10-1-2014

Integrated transcriptome analysis reveals miRNA-mRNA crosstalk in laryngeal squamous cell carcinoma.

Yang Zhang

Yong Chen
Rowan University

Jinhai Yu

Guiming Liu

Zhigang Huang

Recommended Citation

Yang Zhang, Chen Yong, Yu Jinhai, Liu Guiming, Huang Zhigang. (2014). Integrated transcriptome analysis reveals miRNA-mRNA crosstalk in laryngeal squamous cell carcinoma. Genomics 104(4), 249-256.

Genomics 104 (2014) 249–256

Contents lists available at ScienceDirect

Genomics

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

Integrated transcriptome analysis reveals miRNA–mRNA crosstalk in laryngeal squamous cell carcinoma

Yang Zhang [a,1], Yong Chen [b,1], Jinhai Yu [b], Guiming Liu [c,*], Zhigang Huang [a,*]

[a] Department of Otolaryngology Head and Neck Surgery, Beijing Tongren Hospital, Capital Medical University, Key Laboratory of Otolaryngology Head and Neck Surgery, Beijing 100005, China
[b] National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
[c] Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China

ARTICLE INFO

Article history:
Received 30 April 2014
Received in revised form 19 June 2014
Accepted 23 June 2014
Available online 28 June 2014

Keywords:
Laryngeal squamous cell carcinoma
Gene expression
Small RNAs
Data integration

ABSTRACT

Next generation sequencing (NGS) has proven to be a powerful tool in delineating myriads of molecular subtypes of cancer, as well as in revealing accumulation of genomic mutations throughout cancer progression. Whole genome microRNA (miRNA) and mRNA expression profiles were obtained from patients with laryngeal squamous cell carcinoma (LSCC) using deep sequencing technology, and were analyzed by utilizing integrative computational approaches. A large number of protein-coding and non-coding genes were detected to be differentially expressed, indicating a functional switch in LSCC cells. A total of 127 mutated genes were detected to be significantly associated with ectoderm and epidermis development. Eleven miRNAs were found to be differentially expressed, including a potential cancer suppressor miRNA, mir-34c, which was dramatically down-regulated. Integrated analysis of mRNA and miRNA transcriptomes further revealed correlated dynamics among 11 miRNAs and 138 targeted genes, forming a highly dynamical co-regulation network response to LSCC development.

1. Introduction

Laryngeal squamous cell carcinoma (LSCC) is a common malignancy that exhibits a high incidence rate in head and neck squamous cell carcinoma cases (HNSCC). It is highly correlated with aging, smoking, excessive alcohol consumption, poor diet, human papillomavirus (HPV) infection and long-term exposure to specific chemicals, fumes or pollutants. Radiotherapy, surgery, and chemotherapy are used in LSCC therapy. However, the long-term prognosis for intermediate and high-grade cases of LSCC remains low. To improve the chance of effective treatment of LSCC, a more detailed understanding of the molecular mechanisms involved in the development of LSCC is essential, especially for the identification of novel biomarkers that could serve as suitable therapeutic targets.

A multitude of genetic, transcriptomic and epigenetic alterations had been reported previously to be associated with LSCC. For instance, TP53, CDKN2A, PTEN, PIK3CA, HRAS NOTCH1, IRF6, TP63 and FBXW7 have been annotated as genes mutated in HNSCC [1–3]. Recently, the tumor suppressors CTNNA2 and CTTNNA3 were reported to be mutated frequently in LSCC [4]. A microarray approach had been used to scan differential expressed gene that may be involved in LSCC development [5]. One particular class of regulatory molecules, i.e. miRNAs, a class of small non-coding RNA molecules which often function as tumor suppressors or oncogenes, were shown to be aberrantly expressed in many types of human cancers [6–8]. Several miRNAs were identified to be associated with LSCC by either individual miRNA studies or miRNA expression profiling, such as mir-34a [9] and mir-24 [10]. Another phenomenon widely observed in LSCC cells is the dysregulation of epigenetic modifications across regulatory regions. Hypermethylated DNA sequences were found within the promoter regions of miRNA as well as in protein-coding genes [11–13]. Despite these earlier reports on individual gene regulatory alterations, an integrated analysis of any potential modifications in crosstalk between miRNAs and mRNA in LSCC cells remains absent.

Driven by the rapid development of next-generation sequencing (NGS) technologies, the interrogation of cellular properties on a genome-wide scale now offers the possibility of unbiased analysis and exploration of complete sets of specific molecules and pathways. Subsequent integration of these datasets should provide biological insights that would be impossible for isolated data sets. Recently, such integrated approaches were successfully used in investigating the role of genomic alterations in cancer development and metastasis in the context of their biological functions [14,15]. One report describes the successful prediction of several breast cancer subtypes and designed related drug targets by combining signaling network and genomic variations [16]. In cancer system biology, integrating multiple omics data was shown to be useful in identifying both new oncogenes and novel pathways [17]. To develop a comprehensive understanding of the molecular processes involved in the development of LSCC, we utilized an integrated strategy of measuring transcriptome-wide changes in mRNA and

* Corresponding authors.
E-mail addresses: liugm@big.ac.cn (G. Liu), huangzhigangent@163.com (Z. Huang).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.ygeno.2014.06.004
0888-7543/

miRNA levels by combining NGS with systematic analysis. We performed mRNA and miRNA transcriptomes of 12 samples. Integrated analysis described a co-regulated network of 11 miRNAs and their co-regulated 138 genes that are differentially expressed more than 2 fold changes. Our data revealed not only the dynamical landscapes of individual miRNA and mRNA expression, but also more importantly their dynamical crosstalk, providing additional insights into the molecular processes driving LSCC tumorigenesis.

2. Results

2.1. Expression landscape of genes in LSCC samples

In order to comprehensively reveal genes and mechanisms potentially involved in tumorigenesis of LSCC, we used an integrated NGS approach, and examined both mRNA and miRNA transcriptomes of LSCC tissue in comparison to adjacent healthy tissue (Fig. 1A). A total of 10 tumor samples and 2 healthy samples were selected on the basis of their clinical records (Supplementary Table 1). Total RNA was extracted for library construction, and mRNA and small no-coding RNA were sequenced by using an Illumina/Solexa platform. The reads were mapped to the human genome for the identification of genes and noncoding RNAs (ncRNA) for the analysis of differences in expressions patterns.

A total of 1434 genes were found to be significantly up-regulated, and 660 genes were down-regulated (fold change ≥2) in LSCC, representing almost 3.52% of the 59533 gene records presented in Homo sapiens GRCh37.68 (Fig. 1B). Together, the 2094 identified genes cover diverse gene categories of protein-coding genes, pseudogenes, long intergenic non-coding RNAs (lincRNA), antisense, processed_transcript, small nucleolar RNA (snoRNA), miscellaneous RNA (misc_RNA), small nuclear RNA (snRNA), IG_V_gene, miRNA, sense_intronic and other uncategorized genes (Fig. 1C, Supplementary Table 2). The largest gene set represents protein-coding genes, and includes 804 up-regulated as well as 377 down-regulated genes. In comparison, a total of 630 ncRNAs were found to be up-regulated more than 2-fold, and 283 ncRNAs down-regulated more than 2-fold. In conclusion, the large number of genes that exhibit differential expression patterns indicates major changes in the regulation of protein-coding genes as well as ncRNAs in LSCC.

