

Antisense Transcripts With FANTOM2 Clone Set and Their Implications for Gene Regulation

Hidenori Kiyosawa,^{1,2,4} Itaru Yamanaka,¹ Naoki Osato,¹ Shinji Kondo,¹ RIKEN GER Group¹ and GSL Members,^{3,5} and Yoshihide Hayashizaki^{1,2,3,6}

¹Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan; ²Division of Genomic Information Resource Exploration, Science of Biological Supramolecular Systems, Yokohama City University, Graduate School of Integrated Science, Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan; ³Genome Science Laboratory, RIKEN, Hirosawa, Wako, Saitama 351-0198, Japan

We have used the FANTOM2 mouse cDNA set (60,770 clones), public mRNA data, and mouse genome sequence data to identify 2481 pairs of sense–antisense transcripts and 899 further pairs of nonantisense bidirectional transcription based upon genomic mapping. The analysis greatly expands the number of known examples of sense–antisense transcript and nonantisense bidirectional transcription pairs in mammals. The FANTOM2 cDNA set appears to contain substantially large numbers of noncoding transcripts suitable for antisense transcript analysis. The average proportion of loci encoding sense–antisense transcript and nonantisense bidirectional transcription pairs on autosomes was 15.1 and 5.4%, respectively. Those on the X chromosome were 6.3 and 4.2%, respectively. Sense–antisense transcript pairs, rather than nonantisense bidirectional transcription pairs, may be less prevalent on the X chromosome, possibly due to X chromosome inactivation. Sense and antisense transcripts tended to be isolated from the same libraries, where nonantisense bidirectional transcription pairs were not apparently coregulated. The existence of large numbers of natural antisense transcripts implies that the regulation of gene expression by antisense transcripts is more common than previously recognized. The viewer showing mapping patterns of sense–antisense transcript pairs and nonantisense bidirectional transcription pairs on the genome and other related statistical data is available on our Web site.

The level of mRNA in a eukaryotic cell, and its translation into protein, can be controlled at many levels subsequent to transcription initiation. Because mRNA is single stranded, the presence of a complementary antisense strand may alter transcription, elongation, processing, location stability, and translation. Functional antisense RNA has been identified in bacteria (review by Wagner and Simons 1994) and also implicated in gene regulation and differentiation in several eukaryotic organisms (Terry and Rouze 2000; Elmendorf et al. 2001), including mammals (review by Dolnick 1997; Vanhee-Brossollet and Vaquero 1998). Natural antisense transcripts usually arise via separate transcription initiation from the opposite DNA strand at the same genomic locus as the sense strand.

They may be coding or noncoding RNA (ncRNA) complementary to mature processed sense coding mRNA, or they may be complementary only to the primary unprocessed transcript, being contained solely within an intron or overlapping a 5' UTR or 3' UTR. They may or may not be spliced. There are now many examples of functional antisense transcripts in developmental gene regulation (review by Vanhee-Brossollet and Vaquero 1998) and imprinting (Jong et al. 1999b; Mitsuya et al. 1999; Hayward and Bonthonron 2000;

Chamberlain and Brannan 2001; Li et al. 2002; Sleutels et al. 2002), but the number of well-characterized antisense transcripts is still small (Dolnick 1997; Vanhee-Brossollet and Vaquero 1998). In each case, the natural antisense transcript was discovered in the course of studies on the sense RNA.

A genome-wide search for possible antisense transcripts was performed recently (Lehner et al. 2002), and 87 pairs of sense–antisense transcripts that originated from the same chromosomal locus were identified. This search focused mainly on sense–antisense mRNA containing open reading frames, whereas in many known functional examples at least one partner is nonprotein coding.

Another recent search for antisense transcripts has been done on an actual transcript sampling from cDNA libraries from the protozoan parasite *Giardia lamblia* (Elmendorf et al. 2001). Random sampling of 100 cDNA clones from the libraries revealed 23 clones that appeared noncoding, of which three were antisense transcripts complementary to the protein-coding mRNA.

The simultaneous availability of the draft genomic sequence (<http://genome-archive.cse.ucsc.edu/>) and large cDNA resources in the mouse permits the first global analysis of the frequency of genuine antisense transcription in a mammal. In this study, we have searched for pairs of sense–antisense transcripts mainly based upon the FANTOM2 cDNA sequence set, a product of the Mouse Gene Encyclopedia Project in the RIKEN Genomic Sciences Center (The FANTOM Consortium and The RIKEN Genome Exploration Research Group Phase I and II Team 2002; <http://fantom2.gsc.riken.go.jp/db/>). This set contains 60,770 full-length enriched

⁴Present address: RIKEN Tsukuba Institute, BioResource Center (BRC), Tsukuba, Ibaraki, 305-0074, Japan.

⁵Takahiro Arakawa, Piero Carninci, and Jun Kawai.

⁶Corresponding author.

E-MAIL yoshihide@gsc.riken.go.jp; FAX 45-5039216.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.982903>.

		category	description	count (pair)	count from same library (pair)
bidirectional transcription pair	sense-antisense transcript pair	1	overlapping exons, one gene is intronless	1252	480
		2	overlapping exons, both genes with introns	1229	274
	non-antisense bidirectional transcript pair	3	non-overlapping exons, one gene is intronless	311	11
		4	non-overlapping exons, both genes with introns (one gene is in one intron of other gene)	308	12
		5	non-overlapping exons, both genes with introns (exons appear alternately)	280	10

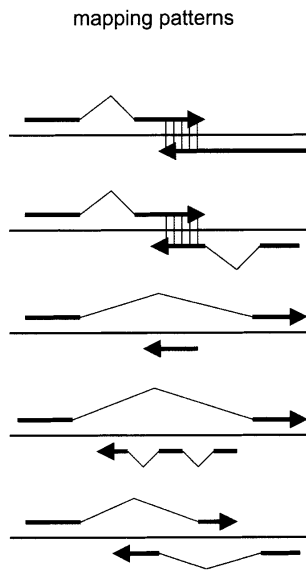


Figure 1 Classification of sense–antisense transcript and nonantisense bidirectional transcription pair patterns. Five categories of bidirectional transcription patterns are shown. The categories are classified according to the patterns of how the two transcripts are mapped on the genome sequences. The schematic examples from each category are shown next to each explanation. Total pair counts as well as pair counts that are from the same cDNA library sources for each category are presented.

to both sense–antisense transcript pair and nonantisense bidirectional transcription pair. The categories and the number of the pairs in each category are summarized in Figure 1. The histogram showing the distribution of the size of overlaps between the pairs of sense–antisense transcript is presented in Figure 2. The number of the pairs mapped on each chromosome is shown in Table 1. The mapping pattern on each chromosome is also shown in Figure 3. One striking feature, which suggests that the mapping pattern of the sense–antisense transcript pairs is nonrandom, is that far fewer pairs were mapped on the X chromosome. The fewer pairs mapped on the X chromosome were specific to sense–antisense transcript pairs, and this phenomenon was not observed in the case of nonantisense bidirectional transcription pairs.

Overall Analysis of the Bidirectional Transcription Pairs

mouse cDNA sequences derived from the libraries of various tissues at various developmental time points. We identified more than 2000 examples of sense–antisense transcript pairs, clearly indicating the prevalence of this mechanism and its potential importance in mammals.

The actual mapping patterns of each bidirectional transcription pair on the genome sequences and accompanying data can be viewed at <http://genome.gsc.riken.go.jp/m/antisense/viewer/>. Examples of the data on this Web site are shown in Figure 4. The types of analyses performed are listed in the field (the most upper line) of Figure 4C. The cDNA sequence

RESULTS

Listing Sense–Antisense Pairs of cDNA Sequences

To list all of the antisense cDNA sequences, the entire set of FANTOM2 cDNA sequences and public mRNA sequences was mapped on the mouse genome sequence draft. The set of the cDNA sequence pairs that originated from the same locus but from opposite strand were selected. These pairs were classified according to the categories depending upon the nature of the overlap. The analysis includes pairs in which the cDNA sequences do not overlap, but they derive from the same locus. We call this type of pair “nonantisense bidirectional transcription pair” (categories 3, 4, and 5 in Fig. 1) in this article. In such cases, we infer that the transcription and/or processing of one member of the pair could interfere with that of the other. We call the pair that shares sequence complementarity with each other “sense–antisense transcript pair” (categories 1 and 2 in Fig. 1). The words “bidirectional transcription pair” (categories 1, 2, 3, 4, and 5 in Fig. 1) refer

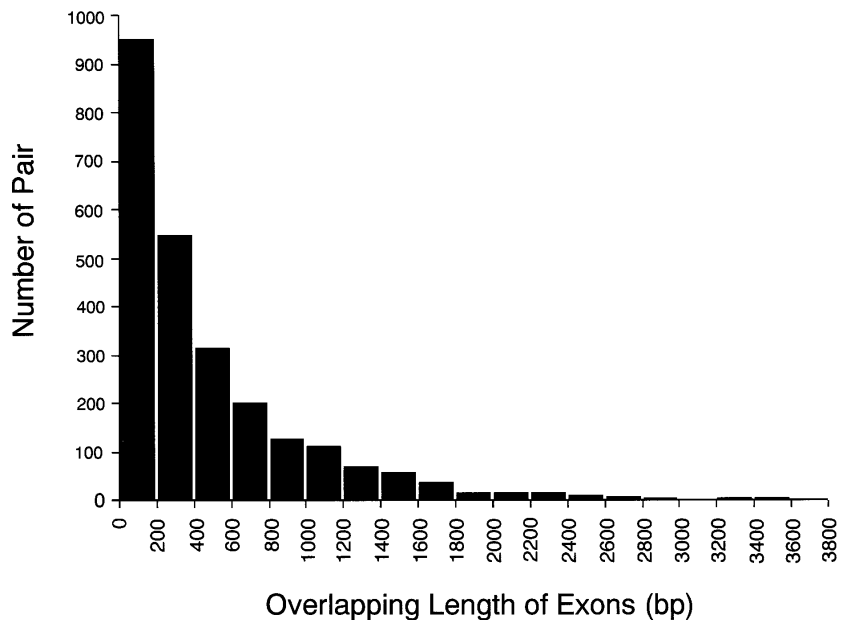


Figure 2 Histogram of overlapping length distribution. The distribution of sense–antisense transcript pairs for their overlapping length of exons is shown. The horizontal axis represents overlapping length (bp) of exons for each sense–antisense transcript pair. The vertical axis represents the number of pair for each bin.

