Evolution of the long non-coding RNAs MALAT1 and MENβ/ϵ

Peter F. Stadler

Bioinformatics Group, Department of Computer Science, and Interdisciplinary Center for Bioinformatics, University of Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany; Max Planck Institute for Mathematics in the Sciences, Inselstraße 22, D-04103 Leipzig, Germany; Fraunhofer Institut für Zelltherapie und Immunologie, Perlickstraße 1, D-04103 Leipzig, Germany; Department of Theoretical Chemistry University of Vienna, Währingerstraße 17, A-1090 Wien, Austria; Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501, USA

Abstract. MALAT1 is one of the best-conserved long ncRNAs in mammals and shares several characteristics, among them nuclear retention and a non-standard processing of its 3' end, with the longer, but less well conserved, adjacent MENβ RNA. We show that MALAT1 is conserved among gnathostomes (with the possible exception of birds), while MENβ likely originated in the mammalian stem lineage. Evolutionary conserved features of both transcripts are discussed, including RNA secondary structure motifs and short RNA processing products.

Key words: MALAT1, MENβ, VINC, NEAT1, TncRNA

1 Introduction

A plethora of diverse non-coding RNAs have been discovered during the last decade, collectively demonstrating that a large fraction of the genomes of higher eukaryotes is transcribed into mRNA-like non-protein-coding transcripts (mlncRNAs) [1, 2]. The evolutionary history of these transcripts is still poorly understood. With very few exceptions, only global statistical information is available to demonstrate that a large number of ncRNAs is under stabilizing selection [3–5]. Nevertheless, most mlncRNAs are poorly conserved at sequence level compared to other functional transcripts [6, 4]. Detailed evolutionary information is available on many families of protein-coding genes and structured "house-keeping" RNAs. For ncRNAs, it is compiled in the `Rfam` database [7] and in specialized data repositories for microRNAs (`miRBase` [8]) and snoRNAs (`snoRNA-LBME-db` [9]). In contrast, evolutionary and phylogenetic information on mlncRNAs is currently neither collected nor organized in a systematic way.

Detailed case-studies are available for only a few prominent transcripts, such as the imprinting-related mammalian H19 ncRNA [10], the *Drosophila roX* RNAs [11], and the eutherian *Xist* transcript [12, 13]. The latter originated by pseudogenization of the protein-coding *Lnx3* gene in the eutherian ancestor [12] under inclusion of repetitive elements [13], which also gave rise to conserved

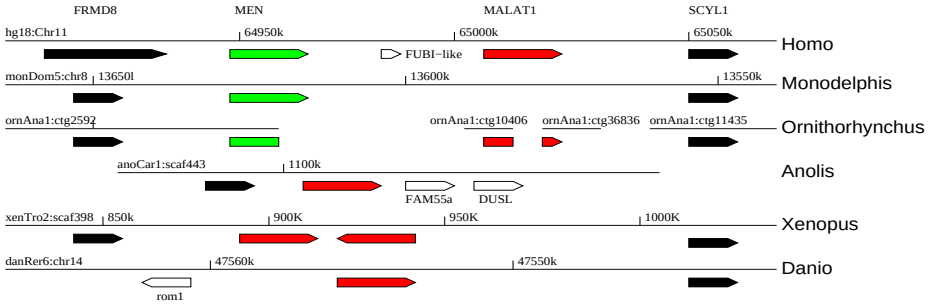

Fig. 1. Overview of the MEN/MALAT locus in different vertebrate species. The non-coding MALAT transcript is linked to at least one of FRM8 and SCYL1 in all species except the stickleback. In *Xenopus*, MALAT1 is duplicated, with one copy arranged in reverse direction.

secondary structure features [14]. *Xist* is one of only three highly expressed poly-adenylated ncRNAs that show strong nuclear retention [15].

The other two transcripts, NEAT1 and NEAT2/MALAT1, are the topic of this contribution. They are located in close genomic proximity at the human 11q13.1 locus. NEAT1 also exists in a longer isoform, known as MENβ [16]. Recent studies showed that MALAT1 and MENβ share a number of peculiar features. Both transcripts are spliced only infrequently [15], a feature that is atypical for transcripts of their size. Most surprisingly, their 3' ends are processed in a non-standard way: RNase P cleaves the primary transcripts before a tRNA-like element [17, 16], which is then processed into an independent cytoplasmic ncRNA. The evolution of these small tRNA-like ncRNAs was studied already in some detail [18].

The $\sim$ 8.7kb MALAT1 transcript (also known as NEAT2 and AlphaTFEB) is overexpressed in a variety of different carcinomas [19–21]. Knockdown of MALAT1 by shRNA implicates the transcript in cell cycle progression [22]. As noted in [15], MALAT1 is exceptionally well-conserved for a long ncRNA. The same study noticed the presence of a homolog in the opossum genome and reported an "apparent absence of the transcript in non-mammalian species". The subnuclear localization of MALAT1 is concentrated in the SC35 splicing domains, indicating a function in pre-mRNA metabolism [15].