2.2. Dynamic switches of cellular functions in LSCC

To analyze these 1181 differentially expressed genes for their functional categories, we performed a systematic functional enrichment analysis. We first clustered genes and samples by applying a two-way hierarchical clustering analysis [18]. Genes clustered together generally have correlated expression patterns, indicating a similar molecular basis, whereas samples clustered together exhibit either similar phenotypes or a significant functional modules (Fig. 1D). Twelve tumor samples were found to be clustered according to their gene expression profiles, while 2 non-tumor samples were found to cluster together. The 1181 genes were clustered into 11 sub-clusters, where D1 and D2 clusters were down-regulated, and D3 to D11 up-regulated in 10 tumor samples. These 11 clusters were further annotated by GO term analysis using the DAVID database [19]. Each of these clusters was significantly enriched in at least one biological process (p-value < 1.0E-3,

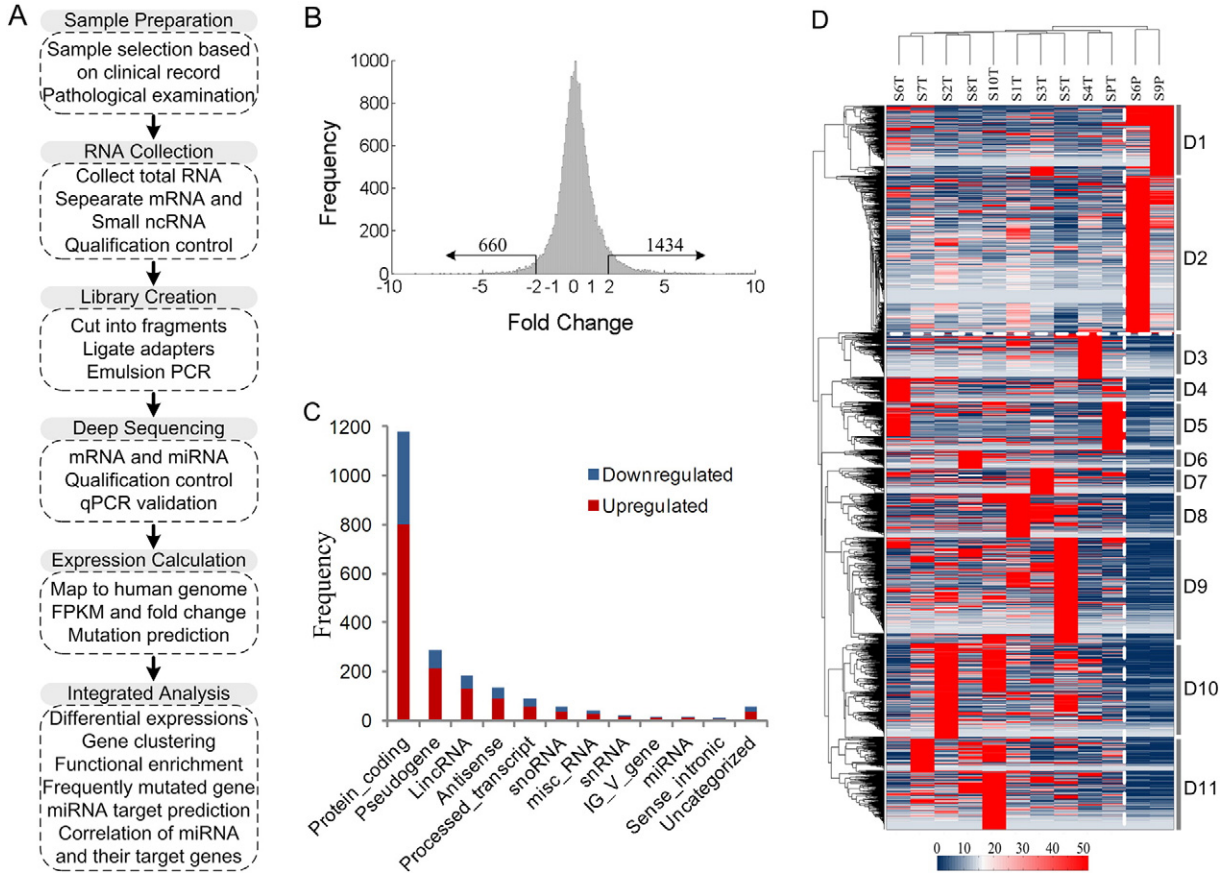

Fig. 1. Summary of mRNA and miRNA transcriptomic study. A) Schematic presentation of the integrated analysis of mRNA and miRNA transcriptomes. B) Distribution of the fold-changes of human genes. 660 down-regulated and 1434 up-regulated genes occupy 1.11% and 2.41% of the 59533 annotated human gene records respectively. C) Genotype distribution of 2094 genes. D) Heatmap of 1181 gene expressions (FPKM) (≥2 fold change). Genes are clustered into 11 groups (D1–D11). The most enriched biological process of each gene group was analyzed by DAVID database. D1: Regulation of hormone levels (1.65E-5). D2: Defense response (2.44E-33). D3: Regulation of growth (7.23E-4). D4: Cell motility (8.81E-4). D5: Epidermis development (1.25E-29). D6: Regulation of transcription (1.48E-4). D7: Collagen fibril organization (1.35E-4). D8: Extracellular structure organization (1.58E-6). D9: Embryonic morphogenesis (5.05E-4). D10: Nucleosome assembly (9.61E-4). D11: Regionalization (4.88E-14).

Fig. 1D). For example, D2 was found to be significantly enriched in defense responses (p-value = 2.44E-33), and D11 was significantly enriched in regionalization (p-value = 4.88E-14).

To systematically interrogate the differentially regulated functional categories involved in LSCC tumor development, we further analyzed the importance those 804 up-regulated and 377 down-regulated genes in specific biological processes. Among the 804 up-regulated genes, we found that epidermis development, epithelium development and ectoderm development were in the top three significant up-regulated categories (p-value < 1.0E-18). With a p-value threshold of 1.0E-5, 30 biological processes were detected, most of which are associated with cell differentiation, biological adhesion and morphogenesis (Fig. 2A1, Supplementary Table 3). Among the 377 down-regulated genes, defense response, inflammatory response and response to wounding were the top three significantly up-regulated biological processes (Fig. 2A2, p-value < 1.0E-22). An analysis of the presence of these genes in particular cellular components showed that the most enriched categories were related to extracellular matrix (up-regulated, Fig. 2B1) and plasma membrane (down-regulated, Fig. 2B2, p-value < 1.0E-5, Supplementary Table 4). We also analyzed the molecular functions that may be dynamically regulated in 804 up-regulated and 377 down-regulated genes. While only the three categories of sequence-specific DNA binding, transcription factor activity and calcium ion binding were found to be enriched for the up-regulated genes, 9 categories were identified for the down-regulated genes (Fig. 2C1–2, Supplementary Table 5). These results reveal a clear dynamic switch in cellular functions in LSCC development compared to healthy tissue.

