Published in final edited form as:

Precis Med. 2015 ; 1: . doi:10.14800/pm.1036.

The mirror RNA expression pattern in human tissues

Tameka N. Bythwood¹, Wei Xu^{1,3}, Wenzhi Li^{1,3}, Weinian Rao^{1,2}, Qiling Li³, Xue Xue³, Jendai Richards¹, Li Ma^{1,2,3}, and Qing Song^{1,2,3}

¹Cardiovascular Research Institute and Department of Medicine, Morehouse School of Medicine, Atlanta, Georgia 30310, USA

²4DGenome Inc, Atlanta, Georgia 30033, USA

³Center for Big Data Science, First Affiliated Hospital of Medical School, Xi'an Jiaotong University, Xi'an, Shaanxi 710061, China

Abstract

It has been realized in recent years that non-coding RNAs are playing important roles in genome functions and human diseases. Here we developed a new technology and observed a new pattern of gene expression. We observed that over 72% of RNAs in human genome are expressed in forward-reverse pairs, just like mirror images of each other between forward expression and reverse expression; the overview showed that it cannot be simply described as transcript overlapping, so we designated it as mirror expression. Furthermore, we found that the mirror expression is gene-specific and tissue-specific, and less common in the proximal promoter regions. The size of the shadows varies between different genes, different tissues and different classes. The shadow expression is most significant in the Alu element, it was also observed among L1, Simple Repeats and LTR elements, but rare in other repeats such as low-complexity, LINE/L2, DNA and MIRs. Although there is no evidence yet about the relationship of this mirror pattern and double-strand RNA (dsRNA), this new striking pattern provides a new clue and a new direction to unveil the role of RNAs in the genome functions and diseases.

Keywords

RNA; antisense; Alu

Introduction

Many non-coding transcripts overlap with known genes in antisense orientation, which is known as natural antisense transcripts (NATs)^[1–5]. Cis-NATs have been well-characterized among prokaryotes^[6]. It has been thought that NATs may play important regulatory roles in

Correspondence: Qing Song or Li Ma, qsong@msm.edu or lma@msm.edu.

Conflict of Interests

The authors declare that they have no competing interests.

Author Contributions

T.N.B. and L.M. performed the experiments. W.R. coded visualization program. T.N.B., W.X., J.R., W.L., L.M., and Q.S. performed bioinformatic analyses. L.M. and Q.S. designed the study. T.N.B. and Q.S. wrote the manuscript.

humans^[7], but relatively little is known about their roles in eukaryotes^[8]. Cis-NATs may function through transcriptional regulation, alternative splicing, RNA duplex formation, double-strand RNA (dsRNA), RNA editing and RNBA interference (RNAi), DNA methylation, genomic imprinting, and X-chromosome inactivation^[9–11]. The molecular mechanisms behind the regulatory roles of cis-NATs are not well understood.

Whole-genome NAT discovery was achieved by in silico methods that can detect the sequence overlapping among expressed sequence tags (ESTs)^[12]. Strand-specific RNAseq has been used to added new discoveries in the NATs collection recently^[13–21]. These studies usually used a cut-off value (generally ~20 nucleotides in most of the studies) to define the sequence overlapping and thus the NATs, and especially paid attention on the overlapping patterns, such as head-to-head, head-to-tail, and tail-to-tail overlapping^[7]. The “mirror” phenomenon has not been reported. In this study, we developed a customized designed method to identify the sense and antisense transcripts and found the “mirror” pattern, in which a large number of expression signals are paired as mirrors or mirror images of each other.

In addition, most of current reports on cis-NATs were focused on the NATs that overlap with known genes on their introns, exons, promoters, enhancers, untranslated regions (UTRs) and flanking sequences^[7], NATs have not been well studied on the repetitive elements. Mirror expression on the Alu elements is one of the findings in our study.

It was estimated that up to 20% of transcripts may contribute to sense-antisense pairs^[1], recently this number was found to be higher (38% of annotated transcripts showed consistent antisense expression in cancers)^[13]. By including the information on the repeats, we found that NATs are significantly much more widespread than current estimates in human tissues.