The NEAT1 transcript, which has a size of $\sim$ 3.2 kb, is also responsive to diverse disease states. It is induced in mouse brain during infection by Japanese encephalitis virus and rabies virus, and hence was termed "Virus Inducible Non-Coding RNA" (VINC) in [23, 24]. It is located at the *Men1* ("multiple endocrine hypoplasia 1") locus, and hence was named MENϵ in [16, 25]. The transcript contains the shorter "trophoblast non-coding RNA" (TncRNA) that suppresses the expression of major histocompatibility antigens [26–28]. The bovine NEAT1 orthologue shows increasing expression levels during development of cattle muscle [29]. The same locus also produces a much longer isoform ($\sim$ 20kb), called

MENβ. Several groups recently reported the involvement of MENϵ and MENβ in the organization of the paraspeckles [30, 31, 25, 16], reviewed in [32]. Protein interaction regions in the VINC/NEAT1/Menϵ RNA are investigated in [33]. The main biological function of NEAT1/MENϵ is the regulation of gene expression by restricting nuclear export [30, 34].

2 Materials & Methods

Genomic sequence data were retrieved from `ensembl` (v.57). In addition, ESTs and unassembled genomic DNA from NCBI GenBank, and high throughput sequencing data from GEO were analyzed. Initial homology searches were performed with `blast` and extended to global alignments using several alignment tools, including `custalw`, `dialign`, `muscle`, and `mafft`. RNA secondary structures were investigated using the `Vienna RNA Package`. The UCSC genome browser was used for visualization. Due to length restrictions on the manuscript, details and references are given throughout the Results section where necessary.

3 Results

Syntenic Conservation. The genomic location of MENβ/MALAT1 is flanked by FRMD8 ("FERM domain containing 8", a.k.a. FKSG44) on the 5' side and by the highly conserved kinase-like gene SCYL1 throughout Eutheria. A small "FUBI-like" gene (AP000769) is located between MENβ and MALAT1. All these transcripts share reading direction, Fig. 1. The MALAT1 homolog is also linked to FRMD8 and/or SCYL1 in other vertebrates, and the arrangement [5'-FRMD8-MALAT1-SCYL1-3'] appears to be the ancestral state. The assembly of the elephant shark genome, however, does not provide sufficient evidence to test this hypothesis directly; no MALAT1 homolog was detectable in the lamprey genome. In teleosts, synteny is broken between FRMD8 and MALAT1, while

Table 1. Approximate locations of MENβ in several mammalian genomes. The coordinates refer to the (mostly unspliced) ESTs located in the approximate region identified by `blastn` as homologous to human MENβ. The 5'-end of the menRNA is also listed. Dots indicate that there are no ESTs near the position of the menRNA.

Species	Assembly	Chr.	±	5'-MENβ	3'-MENβ	5'-menRNA
Homo sapiens	hg19	11	+	65190269	65213007	65213012
Macaca mulatta	rheMac2	14	−	9009052	...	8979130
Mus musculus	mm9	19	−	5845579	5824708	5824707
Rattus norvegicus	rn4	1	−	208481740	208455951	208456537
Canis familiaris	canFam2	18	−	54794495	54775188	54775783
Equus caballus	equCab2	12	+	25591044	...	25613257
Bos taurus	bosTau4	29	+	45474754	45495959	45495960
Ornithorhynchus anatinus	ornAna1	ctg2592	−	13707	...	—

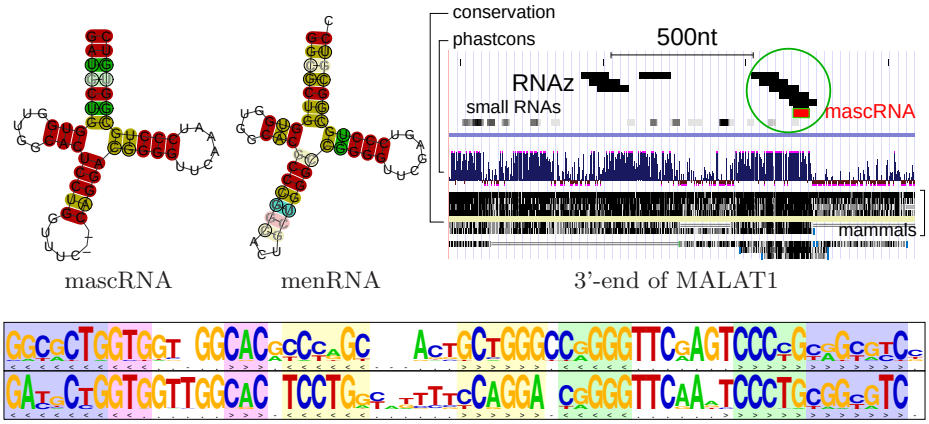

Fig. 2. RNA secondary structure in MALAT and MENβ. By far the best conserved structured signals are mascRNA [17] (upper left) and menRNA [16] (upper middle), respectively. The upper right panel summarizes the `RNAz` predictions of structured RNAs in MALAT1. Besides the mascRNA and the hairpin structure described in [16] at the 3' end (encircled), there is only one additional structured region about 600nt upstream of the mascRNA. Below, the aligned sequence logos of menRNA (above) and mascRNA (below) clearly show that the two ncRNAs are homologous.