2.3. Mutations of differentially expressed genes

Our analysis of transcriptomic data included a systematic search for mutations in exons at the nucleotide level. In order to find potential genes that are highly mutated, we mapped the NGS reads to the human genome by using the Bowtie program [20]. A total of 16,169 genes were detected with mutations by calculating the mutated sites per kilobase nucleotides. Although these mutated genes are almost uniformly spread across the human genome, only few mutations were detected within the genome present along the Y chromosome. Interestingly, most of the mitochondrial genome was found to be highly mutated (Figs. 3A,D). Nine mitochondrial genes, i.e. SFN, CTB-63M22.1,

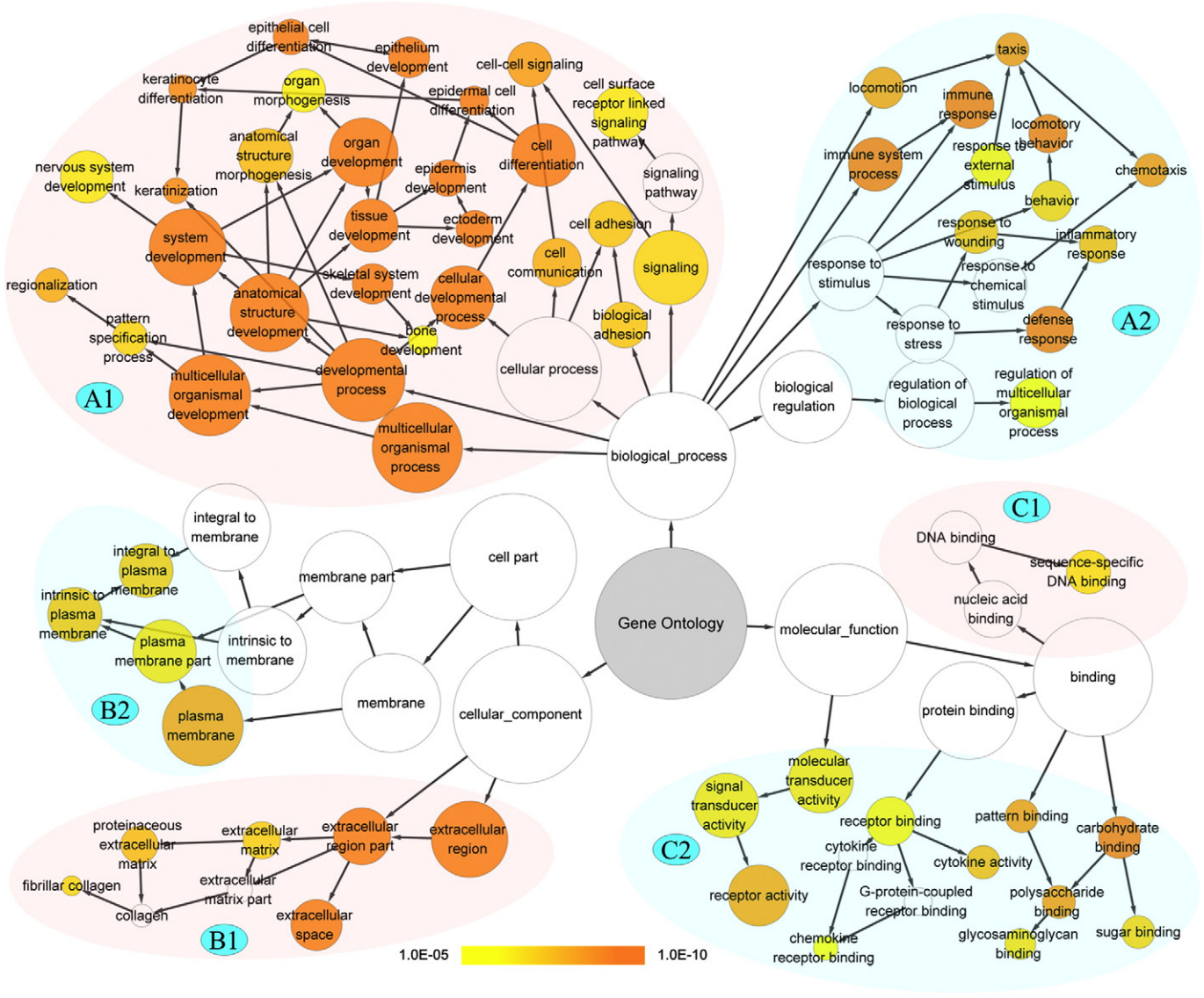

Fig. 2. Functional enrichment of differential expressed genes. Functional enrichment of 804 up-regulated genes and 377 down-regulated genes are analyzed by using the DAVID database with a p-value cutoff of 1.0E-5. The graph was analyzed and printed using BINGO software. Differentially regulated biological processes are shown in A1 and A2. Differentially regulated cellular components are shown in B1 and B2. Differentially regulated molecular functions are shown in C1 and C2. Up-regulated categories are marked in light red, while the down-regulated categories are marked light blue.

HSPB1, KRT5, KRT6A, KRT14, IGHA1, IGLC2 and IGLC3, were found significantly mutated (Mutation ratio > 50).

In order to discover genes that display both a high mutation ratio and significant fold change in their expression, we calculated the product of mutation ratio and fold change for mutated genes exhibiting highly differential expression patterns in tumor compared to healthy samples. A total of 127 interested genes were detected (Fig. 3B; Supplementary Table 6). Our analysis revealed that these genes were enriched for the functional categories of ectoderm development, epidermis development and mitochondrial energy supply (Fig. 3C; p-value < 1.0E-5, Supplementary Table 7). In particular, various mitochondrial genes were found highly mutated and differentially expressed, especially ND4 (MT-ND4) and CO2 (MT-CO2) (Fig. 3D). These findings provide novel evidence for a link between mitochondrial dysfunction and cancer development.

2.4. Expression landscape of miRNAs in LSCC samples

In order to detect differentially expressed miRNAs, we compared miRNA transcriptomes in LSCC tumor and non-tumor tissues. To this end, total miRNA was scanned using the miRDeep2 method [21]. In each sample, hundreds of miRNAs were detected, including many miRNAs not described previously (Fig. 4A). In total, 170 mature miRNAs and 4 novel miRNAs were found to be commonly expressed in all 10 tumor samples and 2 non-tumor samples. In addition, 28 miRNAs were differentially expressed with a change greater than 1.5-fold. Among of them, mir-100, mir-139, mir-34c and mir-375 were found to be down-regulated, while the remaining being up-regulated. Using a more stringent cutoff of 2-fold, 11 miRNAs were identified as differentially expressed, where mir-34c was detected to be down-regulated and others up-regulated (≤ -2 or ≥ 2 fold change, Table 1, Fig. 4B).