Experimental Procedures

Samples and RNA Preparation

Human RNAs from artery, vein, bone marrow, lymph node, small intestine, adrenal glands, pancreas, prostate, thymus, adipose, brain cortex, heart, skin, skeletal muscle, prostate (50 g) were purchased from Clontech and US Biological^[22, 23]. THP-1 cells were grown in RPMI-1640 supplemented with 10% fetal bovine serum (FBS), 2mM L-glutamine, and 1% penicillin-streptomycin. Total RNA was isolated using RNAeasy Mini kit (Qiagen). To further remove the potential DNA contamination in the RNA samples, we incubated the RNA samples at 37°C in a mixture of 1× reaction buffer, DNase I (2U/μl) and DEPC-treated water for approximately 10 minutes before adding 0.5 M EDTA followed by heat inactivation of DNase I at 70°C for 10 minutes. We created a pooled sample by pooling RNAs from bone marrow, lymph node, small intestine, adrenal glands, pancreas, thymus and adipose tissues as a control.

We implemented several controls in this study. To monitor the reproductivity of the results in our tiling arrays, we duplicated the brain cortex RNA samples. To monitor the potential

DNA contaminations, we digested the prostate RNA samples with DNase I (New England BioLabs).

Custom Tiling Array Design and Production

We designed a custom tiling Array using the Agilent eArray platform (Fig. 2). Genomic sequences were downloaded from the UCSC Genome Browser (hg19). These sequences were used to create 60-mer overlapping, end-to-end oligonucleotide tiling array probes (–30 bp separation). Probes complementary to both sense and antisense strands were generated to span the loci of interest on chromosomes 5, 7 and 9. Probes were printed using Agilent's 4 ×44K and 8×15K array formats that include either 4 or 8 arrays per slide. Each array included 581 Agilent negative control probes from randomly selected sequences. Probe sets that overlap with the largest number of distinct mRNA transcripts are considered constitutive and thus most informative for calculating gene expression values.

Microarray Experiments

RNA quality was assessed using an Agilent 2100 Bioanalyzer. Samples with an RNA integrity number of at least 7.0 were further processed. Microarrays were performed in the Microarray Core Facility at Morehouse School of Medicine using the One-Color protocol. Briefly, 0.5 µg of total RNA was labeled using the Agilent QuickAmp Labeling Kit, and then cyanine3 (Cy3) cRNA was purified with RNeasy Mini Spin columns (Qiagen) and quantified with the NanoDrop ND1000 UV-VIS spectrophotometer. Approximately 600 ng of purified Cy3- labeled cRNA was incubated at 60°C for 30 minutes with 10× blocking agent, nuclease-free water and 25× fragmentation buffer to fragment RNA. Then 2× Hi-RPM Hybridization Buffer was added to stop the fragmentation reaction. The fragmented RNA was hybridized onto custom 4×44K or 8×15K microarray slides at 65°C for 17 hours. After hybridization, array slides were washed and then scanned to extract probe features and measure expression signals using Agilent Feature Extraction (FE) software version 9 (Agilent Technologies). Numerical processed signal values (gProcessedSignal) of the Agilent feature extraction file were obtained as representative expression levels for each probe.

Data Integration

Annotations of known gene exons and repetitive elements were downloaded in BED format from the UCSC Genome Table Browser (hg19) from UCSC Genes and RepeatMasker tracks. We also downloaded in BED format genomic coordinates for introns and upstream positions at 1 kb, 2 kb and 10 kb from the transcription start site (TSS). Microarray expression files were integrated with the BED files of known exon, intron and repetitive elements to form one large integrated dataset using several PROC SQL scripts (SAS Enterprise 9.2 Structured Query Language (SQL)). Then we used a threshold of ≤ 50 as a cutoff to filter away the noise expressions. The remaining expression probes were organized by chromosome positions and orientations^[24].

Direct Visualization of Expression Signals

To directly visualize the expressions, we developed a software program GeneFinder. The input includes the logarithmically transformed expression values and orientations of the tiling probes and various datasets downloaded from the UCSC genome browser, such as known genes and repetitive elements. The output displays the logarithmically transformed expression values and orientations along the chromosomal positions. The expression values on the probes of the Watson orientation were displayed above the genomic line in blue color, and the expression values on the probes of the Crick orientation were displayed below the genomic line in the orange color.