SCYL1 is located on a different scaffold in the lizard genome. In *Xenopus* we find two divergent copies of the MALAT1 sequence in tail-to-tail orientation.

Surprisingly, the entire region is missing in all four sequenced bird genomes (chicken, turkey, zebrafinch, and duck). No plausible homolog of SCYL1, FRMD8 (using `tblastn`), or MALAT1 (using `blastn`) are detectable. There are two possible explanations: (1) Birds lost the entire genomic locus. (2) FRMD8-MALAT1-SCYL1 is located on a microchromosome, which are known to be underrepresented in the chicken genome assembly [35]. Given that MALAT1 can be identified in all other sequenced gnathostomes and the high level of sequence conservation of the two flanking genes (whose functions appear to be unrelated to that of MALAT1 and MENβ/ϵ), we suspect that we see a data bias rather than a true loss of the entire locus.

MENβ is clearly present in all mammals. Within eutheria, the homology is easy to establish and the loci can be found by simple `blastn` searches using e.g. the human sequence as query. In several species the presence of the MENβ and/or MENϵ transcripts is supported by (predominantly unspliced) ESTs mapping to the location of the `blastn` hit, see Tab. 1. Due to gaps, break-points between scaffolds, and inaccuracies in the genome assemblies, it is a bit more complicated to trace MENβ in marsupials and in platypus. Unambiguous `blastn` hits to large portions of MENβ are easily obtained, however. No EST support is available in marsupials. The expression of a MENβ transcript in platypus is supported by a handful of ESTs (EY202075, EY201405, EH004653, EG34158) as well as several 454 reads listed in `ensembl` (v.57).

Table 2. Approximate positions of the MALAT1 homologs in non-mammalian vertebrates determined from EST information. Data are from the UCSC genome browser, except for lizard, which was taken from `ensembl` (version 57).

Species	Assembly	Chr.	±	5'-MALAT1	3'-MALAT1	5'-mascRNA
Anolis carolinensis	anoCar1	s.443	+	1097270	1104518	1104209
Xenopus tropicalis +	xenTro2	s.398	+	900641	910280	910287
Xenopus tropicalis -	xenTro2	s.398	−	936780	924118	925349
Danio rerio	Zv8	14	−	47564239	47570301	47564238
Tetraodon nigroviridis	tetNig2	1	−	8135090	...	8130318
Takifugu rubripes	fr2	Un	+	240961971	...	240966770
Gasterosteus aculeatus	gasAcu1	IV	−	5027768	...	5022122
Oryzias latipes	oryLat2	10	+	8151439	8156861	8156576

MascRNA and menRNA. Both MALAT1 and MENβ have a highly structured 3' end, consisting of a hairpin structure, the genomically encoded polyA motive, and the tRNA-like structure that is cleaved off and becomes a stable cytoplasmic ncRNA. This common structure is described in some detail in [17] for mascRNA (MALAT1 associated RNA) and in the supplemental material of [16] for menRNA, see also Fig. 2. The menRNA is by far the best-conserved part of the MENβ transcript. It is easily identified in the two metatheria (*Monodelphis* and *Macropus*) [18]. Although a menRNA homolog is missing from both shotgun traces and the genome assembly of platypus, it is possible to identify other homologous sequences near the 3' end of MENβ. In contrast, no potential ortholog of MENβ/ϵ or menRNA can be found in outside mammalia.

A short region in the lizard genome aligned in the UCSC genome browser to the menRNA region (anoCar1, scaffold 944:1210-1465[-]) cannot be identified unambiguously as the 3'-end of a MENβ ortholog, because a `blastn` search yields 42 similar homologs throughout the lizard genome. Their sequences were retrieved together with about 200nt flanking sequence and aligned (with `clustalw`) to the corresponding regions surrounding mascRNAs and menRNAs. All lizard sequences clearly appear as monophyletic group in this tree (Fig. 3), indicating lineage-specific proliferation of mascRNA. The data are consistent with (but do not provide a conclusive proof for) the origin of MENβ through a duplication of MALAT1, probably in the mammalian stem lineage. The frog genome contains two divergent, and hence ancient, copies of MALAT1 in an unexpected tail-to-tail configurations. The phylogenetic analysis does not provide any evidence that one of these copies might be the ancestor of MENβ.