Most of these 11 miRNAs have been previously reported to be associated with a number of diseases or tumors. For instance, the three mir-34 family members mir-34a, mir-34b and mir-34c have previously been identified as suppressors in LSCC. They are transcriptional targets of TP53, and function in a positive feedback loop to activate TP53 [22,23]. In our LSCC tumor samples, mir-34c was detected to be greatly down-regulated relative to the samples of healthy tissue. Mir-1301 is a newly identified miRNA family that is involved in regulating more than 8000 genes, as predicted in the microRNA database [24]. Among the genes identified, we confirmed 2532 genes by using TargetScan when applying a more critical filtering procedure [25]. The large number of targeted genes suggests that it is widely involved in the regulation of various cellular functions. Our functional enrichment analysis showed that these 2532 genes are involved in at least 30 biological processes (p-value < 1.0E-5, Supplementary Table 8). So far, only a small number of reports indicate that mir-1301 may be an inhibitor of tumorigenesis in HepG2 cells [26]. The differential expression of mir-1301 identified in our analysis suggests that it may be a universal regulator miRNA and plays important roles in the development of LSCC.

2.5. Crosstalk between miRNAs and their target genes

To uncover in a systematic manner the mechanisms underlying tumor development, an integrative approach is essential to discover the new oncogenes, the pathways and the development of novel anticancer therapies [27,28]. With mRNA and miRNA transcriptome integrated, we can systematically investigate the potential functional correlations among miRNA and their target genes. We first tested if the target genes of the 11 miRNA were differentially expressed in LSCC relative to normal tissue samples. For each of the miRNAs identified, the values for the average fold change (AFC) of their targeted genes were calculated. To test if an AFC is significant compared with randomly calculated values, we employed a bootstrapping method that randomly sampled the same number of genes and calculated it as a random AFC (see Materials and methods). A total of 8 AFCs were observed to be

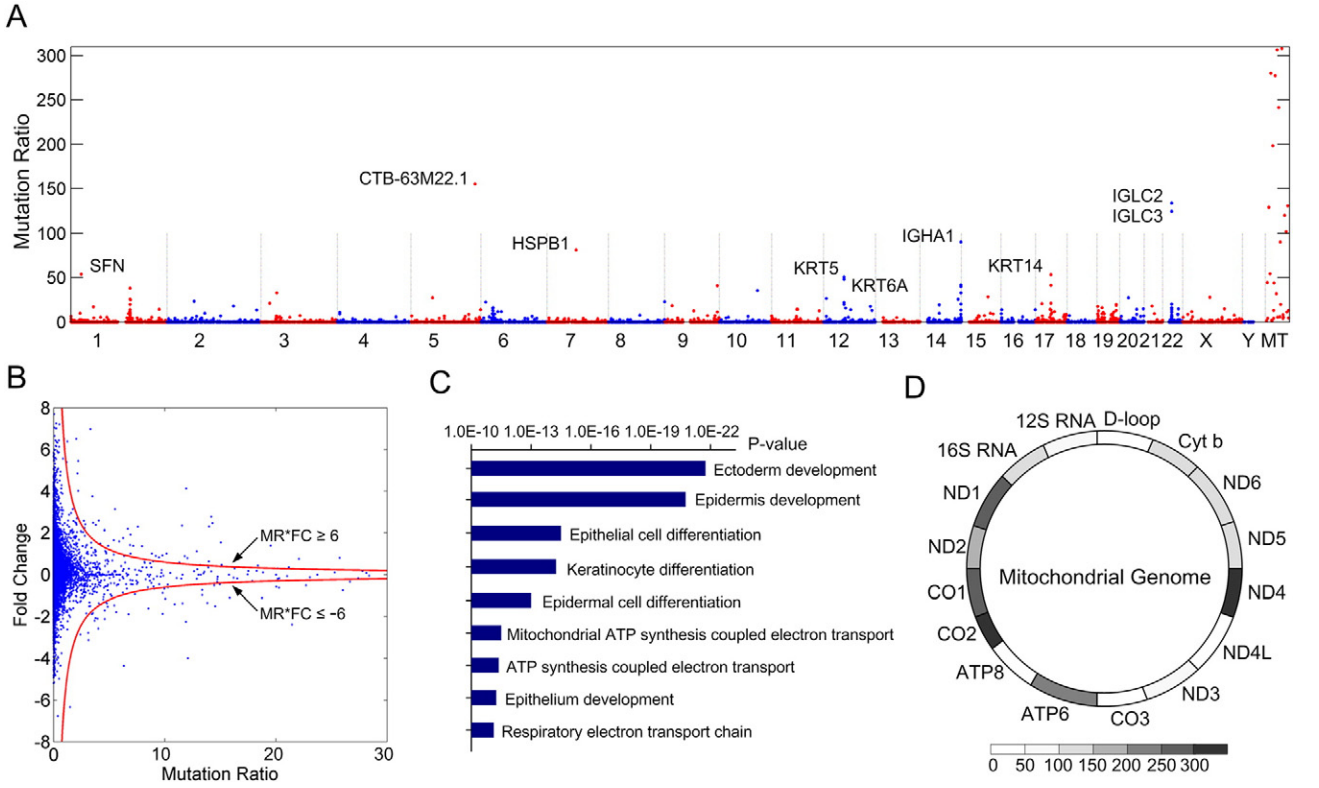

Fig. 3. Functional analysis of frequently mutated genes. A) Mapping of mutated genes on human chromosomes. The genes with a mutation ratio bigger than 50 are annotated (mitochondrial genes are described in D). B) Identifying genes with fold change (FC) and mutation ratio (MR). 127 genes were obtained with a cutoff value of 6. C) Functional enrichment of these genes. The p-values are calculated by using the DAVID database. D) Heatmap of mutated mitochondrial genes.

