NATsDB: Natural Antisense Transcripts DataBase

Tutorial of NATs Database

Yong Zhang
LiPing Wei

July 25, 2006

Center for Bioinformatics
Email: zhangy@mail.cbi.pku.edu.cn
Abstract

This document describes the background of natural antisense transcripts, the introduction to NATsDB, how to use NATsDB.

Content

1. Background
   1.1 What are NATs, cis-NATs and trans-NATs? ................................................................. 3
   1.2 Which biological function do NATs have? ................................................................. 3

2. Introduction
   2.1 What is NATsDB? ....................................................................................................... 5
   2.2 What is NATsDB’s architecture?
      2.2.1 NATsDB’s construction pipeline? ....................................................................... 5
      2.2.2 Data source we used to construction NATsDB? .................................................. 6
      2.2.3 Statistics on natural antisense transcripts we identified? .................................... 6
   2.3 What can NATsDB be used for and what are the main features of NATsDB? ................ 7
   2.4 What differences are there among our NATsDB and other database? ....................... 8

3. Using NATsDB
   3.1 Browse and search NATsDB? .................................................................................. 9
   3.2 Fine-tuning the NATsDB display? ............................................................................ 12
   3.3 Interpreting the NATsDB display?
      3.3.1 In the cluster location page? ............................................................................. 15
      3.3.2 In the sequence page? ..................................................................................... 16

4. Other questions
   4.1 What is a fasta sequence? ....................................................................................... 17
   4.2 How to cite NATsDB? ............................................................................................ 17
Background of natural antisense transcripts

1.1 What are NATs, cis-NATs and trans-NATs?

Natural Antisense Transcripts (NATs) are simply RNAs containing sequences that are complementary to other endogenous RNAs. They can be transcribed in cis from opposing DNA strands at the same genomic locus (cis-NATS), or in trans from separate loci (trans-NATS).

Two other classes of trans-acting noncoding RNA are related to trans-NATs because they recognize their target RNAs by imprecise base-pairing: microRNAs (miRNAs), which inhibit the translation of mRNAs, and small nucleolar RNAs (snoRNAs), which guide the modification of non-coding RNAs.

Two another categories related with cis-NATs are NOB (Non-exon-Overlapping Bidirectional) and NBD (Non-BiDirectional). All of them are depicted in the following figure:

1.2 Which biological function do NATs have?

Pairing of NATs to sense RNAs is known to regulate expression of many different genes in cells and their accessory elements: viruses, plasmids and transposons. In recent years, NATs have been implicated in many aspects of eukaryotic gene expression including genomic imprinting, RNA interference, translational regulation, alternative splicing, RNA editing and X-inactivation. Moreover, there is growing evidence to suggest that antisense transcription might have a key role in a range of human diseases. But the role of NATs, in most cases, is poorly understood in eukaryotic organisms.
NATs as regulators range in size from very small, such as microRNA (miRNA) transcripts which are 21-22 nts in length to extremely large, such as the 108 kb Air transcript associated with the imprinted Igf2r locus.

miRNAs are the best-studied examples of antisense regulators in eukaryotes. Until 2001, hundreds of micro RNAs (miRNAs) have been found in human, mouse and other organisms that closely resemble. Although the physiological function of most microRNAs is unknown, it is likely most if not all exert their effects through base pairing with complementary target sequences.

The role of larger antisense RNAs in the regulation of gene expression is not well established. However, several interesting examples of antisense regulators have been described. One such example involves the early to late phase transition during polyomavirus replication. High levels of readthrough transcription of late viral mRNA suppress expression of complementary early mRNA. These complementary strands form a long, perfectly matched duplex structure that is extensively modified by an enzyme known as ADAR (adenine deaminase acting on double-stranded RNA) and retained in the cell nucleus.

Although at present there are few common themes among possible eukaryotic antisense regulators, it is useful to distinguish between cis-NATs and trans-NATs. While cis_NATs are perfectly complementary to their target, and can potentially form extended regions of perfectly paired duplex, trans_NATs manifest more limited, imperfectly matched base-pairing interactions with their targets. In the relatively few instances where efforts have been made to demonstrate a physiological role for base-pairing interactions between antisense and target, most results have been remained inconclusive.
Introduction to NATsDB

2.1 What is NATsDB?

NATs database: Natural Antisense Transcripts database.