MascRNA and menRNA are clearly homologous [18], Fig. 2, and the similarities of MENβ and MALAT1 extend upstream of the cleavage site to include a hairpin structure and the genomically encoded poly-A tract [16]. At least parts of the MENβ thus may have arisen from a duplication of MALAT1 in the mammalian ancestor. The lack of recognizable homologies further towards the 5' end could be explained by the poor overall conservation of MENβ.

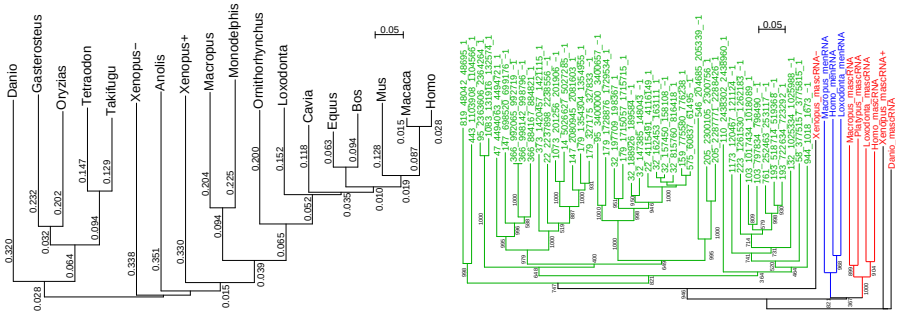

Fig. 3. Left: The vertebrate MALAT1 sequences are alignable and a neighbor-joining tree reproduces the established vertebrate phylogeny quite well, except for the positioning of the two marsupials (*Monodelphis* and *Macropus*) outside of the platypus sequence. **Right:** Neighbor-joining tree of the mascRNA and menRNA loci with about 200nt flanking sequence on both sides. Lizard sequences are shown in green, tetrapod mascRNAs in red, and tetrapod menRNAs blue.

Gene Phylogeny. The MALAT1 transcript is easily recognizable in all mammals [15]. Significant `blastn` hits can also be found in the available genome data of all five sequences teleosts, the elephant shark, the frog, and the lizard. In particular, the mascRNA can be identified unambiguously [18]. In addition to sequence homology, EST data can be used to determine the approximate extent of the MALAT1 transcript in several non-mammalian gnathostomes, see Tab. 2. Approximate full-length sequences were retrieved from the genomic data compiled in the UCSC genome browser and aligned using `clustalw` [36]. Visual inspection of the alignment shows that it indeed consists of homologous sequences. The neighbor-joining tree constructed from this alignment is shown in Fig. 3. It conforms to the established phylogeny of vertebrates with the exception of the relative position of marsupials and platypus, which can be interpret as a long branch attraction artifact.

Promoters. Not much is known about the transcriptional regulation of MALAT1 and MENβ/ϵ. There is evidence for alternative transcription start sites for the human MALAT1 transcript(s): In addition to the longer transcript reported e.g. in [15], a shorter isoform ($\sim$ 7kb) is produced from a CREB-sensitive promoter that can be stimulated by oxytocin [37]. This start sites matches that of mouse *hepcarcin* [20]. Fig. 4 shows that the core promoter region is well conserved within mammals. The alignment of all vertebrate MALAT1 sequences, however, does not provide evidence for a conservation of this feature in other gnathostomes.

Conserved Secondary Structure Elements. Many ncRNAs exhibit evolutionarily conserved secondary structures. Surveys of the human genome [38, 39], for instance, identified tens of thousands of conserved structural motifs. The alignment of the MALAT1 sequences was screened with `RNAz` [38]. As expected, the mascRNA locus and the adjacent conserved hairpin structure just upstream

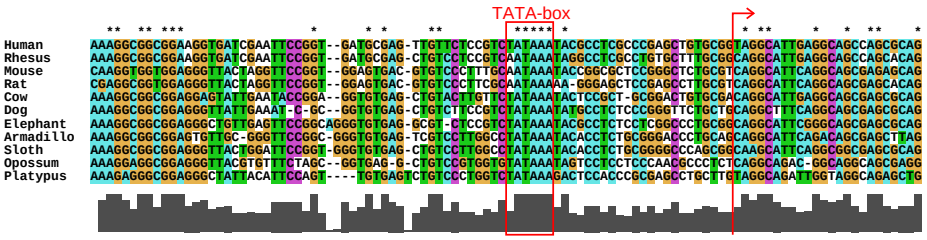

Fig. 4. Conserved promoter of the shorter form of MALAT1 reported in [37], which fits the 5' end of the mouse "hepcarcin" RNA.

of the RNase P processing site [16] was identified as structured region. Despite the size of the MALAT1 transcripts, however, `RNAz` detected only one additional structured location about 600nt upstream of the processing site, see Fig. 2. For MENβ, only the menRNA and a small structured region near the 5'-end of the transcript were detected.