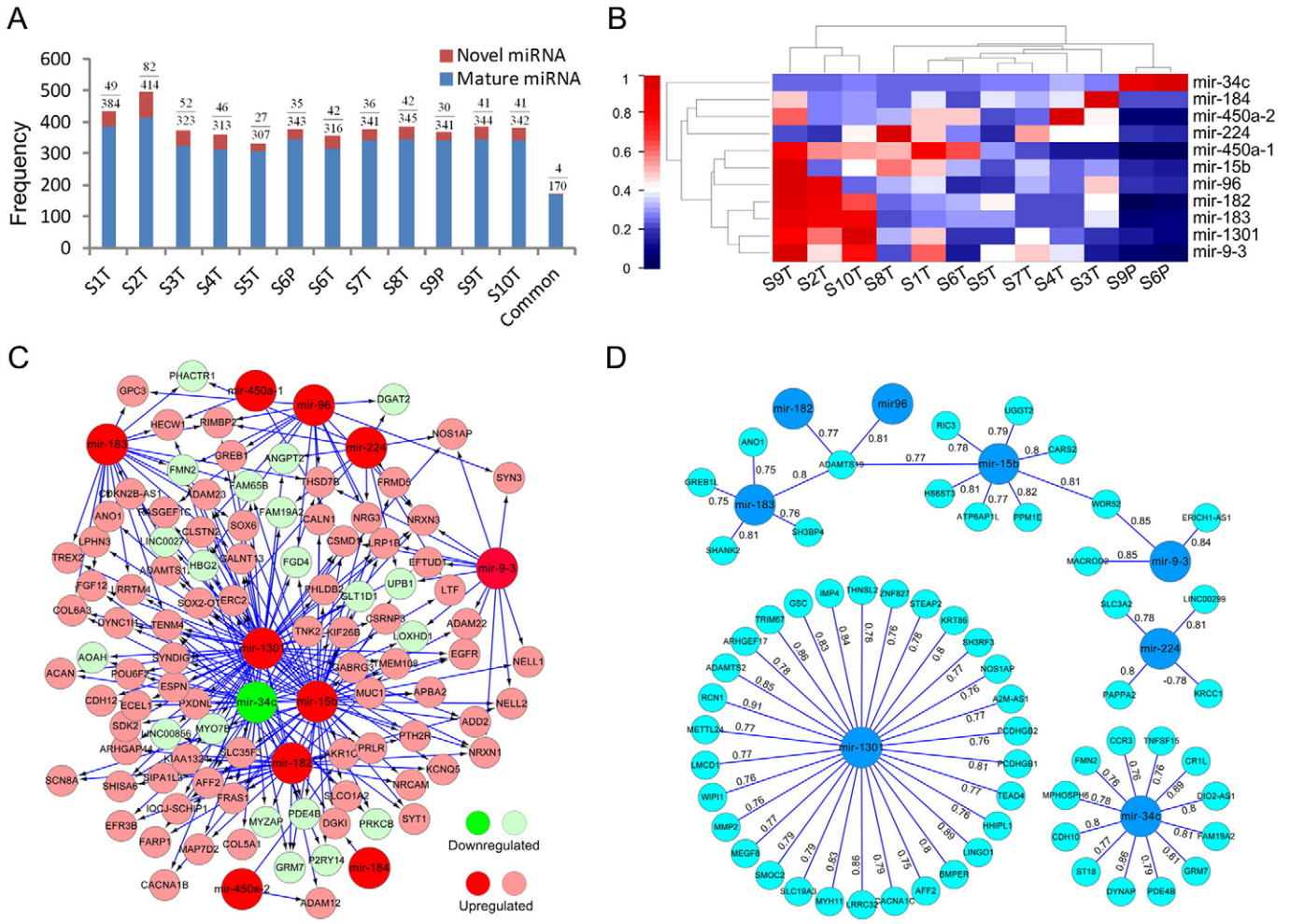

Fig. 4. Expression analysis of miRNA. A) Novel miRNA and miRNA of each sample. Four novel miRNA and 170 mature miRNA were found to be commonly expressed in 12 samples. B) Clustering of 11 significantly expressed miRNAs. The expression values for each miRNA were normalized by its maximal value in all 12 samples. C) The common target genes of the 11 significantly expressed miRNAs. D) 8 miRNA and their highly correlated target genes. Numbers between target genes and their miRNAs indicate the Pearson correlations. A cutoff 0.75 is used to output the highly correlated miRNA-gene pairs (proportion < 1%).

significant (p-value = 1.0E-5, Table 1), suggesting that these targeted genes were indeed differentially expressed in LSCC samples. Second, we tested if the 11 miRNAs and their co-regulated genes were functionally linked as part of a connected network. A total of 1487 genes were identified to be co-regulated by these 11 miRNAs. Functional analysis revealed that they were most enriched in the processes of cell adhesion, cell morphogenesis, as well as neuronal development and differentiation (Supplementary Table 9). When investigating the genes that were differentially expressed for more or less than 2-fold, 339 interactions were found among 138 genes and 11 miRNAs (Fig. 4C). Among the 11 miRNAs, the node degrees of mir-1301, mir-183, mir-96, mir-9-3, mir-182, mir-34c and mir-15b were bigger than 10, while mir-450a-1, mir-450a-2 and mir-184 were all found to be only loosely connected. Among the 138 genes, 118 genes were found to be up-regulated in tumor samples, while 20 genes were down-regulated, suggesting different expression patterns by combinational regulations of miRNAs.

Third, the correlation among miRNAs and their regulated genes were detected by calculating the Pearson correlations, using a cutoff of 0.75 (Fig. 4D). Fifty genes were found to be highly associated with 8 miRNAs. Interestingly, mir-182, mir-96, mir-183, mir-15b and mir-9-3

Table 1
Annotation of 11 significantly expressed miRNAs.

ENSG ID	Name	Mature sequence	FC	NTG	AFC	p-Value
ENSG00000207562	Mir-34c	aggcaguguaguuagcugauugc	−2.38	609	0.71	1.0E-5
ENSG00000221445	Mir-1301	uugcagcugccugggagugacu	2.25	2467	0.65	1.0E-5
ENSG00000207779	Mir-15b	uagcagcacaucaugguuuaca	3.21	2218	0.59	1.0E-5
ENSG00000207990	Mir-182	uuuggcaaugguagaacucacacu	2.62	508	0.58	1.0E-5
ENSG00000207691	Mir-183	uauggcacugguagaauucacu	2.63	560	0.62	1.0E-5
ENSG00000207695	Mir-184	uggacggagaacugauaagg	2.62	22	0.81	2.5E-2
ENSG00000207621	Mir-224	caagucacuaguggguuccguuuag	3.13	289	0.65	1.0E-5
ENSG00000199132	Mir-450a-1	uuuugcgauguguuccuaau	2.39	28	1.07	2.1E-3
ENSG00000207755	Mir-450a-2	uuuugcaauauguuccugaau	2.99	71	0.57	1.1E-2
ENSG00000207819	Mir-9-3	ucuuugguuaucuagcuguauga	3.25	401	0.59	1.0E-5
ENSG00000199158	Mir-96	uuuggcacuagcacauuuuug	2.79	384	0.65	1.0E-5

Note: Fold change (FC). Number of targeted genes (NTG). Average fold change (AFC) of miRNA target genes. The p-value was calculated using a bootstrapping method.

constructed a connected network, indicating that these 5 miRNAs are involved in common biological processes. In this sub-network, two genes, ADAMTS19 and WDR52 were observed as hubs in connecting the 5 miRNAs. ADAMTS19 is a member of ADAMTS (a disintegrin and metalloproteinase containing a thrombospodin motif) gene family that has been recently linked to a variety of physiological and pathological conditions, including arthritis and cancer [29]. Here, its differential expression was observed to be co-regulated by mir-182, mir-183, mir-96 and mir-15b, suggesting that the increase in ADAMTS levels in LSCC is a direct consequence of the dysregulation of a connected miRNA network. Another gene, WDR52, is a connected node of mir-15b and mir-9-3. The precise function of this gene, however, remains unknown. The miRNAs mir-34c, mir-1301 and mir-224 are connected to 12, 30, and 4 genes, respectively. The regulatory connections and the correlation in expression of these 11 miRNAs and their targeted genes all suggest the presence of a convergent miRNA regulatory network that is responsive to the functional switches in the initiation and progression of LSCC.