After developing a fast, integrative pipeline to identify *cis* natural antisense transcript (*cis*-NATs) at genome scale and using transcriptome and genome sequences in UniGene and GoldenPath, we applied the pipeline to identify *cis*-NATs in ten eukaryotic species, screening seven of these species for the first time and bringing the number of candidate SA genes in human to 7,830, a 50% increase from previous studies. We create this free and publicly accessible database that allows researchers to query the ten species's datasets.

2.2 What is NATsDB’s architecture?

The construction pipeline, data source and final results are summarized as the following figures. For more details, please refer to our Nucleic acids research paper (Zhang, et al., 2006).

2.2.1 NATsDB’s construction pipeline
2.2.2 Data source we used to construction NATsDB

<table>
<thead>
<tr>
<th>Species</th>
<th>Version of UniGene Build</th>
<th>Version of Genome Assembly</th>
<th>Number of Usable Transcripts(^a)</th>
<th>Percentage of Entrez Gene Covered(^b)</th>
<th>Total Number of UniDirectional Cluster(^d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hs (human)</td>
<td>184</td>
<td>NCBI 35/UCSC hg17</td>
<td>2,261,824.2.5GB</td>
<td>20,431/25,134 (81.2%)</td>
<td>29,688</td>
</tr>
<tr>
<td>Mm (mouse)</td>
<td>146</td>
<td>NCBI 34/UCSC mm9</td>
<td>1,548,836.2.5GB</td>
<td>23,087/34,603 (66.7%)</td>
<td>27,688</td>
</tr>
<tr>
<td>Dr (fly)</td>
<td>37</td>
<td>UCSC dm2</td>
<td>259,689.0.1GB</td>
<td>13,310/13,836 (96.4%)</td>
<td>11,820</td>
</tr>
<tr>
<td>Gs (worm)</td>
<td>22</td>
<td>UCSC ce2</td>
<td>210,784.0.1GB</td>
<td>18,590/19,170 (97%)</td>
<td>18,771</td>
</tr>
<tr>
<td>St (sea squirt)</td>
<td>17</td>
<td>UCSC c1</td>
<td>428,229.0.15GB</td>
<td>273/345 (99.5%)</td>
<td>10,813</td>
</tr>
<tr>
<td>Gg (chicken)</td>
<td>27</td>
<td>UCSC gallGal2</td>
<td>215,854.1.1GB</td>
<td>5,380/18,289 (29.4%)</td>
<td>15,405</td>
</tr>
<tr>
<td>Rn (rat)</td>
<td>143</td>
<td>UCSC rn3</td>
<td>223,490.2.8GB</td>
<td>9,235/23,457 (38.8%)</td>
<td>19,934</td>
</tr>
<tr>
<td>Sg (frog)</td>
<td>24</td>
<td>UCSC xenTrel</td>
<td>430,633.1.7GB</td>
<td>3,866/4,325 (89.3%)</td>
<td>16,403</td>
</tr>
<tr>
<td>Dr (zebrafish)</td>
<td>84</td>
<td>UCSC danRer</td>
<td>232,436.1.7GB</td>
<td>8,932/17,708 (50.2%)</td>
<td>14,639</td>
</tr>
<tr>
<td>Bt (cow)</td>
<td>69</td>
<td>UCSC boarTus2</td>
<td>47,581.3GB</td>
<td>1,023/37,628 (2.7%)</td>
<td>3,088</td>
</tr>
</tbody>
</table>

I. "Number of usable transcripts" is the total number of mRNAs and ESTs that can be mapped to genome sequences and have reliable orientation.

II. "Percentage of Entrez Gene Covered" is equal to the number of Entrez Gene entries that have usable transcripts divided by the total number of Entrez Gene entries for that species. The total number of Entrez Gene entries for each species is slightly smaller than the number in NCBI because we removed those Entrez Gene entries that were generated only from protein or in silico prediction without any supporting transcripts.

III. "Total Number of UniDirectional Cluster" is equal to 2\*"Number of SA clusters"+2\*"Number of NOB clusters"+"Number of NBD clusters", as the SA clusters and NOB clusters have transcripts from both the forward and backward orientation.

2.2.3 Statistics on natural antisense transcripts we identified

The following figures make statistics on SA pairs, NOB pairs and NBD sequences, such as the overlap median, relative abundance, etc.

