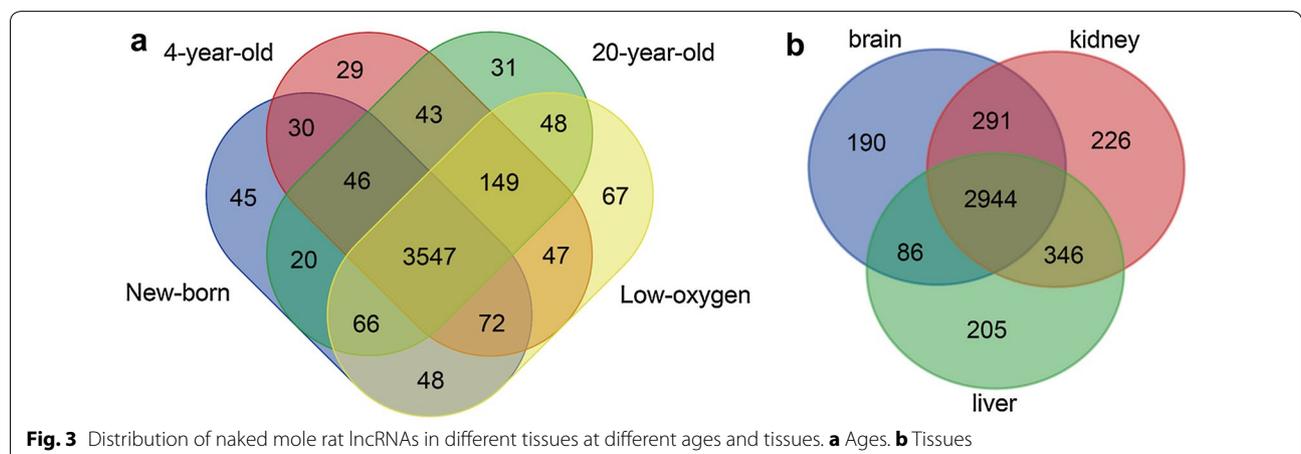
database [25], and then BLASTN [26] homology search with NMR lncRNAs, coverage of query and/or target >40% and *E*-value < 1e-5 were retained. As a result, only 5 NMR lncRNAs showed homology (Table 2), indicating low sequence conservation between NMR lncRNAs and human cancer-related lncRNAs. BLASTN searching of the 5 NMR lncRNAs that exhibited homology with

human cancer-related lncRNAs (NMR-HCRLs) against to the NONCODE database [27] found highly similar sequences not only in human but also in mouse, rat, and etc. (data not shown), indicating that they might be conserved across species.

To determine PCGs in NMR that possibly correlated with the 5 NMR-HCRLs, we conducted coexpression analysis between mRNAs expressed in four or more developmental tissues and the 5 NMR-HCRLs. Consequently, 271 PCGs were intensively positively coexpressed with these 5 lncRNAs (Additional file 6: Table S5), while only 12 PCGs in NMR were negatively coexpressed with them ($r \geq 0.9$ or $r \leq -0.9$, *cP* value ≤ 0.01) (Additional file 6: Table S5). The 271 PCGs were further classified by Gene Ontology (GO) together with KEGG pathway analysis. As a result, we found a significant enrichment of 106 GO terms and 3 KEGG pathways including Neuroactive Ligand Receptor Interaction (*hsa04080*), Maturity Onset Diabetes of the Young (*hsa04950*) and Glycosaminoglycan Biosynthesis Heparan Sulfate (*hsa00534*) ($P < 0.01$) (Additional file 7: Table S6).

Coexpression analysis of NMR lncRNAs with potential tumor-suppressor genes

To further explore potential associations between NMR lncRNAs and cancer resistance, human tumor-suppressor genes were utilized for further analysis. At the first step, a total of 1217 experimentally verified human tumor-suppressor genes were requested from TSGDB database [28], and then BLAST homology search with PCGs annotated from NMR genome [29]. Overall, we found that 901 PCGs (Additional file 8: Table S7) of NMR are homologs of human tumor-suppressor genes, which were considered as potential tumor-suppressor genes of NMR genome (PTSGs-NMR). In order to precisely analyze the relationship between lncRNAs and cancer resistance, PTSGs-NMR and lncRNAs expressed in all