3. Discussion

In order to investigate the interplay of genes and miRNAs in LSCC, we used mRNA and miRNA transcriptomes to detect not only the genomic mutations, expression dynamics and functional switches, but also in fact a correlated effect of these disease factors during the tumorigenic processes. In total, 1181 coding genes miRNAs were found to be differentially expressed. We presented a systematic analysis for the functional enrichment of these coding genes. We also utilized the acquired data sets to discover and validate base-pair mutations that accumulated in these melanomas. Our analysis revealed a surprisingly high rate of somatic mutations in genes of LSCC samples, when compared with healthy tissues surrounding the cancer samples. Interestingly, we observed especially high mutation ratios within the mitochondrial genome. In details, nine mitochondrial genes (SFN, CTB-63M22.1, HSPB1, KRT5, KRT6A, KRT14, IGHA1, IGLC2 and IGLC3) were found significantly mutated. Importantly, most of these genes identified in our work have been reported previously to be involved in the development of LSCC and other cancers. More specifically, SFN is an adapter protein implicated in the regulation of a large spectrum of both general and specialized signaling pathways. The dysregulation of SFN had been correlated with the development of squamous cell carcinoma [30]. CTB-63M22.1 was first reported to be a pseudogene, but subsequently shown to be differentially expressed in multiple diseases, including prostate cancer [31] and acute myeloid leukemia [32]. Differential expression of CTB-63M22.1 was also observed in our study. HSPB1 encodes a heat shock protein that is induced by environmental stress as well as upon developmental changes. HSPB1 protein interacts with TP53, and is involved in stress resistance and actin organization. Defects in this gene are known to cause the onset of carcinogenesis in squamous cells [33,34]. KRT5, KRT6A and KRT14 are members of the keratin protein family, which are co-expressed during the differentiation process of simple and stratified epithelial tissues. Mutations in these keratin proteins have been associated with skin diseases [35,36]. IGHA1, IGLC2 and IGLC3 are members of the immunoglobulin protein family, and are involved in complement activation and regulation of immune response [37]. The high mutation levels of these three immunoglobulin proteins suggest that the dysfunction of the immune response is a likely factor in contributing to LSCC development. In summary, these genomic mutations of mitochondrial genes correspond well with the observed changes in expression dynamics, lending support to the well-established fact that energy production within solid tumors is transformed from mitochondrial oxidative phosphorylation to aerobic glycolysis [38,39]. Thus, the somatic mutations in mtDNA observed in our study provide further confirmation of mitochondrial alterations in cancer.

MiRNAs have been established as key regulators in both normal and pathological cellular processes, which exert both divergent and convergent control of gene expression through miRNA regulatory networks [40]. Cooperative targeting by multiple miRNAs ensures more complex and robust control of gene expression compared to the level of control by a single miRNA. Here, we analyzed a convergent miRNA regulatory network, including 11 miRNAs and 138 genes which were all found to be differentially expressed in LSCC cells. Thus, our results provide supporting evidence of the existence of a p53-miRNA network, where mir-34c is directly regulated by p53 [23,41]. Previously, dynamic miRNA expression in the plasma of LSCC patient was investigated, and 17 miRNAs were found to be up-regulated, and 9 down-regulated [6]. However, none of the miRNAs were identified as one of our 11 significantly expressed miRNAs. A possible reason of this difference may be these miRNAs are tissue-specific. Compared the miRNAs detected in plasma, our analysis present a more directed analysis of miRNA dysfunctions in LSCC cells.

With increasing amounts of cancer-related omics data available, it is essential for a better understanding of cancer to integrate results from across diverse experimental approaches. Here, we presented an integrated analysis of the mRNA–miRNA transcriptome. However, additional omics data should be included once they become available for LSCC samples, in particular large-scale epigenetic and genomics data sets. Integrating more of these omics data should contribute to the identification of effective drug combinations and molecularly targeted therapeutics. Thus, our research presented here not only interrogates transcriptome data sources, but also includes a systematic and integrated analysis that should be useful for further LSCC research.

4. Materials and methods

4.1. Sample preparation and RNA extraction

LSCC tumor tissues were obtained from the tissue bank of Tongren Hospital, Beijing, China. This study was approved by the Ethics Committee of Beijing Tongren Hospital. A total of twelve samples consisting of ten tumor tissues and two non-tumor tissues were selected from ten LSCC patients. Of these patients, 6 were male and 4 were female. Among these patients, 8 were newly diagnosed and 2 were diagnosed with re-ocurring LSCC tumors (Supplementary Table 1). TRIzol Reagent (Invitroen) was used to isolate total RNA for RNA sequencing following manufacturer's instructions. The mRNAs and small RNAs were separated for independent sequencing.

4.2. mRNA sequencing

The quality of mRNA samples was checked using an Agilent 2100 Bioanalyzer total RNA Chip (Agilent) prior to sequencing. We used the Illumina platform for analyzing transcriptomes employing a 100-bp paired-end library according to manufacturer's instructions (Illumina). Libraries were constructed following the Illumina Paired-End Sequencing Library Preparation Protocol. Library quality and concentration were determined using an Agilent 2100 Bioanalyzer (Agilent). Each sample was paired-end sequenced with the Illumina HiSeq 2000 using HiSeq Sequencing kits.

RNA-Seq reads from each mRNA sample were mapped against the human genome by using Bowtie with the 'best' strata option [20]. Reads from each sample were mapped with less than 2 mismatches. To analyze differential expression, FPKM values (fragments per kilobase of transcript per million mapped reads) were calculated by using Cufflinks [42]. The fold change (FC) of a gene is defined as the log2 transformed fold change of the averaged FPKMs that calculated from 10 tumor tissues and 2 non-tumor tissues, $FC = -\log_2 \sum_{i=1}^{10} 0.1 \cdot FPKM(i) / \sum_{j=1}^{2} 0.5 \cdot FPKM(j)$. Deletions and additions were called by using the SAMtools package [43].

4.3. Small RNA sequencing

Libraries of Small RNA cDNA were created using the Illumina Small RNA Sample Preparation Alternative v1.5 Protocol. These cDNA libraries were then sequenced on the Illumina Genome Analyzer IIX with 35 base pair reads according to manufacturer's instructions. Each library was run on a single lane in the flow cell with 36 cycles using Illumina version 5 chemistry. Reads were mapped on known miRNAs using the quantifier script from the miRDeep2 package [21]. Known miRNA sequences were downloaded from miRBase release 18, November 2013 [44]. To discover novel miRNAs, the miRDeep2 algorithm was used with default settings and filtered reads by size 17 nt. Among the new miRNAs discovered using this approach, only high-confidence miRNAs were considered that contained both mature and star sequences complementary with 2-nt 30 overhang detected in multiple samples. Secondary RNA structures of the precursors were directly obtained using miRDeep2, which uses RNAfold as its default setting [45]. The targeted genes of miRNA are predicted by using TargetScan method [25].

4.4. Bioinformatics analysis

Functional enrichment analysis was performed and printed by using DAVID database [19] and BINGO software [46]. Expression matrixes were clustered by using a two-way hierarchical clustering analysis [18]. After calculating the number of mutated sites (insertion and deletion) in gene body, the gene mutation ratio (MR) is defined as MR = (# mutated site) · 1000/gene length. The product of mutation ratio and fold change is then calculated to find mutated genes with highly differential expressions in tumor and non-tumor samples. The correlation networks of miRNA and their targeted genes are performed on statistical and functional analyses. First, we used a Pearson correlation to estimate the expression correlation between a miRNA and its targeted genes (mRNAs). For every miRNA or mRNA, a 12-dimension vector was constructed by using its expression levels for each sample. The Pearson correlation is then calculated as $r(X,Y)=\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})/\left(\sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2}\cdot\sqrt{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}\right)$, where $n=12$ and X and Y are the expression vectors of the miRNA and mRNA. $\overline{X}$ and $\overline{Y}$ are the mean of X and Y. The miRNA co-regulated network is constructed by calculating the miRNA and their targeted genes whose fold changes are bigger than 2. The network was drawn by using Cytoscape software [47].