Table 1. Number of Bidirectional Transcription Pairs per Chromosome

Chromosome	Number of sense-antisense transcript pair	Sense-antisense transcript pair loci/total loci (%)	Number of nonantisense transcription pair	Nonantisense transcription pair loci/total loci (%)	Total number of loci ^a	Length of chromosome ^b (bp)	Number of nonbidirectional transcription loci
1	133	12.5	59	5.5	2116	196842934	1732
2	225	16.9	79	5.9	2651	180335396	2043
3	128	14.5	32	3.6	1756	160564582	1436
4	134	14	48	5	1912	151730910	1548
5	154	14.4	71	6.6	2129	151006098	1679
6	138	13.8	69	6.9	1999	150316567	1585
7	183	15.9	62	5.4	2290	135793178	1800
8	143	16.8	40	4.7	1699	129321983	1333
9	136	15.1	52	5.7	1795	125583845	1419
10	123	14.9	42	5	1648	131187037	1318
11	207	17.4	51	4.3	2371	122883361	1855
12	107	15.6	34	4.9	1371	114251360	1089
13	93	13.4	39	5.6	1382	117315093	1118
14	96	14.6	45	6.8	1310	116006794	1028
15	90	14	34	5.2	1284	104633288	1036
16	75	12.7	33	5.6	1172	99184200	956
17	113	16	38	5.4	1407	94132929	1105
18	65	13.4	23	4.7	969	91189200	793
19	105	19.7	26	4.9	1061	61356199	799
X	33	6.3	22	4.2	1038	147161770	928
Total or average	2481	14.8	899	5.3	33360	2580796724	26600

^aLoci number based on our mapping data of FANTOM2 clone set, RefSeq, and GenBank mRNA sequences.

^bChromosome length based on v3 assembly of MGSC.

counts (Fig. 1; 2481 pairs) belonging to categories 1 and 2, the actual sense-antisense transcript pairs, are striking, considering the small number of sense-antisense gene pairs previously reported (Dolnic 1997; Vanhee-Brossollet and Vaquero 1998; Lehner et al. 2002). Many of the antisense transcripts, especially in category 2, were processed by removal of at least one intron, based upon mapping to the genome sequences. This is one supporting fact that these are genuine transcripts. To confirm that these are reproducibly transcribed, we performed a FASTA-search for these sense-antisense candidates against the EST database. At least one EST sequence hit was obtained in 2265 (category 1), 2246 (category 2), 518 (category 3), 542 (category 4), and 497 (category 5) cDNA sequences (please see our Web site for the exact number of EST support for each cDNA sequence; Fig. 4C). The histogram indicating the overall distribution of EST hits for the bidirectional transcription pairs is also presented in Figure 5.

To function in regulation of its complementary partner, a natural antisense transcript might be expected to be coexpressed, although the alternative would be that the expression is exclusive. To test these alternatives, we traced the library origin of each cDNA sequence. The FANTOM2 cDNA was fully sequenced for the representative clone after clustering with 3' end sequences of over one million clones (phase I sequences; The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium 2001; The FANTOM Consortium and The RIKEN Genome Exploration Research Group Phase I and II Team 2002). When we counted the number of the common libraries for sense and antisense cDNA, we took all the phase I sequences into account and used the library origin of phase I sequences homologous to sense-antisense sequences with at least 96% identity. The count of the common library source for each bidirectional transcription pair is shown on the Web (examples given in

Fig. 4C). The lists of cDNA libraries that produced both members of bidirectional transcription pair were shown on the Web (<http://genome.gsc.riken.go.jp/m/antisense/>). There was no major apparent enrichment of bidirectional transcription pairs in any particular library. The numbers of the bidirectional transcription pairs expressed in the same library were 480, 274, 11, 12, and 10 pairs for the categories 1 through 5, respectively (Fig. 1). Thus, there is a significant bias that the sense-antisense transcript pairs tended to be isolated from the same library sources while nonantisense bidirectional transcription pairs did not.

We investigated CDS (protein coding sequence) in sequences of sense-antisense transcript pairs. Because we used potential CDS data of FANTOM2 collaboration (The FANTOM Consortium and The RIKEN Genome Exploration Research Group Phase I and II Team 2002), first we chose sense-antisense transcript pairs consisting of FANTOM2 sequences out of our bidirectional transcription pairs. There were 1924 sense-antisense transcript pairs that have more than 20 bp of exon overlapping region. Among them, 519 pairs had potential CDS (more than 300 bp) in both sense and antisense strands. In 1054 pairs, only one member of each pair had potential CDS (more than 300 bp). In 351 pairs, no members of each pair had potential CDS (more than 300 bp). Thus, antisense transcripts tended to be ncRNA.

The mapping pattern of sense-antisense transcript and nonantisense bidirectional transcription pairs on the genome can be viewed by clicking "image" or "applet," given as an example at the right end of Figure 4C. The examples of the viewer is shown in Figure 4D. The 5' and 3' directions are represented by the blue and red colors, respectively. The green waveform indicates the position of a CpG island. Other consensus sequences found at the promoter region are also shown.