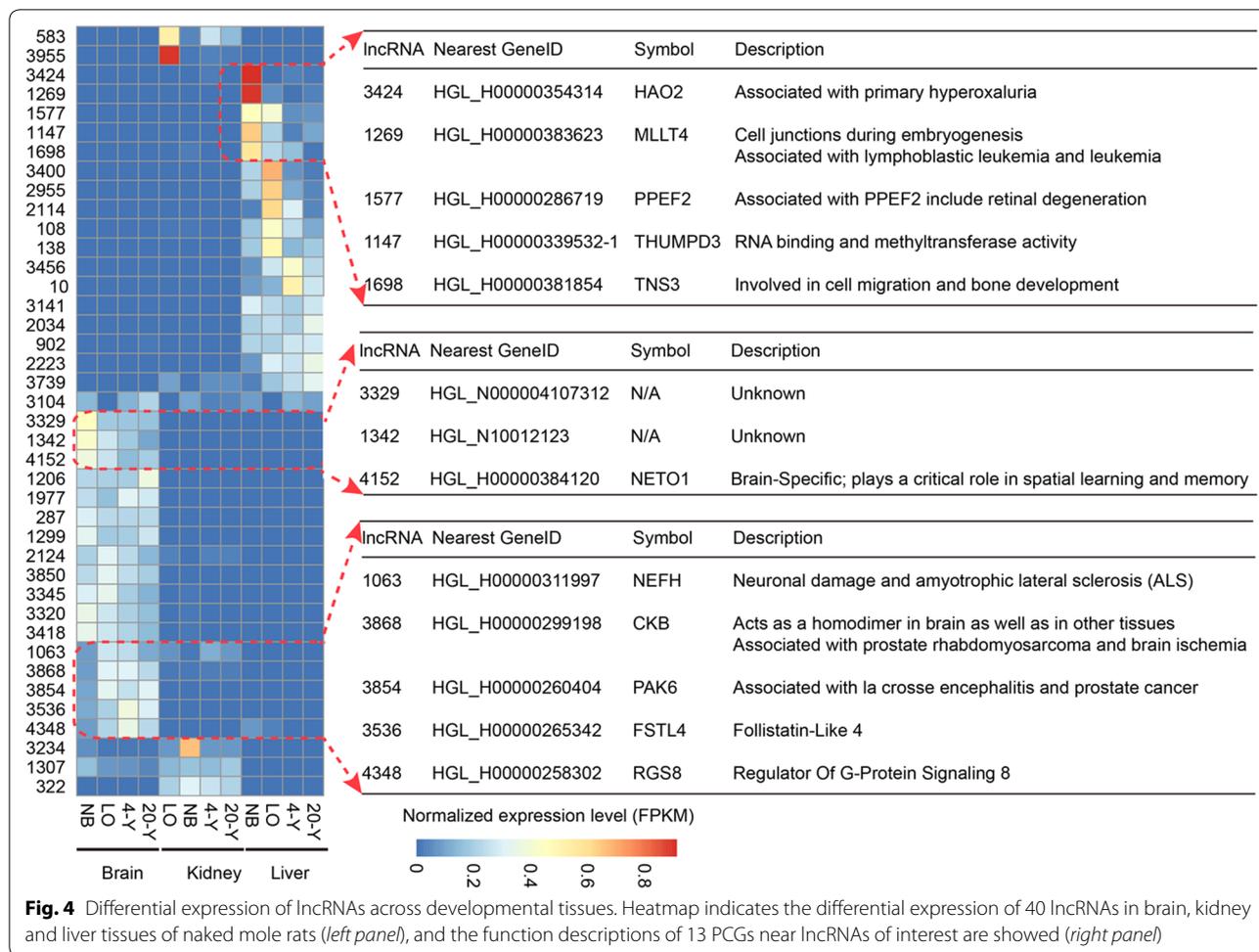


Fig. 4 Differential expression of lncRNAs across developmental tissues. Heatmap indicates the differential expression of 40 lncRNAs in brain, kidney and liver tissues of naked mole rats (left panel), and the function descriptions of 13 PCGs near lncRNAs of interest are showed (right panel)

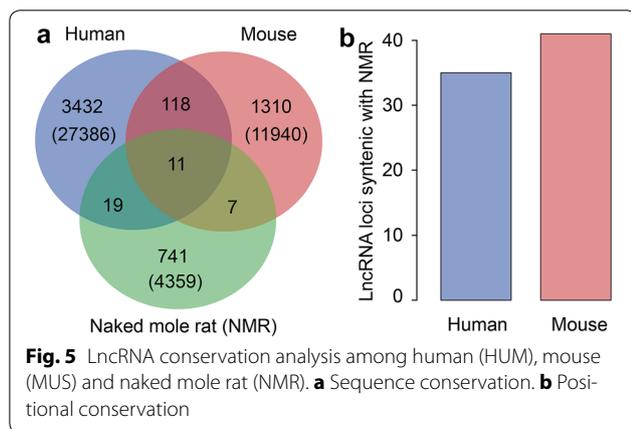


Fig. 5 LncRNA conservation analysis among human (HUM), mouse (MUS) and naked mole rat (NMR). **a** Sequence conservation. **b** Positional conservation

developmental tissues were reserved for next-step analysis. As a result, 2091 lncRNAs and 726 PTSGs-NMR were retained for the ultimate coexpression analysis. Coexpression results show that ~61.93% (1295/2091) lncRNAs are intensively coexpressed with PTSGs-NMR ($r \geq 0.9$