Small Processing Products. A plethora of different types of small RNA products have been detected in eukaryotic genomes, ranging from microRNAs, piRNAs, and endogenous siRNAs [40, 41] to multiple families of small RNAs associated with mRNAs [42, 43]. Several studies using modern high throughput sequencing technologies reported that well-known ncRNA loci are also processed to give rise to small RNAs. MicroRNA precursor hairpins, for instance, are frequently processed to produce additional "off-set RNAs" (moRNAs) that appear to function like mature miRs [44, 45], tRNAs are cleaved to yield multiple shorter products [46–49], snoRNAs frequently give rise to specific miRNA-like short RNAs [50], and a functional short RNA product derives from a vault RNA [51, 52]. The production of small RNA products is a ubiquitous phenomenon that is strongly associated with secondary structure [53].

Here, several published short-read sequencing data sets as well as an extensive library of short RNAs from human brains kindly provided by Philipp Khaitovich [45, 51] is re-evaluated. After mapping the entire dataset to the genome with `segemehl` [54], the subset localized in the MALAT1/MENβ region was extracted. Both MALAT1 (Fig. 5) and MENβ (Fig. 6) give rise to relatively high levels of

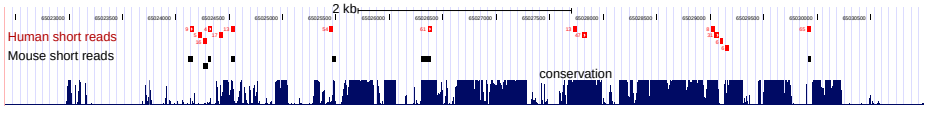

Fig. 5. Conservation of short read expression between human (top) and mouse (below). For comparison, sequence conservation is shown at the bottom of the browser image. Only the most highly expressed blocks of reads are indicated. The genome browser panel covers exactly the annotated human MALAT1 transcript.

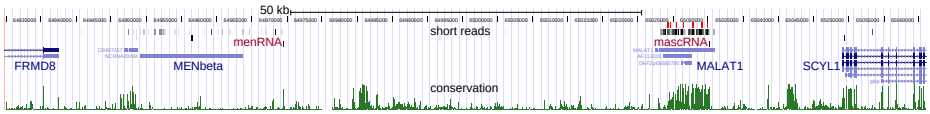

Fig. 6. A diverse set of short-reads is also produced over the complete length of the MENβ transcript. The tRNA-like menRNA is located at small highly conserved locus the very end of the solid EST bar.

short RNA products with a length < 30nt. A comparison of the human and mouse reads shows that several of the most highly expressed locations in the human libraries are also detectable in the much smaller mouse data set GEO: GPL7195 [55]. Surprisingly, this syntenic conservation neither correlates with evolutionary conservation of either sequence or secondary structure. In contrast to MALAT1 and MENβ, most protein-coding transcripts do not give rise to similar patterns of short RNA product. It is unclear whether the processing into small RNAs is a generic feature of nuclear-retained transcripts. The *Xist* transcript, which behaves similarly to MALAT1 and MENβ/MENϵ in several respects [15], does not produce any short reads in any of the investigated libraries. Either *Xist* is simply not expressed under any of the conditions/tissues used here, or its processing is indeed distinct from that MALAT1 and MENβ/MENϵ.

4 Concluding Remarks

The detailed investigation of the MALAT1/MENϵ locus reveals several surprising facts about MALAT1 (conservation at least throughout gnathostomes, the presence of an internal promoter that is conserved across mammals) and MENβ (a probable origin in the mammalian stem lineage) and highlights several commonalities between them: the previously described processing of the 3'-ends by RNase P including the production of small tRNA-like cytoplasmic ncRNA [17, 16], the absence of conserved secondary structures almost everywhere else in the transcript, and the production of many well-defined short RNA products.

On the other hand, this case-study highlighted serious practical difficulties in the comparative analysis of long mlncRNAs. The generally low level of sequence conservation calls for alignment tools that are optimized for this problem. Current alignment editors cannot effectively handle sequences several kb in length and landmarks, such a promoter elements, structured RNA motifs, ESTs, or splice sites cannot be annotated directly in the alignment. Only a few "finished genomes" provide sequences that do not contain gaps or assembly errors over a length of several 10000nt, calling for more efficient ways to explicitly treat missing data in multiple sequence alignents. Thus, detailed case studies are not only of interest in their own right, but are also a necessary prerequisite for the design and development of computational tools that can efficiently assist the analysis of long ncRNAs.

From the biological point of view, the most interesting question concerns the evolutionary origin of mlncRNAs. So far, *Xist* is the only example for which a satisfactory answer — loss of coding capacity of the *Lnx3* transcript and inclusion of adjacent repetitive sequence elements — is known. In the case of MALAT1 and MENβ no candidate for a possible evolutionary precursor could be identified. It seems that mascRNA and menRNA originally derive from tRNAs, similar to, e.g., BC1 and BC200 [56]. MALAT1 and MENβ, like *Xist*, thus are probably composites deriving from several ancestral genomic elements. Interestingly, the large 3' part of MENβ that is not part of the NEAT1/MENϵ transcript consists to a large extent of old SINE (mostly Alu) and a few LINE elements. In contrast, MALAT1 and NEAT1/MENϵ are (nearly) devoid of annotated repeat-derived sequences.