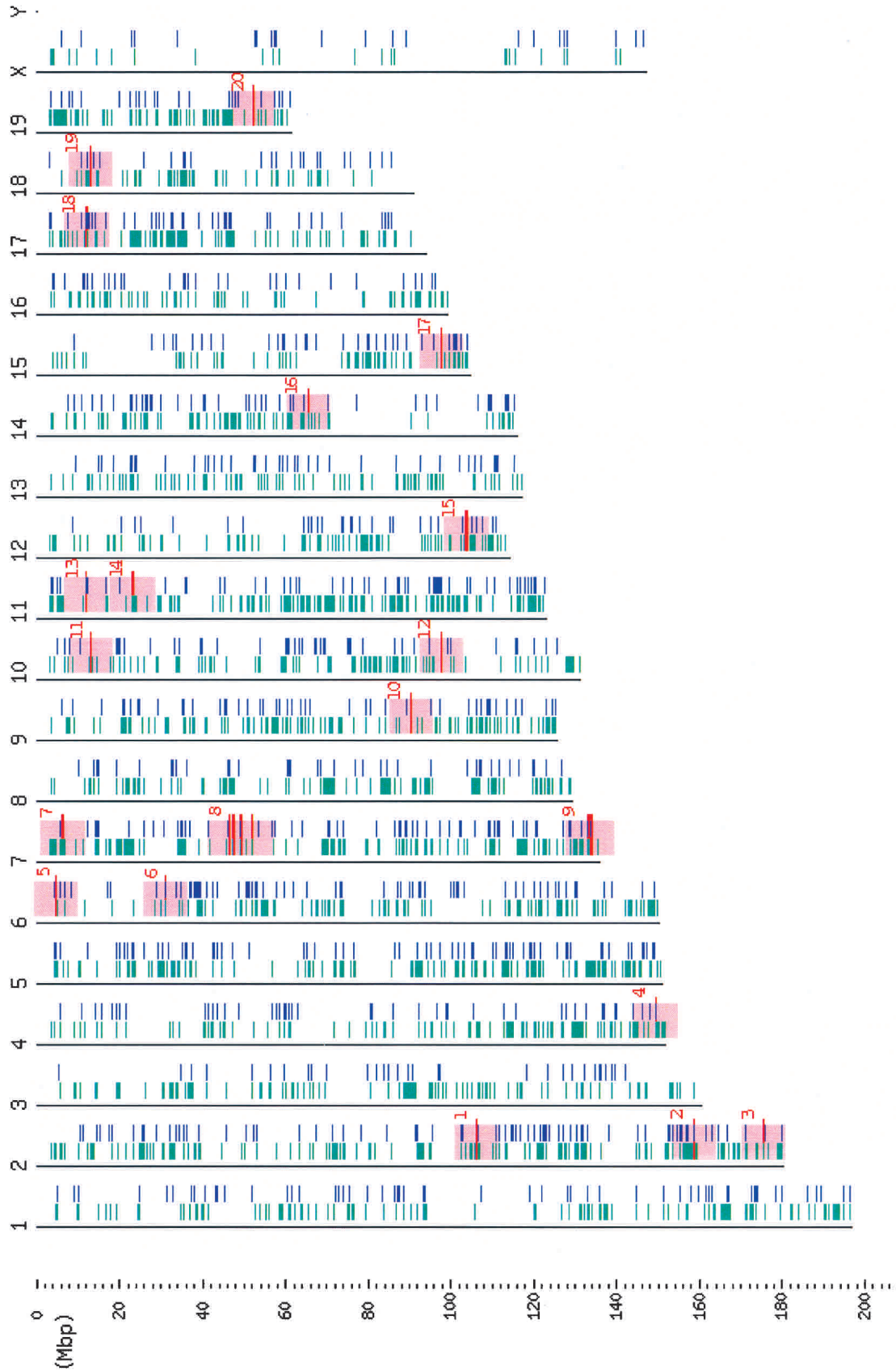


Figure 3 Chromosome map of sense-antisense transcript and nonantisense bidirectional transcription pairs. All of the mapped positions of the bidirectional transcription pairs are schematically presented. The sense-antisense transcript and nonantisense bidirectional transcription pairs are presented in green on the left and in blue on the right, respectively. The red lines indicate the positions of the known imprinted genes. The pink areas are the regions ± 5 Mb of the known imprinted gene positions. Note that there are a few mapped positions of the sense-antisense transcripts on the X chromosome, compared with the autosomes. The names of known imprinted genes mapped in each pink region are: *Wf1* (1); *Nnat* (2); *Gnas*, *Gnasxl*, *Nespas* (3); *p73* (4); *Sage* (5); *Copa2*, *Mit1/lb9* (6); *Peg3/Pw1*, *Usp29*, *Zim1*, *Zim3*, *Zfp264* (7); *Frat3*, *GABRA5*, *GABRB3*, *GABRG3*, *Magel2*, *Mkrr3/Zfp127*, *Ndn*, *Snrpn*, *Snurf*, *Ube3a* (8); *H19*, *Igf2*, *Igf2as*, *Ins2*, *Ipl/Tssc3*, *Kvlt1*, *Mash2*, *Nap1l4*, *Obppl1*, *ps7KIP2/Cdkn1c*, *Slc22a1l*, *Tapat1/Cd81*, *Tssc4* (9); *Rasgrf1* (10); *Zac1* (11); *Dcn* (12); *Meg1/Grb10* (13); *U2af1-rs1 region1*, *U2af1-rs1 region2* (14); *Dlk*, *Meg3/Gtl2* (15); *Htr2a* (16); *Ata3* (17); *Igf2r*, *Mas1*, *Slc22a2*, *Slc22a3* (18); *Impact* (19); *Ins1* (20). Because the genes on the X chromosome are imprinted only on the paternal X chromosome in the extraembryonic tissues (reviewed by Latham 1996), we did not shade the X chromosome with pink color. Although the chromosomal locations of *lpw*, *Kvlt1-as*, *Msuit*, *Peg1/Mest*, and *Pwcr1* are already known, these genes were not computationally mapped by BLAST program. Accordingly, these genes are not shown in Figure 3.

A Antisense Viewer

Sorted by Chromosome

Chromosome: [1](#), [2](#), [3](#), [4](#), [5](#), [6](#), [7](#), [8](#), [9](#), [10](#), [11](#), [12](#), [13](#), [14](#), [15](#), [16](#), [17](#), [18](#), [19](#), [X](#)

Sorted by Category

- Category 1 (Annotation)

Chromosome: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [11](#) [12](#) [13](#) [14](#) [15](#) [16](#) [17](#) [18](#) [19](#) [X](#)

- Category 2 (Annotation)

Chromosome: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [11](#) [12](#) [13](#) [14](#) [15](#) [16](#) [17](#) [18](#) [19](#) [X](#)

- Category 3 (Annotation)

- Category 4 (Annotation)

- Category 5 (Annotation)

B Annotation Category: 1

[B430425B13](#): procollagen, type IV, alpha 3
[C430014B12](#): unknown EST

[9630033H02](#): hypothetical ARM repeat structure containing protein
[9530008A22](#): hypothetical protein

[2310057K05](#): hypothetical protein
[6030459624](#): nucleolin

[D130058F05](#): gastrulation brain homeobox 2
[NM_010262](#): Mus musculus gastrulation brain homeobox 2 (Gbx2), mRNA

[D430044C15](#): unknown EST
[9330186K11](#): hypothetical Arginine-rich region containing protein

[J700030C16](#): hypothetical protein
[A430088H15](#): SIMILAR TO EHM2 GENE homolog [Mus musculus]

[4631402I02](#): hypothetical protein
[1700003B17](#): unclassifiable

[NM_007923](#): Mus musculus ELK4, member of ETS oncogene family (Elk4), mRNA
[D930039D09](#): hypothetical protein

[BC019133](#): Mus musculus, clone IMAGE:5036938, mRNA.
[C130020J17](#): unknown EST

C CHROMOSOME: 1

Category	Overlapping Length of Exons (bp)	Chromosome	Library of Simultaneous Expression	Clone ID	Percent of Overlapping Exons to Total Length (%)	EST Hit (Count)	ORF (> 100 amino acids)	Intron	Genome Start Position	Genome End Position	Links
1	805	1	2	D130041E14	28.4	1	0	0	4505773	4508600	Image
				4930417H09	24.7	8	1	3	4500612	4506577	Applet
1	158	1	0	BC006022	7.1	94	-	9	4866243	4905954	Image
				B430206H13	14.3	1	0	0	4865691	4866788	Applet
4	0	1	0	A930002L11	0	2	0	2	4980449	4986413	Image
				2900073E09	0	8	1	4	4918541	5028424	Applet
3	0	1	0	A430104I20	0	2	0	0	8999054	9001166	Image
				A230066K05	0	8	1	2	8830118	9048010	Applet
2	154	1	0	6330565B14	8.9	49	1	13	9597562	9625892	Image
				2610203C22	23.2	4	1	3	9612996	9680778	Applet
2	385	1	5	A630029A12	12	18	1	4	9650954	9676781	Image
				C230099L04	11	3	0	2	9650101	9680784	Applet
2	409	1	3	1700034P13	30.6	3	0	6	9797185	9841240	Image
				4932442A08	10.2	22	1	2	9773734	9797593	Applet
2	60	1	0	E330040D14	4.2	3	0	1	10020787	10021883	Image
				BC022722	4.2	3	-	8	10021824	10062454	Applet

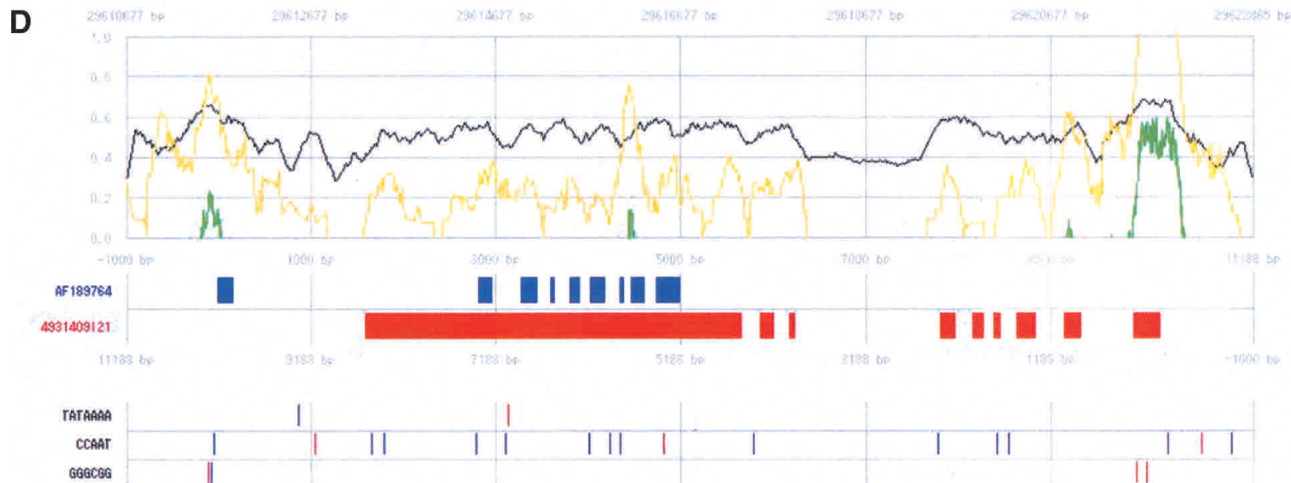


Figure 4 Examples of antisense viewer. All data of the bidirectional transcription pairs can be viewed at <http://genome.gsc.riken.go.jp/m/antisense/viewer/>. (A) Entrance of the viewer. The data are either sorted by chromosome or category. By clicking "Annotation," the FANTOM2 annotation of each cDNA (B) will be shown. By clicking chromosome numbers or category numbers, the pages for the description of the bidirectional transcription pairs will be shown (C). (B) Gene names by FANTOM annotation are shown for each sense-antisense cDNA pairs. (C) Data for each bidirectional transcription pair are shown. By clicking either "Image" or "Applet," graphical mapping patterns of cDNA on the genome sequences (D) will be presented. (D) Mapping patterns of bidirectional transcription cDNA pairs on the genome. Forward and reverse directions on the chromosomes are represented in blue and red colors, respectively. The filled boxes represent the positions of exons. The black line represents GC content along the genome sequences. The yellow and green lines represent a ratio of the observed to expected CpG score, and a final CpG score, respectively (please see the Methods section for a detailed calculation of these scores). The positions of major promoter consensus sequences are shown underneath the positions of exons.

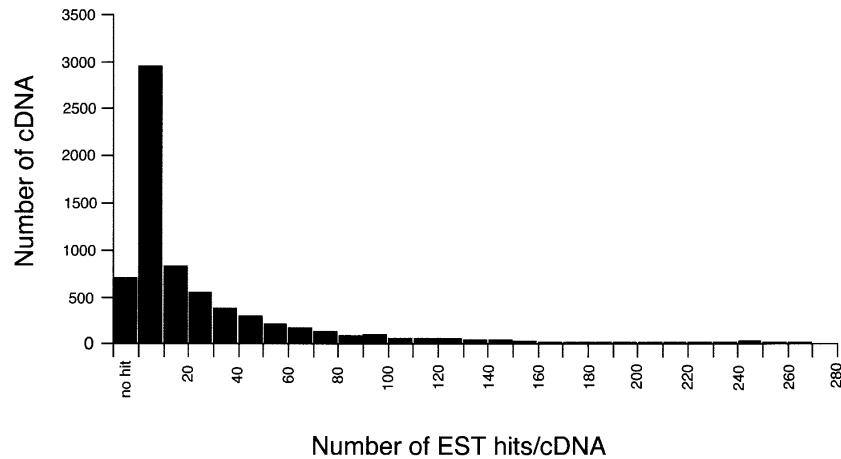


Figure 5 Histogram of number of EST hits distribution. The distribution of bidirectional transcription pair cDNA for the number of EST hits is shown. The horizontal axis represents the number of EST hits for each bidirectional transcription pair cDNA. The vertical axis represents the number of cDNA for each bin. We also analyzed distributions for sense–antisense transcript pairs and nonantisense bidirectional transcription pairs separately, but there was no significant difference between these distributions (data not shown).