Statistical data on SA pairs

<table>
<thead>
<tr>
<th>Species</th>
<th>Number of SA Clusters</th>
<th>Abundance of SA genes (^a)</th>
<th>Mean of length overlap (bp)</th>
<th>Median of length overlap (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hs (human)</td>
<td>3,915</td>
<td>26.3%</td>
<td>318</td>
<td>195</td>
</tr>
<tr>
<td>Mm (mouse)</td>
<td>3,040</td>
<td>21.9%</td>
<td>305</td>
<td>195</td>
</tr>
<tr>
<td>Dr (fly)</td>
<td>997</td>
<td>16.8%</td>
<td>253</td>
<td>116</td>
</tr>
<tr>
<td>Gs (worm)</td>
<td>270</td>
<td>2.8%</td>
<td>128</td>
<td>60</td>
</tr>
<tr>
<td>St (sea squirt)</td>
<td>857</td>
<td>15.8%</td>
<td>242</td>
<td>186</td>
</tr>
<tr>
<td>Gg (chicken)</td>
<td>514</td>
<td>6.6%</td>
<td>209</td>
<td>147</td>
</tr>
<tr>
<td>Rn (rat)</td>
<td>458</td>
<td>4.5%</td>
<td>183</td>
<td>116</td>
</tr>
<tr>
<td>Sg (frog)</td>
<td>354</td>
<td>4.3%</td>
<td>213</td>
<td>138</td>
</tr>
<tr>
<td>Dr (zebrafish)</td>
<td>168</td>
<td>2.2%</td>
<td>217</td>
<td>172</td>
</tr>
<tr>
<td>Bt (cow)</td>
<td>60</td>
<td>3.8%</td>
<td>324</td>
<td>338</td>
</tr>
</tbody>
</table>
2.3 What can NATsDB be used for and what are the main features of NATsDB?

NATsDB can serve as a repository for current knowledge and a starting point for future experimental design or in silico data mining. New technologies such as CAGE, SAGE, and genome tiling array will identify more cis-NATs, as has already been shown to be the case for mouse. The EST-based identification strategy will continue to be useful because for many species, EST data is the only source of transcriptome data available.

### Statistical data on NOB pairs

<table>
<thead>
<tr>
<th>Species</th>
<th>Number of NOB Clusters</th>
<th>Abundance of NOB genes</th>
<th>Mean length overlap (bp)</th>
<th>Median length overlap (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>hs</td>
<td>1,484</td>
<td>9.9%</td>
<td>9,693</td>
<td>2,784</td>
</tr>
<tr>
<td>mm</td>
<td>936</td>
<td>6.7%</td>
<td>6,741</td>
<td>2,468</td>
</tr>
<tr>
<td>dm</td>
<td>360</td>
<td>6.0%</td>
<td>2,672</td>
<td>1,557</td>
</tr>
<tr>
<td>cel</td>
<td>383</td>
<td>4.0%</td>
<td>1,362</td>
<td>974</td>
</tr>
<tr>
<td>cin</td>
<td>150</td>
<td>2.7%</td>
<td>1,781</td>
<td>903</td>
</tr>
<tr>
<td>gga</td>
<td>240</td>
<td>3.1%</td>
<td>5,564</td>
<td>2,071</td>
</tr>
<tr>
<td>rn</td>
<td>325</td>
<td>3.2%</td>
<td>7,345</td>
<td>2,836</td>
</tr>
<tr>
<td>str</td>
<td>205</td>
<td>2.4%</td>
<td>11,613</td>
<td>4,142</td>
</tr>
<tr>
<td>dr</td>
<td>221</td>
<td>3.0%</td>
<td>9,814</td>
<td>4,668</td>
</tr>
<tr>
<td>bt</td>
<td>5</td>
<td>0.3%</td>
<td>874</td>
<td>240</td>
</tr>
</tbody>
</table>