or $r \leq -0.9$, cP value ≤ 0.01). Interestingly, more than 64% (834/1295) lncRNAs are positively coexpressed with a group of PTSGs-NMR and meanwhile exhibit negative coexpression with some other PTSGs-NMR as well, suggesting the possible role of lncRNAs in cancer resistance (Additional file 9: Table S8). To investigate the ratio of lncRNAs that coexpressed with potential tumor-suppressor genes in tumor-prone animals such as rat, we downloaded rat lncRNA sequences from NONCODE database [27] and 12 RNA-seq datasets across three tissues types (kidney, liver and brain) at four developmental stages of rats [30], then performed coexpression analysis between potential tumor-suppressor genes of rat (PTSGs-Rat) and lncRNAs using the same method and standard in NMR. The FPKM value of rat mRNA was requested from Rat body map database [30]. In all, 706 PTSGs-Rat and 16,328 lncRNAs were retained for coexpression analysis. About 60.41% (9864/16,328) lncRNAs are strikingly coexpressed with 670 tumor PTSGs-Rat which is lower than that in NMR (Additional file 10: Table S9).

Table 2 Five NMR lncRNAs show homology with human cancer-related lncRNAs

LncRNA	Human	Qcov	Tcov	Iden	Cancer name
4086	ENST00000602892	40.5	60.5	89.4	Acute megakaryoblastic leukemia
3653	ENST00000540687	27.8	63.8	83.6	Hepatocellular carcinoma
172	ENST00000605417	91.7	9.8	96.3	Lung adenocarcinoma
3712	ENST00000409139	55.8	34.9	81.9	Gastric cancer/hepatocellular carcinoma
1169	ENST00000493116	92.3	73.0	90	Hepatocellular carcinoma/lung squamous cell carcinoma

qcov coverage of query to target, tcov coverage of target to query, iden identity

Coexpression analysis of NMR lncRNAs with four HMM-HA-related genes

HMM-HA, the powerful trigger for the early contact inhibition (ECI) observed in NMR cells [31, 32], mediates the cancer resistance in NMR [33]. A previous study revealed that the HA signaling triggering ECI in NMR is in part transmitted via the CD44 receptor which interacts with neurofibromin 2 (NF2) on the cytoplasmic face [33]. In addition, silencing of hyaluronan synthase 2 (HAS2) or overexpression of HA-degrading enzyme (HYAL2) in NMR cells led to susceptibility to malignant transformation and tumorigenesis in mice [7]. Hence, in order to comprehensively explore lncRNAs that related to HMM-HA metabolism in NMR, we selected four HMM-HA-related genes (CD44, NF2, HYAL2 and HAS2) to perform coexpression analysis with NMR lncRNAs that expressed in all developmental tissues. Consequently, ~9.27% lncRNAs (194/2091) are strongly coexpressed with them (105 positively and 95 negatively, $r \geq 0.8$ or $r \leq -0.8$ with cP value ≤ 0.01) (Additional file 11: Table S10). Among the 194 correlated lncRNAs, three lncRNAs including 2462, 2463 and 2464 are transcribed from one gene locus, and coexpression analysis showed that 2462 is negatively correlated with HAS2 while 2464 is positively coexpressed with HYAL2; meanwhile, both 2463 and 2464 are negatively coexpressed with NF2. Similar to 2464, another two lncRNAs (lncRNA 77 and 471) show the same correlation with HYAL2 and NF2 (Table 3). Interestingly, we found that three lncRNAs (lncRNA 691, 852 and 2980) are positively coexpressed with NF2 and negatively coexpressed with HYAL2, displaying an opposite coexpression relationship compared with lncRNA 2464 (Table 3). In summary, we discovered 3 lncRNAs that transcribed from one gene locus and had close connections with three of the four known HA-related genes of NMR. Besides, 6 lncRNAs were found strongly coexpressed with NF2 as well as HYAL2, negatively or positively. To explore the connection between HA synthesis and lncRNAs of tumor-prone rodent, we analyzed lncRNAs that coexpressed with four HA-related genes in the rat. Around 11.67% (1906/16,328) lncRNAs show strong coexpression ($r \geq 0.8$ or $r \leq -0.8$ with cP value ≤ 0.01)

Table 3 NMR lncRNAs that coexpressed with both NF2 and HYAL2

LncRNA	NF2 (r)	HYAL2 (r)
2464	-0.811	0.839
77	-0.804	0.944
471	-0.804	0.888
691	0.902	-0.811
852	0.839	-0.916
2980	0.825	-0.874

which is a little higher than NMR (~9.27%). Nevertheless, we did not detect any rat lncRNAs coexpressed with both NF2 and HYAL2 like that in NMR, but found some rat lncRNAs simultaneously coexpressed with NF2 and HAS2 (Additional file 12: Table S11). The different coexpression patten between lncRNAs and HA-related genes in NMR and rat may contribute to HA synthesis or tumorigenesis. Collectively, these results suggest that lncRNAs might be involved in HMMA regulation.