Known Sense–Antisense Transcript Pairs Found in Our Search

We searched the known sense–antisense transcript pairs in all categories of bidirectional transcription pair data. First, we chose 19 sense cDNA sequences of protein coding genes that have been reportedly accompanied by antisense transcripts, and searched for them among our bidirectional transcription pairs (see the Methods section for a list of these 19 genes). Seven (*Hif1a*, *Mkrm2*, *Raf1*, *Erc1*, *Cd3z*, *Thra*, and *Hoxa11*) of the 19 genes had antisense sequences in our list of bidirectional transcription.

Sense–Antisense Transcript and Nonantisense Bidirectional Transcription Pairs Related to Known Imprinted Genes

As approximately 15% of the known imprinted genes accompany the antisense transcripts that are suspected to regulate imprinting of the sense gene (Reik and Walter 2001), we investigated the relationship between the imprinted regions on the chromosomes and the chromosomal positions of the cDNA sequences of all bidirectional transcription pairs. The positions of known imprinted genes and the chromosome regions within ± 5 Mb are highlighted in Figure 3. The cDNA mapped in the pink regions (± 5 Mb of known imprinted genes) is listed on the Web (http://genome.gsc.riken.go.jp/m/antisense/imprinted_genes/). There was no apparent relationship between the mapping positions of imprinted genes and the bidirectional transcription pairs. We also searched antisense sequences for the known imprinted genes in our bidirectional transcription pair sequences. Out of 58 imprinted genes (see the Methods section for the names of these 58 genes), 22 genes (*Copg2*, *GABRA5*, *Gnas*, *Gnasxl*, *Htr2a*, *Igf2*, *Igf2r*, *Kvlqt1*, *Magel2*, *Mas1*, *Mit1/lb9*, *Ndn*, *Nespas*, *Nnat*, *Slc22a3*, *Tapa1/Cd81*, *U2af1-rs1 region1*, *U2af1-rs1 region2*, *Ube3a*, *Usp29*, *Zfp264*, and *Zim3*) had bidirectional sequences.

DISCUSSION

This is the first systematic analysis of sense–antisense transcripts using a comprehensive full-length transcript sequence set. Major findings of this analysis are: (1) As many as 2481 pairs of sense–antisense transcript were identified; (2) there is a strong bias in the frequency of the mapping patterns of the sense–antisense transcript pairs (few only on the X chromosome); (3) cDNA clones of the sense–antisense transcript pairs tended to be isolated from the same cDNA library sources. These findings are made possible with the large-scale isolation of full-length cDNA from the RIKEN Mouse Gene Encyclopedia Project. (The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium 2001; The FANTOM Consortium and The RIKEN Genome Exploration Research Group Phase I and II Team, 2002.) Although the mouse genome sequence draft is now available for extensive bioinformatic analyses (Mouse Genome Sequencing Consortium

2002), predicting noncoding antisense transcripts on the genome sequences by computer programs is almost impossible because virtually all the existing programs are designed to predict protein-coding regions (exons). Recently, trials to predict ncRNA were attempted using computational algorithms (Argaman et al. 2001; Carter et al. 2001; Rivas et al., 2001; Wassarman et al. 2001), but these attempts were possible only for ncRNA that shares structural similarities. The antisense RNA is expected to function through sequence complementarity to the sense strand. This type of ncRNA cannot be predicted by the methods based on RNA sequence or structural similarities. The most recent review of natural antisense transcripts listed 20 pairs of sense–antisense transcript in mammals (Vanhee-Brossollet and Vaquero 1998). The number we counted for sense–antisense transcript pair is 2481. The gene expression regulation by the natural antisense transcripts is now recognized as common in mammals (Dolnick 1997; Vanhee-Brossollet and Vaquero 1998), but this number is still much greater than previously envisaged.

Although the numbers of the cDNA sequences mapped on the X chromosome were relatively small compared to those on similar-length autosomes (see Table 1), the ratio of sense–antisense transcript pairs mapped on the X chromosome was particularly low. By contrast, the percentage of nonantisense bidirectional transcription pairs that do not generate complementary products, mapped on the X chromosome, was similar to those on other chromosomes. These differences in the numbers between sense–antisense transcript pairs and nonantisense bidirectional transcription pairs were also found in the library origins as mentioned in the Results section (Fig. 1), and indicate that there is a basic, biologic difference in the nature of the two types of transcription. Precisely, the fact that both cDNA of the sense–antisense transcript pair were isolated from the same cDNA library source does not necessarily mean that the pairs of the transcripts existed in the same single cell. Expression analyses such as RT-PCR with the RNA from a single cell, or RNA-FISH may be necessary to confirm the existence of both of the pair in a

single cell. Yet the bias found in the cDNA library source differences between two types of bidirectional transcription pairs (Fig. 1) is still significant.

The small number of antisense transcripts on the X chromosome argues that the regulation of gene expression by antisense transcripts may have something to do with X chromosome inactivation. If the sense and antisense transcripts are expressed in a mono-allelic manner, each of which is expressed only from either the paternal or maternal chromosome, and if both sense and antisense transcripts are necessary for the regulation of those sense and antisense transcripts' expression, loci encoding sense-antisense transcript pairs may have been excluded on the X chromosome during evolution.

The actual biologic functions of these natural antisense transcripts in living organisms are hardly known, leaving a reasonable speculation that they form a double-stranded RNA (dsRNA) to downregulate the expression of sense RNA molecules. The dsRNA would prevent single-stranded mRNA from interacting with cellular components for normal gene expression. Alternatively, the resultant dsRNA could be a target for RNA interference (RNAi). The molecular mechanisms of RNAi began to be revealed (recent reviews by Hutvagner and Zamore 2002; Zamore 2002), and the target genes can be efficiently knocked out with dsRNA in model organisms such as worms and flies (Fire et al. 1998; Caplen et al. 2001; Elbashir et al. 2001), as well as mammalian cells (Brummelkamp et al. 2002; Paddison et al. 2002; Sui et al. 2002). However, the native biologic function and meaning of RNAi are still unknown, except for a few examples, such as RNAi-like, antiviral capability in plants (Kasschau and Carrington 1998) or degradation of *Stellate* mRNA by a RNAi-like mechanism in *Drosophila* (Aravin et al. 2001). Virtually all of the sense-antisense transcripts found in our analysis could be the target of a dsRNA-dependent RNAi mechanism, if both sense and antisense transcripts were produced in the same single cell and the concentration of the transcripts were high enough for the RNAi mechanism to proceed with its function. The abundance of such sense and antisense transcripts in living mice indicates that gene expression regulation by sense-antisense transcripts, possibly via a RNAi-like mechanism, might be fairly common.

The number of genes or transcripts on the genome has always been a hot area of scientific concern, especially in the genome era (Ewing and Green 2000; Liang et al. 2000; Roest Crolius et al. 2000). An estimation of genes from the human genome sequences is roughly between 30,000 and 40,000 (International Human Genome Sequencing Consortium 2001; Venter et al. 2001). However, these numbers are given based on the protein-coding RNA. Recently, the importance of ncRNA began to be widely accepted (review by Eddy 2001; Storz 2002). A subset of very small ncRNA exists among ncRNA, for example, in microRNA (miRNA), *lin-4* (22 bp) (Lee et al. 1993) and in *Caenorhabditis elegans*, *let-7* (21 bp) (Reinhart et al. 2000). The FANTOM cDNA set does not contain such small transcripts, but appears to contain substantial numbers of noncoding RNA (The FANTOM Consortium and The RIKEN Genome Exploration Research Group Phase I and II Team 2002), which is an excellent source for analyzing genome-wide, noncoding transcripts. The transcript number on the genome might be much higher because these noncoding RNAs also have important functions. A recent work analyzing actual transcriptional activity in the selected regions of chromosomes 21 and 22 revealed as much as an order of magni-

tude and more genomic sequences are being transcribed (Kapranov et al. 2002). As we presented in the Results section, the antisense transcripts are not always ncRNA but tend to be ncRNA. As one type of ncRNA, antisense transcripts may play essential roles in cellular gene expression regulation. Confirmation of the actual expression and function of these antisense transcripts will contribute to evaluation of the total number of genes on the genome.

METHODS

Mapping cDNA Sequences Onto the Mouse Genome Sequences

We used the mapping data calculated in the FANTOM2 collaboration. (The FANTOM Consortium and The RIKEN Genome Exploration Research Group Phase I and II Team 2002) In the data, the FANTOM2 cDNA sequence set (60,770 sequences), RefSeq mouse sequences (<http://www.ncbi.nih.gov/RefSeq/>) and GenBank mRNA sequences were attempted to map on the mouse genome sequence (MGSCv3) assembly (Mouse Genome Sequencing Consortium 2002). In this article, we refer to the region where cDNA sequences align with the genome sequences' "exon" expedient. We extracted sense-antisense transcript pairs and nonantisense bidirectional transcription pairs from the map data based on loci overlapping. We call the list of the extracted pairs as redundant version of our pair list. We eliminated redundancy with the following procedure: First, we split cDNA sequences into exon-overlapping and exon-nonoverlapping groups. In the exon-overlapping group, we made clusters based on exon overlapping. For each cluster, we chose a pair having the longest exon overlapping as a representative pair. In the exon-nonoverlapping group, we made clusters based on loci overlapping. In this case, we chose a pair that has the longest loci overlapping as a representative pair for each cluster. Then we split the exon-overlapping group into two categories and the exon-nonoverlapping group into three categories. Definitions and the number of pairs of these five categories are as follows. Category 1—exons overlapping: one of the genes is intronless; 1252 pairs were found. Category 2—exons overlapping: both genes have introns; 1229 pairs were found. Category 3—no exons overlapping: one of the genes is intronless; 311 pairs were found. Category 4—no exons overlapping: both genes have introns. One of the genes is within one intron of the other gene; 308 pairs were found. Category 5—no exons overlapping: both genes have introns. Several exons of two genes appear alternately; 280 pairs were found.