Acknowledgements. Thanks to Phillip Khaitovich for access to short-read sequencing data, to David Langenberger and Steve Hoffmann for access to their short-read maps, and to Manja Marz for comments on the manuscript.

References

1. The ENCODE Project Consortium: Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature **447** (2007) 799–816
2. Maeda, N., Kasukawa, T., Oyama, R., Gough, J., Frith, M., Engström, P.G., Lenhard, B., Aturaliya, R.N., Batalov, S., Beisel, K.W., Bult, C.J., Fletcher, C.F., Forrest, A.R., Furuno, M., Hill, D., Itoh, M., Kanamori-Katayama, M., Katayama, S., Katoh, M., Kawashima, T., Quackenbush, J., Ravasi, T., Ring, B.Z., Shibata, K., Sugiura, K., Takenaka, Y., Teasdale, R.D., Wells, C.A., Zhu, Y., Kai, C., Kawai, J., Hume, D.A., Carninci, P., Hayashizaki, Y.: Transcript annotation in FANTOM3: Mouse Gene Catalog based on physical cDNAs. PLoS Genetics **2** (2006) e62
3. Ponjavic, J., Ponting, C.P., Lunter, G.: Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. Genome Res **17** (2007) 556–565
4. Marques, A.C., Ponting, C.P.: Catalogues of mammalian long noncoding RNAs: modest conservation and incompleteness. Genome Biol. **10** (2009) R124
5. Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., Cabili, M.N., Jaenisch, R., Mikkelsen, T.S., Jacks, T., Hacohen, N., Bernstein, B.E., Kellis, M., Regev, A., Rinn, J.L., Lander, E.S.: Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature **458** (2009) 223–227
6. Pang, K.C., Frith, M.C., Mattick, J.S.: Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. Trends Genetics **22** (2006) 1–5
7. Gardner, P.P., Daub, J., Tate, J.G., Nawrocki, E.P., Kolbe, D.L., Lindgreen, S., Wilkinson, A.C., Finn, R.D., Griffiths-Jones, S., Eddy, S.R., Bateman, A.: Rfam: updates to the RNA families database. Nucleic Acids Res. **37** (2009) D136–D140
8. Griffiths-Jones, S., Saini, H.K., van Dongen, S., Enright, A.J.: miRBase: tools for microRNA genomics. Nucleic Acids Res. **36** (2008) D154–D158
9. Lestrade, L., Weber, M.J.: snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. Nucleic Acids Res. **34** (2006) D158–D162

10. Smits, G., Mungall, A.J., Griffiths-Jones, S., Smith, P., Beury, D., Matthews, L., Rogers, J., Pask, A.J., Shaw, G., VandeBerg, J.L., McCarrey, J.R., SAVOIR Consortium, Renfree, M.B., Reik, W., Dunham, I.: Conservation of the H19 noncoding RNA and H19-IGF2 imprinting mechanism in therians. Nat Genet. **40** (2008) 971–976
11. Park, S.W., Kang, Y.I., Sypula, J.G., Choi, J., Oh, H., Park, Y.: An evolutionarily conserved domain of roX2 RNA is sufficient for induction of H4-Lys16 acetylation on the Drosophila X chromosome. Genetics **177** (2007) 1429–1437
12. Duret, L., Chureau, C., Samain, S., Weissenbach, J., Avner, P.: The *Xist* RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. Science **312** (2006) 1653–1655
13. Elisaphenko, E.A., Kolesnikov, N.N., Shevchenko, A.I., Rogozin, I.B., Nesterova, T.B., Brockdorff, N., Zakian, S.M.: A dual origin of the *Xist* gene from a protein-coding gene and a set of transposable elements. PLoS One **3** (2008) e2521
14. Maenner, S., Blaud, M., Fouillen, L., Savoye, A., Marchand, V., Dubois, A., Sanglier-Cianférani, S., Van Dorsselaer, A., Clerc, P., Avner, P., Visvikis, A., Branlant, C.: 2-D structure of the A region of *Xist* RNA and its implication for PRC2 association. PLoS Biol. **8** (2010) e1000276
15. Hutchinson, J., Ensminger, A.W., Clemson, C.M., Lynch, C.R., Lawrence, J.B., Chess, A.: A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. BMC Genomics **8** (2007) 39
16. Sunwoo, H., Dinger, M.E., Wilusz, J.E., Amaral, P.P., Mattick, J.S., Spector, D.L.: MEN ϵ/β nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. Genome Res. **19** (2009) 347–359
17. Wilusz, J.E., Freier, S.M., L., S.D.: 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. Cell **135** (2008) 919–932
18. Marz, M., Stadler, P.F.: Homology search for small structured ncRNAs. In Hartmann, R., Bindereif, A., Schön, A., Westhof, E., eds.: Handbook of RNA Biochemistry. 2nd edn. Wiley VCH, Weinheim, D (2010) in press.
19. Ji, P., Diederichs, S., Wang, W., Böing, S., Metzger, R., Schneider, P.M., Tidow, N., Brandt, B., Buerger, H., Bulk, E., Thomas, M., Berdel, W.E., Serve, H., Müller-Tidow, C.: MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. Oncogene **22** (2003) 8031–8041
20. Lin, R., Maeda, S., Liu, C., Karin, M., Edgington, T.S.: A large noncoding RNA is a marker for murine hepatocellular carcinomas and a spectrum of human carcinomas. Oncogene **26** (2007) 851–858
21. Guffanti, A., Iacono, M., Pelucchi, P., Kim, N., Soldà, G., Croft, L.J., Taft, R.J., Rizzi, E., Askarian-Amiri, M., Bonnal, R.J., Callari, M., Mignone, F., Pesole, G., Bertalot, G., Rossi Bernardi, L., Albertini, A., Lee, C., Mattick, J.S., Zucchi, I., De Bellis, G.: A transcriptional sketch of a primary human breast cancer by 454 deep sequencing. BMC Genomics **10** (2009) 163
22. Guo, F., Li, Y., Liu, Y., Wang, J., Li, Y., Li, G.: Inhibition of metastasis-associated lung adenocarcinoma transcript 1 in CaSki human cervical cancer cells suppresses cell proliferation and invasion. Acta. Biochim. Biophys. Sin. (Shanghai) **42** (2010) 224–229
23. Saha, S., Rangarajan, P.N.: Common host genes are activated in mouse brain by Japanese encephalitis and rabies viruses. J. Gen. Virol. **84** (2003) 1729–1735