Discussion

In this study, 4422 lncRNAs corresponding to 2946 loci were identified from the reported NMR RNA-seq datasets including three different tissues from newborn, young adult (4-year-old) and old adult (20-year-old) NMRs. Orthologous analysis by tool OrthoMCL indicated that most of the NMR lncRNAs have detectable homology with neither human or mouse lncRNAs, demonstrating that lncRNAs lack sequence conservation. However, five human lncRNAs which were previously reported to be cancer-related lncRNAs displayed high sequence conservation with NMR lncRNAs, especially the human lncRNA ENST00000493116 (also known as SOX2 overlapping transcript, SOX2OT). SOX2OT is a lncRNA that harbors in the intronic region of SOX2 gene which is one of the major regulators of pluripotency [34]. In human cancers, SOX2OT is co-upregulated with SOX2 and OCT4 in esophageal squamous cell carcinoma [35] and was also suggested to play key roles in the induction

and/or maintenance of *SOX2* expression in breast cancer [36]. In NMR genome, *SOX2OT* has high sequence homology with lncRNA *1169* which is expressed only in brains and kidney with low-oxygen treatment but not in livers of NMR. The function of lncRNA *1169* in the NMR is not clear, but BLASTN searching of lncRNA *1169* as well as the other four NMR-HCRLs against the NON-CODE database found BLAST hits on the queries, not only in human but also in mouse, rat, and etc., suggesting that they are possibly conserved across species. LncRNA evolution analysis by Hezroni et al. [18] revealed that *SOX2OT* is a bona fide highly conserved lncRNAs which further strengthened our results.

LncRNA expression displays spatiotemporal and tissue-specific characteristics [37, 38] which were also observed in NMR. We found some lncRNAs were particularly expressed in newborn, 4-year-old or 20-year-old NMR, exhibiting age-specific expression. The specific expression of 31 lncRNAs that merely detected in 20-year-old NMR implies their potential connection with aging or senescence due to the important roles of lncRNAs in senescence [39]. RNA-seq and microarray studies have identified altered lncRNAs during aging and in response to various types of senescence stimuli, such as *ANRIL* [40–42], *MIR31HG* [43], *PANDA* [44, 45] and *lincRNA-p21* [46, 47]. Hence, our analysis of NMR lncRNAs will provide new information for senescence study.

It was reported that functions of lncRNAs can be inferred by coexpression analysis and their genome locations [48, 49]. In NMR genome, we performed coexpression analysis of lncRNAs with PTSGs-NMR and four HA-related genes. In spite of the unknown functions of NMR lncRNAs, coexpression analysis showed ~61.93% (1295/2091) of NMR lncRNAs were intensively coexpressed with PTSGs-NMR ($r \geq 0.9$ or $r \leq -0.9$, cP value ≤ 0.01). This ratio is slightly higher than that in rat, demonstrating the potential role of lncRNAs in cancer resistance of NMR. HA was verified to be involved in regulating anticancer mechanism in NMR, and we found that three lncRNAs transcribed from one gene locus are closely related to three of the four HA-related genes, especially lncRNA *2464* which coexpressed with both *NF2* and *HYAL2*. As a result, a total of six lncRNAs that coexpressed with *NF2* as well as *HYAL2* were found. Unlike the NMR, in the rat, a tumor-prone rodent, we found some lncRNAs coexpressed with both *NF2* and *HAS2*. *HYAL2* is an enzyme responsible for regulating the degrading of HMM-HA which is the powerful trigger for the contact inhibition [7, 49], while *NF2* (merlin) interacts with CD44 receptor and mediates the contact inhibition [50]; both of them were verified to play crucial

roles in cancer resistance of NMR. Comparing with other mammalian, NMR has2 protein has some unique amino acid changes which may be partially responsible for the NMR's unusual function of HMM-HA [33]. LncRNAs expression in NMR and rat shows different coexpression characters with HA-related genes and highlights the potential function of lncRNAs in HMM-HA regulation and cancer resistance.

Conclusion

In summary, we first identified and featured lncRNAs across NMR genome and our integrated analysis of NMR lncRNAs suggests the high potential of these lncRNAs in regulating cancer resistance in the NMR, therefore providing new insight into understanding human cancer biology as well as promising targets of cancer treatment and anticancer drug development.

Methods

Data accessibility

The raw NMR and rat transcriptome datasets are available at *National Center for Biotechnology Information* sra database (<http://www.ncbi.nlm.nih.gov/sra/>). The assembled genome sequences and annotation files of NMR were requested from GIGA database (<http://gigadb.org/dataset/100022>). LncRNA sequences of human and mouse were downloaded from GENCODE database (<http://www.genencodegenes.org/>). LncRNA sequences of rat were downloaded from NONCODE database [27].