Identification of CpG Islands

A CpG island is a region of vertebrate DNA with a high GC content and a high frequency of CpG. Three criteria for the CpG island are defined as follows (Gardiner-Garden and Frommer 1987):

1. "A region of DNA used for evaluation of its base content should be longer than 200 bp in length." Our computer program with a window size of 200 bp met this condition. In Figure 4D, for all bases between 900 bp upstream of genes and 900 bp downstream, the following quantities were calculated by shifting the window one base pair per each calculation.
2. "GC content should be over 0.5." GC content = [(number of G) + (number of C)]/(window size). In Figure 4D, a black line represents GC content.
3. "CpG score defined below should be over 0.6." (observed CpG) = (number of CpG)/(window size)

$$\text{(expected CpG)} = \frac{[(\text{number of C})/(\text{window size})]}{* [(\text{number of G})/(\text{window size})]}$$

$$\text{(CpG score)} = \text{observed/expected} = \frac{[(\text{number of CpG}) * (\text{window size})]}{[(\text{number of C}) * (\text{number of G})]}$$

CpG score is represented as a yellow line in Figure 4D. Because the threshold of the CpG score is 0.6, a final CpG island score is defined as the CpG score minus 0.6, and the criterion (3) is equivalent to a positive final CpG island score. In Figure 4D, the positive final CpG island score is indicated by a green line only in a region of DNA that satisfies the criterion (2). Therefore, the green line represents the existence of a CpG island in the DNA region, because the region satisfies all three criteria explained above.

Search for EST Sequences Homologous to Bidirectional Transcription Pairs

To confirm that sense–antisense candidates are derived from natural transcripts, we performed a FASTA search (Pearson and Lipman 1988) for these sense–antisense candidates against approximately one million EST sequences annotated as mouse or *Mus musculus*, and a 5' end in the EST division of the GenBank database. The number of EST sequences that match a sense–antisense candidate with $\geq 94\%$ identical with $\geq 80\%$ overall length was counted for each candidate.

Library Origin of Bidirectional Transcription Pairs

To analyze the library origin of the sequences from bidirectional transcription, we used the library origin of not only the full-length sequences, but also the mouse end sequences highly homologous to the full-length sequences. After we masked repeat sequences in all the sequences from the redundant version of our pair list using RepeatMasker software (<http://ftp.genome.washington.edu/RM/RepeatMasker.html>), we searched the sequences against approximately 1.3 million 3' end sequences using BLAST software (Altschul et al. 1990). The library origins of 3' end sequences with $\geq 94\%$ identity were added to the library origin of the matched sense or antisense sequences. Next, we counted the frequency of libraries in which both sense and antisense clones expressed simultaneously as follows. For each library, we counted the number of FANTOM clones expressed in the library and the number of sense and antisense clones expressed in the library, and then calculated the percentage of the number of the sense and antisense clones to the number of the FANTOM clones. We sorted the names of the library origins included in each category of bidirectional transcription pairs by their ratios and listed them (<http://genome.gsc.riken.go.jp/m/antisense/>). We also counted the numbers of sense and antisense clones expressed only in a unique library and listed them (<http://genome.gsc.riken.go.jp/m/antisense/>).

A List of Known Sense Transcripts With Antisense Transcripts and a Search of Antisense Transcripts

The following are known mammal cDNA sequences reported to accompany antisense transcripts selected from the recent literature and used in this study: *Fgf2* (Knee et al. 1997), *Hif1a* (Thrash-Bingham and Tartof 1999), *Tgfb2* (Coker et al. 1998), *Tnfrsf17* (Hatzoglou et al. 2002), *Mkx2* (Gray et al. 2001), *Hfe* (Thenie et al. 2001), *Klhl1* (Benzow and Koob 2002), *Ucn* (Shi et al. 2000), *Wt1* (Moorwood et al. 1998), *Tmi1* (Podlowski et al. 2002), *Thra* (Hastings et al. 2000), *Hsp70.2* (Murashov and Wolgemuth 1996), *Hoxd-3* (Bedford et al. 1995), *Tyms* (Dolnick 1993), *CD3z* (Lerner et al. 1993), *N-myc* (Krystal et al. 1990), *Trp53* (Khochbin and Lawrence 1989), *Ercc1* (van Duin et al. 1989), and *Hoxa11* (Hsieh-Li et al. 1995). To identify the known sense–antisense transcripts in the bidirectional tran-

scription pairs calculated in this article, we looked up the GenBank# or Refseq# of the known transcripts in the redundant version of our pair list. In addition, we performed a FASTA search (Pearson and Lipman 1988) for the known sense–antisense genes against bidirectional transcription pair sequences. The sequences $\geq 96\%$ identical with $\geq 80\%$ overall length were selected.

A List of Known Imprinted Genes and a Search for Antisense Transcripts of Known Imprinted Genes

Symbols used as known imprinted gene references are as follows; *Ata3* (Mizuno et al. 2002), *Copg2* (Lee et al. 2000), *Dcn* (Mizuno et al. 2002), *Dlk* (Takada et al. 2000), *Frat3* (Chai J-H et al. 2001), *Gabra5* (Knoll et al. 1993), *Gabrb3* (Wagstaff et al. 1991), *Gabrg3* (Greger et al. 1995), *Gnas* (Williamson et al. 1996), *Gnasxl* (Peters et al. 1999), *H19* (Leighton et al. 1995), *Htr2a* (Kato et al. 1998), *Igf2* (De Chiara et al. 1991), *Igf2as* (Moore et al. 1997), *Igf2r* (Barlow et al. 1991), *Impact* (Hagiwara et al. 1997), *Ins1* (Giddings et al. 1994), *Ins2* (Leighton et al. 1995), *Ipl/Tssc3* (Qian et al. 1997), *Ipw* (Wevrick and Francke 1997), *Kvlqt1* (Gould and Pfeifer 1998), *Kvlqt1-as* (Smilnich et al. 1999), *Magel2* (Boccaccio et al. 1999), *Mas1* (Villar and Pedersen 1994), *Mash2* (Guillemot et al. 1995), *Meg1/Grb10* (Miyoshi et al. 1998), *Meg3/Gtl2* (Miyoshi et al. 2000), *Mit1/lb9* (Lee et al. 2000), *Mkx3/Zfp127* (Jong et al. 1999a), *Msu1* (Onyango et al. 2000), *Nap114* (Paulsen et al. 1998), *Ndn* (MacDonald and Wevrick 1997), *Nespas* (Wroe et al. 2000), *Nnat* (Kagitani et al. 1997), *Obph1* (Engemann et al. 2000), *p57KIP2/Cdkn1c* (Hatada and Mukai 1995), *p73* (Kaghad et al. 1997), *Peg1/Mest* (Kaneko-Ishino et al. 1995), *Peg3/Pw1* (Kaneko-Ishino et al. 1995), *Pwcr1* (de Los Santos et al. 2000), *Rasgrf1* (Plass et al. 1996), *Sgce* (Piras et al. 2000), *Slc22a11* (Cooper et al. 1998), *Slc22a2* (Zwart et al. 2001), *Slc22a3* (Zwart et al. 2001), *Srrpn* (Leff et al. 1992), *Snruf* (Gray et al. 1999), *Tapa1/Cd81* (Andria et al. 1991), *Tssc4* (Paulsen et al. 2000), *U2af1-rs1* region1 (Nabetani et al. 1997), *U2af1-rs1* region2 (Nabetani et al. 1997), *Ube3a* (Albrecht et al. 1999), *Usp29* (Kim et al. 2000), *Wt1* (Rainier et al. 1993), *Zac1* (Piras et al. 2000), *Zfp264* (Kim et al. 2001), *Zim1* (Kim et al. 1999), and *Zim3* (Kim et al. 2001). These genes were chosen based on the data at <http://www.mgu.har.mrc.ac.uk/imprinting/imprinting.html> and <http://www.geneimprint.com/>. To search for antisense transcripts of known imprinted genes, we looked up the GenBank# or Refseq# of the known imprinted genes in the redundant version of our pair list. In addition, we performed a FASTA search (Pearson and Lipman 1988) for the known imprinted genes against bidirectional transcription pair sequences. The sequences $\geq 96\%$ identical over $\geq 80\%$ overall length were selected.

ACKNOWLEDGMENTS

We thank A. Hasegawa for Web interface. This study was supported by a Research Grant for the RIKEN Genome Exploration Research Project from the Ministry of Education, Culture, Sports, Science and Technology of the Japanese Government to Y.H.

REFERENCES

- Albrecht, U., Sutcliffe, J.S., Cattanch, B.M., Beechey, C.V., Armstrong, D., Eichele, G., and Beaudet, A.L. 1997. Imprinted expression of the murine Angelman syndrome gene, *Ube3a*, in hippocampal and Purkinje neurons. *Nat. Genet.* **17**: 75–78.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**: 403–410.
- Andria, M.L., Hsieh, C.L., Oren, R., Francke, U., and Levy, S. 1991. Genomic organization and chromosomal localization of the TAPA-1 gene. *Immunology* **147**: 1030–1036.