For every miRNA, the average fold change (AFC) value of its targeted genes was calculated as $\mathrm{AFC}=\sum_{i=1}^{N}|\mathrm{FC}(i)|/N$, where $\mathrm{FC}(i)$ is the fold change value of i-th gene, and N is the number of targeted genes of the miRNA. For this AFC, a bootstrapping method was used to test if its change is significant compared with randomly calculated values. The method was implemented by randomly selecting N genes from human genome, and then calculating the AFC values for these N genes. To obtain a bootstrapping distribution of AFC values, the procedure was repeated 100,000 times.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ygeno.2014.06.004.

Acknowledgments

The authors would like to thank Dr. Torsten Juelich for the critical reading of this manuscript and useful suggestions. This work was supported by the grants from the National Natural Science Foundation of China (81302374, 81241084 and 61273228), Beijing Natural Science Foundation (7121005 and 5122016), and Beijing Nova Program (xx2013043).

References

[1] N. Agrawal, M.J. Frederick, C.R. Pickering, C. Bettegowda, K. Chang, R.J. Li, C. Fakhry, T.X. Xie, J. Zhang, J. Wang, N. Zhang, A.K. El-Naggar, S.A. Jasser, J.N. Weinstein, L. Trevino, J.A. Drummond, D.M. Muzny, Y. Wu, L.D. Wood, R.H. Hruban, W.H. Westra, W.M. Koch, J.A. Califano, R.A. Gibbs, D. Sidransky, B. Vogelstein, V.E. Velculescu, N. Papadopoulos, D.A. Wheeler, K.W. Kinzler, J.N. Myers, Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1, Science 333 (2011) 1154–1157.
[2] N. Stransky, A.M. Egloff, A.D. Tward, A.D. Kostic, K. Cibulskis, A. Sivachenko, G.V. Kryukov, M.S. Lawrence, C. Sougnez, A. McKenna, E. Shefler, A.H. Ramos, P. Stojanov, S.L. Carter, D. Voet, M.L. Cortes, D. Auclair, M.F. Berger, G. Saksena, C. Guiducci, R.C. Onofrio, M. Parkin, M. Romkes, J.L. Weissfeld, R.R. Seethala, L. Wang, C. Rangel-Escareno, J.C. Fernandez-Lopez, A. Hidalgo-Miranda, J. Melendez-Zajgla, W. Winckler, K. Ardlie, S.B. Gabriel, M. Meyerson, E.S. Lander, G. Getz, T.R. Golub, L.A. Garraway, J.R. Grandis, The mutational landscape of head and neck squamous cell carcinoma, Science 333 (2011) 1157–1160.
[3] C.R. Leemans, B.J. Braakhuis, R.H. Brakenhoff, The molecular biology of head and neck cancer, Nat. Rev. Cancer 11 (2011) 9–22.
[4] M. Fanjul-Fernandez, V. Quesada, R. Cabanillas, J. Cadinanos, T. Fontanil, A. Obaya, A. J. Ramsay, J.L. Llorente, A. Astudillo, S. Cal, C. Lopez-Otin, Cell–cell adhesion genes CTNNA2 and CTNNA3 are tumour suppressors frequently mutated in laryngeal carcinomas, Nat. Commun. 4 (2013) 2531.
[5] L.J. Ma, W. Li, X. Zhang, D.H. Huang, H. Zhang, J.Y. Xiao, Y.Q. Tian, Differential gene expression profiling of laryngeal squamous cell carcinoma by laser capture microdissection and complementary DNA microarrays, Arch. Med. Res. 40 (2009) 114–123.
[6] L. Ayaz, A. Gorur, H.Y. Yaroglu, C. Ozcan, L. Tamer, Differential expression of microRNAs in plasma of patients with laryngeal squamous cell carcinoma: potential early-detection markers for laryngeal squamous cell carcinoma, J. Cancer Res. Clin. Oncol. 139 (2013) 1499–1506.
[7] M. Esteller, Non-coding RNAs in human disease, Nat. Rev. Genet. 12 (2011) 861–874.
[8] C.M. Croce, G.A. Calin, miRNAs, cancer, and stem cell division, Cell 122 (2005) 6–7.
[9] Z. Shen, G. Zhan, D. Ye, Y. Ren, L. Cheng, Z. Wu, J. Guo, MicroRNA-34a affects the occurrence of laryngeal squamous cell carcinoma by targeting the antiapoptotic gene survivin, Med. Oncol. 29 (2012) 2473–2480.
[10] Y. Guo, W. Fu, H. Chen, C. Shang, M. Zhong, miR-24 functions as a tumor suppressor in Hep2 laryngeal carcinoma cells partly through down-regulation of the S100A8 protein, Oncol. Rep. 27 (2012) 1097–1103.
[11] S.M. Langevin, R.A. Stone, C.H. Bunker, M.A. Lyons-Weiler, W.A. LaFramboise, L. Kelly, R.R. Seethala, J.R. Grandis, R.W. Sobol, E. Taioli, MicroRNA-137 promoter methylation is associated with poorer overall survival in patients with squamous cell carcinoma of the head and neck, Cancer 117 (2011) 1454–1462.
[12] S. Pierini, S.H. Jordanov, A.V. Mitkova, I.J. Chalakov, M.B. Melnicharov, V. Kunev, V.I. Mitev, R.P. Kaneva, T.E. Goranova, Promoter hypermethylation of CDKN2A, MGMT, MLH1, and DAPK genes in laryngeal squamous cell carcinoma and their associations with clinical profiles of the patients, Head Neck (2013), http://dx.doi.org/10.1002/hed.23413.
[13] T.S. Wong, W. Gao, Z.H. Li, J.Y. Chan, W.K. Ho, Epigenetic dysregulation in laryngeal squamous cell carcinoma, J. Oncol. 2012 (2012) 739461.
[14] E. Wang, Understanding genomic alterations in cancer genomes using an integrative network approach, Cancer Lett. 340 (2013) 261–269.
[15] Y. Chen, J. Hao, W. Jiang, T. He, X. Zhang, T. Jiang, R. Jiang, Identifying potential cancer driver genes by genomic data integration, Sci. Rep. 3 (2013) 3538.
[16] N. Zaman, L. Li, M.L. Jaramillo, Z. Sun, C. Tibiche, M. Banville, C. Collins, M. Trifiro, M. Paliouras, A. Nantel, M. O'Connor-McCourt, E. Wang, Signaling network assessment of mutations and copy number variations predict breast cancer subtype-specific drug targets, Cell Rep. 5 (2013) 216–223.
[17] K. Wang, S.J. Diskin, H. Zhang, E.F. Attiyeh, C. Winter, C. Hou, R.W. Schnepp, M. Diamond, K. Bosse, P.A. Mayes, J. Glessner, C. Kim, E. Frackelton, M. Garris, Q. Wang, W. Glaberson, R. Chiavacci, L. Nguyen, J. Jagannathan, N. Saeki, H. Sasaki, S. F. Grant, A. Iolascon, Y.P. Mosse, K.A. Cole, H. Li, M. Devoto, P.W. McGrady, W.B. London, M. Capasso, N. Rahman, H. Hakonarson, J.M. Maris, Integrative genomics identifies LMO1 as a neuroblastoma oncogene, Nature 469 (2011) 216–220.
[18] M.B. Eisen, P.T. Spellman, P.O. Brown, D. Botstein, Cluster analysis and display of genome-wide expression patterns, Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 14863–14868.
[19] G. Dennis Jr., B.T. Sherman, D.A. Hosack, J. Yang, W. Gao, H.C. Lane, R.A. Lempicki, DAVID: database for annotation, visualization, and integrated discovery, Genome Biol. 4 (2003) P3.
[20] B. Langmead, C. Trapnell, M. Pop, S.L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, Genome Biol. 10 (2009) R25.
[21] M.R. Friedlander, W. Chen, C. Adamidi, J. Maaskola, R. Einspanier, S. Knespel, N. Rajewsky, Discovering microRNAs from deep sequencing data using miRDeep, Nat. Biotechnol. 26 (2008) 407–415.
[22] D.C. Corney, A. Flesken-Nikitin, A.K. Godwin, W. Wang, A.Y. Nikitin, MicroRNA-34b and MicroRNA-34c are targets of p53 and cooperate in control of cell proliferation and adhesion-independent growth, Cancer Res. 67 (2007) 8433–8438.
[23] H. Hermeking, MicroRNAs in the p53 network: micromanagement of tumour suppression, Nat. Rev. Cancer 12 (2012) 613–626.
[24] D. Betel, M. Wilson, A. Gabow, D.S. Marks, C. Sander, The microRNA.org resource: targets and expression, Nucleic Acids Res. 36 (2008) D149–D153.
[25] R.C. Friedman, K.K. Farh, C.B. Burge, D.P. Bartel, Most mammalian mRNAs are conserved targets of microRNAs, Genome Res. 19 (2009) 92–105.
[26] L. Fang, N. Yang, J. Ma, Y. Fu, G.S. Yang, microRNA-1301-mediated inhibition of tumorigenesis, Oncol. Rep. 27 (2012) 929–934.