LncRNA identification pipeline

A stepwise filtering pipeline (Fig. 1) was used to identify putative lncRNAs from 12 RNA-seq datasets. (1) Low-quality (Phred score < 20) and short (length < 30 bp) reads were trimmed using SolexaQA [14]. The trimmed and size selected reads were then mapped to the NMR's genome using Tophat [50]. (2) Aligned reads were assembled and merged by Cufflinks and Cuffmerge, and then noncoding transcripts for each sample were obtained by utilizing Cuffcompare [51]. Transcripts shorter than 200 bp were excluded as putative long noncoding RNAs which were commonly defined as transcripts with length longer than 200 bp. (3) The tool coding potential calculator (CPC) was employed to assess the protein-coding potential of a transcript, and transcripts with CPC ≥ -1 were eliminated [52]. (4) To evaluate which of the remaining transcripts contains a known protein-coding domain, transcripts are BLAST to Pfam database [53] and nonredundant protein database (NR) and those that with a Pfam or NR hit are excluded. (5) Transcripts with FPKM value lower than 0.3 were removed.

OrthoMCL, BLAST and positional conservation analyses

According to method reported previously [23], here we adopted the OrthoMCL [21] pipeline to compare the sequence similarity of lncRNAs among NMR, mouse and human using the BLASTN program. The BLASTN hits with coverage $\geq 50\%$ and E -value $\leq 1E-5$ were retained and applied to assign putative orthologous groups using Markov Cluster (MCL) algorithm. BLASTN [26] homology search was conducted between NMR lncRNA and human cancer-related lncRNAs. lncRNAs with coverage of query and/or target $>40\%$ and E -value $< 1e-5$ were retained.

To perform positional conservation analyses, we first retrieved those lncRNAs pairs with E -value $\leq 1E-5$ from BLAST results and regarded them as putative homologous sequences. We then constructed the syntenic blocks between NMR, mouse and human using MCScanX with default parameters [54]. When considering two orthologous protein-coding genes of G1 and G2 between naked mole rat and human, we detected lncRNAs within 815 kb of G1 in NMR and within 894 kb of G2 in human as previously suggested [18]. An lncRNA was considered to be found “upstream” of the protein-coding gene when it overlapped or ended 5' end, and “downstream” when it overlapped or started the 3' end of the protein-coding gene. Two lncRNAs of L1 and L2 from NMR and human were considered syntenic, if they were both upstream or both downstream of G1 and G2, with the same relative orientations. Similar standards and methods were also used to identify syntenic lncRNAs between naked mole rat and mouse.

Coexpression analysis

To investigate the possible relationship between lncRNA and cancer-resistant, we performed coexpression analysis in both NMR and rat by testing FPKM values of their 12 transcriptomes, respectively. Tophat and cufflinks were used to obtain FPKM values of lncRNAs and mRNAs [51]. PCGs expressed in at least four developmental tissues were retained for coexpression analysis with 5 NMR-HCRLs. In coexpression analysis between lncRNAs and candidate tumor-suppressor mRNAs/HA-related genes, lncRNAs and mRNAs that expressed in all tissues were reserved for analysis in both NMR and rat.

In this study, Pearson correlation coefficient (r) and correlation P value (cP value) were used to assess the coexpression relationship by an in-house matlab script. $r \geq 0.8$ or $r \leq -0.8$ with cP value ≤ 0.01 was considered as strong correlation, while $r \geq 0.9$ or $r \leq -0.9$ with cP value ≤ 0.01 was deemed as intensively correlation.

Functional enrichment analysis

To classify protein-coding genes that correlated with the 5 NMR-HCRLs, Gene Ontology and KEGG

pathway enrichment analysis was performed using the hypergeometric distribution and Bonferroni correction for multiple hypotheses testing with a cutoff P value of 0.01.

Additional files

Additional file 1: Table S1. Summary of transcriptome sequencing data of the naked mole rat used in this study.

Additional file 2: Table S2. Expression levels of NMR lncRNAs among different tissues.

Additional file 3: Figure S1. Heatmap of expression profiles of NMR lncRNAs across 12 developmental tissues.

Additional file 4: Table S3. Differential expressed lncRNAs identified in naked mole rat (*Heterocephalus glaber*) genome.

Additional file 5: Table S4. List of homologous lncRNA groups between human, mouse and naked mole rat.

Additional file 6: Table S5. List of NMR lncRNAs that intensively coexpressed with five NMR-HCRLs.

Additional file 7: Table S6. Functional enrichment of PCGs that intensively coexpressed with five NMR-HCRLs.

Additional file 8: Table S7. List of potential tumor-suppressor genes of NMR genome (PTSGs-NMR).

Additional file 9: Table S8. List of NMR lncRNAs that intensively coexpressed with PTSGs-NMR.