### Statistical data on NBD sequences

<table>
<thead>
<tr>
<th>Species</th>
<th>Number of NBD Clusters</th>
<th>Abundance of NBD genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>hs</td>
<td>18,890</td>
<td>63.6%</td>
</tr>
<tr>
<td>mm</td>
<td>19,736</td>
<td>71.2%</td>
</tr>
<tr>
<td>dm</td>
<td>9,106</td>
<td>77.0%</td>
</tr>
<tr>
<td>cel</td>
<td>17,465</td>
<td>93.0%</td>
</tr>
<tr>
<td>cin</td>
<td>8,799</td>
<td>81.3%</td>
</tr>
<tr>
<td>gga</td>
<td>13,897</td>
<td>90.2%</td>
</tr>
<tr>
<td>rn</td>
<td>18,868</td>
<td>92.1%</td>
</tr>
<tr>
<td>str</td>
<td>15,285</td>
<td>93.1%</td>
</tr>
<tr>
<td>dr</td>
<td>13,856</td>
<td>94.6%</td>
</tr>
<tr>
<td>bt</td>
<td>2,958</td>
<td>95.7%</td>
</tr>
</tbody>
</table>
NATsDB offers the following features in one unified web-based user interface:

I. *cis*-NATs identified in 10 genomes including human, mouse, fly, worm, sea squirt, chicken, rat, frog, zebrafish and cow, the largest collection so far. In addition, non-exon-overlapping bi-directional clusters and non-bidirectional clusters are also included.

II. A web-based graphical interface we developed for NATsDB that shows the alignment of all sense transcripts, antisense transcripts, and genomic sequences, with hyperlinks to related databases. It also contains many features of the sense and antisense transcripts such as phastCons conservation. Sense-antisense pairs were divided into six sub-groups according to their overlapping patterns.

III. A web-based graphical interface for browsing by species or by chromosomes. It also allows users to easily select subsets of the data such as ESTs with polyA signals and tails or only ESTs with splicing sites.

2.4 What differences are there among our NATsDB and other database?

Currently, there are only a few databases on *cis*-NATs except NATsDB, such as SADB and Sense/Antisense Database. Compared to these database, our NATsDB contains more species and larger datasets, provides more powerful search/browse function, more information, such as phastCons conservation.
3.1 Browse and search NATsDB

There are two main approaches to search your interested natural antisense transcripts.

Ⅰ. In the Browse page, you can use dropdownlist or mouse over the figure to get information.

As for dropdownlist,

ⅰ. you should specify a type of cluster:
   - SA, Sense-Antisense pair cluster
   - NOB, Non-exon-Overlapping Bidirectional cluster
   - NBD, Non-BiDirectional cluster

ⅱ. you can specify the "species" (seven species from human to zebrafish with complete genome), the "Overlapping pattern" of SA pairs and the figure configuration "Height".
The "Overlapping pattern" of SA pairs:
55: head-head (divergent), SA gene pairs with first exon of both partners involved in the overlap;
33: tail-tail (convergent), SA gene pairs with last exon of both partners involved in the overlap;
complete: one gene sequence completely covered by an exon of the other;
contained: one gene sequence completely covered by the intron and exon of the other;
intronic: one gene starting within an intron of the other and transcribing within and across the exons
others: all other SA pairs.

The corresponding schema is shown in the following figure:

The “coding potential” specify the SA pairs in terms of the coding potential of representative sequences. For example, coding/ noncoding indicates one gene has CDS (coding sequence) annotation, while the other does not.

And, if users input “apoptosis” in query box, only clusters including at least one sequence with description containing ‘apoptosis’ will be displayed. As for overlap box, users specify the minimum overlapping length for SA pairs or NOB pairs. As for mousing over the figure, you can click "+" to get information.
In the search page, we provide text search, chromosome location search and sequence search.

Text search mode supports Boolean mode. You can fill the “Gene search” text box with Entrez Gene name, synonyms and description, such as THRA or the “Transcript search” text box with mRNA/EST accession number and description, such as X55005.

Especially, for exactness and acceleration of search, you can click the “Name only” checkbox and choose one of seven species in the “Species” dropdownlist. Click the "Name only" checkbox, and you only type gene name other than gene description to get information. You also can choose one of seven species in the "Species" dropdownlist in Transcript search.

The image below is search results via "Gene search" or "Transcript search". Click the "ClusterID" or "Name/Accession", then go into sequence detailed information page.
ii. Chromosome location search.