[27] D.R. Rhodes, A.M. Chinnaiyan, Integrative analysis of the cancer transcriptome, Nat. Genet. 37 (2005) S31–S37 (Suppl.).
[28] A.L. Barabasi, N. Gulbahce, J. Loscalzo, Network medicine: a network-based approach to human disease, Nat. Rev. Genet. 12 (2011) 56–68.
[29] E.A. Knauff, L. Franke, M.A. van Es, L.H. van den Berg, Y.T. van der Schouw, J.S. Laven, C.B. Lambalk, A. Hoek, A.J. Goverde, S. Christin-Maitre, A.J. Hsueh, C. Wijmenga, B.C. Fauser, Genome-wide association study in premature ovarian failure patients suggests ADAMTS19 as a possible candidate gene, Hum. Reprod. 24 (2009) 2372–2378.
[30] Y. Qi, J.F. Chiu, L. Wang, D.L. Kwong, Q.Y. He, Comparative proteomic analysis of esophageal squamous cell carcinoma, Proteomics 5 (2005) 2960–2971.
[31] U.R. Chandran, C. Ma, R. Dhir, M. Bisceglia, M. Lyons-Weiler, W. Liang, G. Michalopoulos, M. Becich, F.A. Monzon, Gene expression profiles of prostate cancer reveal involvement of multiple molecular pathways in the metastatic process, BMC Cancer 7 (2007) 64.
[32] Y. Oshima, M. Ueda, Y. Yamashita, Y.L. Choi, J. Ota, S. Ueno, R. Ohki, K. Koinuma, T. Wada, K. Ozawa, A. Fujimura, H. Mano, DNA microarray analysis of hematopoietic stem cell-like fractions from individuals with the M2 subtype of acute myeloid leukemia, Leukemia 17 (2003) 1990–1997.
[33] H. Mese, A. Sasaki, S. Nakayama, N. Yoshioka, Y. Yoshihama, K. Kishimoto, T. Matsumura, Prognostic significance of heat shock protein 27 (HSP27) in patients with oral squamous cell carcinoma, Oncol. Rep. 9 (2002) 341–344.
[34] F. Trautinger, C. Kokesch, I. Herbacek, R.M. Knobler, I. Kindas-Mugge, Overexpression of the small heat shock protein, hsp27, confers resistance to hyperthermia, but not to oxidative stress and UV-induced cell death, in a stably transfected squamous cell carcinoma cell line, J. Photochem. Photobiol. B Biol. 39 (1997) 90–95.
[35] C.B. Sorensen, A.S. Ladekjaer-Mikkelsen, B.S. Andresen, F. Brandrup, N.K. Veien, S.K. Buus, I. Anton-Lamprecht, T.A. Kruse, P.K. Jensen, H. Eiberg, L. Bolund, N. Gregersen, Identification of novel and known mutations in the genes for keratin 5 and 14 in Danish patients with epidermolysis bullosa simplex: correlation between genotype and phenotype, J. Invest. Dermatol. 112 (1999) 184–190.
[36] B. Jerabkova, J. Marek, H. Buckova, L. Kopeckova, K. Vesely, J. Valickova, J. Fajkus, L. Fajkusova, Keratin mutations in patients with epidermolysis bullosa simplex: correlations between phenotype severity and disturbance of intermediate filament molecular structure, Br. J. Dermatol. 162 (2010) 1004–1013.
[37] C.T. Watson, F. Breden, The immunoglobulin heavy chain locus: genetic variation, missing data, and implications for human disease, Genes Immun. 13 (2012) 363–373.
[38] D.C. Wallace, Mitochondria and cancer, Nat. Rev. Cancer 12 (2012) 685–698.
[39] O. Warburg, On the origin of cancer cells, Science 123 (1956) 309–314.
[40] N. Pencheva, S.F. Tavazoie, Control of metastatic progression by microRNA regulatory networks, Nat. Cell Biol. 15 (2013) 546–554.
[41] Y.H. Cha, N.H. Kim, C. Park, I. Lee, H.S. Kim, J.I. Yook, MiRNA-34 intrinsically links p53 tumor suppressor and Wnt signaling, Cell Cycle 11 (2012) 1273–1281.
[42] C. Trapnell, B.A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M.J. van Baren, S.L. Salzberg, B.J. Wold, L. Pachter, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation, Nat. Biotechnol. 28 (2010) 511–515.
[43] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The sequence alignment/map format and SAMtools, Bioinformatics 25 (2009) 2078–2079.
[44] A. Kozomara, S. Griffiths-Jones, miRBase: integrating microRNA annotation and deep-sequencing data, Nucleic Acids Res. 39 (2011) D152–D157.
[45] I.L. Hofacker, P.F. Stadler, Memory efficient folding algorithms for circular RNA secondary structures, Bioinformatics 22 (2006) 1172–1176.
[46] S. Maere, K. Heymans, M. Kuiper, BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks, Bioinformatics 21 (2005) 3448–3449.
[47] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Res. 13 (2003) 2498–2504.