Additional file 10: Table S9. List of rat lncRNAs that intensively coexpressed with PTSGs-Rat.

Additional file 11: Table S10. List of NMR lncRNAs that strongly coexpressed with four HA-related genes.

Additional file 12: Table S11. List of rat lncRNAs that strongly coexpressed with four HA-related genes.

Abbreviations

NMR: naked mole rat; lncRNA: long noncoding RNA; ncRNA: noncoding RNA; lincRNAs: long intergenic noncoding RNAs; anti-lncRNA: antisense lncRNAs; in-lncRNA: lncRNAs transcribed from intronic regions; PCG: protein-coding gene; HMM-HA: high-molecular-mass hyaluronan; FPKM: fragments per kilobase of exon per million fragments mapped; PTSGs-NMR: potential tumor-suppressor genes of NMR genome; PTSGs-Rat: potential tumor-suppressor genes of rat; NMR-HCRLs: NMR lncRNAs that exhibited homology with human cancer-related lncRNAs; FDR: false discovery rate; CPC: coding potential calculator.

Authors' contributions

KQP and HYH conceived and designed the study. JJJ, CLH and WH performed all analysis and drafted the manuscript. KQP, HYH and JJJ revised the manuscript. All authors read and approved the final manuscript.

Author details

¹ State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, the Chinese Academy of Sciences, Kunming 650223, China. ² Kunming College of Life Science, University of Chinese Academy of Sciences, Beijing 100049, China. ³ The School of Life Science, University of Science and Technology of China, Hefei 230027, China.

Acknowledgements

We would like to acknowledge other members of the Kong laboratory for helpful discussions and comments.

Competing interests

The authors declare that they have no competing interests

Availability of supporting data

Requests for the datasets supporting the conclusions of this article should be directed to the corresponding author.

Consent for publication

Yes.

Funding

This work was supported by grants from the Chinese Academy of Sciences (No. KJZD-EW-L14, No. QYZDB-SSW-SMC020), National Basic Research Program of China (No. 2012CB518205), the National Natural Science Foundation of China (Nos. 81671404, 81272309, 31322029), and the Youth Innovation Promotion Association of the Chinese Academy of Sciences.