Users could specify special genomic location and retrieve clusters derived from this region. UCSC-like chromosome location format is supported, for example, chr17:35471999-35510504. "chr17", "35471999" and "35510504" indicate chromosome 17, chromosome beginning coordinate, and chromosome end coordinate, respectively. For versions of chromosomes used in current NATsDB, please see “2.2.3 Statistics on natural antisense transcripts we identified” section.

iii. Sequence search way via blast.

Enter your sequence in FASTA format into the textbox, then choose programs based on your sequence. If it is protein sequence, program "tblastn" is suggested; Otherwise "blastn" is suggested.

You also can choose one of seven species in the second dropdownlist to acceleration of alignment. See an example. Click the high score alignment, then go into the detailed information page of this hit.

3.2 Fine-tuning the NATs display

The NATs annotation tracks page displays a sequence detailed information page and a cluster location specified through a text search, a sequence search, or clicking "+" on chromosome which we described above and a sequence information.

Two main features in cluster location page: The annotations tracks image and a set of controls including navigation controls, display configuration buttons and display controls.
The first time you open the NATsDB, it will use the application default values to configure the annotation tracks display which only show Genome location of this cluster, mRNA the cluster contain and representative sequences.

The track display controls are gathered together that reflect the type of data in the track, e.g. Isoform Prediction Tracks, mRNA and EST tracks.

The cluster annotation tracks displayed in the NATs use a default set of display conventions: Genome location, mRNAs which this cluster contain and representative sequences of this clusters etc.

To change the display mode for a track, find the track’s controller at the top of the cluster information page, select the desired mode from the control's display menu, and then click the GO button. These options let the user restrict the data displayed within an annotation track.

I. Changing the text size in the annotation track image

The annotation track image may be adjusted to display text in a range of fonts from "small" to "large". To change the size of the text, select an option from the text size pull-down menu, then click Submit. The text size is set to "small" by default.

II. Changing the width of the annotation track image

By default, the width of the annotation track window is set to 900 pixels. Notice that 900 pixels are also the least pixels. To modify the width to best suit the display capabilities of your monitor, enter a new value in the image width text box, then click the GO button. For example, setting the display to 1100 pixels on a 19" monitor will increase the visible portion of the cluster and reduce the need for redraws.
Annotation track descriptions controls:

Each annotation track has an associated control to make it hidden or shown.

i. Conservation score track:
   "Show phastcons" control restricts the data "conservation score" displayed within an annotation track. Chromosome region in this cluster is calculated by PhastCon. And the "Conservation Score" represents the degree of region's conservation.

ii. Repeat regions track
    This track, associated with the "Show repeat" control, displays whether repeats exist in the chromosome region or not. If the region you select doesn't has any repeats, even if the "show repeat" control has been clicked, there will be none shown!

iii. Isoform prediction tracks (under development):
    The alternative splicing isoforms are predicted by the ECgene's prediction algorithm (Genome Res. 2005 15: 566-576). If you click the "show isoform" box, all the isoforms in this cluster will be listed. Due to immense computation cost, this track will be released in one week.

iv. Transcript box
    If users input one accession number in the 'Transcript' box, for example, X72304, the browser will show the corresponding genomic region and all the sequences derived from this region. A noteworthy point is the retrieved sequence must meet with the other control criteria too, such as "Subset", "Show 'hidden' transcripts", etc.

v. Subset controls:
    We collect orientation-reliable sequences in UniGene, including Refseq sequences, mRNA which CDS annotation, mRNA without CDS annotation, spliced ESTs, ESTs with both polyA signal and polyA tail. Every kind of transcripts can be shown or hidden via corresponding controls box in the "subset" row.

vi. Changing the strand in the sequence track image
    Cis-NATs describe RNAs containing sequences that are complementary to other endogenous RNA from opposing DNA strands at the same genomic locus. In a genome cluster, the sequence track image may be adjusted to display text from "minus", "plus" to "both".

vii. Other controls:
    A cluster may cover a long region of the chromosome. To display a completely different position in the cluster, enter the new query in the "start/end" text box, then click the GO button. Besides, by default, only representative sequences are listed in a cluster. Remove the "show representative" box, and all the sequences in the cluster will be shown. And if "show hidden gene" control is clicked, the gene which has not complementary gene will be shown.

NOTICE:

For we set the "show representative" control the highest priority, in order to make other controls useful, the control should not be clicked!
3.3. Interpreting the NATsDB display

3.3.1 In the cluster location page

Four tracks: Genome location tracks, Conservation score track, Repeat regions track, Sequence tracks, are shown above.