Analysis of Mirror Transcription

Unidirectional expression is defined as having an expression value >50 in either the Watson or Crick orientation. Mirror expression is defined as having expression values >50 in both orientations. The direction of expression for a probe pair was assigned as 0, 1, or 2, where 0 signifies no expression. We calculated the relative ratio within the mirror expression pairs using the following equation:

$$= \text{MIN} \left(\text{Relative Ratio}, \frac{1}{\text{Relative Ratio}} \right)$$

Next, we calculated the overall number, percentage and mean expression of unidirectional and bidirectional expressions in each of the following categories: (1) Known gene exons, (2) introns, (3) intergenic, (4) various repetitive elements, and (5) TSS/promoter region. To evaluate the specificity and reproducibility, we performed Pearson's correlation analysis on prostate and brain tissues treated with DNase I. We also used correlation analysis to examine the tissue specificity. All data are expressed as mean $\pm$ standard deviation (SD). We used Student's t-test, ANOVA, Fisher's exact test and Pearson's correlation test in our analysis where appropriate to determine significance. Statistical significance was defined as $p < 0.05$.

Results

Discovery of mirror expressions

After we aligned the probe expression signals, we observed a striking pattern, in which a large number of expression signals are paired asymmetrically as mirrors or mirror images of each other (Fig. 1 & 2). Bidirectional transcription represented a significant proportion of the human tissue transcriptome (Fig. 3). We observed that 53–67% is mirror expressions at chr5q12 and similar results at Chr7q11.23 (72%) and Chr9p21 (74%). Comparative analysis of expression levels based on signal intensity revealed that mean unidirectional expression level was significantly lower than mean bidirectional expression level ($p < 0.0001$).

Mirror RNAs across various tissues

The presence of mirror RNAs are related to gene expression levels. As shown in Fig. 4, probes with higher expression values tend to have mirrors, and probes with lower expressions tend to be unidirectional. Mirror expression was strongly biased towards higher

level of expressions. We further calculated the statistics how many mirror RNAs were present in multiple tissues, we found that as many as 72% of mirror RNAs are present in all 8 tissues (Fig. 5), indicating that most of the mirror RNAs are consistently present across many tissues instead of being specific in a particular tissue.

Mirror expression in known genes

We investigated the presence of mirror expression at 8 known protein-coding genes located in three genomic loci (Fig. 6). We noticed that transcripts produced from the DEPDC1B and PDE4D genes were primarily expressed in a unidirectional manner. The ELOVL7 and GTF2IRD2 genes demonstrated a little higher level of mirror RNA expressions, which is still substantially lower compared with other genomic elements. Substantial variations were also observed among different tissues on the percentage of mirror expressions (Fig. 6).

Mirror expression in non-coding regions

We explored the mirror transcription in non-coding intronic and intergenic regions of the genome (Fig. 7). We found no significant difference between unidirectional expression and mirror expression in the intronic regions ($p>0.1$), but we observed a statistically significant difference between unidirectional expression and mirror expression in the intergenic regions ($p=0.0005$). The highest mirror expression among our test tissues were brain and prostate tissue, 40% and 35% respectively.

Mirror RNAs in the promoters

We analyzed the presence of mirror expressions within 1-kb, 2-kb and 10-kb upstream positions upstream to the transcription start site (TSS) (Fig. 8). The results showed that mirror RNAs were substantially higher at the distal regions rather than the proximal promoter regions. The mirror expression was very low within 1-kb and 2-kb upstream to the TSS.

Mirror expression in the repetitive elements

Interspersed transposable elements are ubiquitous in the mammalian genome. We evaluated the mirror expression in different types of repetitive elements (Fig. 9). Low-complexity, LINE/L2, DNA and MIRs showed very low amounts of mirror expression. On the contrary, the highest mirror expression was observed among the *Alu* repeats within all tissues. The mirror expression on the *Alu* elements represented 69–82% of the total repeat population and only 16–26% of *Alu*-repeats were associated with unidirectional transcripts ($p<0.0001$). In addition, mirror expression was also observed among L1, Simple repeats and LTR elements at lower frequencies (14–40%). *Alu*-associated mirror expression was abundant in all tissues.