Received: 13 July 2016 Accepted: 25 October 2016

Published online: 10 November 2016

References

- Edrey YH, Park TJ, Kang H, Biney A, Buffenstein R. Endocrine function and neurobiology of the longest-living rodent, the naked mole-rat. *Exp Gerontol*. 2011;46(2–3):116–23.
- Crish SD, Rice FL, Park TJ, Comer CM. Somatosensory organization and behavior in naked mole-rats I: vibrissa-like body hairs comprise a sensory array that mediates orientation to tactile stimuli. *Brain Behav Evol*. 2003;62(3):141–51.
- Kim EB, Fang X, Fushan AA, Huang Z, Lobanov AV, Han L, Marino SM, Sun X, Turanov AA, Yang P, et al. Genome sequencing reveals insights into physiology and longevity of the naked mole rat. *Nature*. 2011;479(7372):223–7.
- Buffenstein R, Jarvis JUM. The naked mole rat—a new record for the oldest living rodent. *Sci Aging Knowl Environ SAGE KE*. 2002;2002(21):pe7.
- Buffenstein R. Negligible senescence in the longest living rodent, the naked mole-rat: insights from a successfully aging species. *J Comp Physiol B Biochem Syst Environ Physiol*. 2008;178(4):439–45.
- Edrey YH, Hanes M, Pinto M, Mele J, Buffenstein R. Successful aging and sustained good health in the naked mole rat: a long-lived mammalian model for biogerontology and biomedical research. *ILAR J*. 2011;52(1):41–53.
- Tian X, Azpurua J, Hine C, Vaidya A, Myakishev-Rempel M, Ablavaeva J, Mao Z, Nevo E, Gorbunova V, Seluanov A. High-molecular-mass hyaluronan mediates the cancer resistance of the naked mole rat. *Nature*. 2013;499(7458):346–9.
- Tian X, Azpurua J, Ke ZH, Augereau A, Zhang ZDD, Vijj J, Gladyshev VN, Gorbunova V, Seluanov A. INK4 locus of the tumor-resistant rodent, the naked mole rat, expresses a functional p15/p16 hybrid isoform. *Proc Natl Acad Sci USA*. 2015;112(4):1053–8.
- Miyawaki S, Kawamura Y, Oiwa Y, Shimizu A, Hachiya T, Bono H, Koya I, Okada Y, Kimura T, Tsuchiya Y, et al. Tumour resistance in induced pluripotent stem cells derived from naked mole-rats. *Nat Commun*. 2016;7:11471. doi:10.1038/ncomms11471.
- Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. *Annu Rev Biochem*. 2012;81:145–66.
- Zhang H, Chen Z, Wang X, Huang Z, He Z, Chen Y. Long non-coding RNA: a new player in cancer. *J Hematol Oncol*. 2013;6:1.
- Prensner JR, Chinnaiyan AM. The emergence of lincRNAs in cancer biology. *Cancer Discov*. 2011;1(5):391–407.
- Qi P, Du X. The long non-coding RNAs, a new cancer diagnostic and therapeutic gold mine. *Mod Pathol*. 2013;26(2):155–65.
- Cox MP, Peterson DA, Biggs PJ. SolexaQA: at-a-glance quality assessment of illumina second-generation sequencing data. *BMC Bioinform*. 2010;11:485.
- Wang J, Fu L, Koganti PP, Wang L, Hand JM, Ma H, Yao J. Identification and functional prediction of large intergenic noncoding RNAs (lincRNAs) in rainbow trout (*Oncorhynchus mykiss*). *Mar Biotechnol*. 2016;18(2):271–82.
- Yu H, Zhao X, Li Q. Genome-wide identification and characterization of long intergenic noncoding RNAs and their potential association with larval development in the Pacific oyster. *Sci Rep*. 2016;6:20796. doi:10.1038/srep20796.
- Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, Baker JC, Gruetznert F, Kaessmann H. The evolution of lincRNA repertoires and expression patterns in tetrapods. *Nature*. 2014;505(7485):635.
- Hezroni H, Koppstein D, Schwartz MG, Avrutin A, Bartel DP, Ulitsky I. Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep*. 2015;11(7):1110–22.
- Wilusz JE, Sunwoo H, Spector DL. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev*. 2009;23(13):1494–504.
- Orom UA, Derrien T, Berlinger M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytnicki M, Notredame C, Huang Q, et al. Long noncoding RNAs with enhancer-like function in human cells. *Cell*. 2010;143(1):46–58.
- Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res*. 2003;13(9):2178–89.
- Amaral PP, Leonardi T, Han N, Vire E, Gascoigne DK, Arias-Carrasco R, Buscher M, Zhang A, Pluchino S, Maracaja-Coutinho V, et al. Genomic positional conservation identifies topological anchor point (tap)RNAs linked to developmental loci. *Biorxiv*. 2016. doi:10.1101/051052.
- Mohammadin S, Edger PP, Pires JC, Schranz ME. Positionally-conserved but sequence-diverged: identification of long non-coding RNAs in the Brassicaceae and Cleomaceae. *BMC Plant Biol*. 2015;15:217.
- Liang S, Mele J, Wu Y, Buffenstein R, Hornsby PJ. Resistance to experimental tumorigenesis in cells of a long-lived mammal, the naked mole-rat (*Heterocephalus glaber*). *Aging Cell*. 2010;9(4):626–35.
- Ning S, Zhang J, Wang P, Zhi H, Wang J, Liu Y, Gao Y, Guo M, Yue M, Wang L, et al. Lnc2Cancer: a manually curated database of experimentally supported lincRNAs associated with various human cancers. *Nucleic Acids Res*. 2016;44(D1):D980–5.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403–10.
- Zhao Y, Li H, Fang S, Kang Y, Wu W, Hao Y, Li Z, Bu D, Sun N, Zhang MQ, et al. NONCODE 2016: an informative and valuable data source of long non-coding RNAs. *Nucleic Acids Res*. 2016;44(D1):D203–8.
- Yang YM, Fu LM. TSGDB: a database system for tumor suppressor genes. *Bioinformatics*. 2003;19(17):2311–2.
- Kim EB, Fang XD, Fushan AA, Huang ZY, Lobanov AV, Han LJ, Marino SM, Sun XQ, Turanov AA, Yang PC, et al. Genome sequencing reveals insights into physiology and longevity of the naked mole rat. *Nature*. 2011;479(7372):223–7.
- Yu Y, Fuscoe JC, Zhao C, Guo C, Jia MW, Qing T, Bannon DI, Lancashire L, Bao WJ, Du TT, et al. A rat RNA-Seq transcriptomic BodyMap across 11 organs and 4 developmental stages. *Nat Commun*. 2014;5:3230. doi:10.1038/ncomms4230.
- Gorbunova V, Seluanov A, Zhang Z, Gladyshev VN, Vijj J. Comparative genetics of longevity and cancer: insights from long-lived rodents. *Nat Rev Genet*. 2014;15(8):531–40.
- Tian X, Azpurua J, Ke Z, Augereau A, Zhang ZD, Vijj J, Gladyshev VN, Gorbunova V, Seluanov A. INK4 locus of the tumor-resistant rodent, the naked mole rat, expresses a functional p15/p16 hybrid isoform. *Proc Natl Acad Sci USA*. 2015;112(4):1053–8.
- Hadoux J. High-molecular-mass hyaluronan mediates the cancer resistance of the naked mole rat. *Oncologie*. 2013;15(10–11):546–7.
- Shahryari A, Jazi MS, Samaei NM, Mowla SJ. Long non-coding RNA SOX2OT: expression signature, splicing patterns, and emerging roles in pluripotency and tumorigenesis. *Front Genet*. 2015;6:196.
- Shahryari A, Rafiee MR, Fouani Y, Oliaie NA, Samaei NM, Shafiee M, Semnani S, Vasei M, Mowla SJ. Two novel splice variants of SOX2OT, SOX2OT-S1, and SOX2OT-S2 are coexpressed with SOX2 and OCT4 in esophageal squamous cell carcinoma. *Stem Cells*. 2014;32(1):126–34.
- Askarian-Amiri ME, Seyfoddin V, Smart CE, Wang J, Kim JE, Hansji H, Baguley BC, Finlay GJ, Leung EY. Emerging role of long non-coding RNA SOX2OT in SOX2 regulation in breast cancer. *PLOS ONE*. 2014;9(7):e102140.
- Kaushik K, Leonard VE, Shamsudheen KV, Lalwani MK, Jalali S, Patowary A, Joshi A, Scaria V, Sivasubbu S. Dynamic expression of long non-coding RNAs (lincRNAs) in adult zebrafish. *PLOS ONE*. 2013;8(12):e83616.
- Tsoi LC, Iyer MK, Stuart PE, Swindell WR, Gudjonsson JE, Tejasvi T, Sarkar MK, Li BS, Ding J, Voorhees JJ, et al. Analysis of long non-coding RNAs highlights tissue-specific expression patterns and epigenetic profiles in normal and psoriatic skin. *Genome Biol*. 2015;16:1.
- Grammatikakis I, Panda AC, Abdelmohsen K, Gorospe M. Long noncoding RNAs (lincRNAs) and the molecular hallmarks of aging. *Aging-U.S*. 2014;6(12):992–1009.