I. Genome location tracks:

The genome location tracks, located just above the conservation score tracks image, provides a graphical overview of chromosome coordinate, including an indication of the region currently displayed in the annotation tracks image. Click the red description above the chromosome coordinate base line, then your browser will turn to the corresponding regions in UCSC genome browser.

II. Sequence tracks:

Every sequence description have three fields in order: display id, sequence structure, links to detailed information page.

i. Display id and Links

If you click the box before the "display id", the sequences are chosen and displayed in the refreshed page. Click the display id in the first field or, the browser will turn to the sequence's detailed information which we discuss below.

ii. Sequence structure

In the sequence structure, three rows describe sequence. The type of transcriptome (for instance, mRNA with CDS, ployA-EST, et al), sequence' function description, sequence length (for instance, ~1k), exon numbers (for instance, 3 blocks), standard splicing site number (for instance, IO=2) are all shown in the first row. Coding exons are represented by blocks connected by horizontal lines representing introns. The number inserted into the horizontal lines represents the intron's length.

The 5' and 3' untranslated regions (UTRs) are displayed as green blocks on the leading and
trailing ends of the aligning regions. Arrowheads on the connecting intron lines indicate the direction of transcription. In situations where no intron is visible (e.g. single-exon genes, extremely zoomed-in displays), the arrowheads are displayed on the exon block itself. Click the coding exon blocks, then the browser will also turn to the corresponding regions in UCSC genome browser. As for some sequences, polyA tail length and potential polyA signals are shown in green. Here, polyA tail was defined as a stretch of at least 10 As at 3’ end of a sequence and PolyA signal was defined as hexanucleotide ‘AATAA’, ‘ATTAAA’, ‘AATTTA’, ‘AATAAT’, ‘CATAAA’ or ‘AGTAAA’ within the last 50 bp of 3’ end of a sequence after the polyA tail was trimmed. Possible polyA tail or polyA signal predicted in the reverse complement strand will be shown in red.

iii. Representative sequence

The red information in the first row shows the sequence has been chosen as the representative sequence in this cluster.

3.3.2. In the sequence page

Click the display id or the links of sequence tracks in the cluster information page, then browser goes into sequence detail information page. General information (Unigene Cluster, Description, et al), Local annotation, Cross reference (Gene, Homologene) are given to complementarily describe the sequence. The items in the cross reference are linked to NCBI Gene, UniGene, HomoloGene and other homolog sequences in our dataset.
CHAPTER
FOUR

Other questions

4.1 What is a FASTA sequence?

A sequence in FASTA format begins with a single-line description, followed by lines of sequence data. The description line is distinguished from the sequence data by a greater-than (">") symbol in the first column. It is recommended that all lines of text be shorter than 80 characters in length.

See an example:

>BU938537 AGENCOURT_10517608 NIH_MGC_169 Mus musculus cDNA clone IMAGE:6706694 5', mRNA sequence
GGGACAGGCAGTTAGTGCTCAGAACTGTGCTGTTTCTGCTGCCCGAGGAGGAG
CTCTTGCGCCTCCATCTCTCGCCGACCAAAAATGGGCTCTGCAGGGCGAGTGCC
CTCTGGTCTCGGGCTTCCCACAAATCTCTCTCTTCGTTTAAGAGAAGCTTGCTGCTGCT
GCCAGGCTCTTGGCAGAGATGGTGAGGAGCAGCAGATGAGAAGACACCACACGACAC
CTGACAGGAGAGAATCTCAGTCGTTGCTGCTGTGCTAGGAGAACGTCTCTGCTTCGCAAGAGAG
GGAAGACCCAGTGAGTCTCAGCCAGCAGACATCTACACACACCTGCAAGACGACAAGTGCTCTC
ATCACCCTTGAAAGTGATGACTCTGATGCAGACACCGTCCAGACTCCACACCCACATTCC

4.2 How to cite NATsDB?

Please cite the following article:
Yong Zhang, XS Liu, Qing-Rong Liu and Liping Wei. Genome-wide in silico identification and analysis of cis natural antisense transcripts (cis-NATs) in ten species. *Nucleic Acids Res.*, 34: 3465-3475