- Aravin, A.A., Naumova, N.M., Tulin, A.V., Vagin, V.V., Rozovsky, Y.M., and Gvozdev, V.A. 2001. Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. *Curr. Biol.* **11**: 1017–1027.
- Argaman, L., Hershberg, R., Vogel, J., Bejerano, G., Wagner, E.G., Margalit, H., and Altuvia, S. 2001. Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr. Biol.* **11**: 941–950.
- Barlow, D.P., Stoger, R., Herrmann, B.G., Saito, K., and Schweifer, N. 1991. The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. *Nature* **349**: 84–87.
- Bedford, M., Arman, E., Orr-Urtreger, A., and Lonai, P. 1995. Analysis of the *Hoxd-3* gene: Structure and localization of its sense and natural antisense transcripts. *DNA Cell Biol.* **14**: 295–304.
- Benzow, K.A. and Koob, M.D. 2002. The *KLHL1*-antisense transcript (*KLHL1AS*) is evolutionarily conserved. *Mamm. Genome* **13**: 134–141.
- Boccaccio, I., Glatt-Deeley, H., Watrin, F., Roedel, N., Lalande, M., and Muscatelli, F. 1999. The human *MAGE2* gene and its mouse homologue are paternally expressed and mapped to the Prader-Willi region. *Hum. Mol. Genet.* **8**: 2497–2505.
- Brummelkamp, T.R., Bernards, R., and Agami, R. 2002. A system for stable expression of short interfering RNAs in mammalian cells. *Science* **296**: 550–553.
- Caplen, N.J., Parrish, S., Iman, F., Fire, A., and Morgan, R.A. 2001. Specific inhibition of gene expression by small double-stranded RNAs in invertebrate and vertebrate systems. *Proc. Natl. Acad. Sci.* **98**: 9742–9747.
- Carter, R.J., Dubchak, I., and Holbrook, S.R. 2001. A computational approach to identify genes for functional RNAs in genomic sequences. *Nucleic Acids Res.* **29**: 3928–3938.
- Chai, J.H., Locke, D.P., Ohta, T., Grealley, J.M., and Nicholls, R.D. 2001. Retrotransposed genes such as *Frat3* in the mouse Chromosome 7C Prader-Willi syndrome region acquire the imprinted status of their insertion site. *Mamm. Genome* **12**: 813–821.
- Chamberlain, S.J. and Brannan, C.I. 2001. The Prader-Willi syndrome imprinting center activates the paternally expressed murine *Ube3a* antisense transcript but represses paternal *Ube3a*. *Genomics* **73**: 316–322.
- Coker, R.K., Laurent, G.J., Dabbagh, K., Dawson, J., and McNulty, R.J. 1998. A novel transforming growth factor $\beta 2$ antisense transcript in mammalian lung. *Biochem. J.* **332**: 297–301.
- Cooper, P.R., Smilnich, N.J., Day, C.D., Nowak, N.J., Reid, L.H., Pearsall, R.S., Reece, M., Prawitt, D., Landers, J., Housman, D.E., et al. 1998. Divergently transcribed overlapping genes expressed in liver and kidney and located in the 11p15.5 imprinted domain. *Genomics* **49**: 38–51.
- DeChiara, T.M., Robertson, E.J., and Efstratiadis, A. 1991. Parental imprinting of the mouse insulin-like growth factor II gene. *Cell* **64**: 849–859.
- de los Santos, T., Schweizer, J., Rees, C.A., and Francke, U. 2000. Small evolutionarily conserved RNA, resembling C/D box small nucleolar RNA, is transcribed from *PWCR1*, a novel imprinted gene in the Prader-Willi deletion region, which is highly expressed in brain. *Am. J. Hum. Genet.* **67**: 1067–1082.
- Dolnick, B.J. 1993. Cloning and characterization of a naturally occurring antisense RNA to human thymidylate synthase mRNA. *Nucleic Acids Res.* **21**: 1747–1752.
- Dolnick, B.J. 1997. Naturally occurring antisense RNA. *Pharmacol. Ther.* **75**: 179–184.
- Eddy, S.R. 2001. Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.* **2**: 919–929.
- Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W., and Tuschl, T. 2001. Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *EMBO J.* **20**: 6877–6888.
- Elmendorf, H.G., Singer, S.M., and Nash, T.E. 2001. The abundance of sterile transcripts in *Giardia lamblia*. *Nucleic Acids Res.* **29**: 4674–4683.
- Engemann, S., Strodicke, M., Paulsen, M., Franck, O., Reinhardt, R., Lane, N., Reik, W., and Walter, J. 2000. Sequence and functional comparison in the Beckwith-Wiedemann region: Implications for a novel imprinting centre and extended imprinting. *Mol. Genet.* **9**: 2691–2706.
- Ewing, B. and Green, P. 2000. Analysis of expressed sequence tags indicates 35,000 human genes. *Nat. Genet.* **25**: 232–234.
- The FANTOM Consortium and The RIKEN Genome Exploration Research Group Phase I and II Team. 2002. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**: 563–573.
- Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**: 806–811.
- Gardiner-Garden, M. and Frommer, M. 1987. CpG islands in vertebrate genomes. *J. Mol. Biol.* **196**: 261–282.
- Giddings, S.J., King, C.D., Harman, K.W., Flood, J.F., and Carnaghi, L.R. 1994. Allele specific inactivation of insulin 1 and 2, in the mouse yolk sac, indicates imprinting. *Nat. Genet.* **6**: 310–313.
- Gould, T.D. and Pfeifer, K. 1998. Imprinting of mouse *Kvlqt1* is developmentally regulated. *Hum. Mol. Genet.* **7**: 483–487.
- Gray, T.A., Azama, K., Whitmore, K., Min, A., Abe, S., and Nicholls, R.D. 2001. Phylogenetic conservation of the *makorin-2* gene, encoding a multiple zinc-finger protein, antisense to the *RAFI* proto-oncogene. *Genomics* **77**: 119–126.
- Gray, T.A., Saitoh, S., and Nicholls, R.D. 1999. An imprinted, mammalian bicistronic transcript encodes two independent proteins. *Proc. Natl. Acad. Sci.* **96**: 5616–5621.
- Greger, V., Knoll, J.H., Woolf, E., Glatt, K., Tyndale, R.F., DeLorey, T.M., Olsen, R.W., Tobin, A.J., Sikela, J.M., Nakatsu, Y., et al. 1995. The γ -aminobutyric acid receptor $\gamma 3$ subunit gene (*GABRG3*) is tightly linked to the $\alpha 5$ subunit gene (*GABRA5*) on human chromosome 15q11-q13 and is transcribed in the same orientation. *Genomics* **26**: 258–264.
- Guillemot, F., Caspary, T., Tilghman, S.M., Copeland, N.G., Gilbert, D.J., Jenkins, N.A., Anderson, D.J., Joyner, A.L., Rossant, J., and Nagy, A. 1995. Genomic imprinting of *Mash2*, a mouse gene required for trophoblast development. *Nat. Genet.* **9**: 235–242.
- Hagiwara, Y., Hirai, M., Nishiyama, K., Kanazawa, I., Ueda, T., Sakaki, Y., and Ito, T. 1997. For imprinted genes by allelic message display: Identification of a paternally expressed gene impact on mouse chromosome 18. *Proc. Natl. Acad. Sci.* **94**: 9249–9254.
- Hastings, M.L., Ingle, H.A., Lazar, M.A., and Munroe, S.H. 2000. Post-transcriptional regulation of thyroid hormone receptor expression by *cis*-acting sequences and a naturally occurring antisense RNA. *J. Biol. Chem.* **275**: 11507–11513.
- Hatada, I. and Mukai, T. 1995. Genomic imprinting of *p57KIP2*, a cyclin-dependent kinase inhibitor, in mouse. *Nat. Genet.* **11**: 204–206.
- Hatzoglou, A., Deshayes, F., Madry, C., Lapree, G., Castanas, E., and Tsipa, A. 2002. Natural antisense RNA inhibits the expression of *BCMA*, a tumour necrosis factor receptor homologue. *BMC Mol. Biol.* **3**: 4.
- Hayward, B.E. and Bonthron, D.T. 2000. An imprinted antisense transcript at the human *GNAS1* locus. *Hum. Mol. Genet.* **9**: 835–841.
- Hsieh-Li, H.M., Witte, D.P., Weinstein, M., Branford, W., Li, H., Small, K., and Potter, S.S. 1995. *Hoxa 11* structure, extensive antisense transcription, and function in male and female fertility. *Development* **121**: 1373–1385.
- Hutvagner, G. and Zamore, P.D. 2002. RNAi: Nature abhors a double-strand. *Curr. Opin. Genet. Dev.* **12**: 225–232.
- International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.
- Jong, M.T., Carey, A.H., Caldwell, K.A., Lau, M.H., Handel, M.A., Driscoll, D.J., Stewart, C.L., Rinchik, E.M., and Nicholls, R.D. 1999a. Imprinting of a RING zinc-finger encoding gene in the mouse chromosome region homologous to the Prader-Willi syndrome genetic region. *Hum. Mol. Genet.* **8**: 795–803.
- Jong, M.T., Gray, T.A., Ji, Y., Glenn, C.C., Saitoh, S., Driscoll, D.J., and Nicholls, R.D. 1999b. A novel imprinted gene, encoding a RING zinc-finger protein, and overlapping antisense transcript in the Prader-Willi syndrome critical region. *Hum. Mol. Genet.* **8**: 783–793.
- Kaghad, M., Bonnet, H., Yang, A., Creancier, L., Biscan, J.C., Valent, A., Minty, A., Chalou, P., Lelias, J.M., Dumont, X., et al. 1997. Monoallelically expressed gene related to *p53* at 1p36, a region frequently deleted in neuroblastoma and other human cancers. *Cell* **90**: 809–819.
- Kagitani, F., Kuroiwa, Y., Wakana, S., Shiroishi, T., Miyoshi, N., Kobayashi, S., Nishida, M., Kohda, T., Kaneko-Ishino, T., and Ishino, F. 1997. *Peg5/Neuronatin* is an imprinted gene located on sub-distal chromosome 2 in the mouse. *Nucleic Acids Res.* **25**: 3428–3432.
- Kaneko-Ishino, T., Kuroiwa, Y., Miyoshi, N., Kohda, T., Suzuki, R., Yokoyama, M., Viville, S., Barton, S.C., Ishino, F., and Surani,