24. Saha, S., Murthy, S., Rangarajan, P.N.: Identification and characterization of a virus-inducible non-coding RNA in mouse brain. J. Gen. Virol. **87** (2006) 1991–1995
25. Sasaki, Y.T.F., Ideue, T., Sano, M., Mituyama, T., Hirose, T.: MENϵ/β noncoding RNAs are essential for structural integrity of nuclear paraspeckles. Proc Natl Acad Sci USA **106** (2009) 2525–2530
26. Peyman, J.A.: Mammalian expression cloning of two human trophoblast suppressors of major histocompatibility complex genes. Am. J. Reprod. Immunol. **45** (2001) 382–392
27. Geirsson, A., Paliwal, I., Lynch, R.J., Bothwell, A.L., Hammond, G.L.: Class II transactivator promoter activity is suppressed through regulation by a trophoblast noncoding RNA. Transplantation **76** (2003) 387–394
28. Geirsson, A., Lynch, R.J., Paliwal, I., Bothwell, A.L., Hammond, G.L.: Altered nuclear retention of Human trophoblast noncoding RNA suppresses CIITA promoter III activity in murine B-lymphocytes. Biochem. Biophys. Res. Commun. **301** (2003) 718–724
29. Lehnert, S.A., Reverter, A., Byrne, K.A., Wang, Y., Nattrass, G.S., Hudson, N.J., Greenwood, P.L.: Gene expression studies of developing bovine longissimus muscle from two different beef cattle breeds. BMC Dev. Biol. **7** (2007) 95
30. Chen, L., Carmichael, G.: RNAs containing inverted repeats in human embryonic stem cells: Functional role of a nuclear noncoding RNA. Mol. Cell **35** (2009) 467–478
31. Clemson, C.M., Hutchinson, J.N., Sara, S.A., Ensminger, A.W., Fox, A.H., Chess, A., Lawrence, J.B.: An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. Mol. Cell **33** (2009) 717–726
32. Bond, C.S., Fox, A.H.: Paraspeckles: nuclear bodies built on long noncoding RNA. J. Cell Biol. **186** (2009) 637–644
33. Murthy, S.U.M., Rangarajan, P.N.: Identification of protein interaction regions of VINC/NEAT1/Men epsilon RNA. FEBS Lett. **584** (2010) 1531–1535
34. Scadden, D.: A NEAT way of regulating nuclear export of mRNAs. Mol. Cell **35** (2009) 395–396
35. Douaud, M., Fève, K., Gerus, M., Fillon, V., Bardes, S., Gourichon, D., Dawson, D.A., Hanotte, O., Burke, T., Vignoles, F., Morisson, M., Tixier-Boichard, M., Vignal, A., Pitel, F.: Addition of the microchromosome GGA25 to the chicken genome sequence assembly through radiation hybrid and genetic mapping. BMC Genomics **9** (2008) 129
36. Thompson, J.D., Higgs, D.G., Gibson, T.J.: CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties, and weight matrix choice. Nucl. Acids Res. **22** (1994) 4673–4680
37. Koshimizu, T.A., Fujiwara, Y., Sakai, N., Shibata, K., Tsuchiya, H.: Oxytocin stimulates expression of a noncoding RNA tumor marker in a human neuroblastoma cell line. Life Sci. **86** (2010) 455–460
38. Washietl, S., Hofacker, I.L., Stadler, P.F.: Fast and reliable prediction of noncoding RNAs. Proc. Natl. Acad. Sci. USA **102** (2005) 2454–2459
39. Pedersen, J.S., Bejerano, G., Siepel, A., Rosenbloom, K., Lindblad-Toh, K., Lander, E.S., Kent, J., Miller, W., Haussler, D.: Identification and classification of conserved RNA secondary structures in the human genome. PLoS Comput. Biol. **2** (2006) e33
40. Moazed, D.: Small RNAs in transcriptional gene silencing and genome defence. Nature **457** (2009) 413–420