Microarray Performance, Reproducibility and Probe Hybridization

We investigated the correlations of expression levels of all probes between two specimens to see if the data is reproducible between independent experiments and if there is any correlation among different tissues (Fig. 10). The results showed that the expression levels of probes are very similar (high correlation) between two brain specimens and two prostate

samples, indicating a high agreement between technical duplicates. Next we assessed the correlation between different tissues. We observed low correlations between various tissues, demonstrating the presence of substantial differences on the mirror RNA expression between various tissues.

Discussion

In this study, we discovered that human genome are expressed in a mirror pattern, in which the majority of transcripts are accompanied by their shadow RNAs like mirrors when they are aligned along the genomic reference backbone (Fig. 1), we name this phenomenon the “mirror RNAs”. We suspect that this mirror pattern may be related to their functionalities and molecular mechanisms^[10,11].

Our data showed that the dominant expression pattern of human genome is not single-orientation transcription; instead, it is mainly expressed in the mirror pattern on at least 53–74% of its transcripts. In addition, 90% of these mirror RNAs are observed in multiple tissues. These results suggested that the dsRNA may be very important in the genome functionality.

The transcript overlapping has been described a decade ago when people were analyzing in silico the EST (expression sequence tags) datasets, and the overlapping transcripts have been given a name NATs (natural antisense transcripts). The molecular mechanisms behind the functional role of NATs are not well understood in eukaryotes. SO far there are no clear sequence motifs identified so far identified among NATs, and it has been observed that NAT sequences are poorly conserved among species^[25]. Three models have been proposed to explain the regulatory effects of cis-NATs on gene expression. The first model attributes RNA duplex formation between the cis-NATs that may affect splicing, translation and degradation, RNA editing and RNAi pathways^[26]. For example, ADAR editing is a mechanism for Alu elements to contribute to gene regulation and occurs when two Alu elements in opposite direction form hairpins^[27]. The second model is that the reverse transcript may guide the DNA methylation and chromatin remodeling through sense-antisense pairing^[28]. Third, some antisense RNAs may serve as self-regulating RNAs and compete with their transcript partner for binding to regulate its own gene^[29]. All of these functional models may depend on the double-stranded RNA nature, rather than the overlapping between two opposite transcripts. Although there is no evidence yet about the relationship of this mirror pattern and double-strand RNA (dsRNA), this new striking pattern provides a new clue and a new direction to unveil the role of RNAs in the genome functions and diseases.

Acknowledgments

This work was supported by National Institutes of Health (R21HG006173, R43HG007621, HL117929, RC4MD005964, MD007602, MD005964, RR003034, U54MD07588); and the American Heart Association grant (09GRNT2300003).