- M.A. 1995. Peg1/Mest imprinted gene on chromosome 6 identified by cDNA subtraction hybridization. *Nat. Genet.* **11**: 52–59.
- Kapranov, P., Cawley, S.E., Drenkow, J., Bekiranov, S., Strausberg, R.L., Fodor, S.P., and Gingeras, T.R. 2002. Large-scale transcriptional activity in chromosomes 21 and 22. *Science* **296**: 916–919.
- Kasschau, K.D. and Carrington, J.C. 1998. A counter-defensive? Strategy of plant viruses: Suppression of posttranscriptional gene silencing. *Cell* **95**: 461–470.
- Kato, M.V., Ikawa, Y., Hayashizaki, Y., and Shibata, H. 1998. Paternal imprinting of mouse serotonin receptor 2A gene Htr2 in embryonic eye: A conserved imprinting regulation on the RB/Rb locus. *Genomics* **47**: 146–148.
- Khochbin, S. and Lawrence, J.J. 1989. An antisense RNA involved in p53 mRNA maturation in murine erythroleukemia cells induced to differentiate. *EMBO J.* **8**: 4107–4114.
- Kim, J., Bergmann, A., Wehri, E., Lu, X., and Stubbs, L. 2001. Imprinting and evolution of two Kruppel-type zinc-finger genes, ZIM3 and ZNF264, located in the PEG3/USP29 imprinted domain. *Genomics* **77**: 91–98.
- Kim, J., Lu, X., and Stubbs, L. 1999. Zim1, a maternally expressed mouse Kruppel-type zinc-finger gene located in proximal chromosome 7. *Hum. Mol. Genet.* **8**: 847–854.
- Kim, J., Noskov, V.N., Lu, X., Bergmann, A., Ren, X., Warth, T., Richardson, P., Kouprina, N., and Stubbs, L. 2000. Discovery of a novel, paternally expressed ubiquitin-specific processing protease gene through comparative analysis of an imprinted region of mouse chromosome 7 and human chromosome 19q13.4. *Genome Res.* **10**: 1138–1147.
- Knee, R., Li, A.W., and Murphy, P.R. 1997. Characterization and tissue-specific expression of the rat basic fibroblast growth factor antisense mRNA and protein. *Proc. Natl. Acad. Sci.* **94**: 4943–4947.
- Knoll, J.H., Sinnott, D., Wagstaff, J., Glatt, K., Wilcox, A.S., Whiting, P.M., Wingrove, P., Sikela, J.M., and Lalande, M. 1993. FISH ordering of reference markers and of the gene for the α 5 subunit of the gamma-aminobutyric acid receptor (*GABRA5*) within the Angelman and Prader-Willi syndrome chromosomal regions. *Hum. Mol. Genet.* **2**: 183–189.
- Krystal, G.W., Armstrong, B.C., and Battey, J.F. 1990. *N-myc* mRNA forms an RNA-RNA duplex with endogenous antisense transcripts. *Mol. Cell. Biol.* **10**: 4180–4191.
- Latham, K.E. 1996. X chromosome imprinting and inactivation in the early mammalian embryo. *Trends Genet.* **12**: 134–138.
- Lee, R.C., Feinbaum, R.L., and Ambros, V. 1993. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**: 843–854.
- Lee, Y.J., Park, C.W., Hahn, Y., Park, J., Lee, J., Yun, J.H., Hyun, B., and Chung, J.H. 2000. *Mit1/Lb9* and *Copg2*, new members of mouse imprinted genes closely linked to *Peg1/Mest*. *FEBS Lett.* **472**: 230–234.
- Leff, S.E., Brannan, C.I., Reed, M.L., Ozcelik, T., Francke, U., Copeland, N.G., and Jenkins, N.A. 1992. imprinting of the mouse *Snrpn* gene and conserved linkage homology with the human Prader-Willi syndrome region. *Nat. Genet.* **2**: 259–264.
- Lehner, B., Williams, G., Campbell, R.D., and Sanderson, C.M. 2002. Antisense transcripts in the human genome. *Trends Genet.* **18**: 63–65.
- Leighton, P.A., Saam, J.R., Ingram, R.S., Stewart, C.L., and Tilghman, S.M. 1995. An enhancer deletion affects both H19 and *Igf2* expression. *Genes Dev.* **9**: 2079–2089.
- Lerner, A., D'Adamio, L., Diener, A.C., Clayton, L.K., and Reinherz, E.L. 1993. *CD3 $\zeta/\eta/\theta$* locus is collinear with and transcribed antisense to the gene encoding the transcription factor *Oct-1*. *J. Immunol.* **151**: 3152–3162.
- Li, T., Vu, T.H., Lee, K.O., Yang, Y., Nguyen, C.V., Bui, H.Q., Zeng, Z.L., Nguyen, B.T., Hu, J.F., Murphy, S.K., et al. 2002. An imprinted *PEG1/MEST* antisense expressed predominantly in human testis and in mature spermatozoa. *J. Biol. Chem.* **277**: 13518–13527.
- Liang, F., Holt, I., Perteau, G., Karamycheva, S., Salzberg, S.L., and Quackenbush, J. 2000. Gene index analysis of the human genome estimates approximately 120,000 genes. *Nat. Genet.* **25**: 239–420.
- MacDonald, H.R. and Wevrick, R. 1997. The *necdin* gene is deleted in Prader-Willi syndrome and is imprinted in human and mouse. *Hum. Mol. Genet.* **6**: 1873–1878.
- Mitsuya, K., Meguro, M., Lee, M.P., Katoh, M., Schulz, T.C., Kugoh, H., Yoshida, M.A., Niikawa, N., Feinberg, A.P., and Oshimura, M. 1999. *LIT1*, an imprinted antisense RNA in the human *KvLQT1* locus identified by screening for differentially expressed transcripts using monochromosomal hybrids. *Hum. Mol. Genet.* **8**: 1209–1217.
- Miyoshi, N., Kuroiwa, Y., Kohda, T., Shitara, H., Yonekawa, H., Kawabe, T., Hasegawa, H., Barton, S.C., Surani, M.A., Kaneko-Ishino, T., et al. 1998. Identification of the *Meg1/Grb10* imprinted gene on mouse proximal chromosome 11, a candidate for the Silver-Russell syndrome gene. *Proc. Natl. Acad. Sci.* **95**: 1102–1107.
- Miyoshi, N., Wagatsuma, H., Wakana, S., Shiroishi, T., Nomura, M., Aisaka, K., Kohda, T., Surani, M.A., Kaneko-Ishino, T., and Ishino, F. 2000. Identification of an imprinted gene, *Meg3/Gtl2* and its human homologue *MEG3*, first mapped on mouse distal chromosome 12 and human chromosome 14q. *Genes Cells* **5**: 211–220.
- Mizuno, Y., Sotomaru, Y., Katsuzawa, Y., Kono, T., Meguro, M., Oshimura, M., Kawai, J., Tomaru, Y., Kiyosawa, H., Nikaido, I., et al. 2002. *Asb4*, *Ata3*, and *Dcn* are novel imprinted genes identified by high-throughput screening using RIKEN cDNA microarray. *Biochem. Biophys. Res. Commun.* **290**: 1499–1505.
- Moore, T., Constancia, M., Zubair, M., Bailleul, B., Feil, R., Sasaki, H., and Reik, W. 1997. Multiple imprinted sense and antisense transcripts, differential methylation and tandem repeats in a putative imprinting control region upstream of mouse *Igf2*. *Proc. Natl. Acad. Sci.* **94**: 12509–12514.
- Moorwood, K., Charles, A.K., Salpekar, A., Wallace, J.I., Brown, K.W., and Malik, K. 1998. Antisense *WT1* transcription parallels sense mRNA and protein expression in fetal kidney and can elevate protein levels in vitro. *J. Pathol.* **185**: 352–359.
- Mouse Genome Sequencing Consortium. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**: 520–562.
- Murashov, A.K. and Wolgemuth, D.J. 1996. Sense and antisense transcripts of the developmentally regulated murine *hsp70.2* gene are expressed in distinct and only partially overlapping areas in the adult brain. *Brain Res. Mol. Brain Res.* **37**: 85–95.
- Nabetani, A., Hatada, I., Morisaki, H., Oshimura, M., and Mukai, T. 1997. Mouse *U2af1-rs1* is a neomorphic imprinted gene. *Mol. Cell. Biol.* **17**: 789–798.
- Onyango, P., Miller, W., Lehoczy, J., Leung, C.T., Birren, B., Wheelan, S., Dewar, K., and Feinberg, A.P. 2000. Sequence and comparative analysis of the mouse 1-megabase region orthologous to the human 11p15 imprinted domain. *Genome Res.* **10**: 1697–1710.
- Paddison, P.J., Caudy, A.A., and Hannon, G.J. 2002. Stable suppression of gene expression by RNAi in mammalian cells. *Proc. Natl. Acad. Sci.* **99**: 1443–1448.
- Paulsen, M., Davies, K.R., Bowden, L.M., Villar, A.J., Franck, O., Fuermann, M., Dean, W.L., Moore, T.F., Rodrigues, N., Davies, K.E., et al. 1998. Syntenic organization of the mouse distal chromosome 7 imprinting cluster and the Beckwith-Wiedemann syndrome region in chromosome 11p15.5. *Hum. Mol. Genet.* **7**: 1149–1159.
- Paulsen, M., El-Maarri, O., Engemann, S., Stroedicke, M., Franck, O., Davies, K., Reinhardt, R., Reik, W., and Walter, J. 2000. Sequence conservation and variability of imprinting in the Beckwith-Wiedemann syndrome gene cluster in human and mouse. *Hum. Mol. Genet.* **9**: 1829–1841.
- Pearson, W.R. and Lipman, D.J. 1988. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci.* **85**: 2444–2448.
- Peters, J., Wroe, S.F., Wells, C.A., Miller, H.J., Bodle, D., Beechey, C.V., Williamson, C.M., and Kelsey, G. 1999. A cluster of oppositely imprinted transcripts at the *Gnas* locus in the distal imprinting region of mouse chromosome 2. *Proc. Natl. Acad. Sci.* **96**: 3830–3835.
- Piras, G., El Kharroubi, A., Kozlov, S., Escalante-Alcalde, D., Hernandez, L., Copeland, N.G., Gilbert, D.J., Jenkins, N.A., and Stewart, C.L. 2000. *Zac1* (*Lot1*), a potential tumor suppressor gene, and the gene for ϵ -sarcoglycan are maternally imprinted genes: Identification by a subtractive screen of novel uniparental fibroblast lines. *Mol. Cell. Biol.* **20**: 3308–3315.
- Plass, C., Shibata, H., Kalcheva, I., Mullins, L., Kotelevtseva, N., Mullins, J., Kato, R., Sasaki, H., Hirotsune, S., Okazaki, Y., et al. 1996. Identification of *Grf1* on mouse chromosome 9 as an imprinted gene by RLGS-M. *Nat. Genet.* **14**: 106–109.
- Podlowski, S., Bramlage, P., Baumann, G., Morano, I., and Luther, H.P. 2002. Cardiac troponin I sense-antisense RNA duplexes in the myocardium. *J. Cell. Biochem.* **85**: 198–207.
- Qian, N., Frank, D., O'Keefe, D., Dao, D., Zhao, L., Yuan, L., Wang,