41. Tanzer, A., Riester, M., Hertel, J., Bermudez-Santana, C.I., Gorodkin, J., Hofacker, I.L., Stadler, P.F.: Evolutionary genomics of microRNAs and their relatives. In Caetano-Anolles, G., ed.: Evolutionary Genomics. Wiley (2010) in press.
42. Kapranov, P., Cheng, J., Dike, S., Nix, D., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermüller, J., Hofacker, I.L., Bell, I., Cheung, E., Drenkow, J., Dumais, E., Patel, S., Helt, G., Madhavan, G., Piccolboni, A., Sementchenko, V., Tammana, H., Gingeras, T.R.: RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science **316** (2007) 1484–1488
43. Taft, R.J., Kaplan, C.D., Simons, C., Mattick, J.S.: Evolution, biogenesis and function of promoter-associated RNAs. Cell Cycle **8** (2009) 2332–2338
44. Shi, W., Hendrix, D., Levine, M., Haley, B.: A distinct class of small RNAs arises from pre-miRNA-proximal regions in a simple chordate. Nat Struct Mol Biol. **16** (2009) 183–189
45. Langenberger, D., Bermudez-Santana, C., Hertel, J., Hoffmann, S., Khaitovich, S., Stadler, P.F.: Evidence for human microRNA-offset RNAs in small RNA sequencing data. Bioinformatics **25** (2009) 2298–2301
46. Jöchl, C., Rederstorff, M., Hertel, J., Stadler, P.F., Hofacker, I.L., Schrettl, M., Haas, H., Hüttenhofer, A.: Small ncRNA transcriptome analysis from *Aspergillus fumigatus* suggests a novel mechanism for regulation of protein-synthesis. Nucleic Acids Res. **36** (2008) 2677–2689
47. Kawaji, H., Hayashizaki, Y.: Exploration of small RNAs. PLoS Genet **4** (2008) e22
48. Cole, C., Sobala, A., Lu, C., Thatcher, S.R., Bowman, A., Brown, J.W., Green, P.J., Barton, G.J., Hutvagner, G.: Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. RNA **15** (2009) 2147–2160
49. Lee, Y.S., Shibata, Y., Malhotra, A., Dutta, A.: A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). Genes Dev. **23** (2009) 2639–2649
50. Ender, C., Krek, A., Friedländer, M.R., Beitzinger, M., Weinmann, L., Chen, W., Pfeffer, S., Rajewsky, N., Meister, G.: A human snoRNA with microRNA-like functions. Mol. Cell **32** (2008) 519–528
51. Stadler, P.F., Chen, J.J.L., Hackermüller, J., Hoffmann, S., Horn, F., Khaitovich, P., Kretzschmar, A.K., Mosig, A., Prohaska, S.J., Qi, X., Schutt, K., Ullmann, K.: Evolution of vault RNAs. Mol. Biol. Evol. **26** (2009) 1975–1991
52. Persson, H., Kvist, A., Vallon-Christersson, J., Medstrand, P., Borg, A., Rovira, C.: The non-coding RNA of the multidrug resistance-linked vault particle encodes multiple regulatory small RNAs. Nat. Cell Biol. **11** (2009) 1268–1271
53. Langenberger, D., Bermudez-Santana, C., Stadler, P.F., Hoffmann, S.: Identification and classification of small RNAs in transcriptome sequence data. Pac. Symp. Biocomput. **15** (2010) 80–87
54. Hoffmann, S., Otto, C., Kurtz, S., Sharma, C., Khaitovich, P., Stadler, P.F., Hackermüller, J.: Fast mapping of short sequences with mismatches, insertions and deletions using index structures. PLoS Comp. Biol. **5** (2009) e1000502
55. Babiarz, J.E., Ruby, J.G., Wang, Y., Bartel, D.P., Blelloch, R.: Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. Genes Dev. **22** (2008) 2773–2785
56. Volff, J.N., Brosius, J.: Modern genomes with retro-look: Retrotransposed elements, retroposition and the origin of new genes. Genome Dyn. **3** (2007) 175–190