References

1. Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M, et al. Antisense transcription in the mammalian transcriptome. *Science*. 2005; 309:1564–1566. [PubMed: 16141073]
2. David L, Huber W, Granovskaia M, Toedling J, Palm CJ, Bofkin L, et al. A high-resolution map of transcription in the yeast genome. *Proc Natl Acad Sci U S A*. 2006; 103:5320–5325. [PubMed: 16569694]
3. Yu X, Yang J, Li X, Liu X, Sun C, Wu F, et al. Global analysis of cis-natural antisense transcripts and their heat-responsive nat-siRNAs in *Brassica rapa*. *BMC Plant Biol*. 2013; 13:208. [PubMed: 24320882]
4. Smith CA, Robertson D, Yates B, Nielsen DM, Brown D, Dean RA, et al. The effect of temperature on Natural Antisense Transcript (NAT) expression in *Aspergillus flavus*. *Curr Genet*. 2008; 54:241–269. [PubMed: 18813928]
5. Lu C, Jeong DH, Kulkarni K, Pillay M, Nobuta K, German R, et al. Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs). *Proc Natl Acad Sci U S A*. 2008; 105:4951–4956. [PubMed: 18353984]
6. Spiegelman WG, Reichardt LF, Yaniv M, Heinemann SF, Kaiser AD, Eisen H. Bidirectional transcription and the regulation of Phage lambda repressor synthesis. *Proc Natl Acad Sci U S A*. 1972; 69:3156–3160. [PubMed: 4508309]
7. Khorkova O, Myers AJ, Hsiao J, Wahlestedt C. Natural antisense transcripts. *Hum Mol Genet*. 2014; 23:R54–63. [PubMed: 24838284]
8. Xu W, Ma L, Li W, Brunson TA, Tian X, Richards J, et al. Functional pseudogenes inhibit the superoxide production. *Precis Med*. 2015; 1
9. Zhang Y, Liu XS, Liu QR, Wei L. Genome-wide in silico identification and analysis of cis natural antisense transcripts (cis-NATs) in ten species. *Nucleic Acids Res*. 2006; 34:3465–3475. [PubMed: 16849434]
10. Zhong J, Lei XH, Zhu JZ, Song G, Zhang YD, Chen Y, et al. Detection and sequence analysis of two novel co-infecting double-strand RNA mycoviruses in *Ustilagoidea vires*. *Arch Virol*. 2014; 159:3063–3070. [PubMed: 24938486]
11. Wen T, Li H, Song H, Chen F, Zhao C, Lu W, et al. Down-regulation of specific gene expression by double-strand RNA induces neural stem cell differentiation in vitro. *Mol Cell Biochem*. 2005; 275:215–221. [PubMed: 16335801]
12. Lavorgna G, Dahary D, Lehner B, Sorek R, Sanderson CM, Casari G. In search of antisense. *Trends Biochem Sci*. 2004; 29:88–94. [PubMed: 15102435]
13. Balbin OA, Malik R, Dhanasekaran SM, Prensner JR, Cao X, Wu YM, et al. The landscape of antisense gene expression in human cancers. *Genome Res*. 2015; 25:1068–1079. [PubMed: 26063736]
14. Tindwa H, Jo YH, Patnaik BB, Noh MY, Kim DH, Kim I, et al. Depletion of autophagy-related genes ATG3 and ATG5 in *Tenebrio molitor* leads to decreased survivability against an intracellular pathogen, *Listeria monocytogenes*. *Arch Insect Biochem Physiol*. 2015; 88:85–99. [PubMed: 25403020]
15. Oszolak F, Goren A, Gymrek M, Guttman M, Regev A, Bernstein BE, et al. Digital transcriptome profiling from attomole-level RNA samples. *Genome Res*. 2010; 20:519–525. [PubMed: 2013332]
16. Demasius W, Weikard R, Hadlich F, Muller KE, Kuhn C. Monitoring the immune response to vaccination with an inactivated vaccine associated to bovine neonatal pancytopenia by deep sequencing transcriptome analysis in cattle. *Vet Res*. 2013; 44:93. [PubMed: 24099437]
17. Alabady MS, Rogers WL, Malmberg RL. Development of Transcriptomic Markers for Population Analysis Using Restriction Site Associated RNA Sequencing (RARseq). *PLoS One*. 2015; 10:e0134855. [PubMed: 26241739]
18. Li W, Xu W, He S, Ma L, Song Q. Data supporting the high-accuracy haplotype imputation using unphased genotype data as the references. *Data in Brief*. 2015