- Q., Keating, M., Walsh, C., and Tycko, B. 1997. The IPL gene on chromosome 11p15.5 is imprinted in humans and mice and is similar to TDAG51, implicated in Fas expression and apoptosis. *Hum. Mol. Genet.* **6**: 2021–2029.
- Rainier, S., Johnson, L.A., Dobry, C.J., Ping, A.J., Grundy, P.E., and Feinberg, A.P. 1993. Relaxation of imprinted genes in human cancer. *Nature* **362**: 747–749.
- Reik, W. and Walter, J. 2001. Genomic imprinting: Parental influence on the genome. *Nat. Rev. Genet.* **2**: 21–32.
- Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. 2000. The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**: 901–906.
- The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. 2001. Functional annotation of a full-length mouse cDNA collection. *Nature* **409**: 685–690.
- Rivas, E., Klein, R.J., Jones, T.A., and Eddy, S.R. 2001. Computational identification of noncoding RNAs in *E. coli* by comparative genomics. *Curr. Biol.* **11**: 1369–1373.
- Roest Crolius, H., Jaillon, O., Bernot, A., Dasilva, C., Bouneau, L., Fischer, C., Fizames, C., Wincker, P., Brottier, P., Quetier, F., et al. 2000. Estimate of human gene number provided by genome-wide analysis using *Tetraodon nigroviridis* DNA sequence. *Nat. Genet.* **25**: 235–238.
- Shi, M., Yan, X., Ryan, D.H., and Harris, R.B. 2000. Identification of urocortin mRNA antisense transcripts in rat tissue. *Brain Res. Bull.* **53**: 317–324.
- Sleutels, F., Zwart, R., and Barlow, D.P. 2002. The non-coding *Air* RNA is required for silencing autosomal imprinted genes. *Nature* **415**: 810–813.
- Smilnich, N.J., Day, C.D., Fitzpatrick, G.V., Caldwell, G.M., Lossie, A.C., Cooper, P.R., Smallwood, A.C., Joyce, J.A., Schofield, P.N., Reik, W., et al. 1999. A maternally methylated CpG island in KvLQT1 is associated with an antisense paternal transcript and loss of imprinting in Beckwith-Wiedemann syndrome. *Proc. Natl. Acad. Sci.* **96**: 8064–8069.
- Storz, G. 2002. An expanding universe of noncoding RNAs. *Science* **296**: 1260–1263.
- Sui, G., Soohoo, C., el Affar, B., Gay, F., Shi, Y., Forrester, W.C., and Shi, Y. 2002. A DNA vector-based RNAi technology to suppress gene expression in mammalian cells. *Proc. Natl. Acad. Sci.* **99**: 5515–5520.
- Takada, S., Tevendale, M., Baker, J., Georgiades, P., Campbell, E., Freeman, T., Johnson, M.H., Paulsen, M., and Ferguson-Smith, A.C. 2000. *delta*-like and *gtl2* are reciprocally expressed, differentially methylated linked imprinted genes on mouse chromosome 12. *Curr. Biol.* **10**: 1135–1138.
- Terry, N. and Rouze, P. 2000. The sense of naturally transcribed antisense RNAs in plants. 2000. *Trends Plant Sci.* **5**: 394–396.
- Thenie, A.C., Gicquel, I.M., Hardy, S., Ferran, H., Fergelot, P., Le Gall, J.Y., and Mosser, J. 2001. Identification of an endogenous RNA transcribed from the antisense strand of the *HFE* gene. *Hum. Mol. Genet.* **10**: 1859–1866.
- Thrash-Bingham, C.A. and Tartof, K.D. 1999. *aHIF*: A natural antisense transcript overexpressed in human renal cancer and during hypoxia. *J. Natl. Cancer Inst.* **91**: 143–151.
- van Duin, M., van Den Tol, J., Hoeijmakers, J.H., Bootsma, D., Rupp, I.P., Reynolds, P., Prakash, L., and Prakash, S. 1989. Conserved pattern of antisense overlapping transcription in the homologous human *ERCC-1* and yeast *RAD10* DNA repair gene regions. *Mol. Cell. Biol.* **9**: 1794–1798.
- Vanhee-Brossollet, C. and Vaquero, C. 1998. Do natural antisense transcripts make sense in eukaryotes? *Gene* **211**: 1–9.
- Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans C.A., Holt, R.A., et al. 2001. The sequence of the human genome. *Science* **291**: 1304–1351.
- Villar, A.J. and Pedersen, R.A. 1994. Parental imprinting of the *Mas* protooncogene in mouse. *Nat. Genet.* **8**: 373–379.
- Wagner, E.G. and Simons, R.W. 1994. Antisense RNA control in bacteria, phages, and plasmids. *Annu. Rev. Microbiol.* **48**: 713–742.
- Wagstaff, J., Knoll, J.H., Fleming, J., Kirkness, E.F., Martin-Gallardo, A., Greenberg, F., Graham Jr., J.M., Menninger, J., Ward, D., Venter, J.C., et al. 1991. Localization of the gene encoding the GABAA receptor β 3 subunit to the Angelman/Prader-Willi region of human chromosome 15. *Am. J. Hum. Genet.* **49**: 330–337.
- Wassarman, K.M., Repoila, F., Rosenow, C., Storz, G., and Gottesman, S. 2001. Identification of novel small RNAs using comparative genomics and microarrays. *Genes & Dev.* **15**: 1637–1651.
- Wevrick, R. and Francke, U. 1997. An imprinted mouse transcript homologous to the human imprinted in Prader-Willi syndrome (IPW) gene. *Hum. Mol. Genet.* **6**: 325–332.
- Williamson, C.M., Schofield, J., Dutton, E.R., Seymour, A., Beechey, C.V., Edwards, Y.H., and Peters, J. 1996. Glomerular-specific imprinting of the mouse *gsa* gene: How does this relate to hormone resistance in albright hereditary osteodystrophy? *Genomics* **36**: 280–287.
- Wroe, S.F., Kelsey, G., Skinner, J.A., Bodle, D., Ball, S.T., Beechey, C.V., Peters, J., and Williamson, C.M. 2000. An imprinted transcript, antisense to *Nesp*, adds complexity to the cluster of imprinted genes at the mouse *Gnas* locus. *Proc. Natl. Acad. Sci.* **97**: 3342–3346.
- Zamore, P.D. 2002. Ancient pathways programmed by small RNAs. *Science* **296**: 1265–1269.
- Zwart, R., Sleutels, F., Wutz, A., Schinkel, A.H., and Barlow, D.P. 2001. Bidirectional action of the *Igf2r* imprint control element on upstream and downstream imprinted genes. *Genes & Dev.* **15**: 2361–2366.

WEB SITE REFERENCES

- <http://fantom2.gsc.riken.go.jp/db/>; FANTOM2 database Web site.
- <http://ftp.genome.washington.edu/RM/RepeatMasker.html>; RepeatMasker Web site.
- <http://genome.gsc.riken.go.jp/m/antisense/>; Entrance of the data for this article.
- http://genome.gsc.riken.go.jp/m/antisense/imprinted_genes/; list of sense-antisense transcript/transcription mapped +/- 5Mb region of known imprinted genes.
- <http://genome.gsc.riken.go.jp/m/antisense/viewer/>; Entrance of antisense viewer of this article.
- <http://genome-archive.cse.ucsc.edu/>; University of California Santa Cruz working draft archive site.
- <http://www.geneimprint.com/>; Duke University, Jirtle Laboratory.
- <http://www.mgu.har.mrc.ac.uk/imprinting/imprinting.html>; The Harwell (Mammalian Genetics Unit) Mouse Imprinting Site.
- <http://www.ncbi.nih.gov/RefSeq/>; NCBI Reference Sequences home page.

Received December 3, 2002; accepted in revised form February 25, 2003.