40. Serrano M, Lin AW, McCurrach ME, Beach D, Lowe SW. Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16(INK4a). *Cell*. 1997;88(5):593–602.
41. Pasmant E, Laurendeau I, Heron D, Vidaud M, Vidaud D, Bieche I. Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF. *Cancer Res*. 2007;67(8):3963–9.
42. Yap KL, Li SD, Munoz-Cabello AM, Raguz S, Zeng L, Mujtaba S, Gil J, Walsh MJ, Zhou MM. Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol Cell*. 2010;38(5):662–74.
43. Montes M, Nielsen MM, Maglieri G, Jacobsen A, Hojfeldt J, Agrawal-Singh S, Hansen K, Helin K, de Werken H, Pedersen JS, et al. The lncRNA MIR31HG regulates p16^{INK4A} expression to modulate senescence. *Nat Commun*. 2015;6:6967. doi:10.1038/ncomms7967.
44. Hung T, Wang YL, Lin MF, Koegel AK, Kotake Y, Grant GD, Horlings HM, Shah N, Umbricht C, Wang P, et al. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat Genet*. 2011;43(7):621–9.
45. Puvvula PK, Desetty RD, Pineau P, Marchio A, Moon A, Dejean A, Bischof O. Long noncoding RNA PANDA and scaffold-attachment-factor SAFA control senescence entry and exit. *Nat Commun*. 2014;5:5323. doi:10.1038/ncomms6323.
46. Dimitrova N, Zamudio JR, Jong RM, Soukup D, Resnick R, Sarma K, Ward AJ, Raj A, Lee JT, Sharp PA, et al. LincRNA-p21 activates p21 in cis to promote polycomb target gene expression and to enforce the G1/S checkpoint. *Mol Cell*. 2014;54(5):777–90.
47. Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, Khalil AM, Zuk O, Amit I, Rabani M, et al. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell*. 2010;142(3):409–19.
48. Guo X, Gao L, Liao Q, Xiao H, Ma X, Yang X, Luo H, Zhao G, Bu D, Jiao F, et al. Long non-coding RNAs function annotation: a global prediction method based on bi-colored networks. *Nucleic Acids Res*. 2013;41(2):967.
49. Liao Q, Liu C, Yuan X, Kang S, Miao R, Xiao H, Zhao G, Luo H, Bu D, Zhao H, et al. Large-scale prediction of long non-coding RNA functions in a coding-non-coding gene co-expression network. *Nucleic Acids Res*. 2011;39(9):3864–78.
50. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009;25(9):1105–11.
51. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks (vol 7, pg 562, 2012). *Nat Protoc*. 2014;9(10):2513.
52. Kong L, Zhang Y, Ye Z-Q, Liu X-Q, Zhao S-Q, Wei L, Gao G. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res*. 2007;35:W345–9.
53. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 2016;44(D1):D279–85.
54. Wang YP, Tang HB, DeBarry JD, Tan X, Li JP, Wang XY, Lee TH, Jin HZ, Marler B, Guo H, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res*. 2012;40(7):e49.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