19. Li W, Xu W, Fu G, Ma L, Richards J, Rao W, et al. High-accuracy haplotype imputation using unphased genotype data as the references. *Gene*. 2015; 572:279–284. [PubMed: 26232609]
20. Li W, Fu G, Rao W, Xu W, Ma L, Guo S, et al. GenomeLaser: fast and accurate haplotyping from pedigree genotypes. *Bioinformatics*. 2015
21. Li Q, Ma Y, Li W, Xu W, Ma L, Fu G, et al. A promoter that drives gene expression preferentially in male transgenic rats. *Transgenic Res*. 2014; 23:341–349. [PubMed: 24338332]
22. Li Q, Xu W, Cui Y, Ma L, Richards J, Li W, et al. A preliminary exploration on DNA methylation of transgene across generations in transgenic rats. *Sci Rep*. 2015; 5:8292. [PubMed: 25659774]
23. Li Q, Li M, Ma L, Li W, Wu X, Richards J, et al. A method to evaluate genome-wide methylation in archival formalin-fixed, paraffin-embedded ovarian epithelial cells. *PLoS One*. 2014; 9:e104481. [PubMed: 25133528]
24. Ma Y, Zhao J, Wong JS, Ma L, Li W, Fu G, et al. Accurate inference of local phased ancestry of modern admixed populations. *Sci Rep*. 2014; 4:5800. [PubMed: 25052506]
25. Johnsson P, Lipovich L, Grander D, Morris KV. Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochim Biophys Acta*. 2014; 1840:1063–1071. [PubMed: 24184936]
26. Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu JK. Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in *Arabidopsis*. *Cell*. 2005; 123:1279–1291. [PubMed: 16377568]
27. Nishikura K. Functions and regulation of RNA editing by ADAR deaminases. *Annu Rev Biochem*. 2010; 79:321–349. [PubMed: 20192758]
28. Osato N, Suzuki Y, Ikeo K, Gojobori T. Transcriptional interferences in cis natural antisense transcripts of humans and mice. *Genetics*. 2007; 176:1299–1306. [PubMed: 17409075]
29. Zhou B, Wei FY, Kanai N, Fujimura A, Kaitsuka T, Tomizawa K. Identification of a splicing variant that regulates type 2 diabetes risk factor CDKAL1 level by a coding-independent mechanism in human. *Hum Mol Genet*. 2014; 23:4639–4650. [PubMed: 24760768]

Figure 1. A global view of mirror expression

Expression signal intensities at chromosomal 7q11.23 between 74,178,309–74,213,659 nucleotides (nt) are shown. The height of the vertical bars indicates the logarithmically transformed expression values; the bars above the line in blue indicate the expression values on the Watson orientation, and the bars below the line in orange indicate the expression values on the Crick orientation. Various annotations, such as known gene exons and repetitive elements are shown on the top of this figure.

Figure 2. The design of tiling array probes

There are 4 groups of probes, 60-nt each, in two orientations (Watson, blue arrows; Crick, yellow arrows). The tiling interval is zero, the overlapping length between A and B and between C and D is 30-nt.

Figure 3. The percentage of mirror expression in various tissues

The data at Chr5q12 is shown. The percentages of unidirectional expression are shown by red bars; the percentages of mirror expression are shown by blue bars; the noise expression (< 50) is shown by green bars.

Figure 4. The relationship between mirror RNAs and expression levels

The percentages of mirror RNAs in eight tissues are shown at Chr7q11.23. (a) mirror, (b) unidirectional expression. The expression activities in various tissues are grouped by the expression levels.

Figure 5. Mirror expression is ubiquitous throughout various tissues

The results at chromosomal loci 7q11.23 are shown. For example, 72% of mirror RNAs was expressed in all of the eight tissues under the study, and only 10% of these mirror RNAs was observed in only one tissue.

Figure 6. Gene-to-gene variation and tissue-to-tissue variation on the mirror RNA expression
Four known protein-coding genes are shown. (a) PDE4D, (b) ELOVL7, (c) DEPDC1B, and (d) GTF2IRD2. Each bar represents a percentage of mirror (blue) or unidirection (red) in total RNAs.

Figure 7. Mirror expression in non-coding regions

(a) introns; **(b)** intergenic regions. Each bar represents a percentage of mirror (blue) or unidirection (red) in total RNAs.

Figure 8. Mirror expression in the promoters

Each bar represents a percentage of mirror (blue) or unidirection (red) in total RNAs in the 5'-flanking regions of 1-kb, 2-kb and 10-kb upstream to the transcription start site (TSS).

Figure 9. Mirror expression in the promoters

Each bar represents a percentage of mirror (blue) or unidirection (red) in total RNAs in the repetitive elements in human genome. The types of repeats are labeled above each figure.

Figure 10. Pearson's correlation (r^2) analysis among various tissues

(a) untreated brain cortex and DNase I treated brain cortex; (b) untreated brain replicates (n=2); (c) brain vs. prostate; (d) artery vs. brain; (e) artery vs. vein; (f) untreated prostate vs. DNase I treated prostate; (g) prostate vs. artery; (h) prostate vs. vein.