(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 March 2007 (08.03.2007)

PCT

(10) International Publication Number
WO 2007/028030 A2

(51) International Patent Classification:
C07C 4/02 (2006.01)

(21) International Application Number:
PCT/US2006/034250

(22) International Filing Date:
1 September 2006 (01.09.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/713,674 2 September 2005 (02.09.2005) US

(71) Applicant *(for all designated States except US)*: **PICOBELLA LP** [US/US]; 863 Mitten Road, Suite 102, Burlingame, CA 94010 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **WABL, Matthias** [US/US]; 1515 Fifth Avenue, San Francisco, CA 94122 (US). **WANG, Bruce** [US/US]; 2408 Villa Nueva Way, Mountain View, CA 94040 (US).

(74) Agents: **MAHONEY, Jacqueline, F.** et al.; Perkins Coie LLP, P.O. Box 2168, Menlo Park, CA 94026 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2007/028030 A2

(54) **Title:** ONCOGENIC REGULATORY RNAS FOR DIAGNOSTICS AND THERAPEUTICS

(57) **Abstract:** A method of identifying regulatory RNAs, including miRNAs, using insertional mutagenesis to generate tumors in mice and determining the human orthologs is disclosed. Further, specific miRNA sequences are identified. The causal nature and expression patterns of these regulatory RNAs and miRNAs in human tumors demonstrate their utility in diagnosis and therapy of cancer. Furthermore, a set of co-mutations that act in conjunction with miRNAs in tumor formation is disclosed.

ONCOGENIC REGULATORY RNAs FOR DIAGNOSTICS AND THERAPEUTICS

TABLES 1A, 1B, IC, 2A AND 1B

[0001] The present application incorporates by reference Tables 1A, 1B, 2A, and 2B contained on one compact disc filed concurrently herewith, which compact disc is labeled "Copy 1- Tables 1A-2B". The details of Tables 1A-2B are further described later in this disclosure. This compact disc was created on 2 September 2005 and is 680 MB in size. The CD contains three files labeled Table 1A.doc (88 KB), Table 1B.doc (5721 KB), and Table 2A-2B.doc (223 KB). These files are expressly incorporated herein by reference.

I. REFERENCES

The following references are cited below in support of the background of the invention or methods employed in practicing the invention.

1. McManus, Immunity, 21:747-756 (2004).
2. Bartel, Cell, 116:281-297 (2004).
3. Cai et al., Rna, 10:1957-1966 (2004).
4. Lee et al., Embo J, 23:4051-4060 (2004).
5. Lee et al., Nature, 425:415-419 (2003).
6. Bernstein et al., Nature, 409:363-366 (2001).
7. Calin et al., Proc Natl Acad Sci USA, 99:15524-15529 (2002).
8. Calin et al., Proc Natl Acad Sci USA, 101:2999-3004 (2004).
9. Calin et al., Proc Natl Acad Sci USA, 101:11755-11760 (2004).
10. Griffiths-Jones, Nucleic Acids Res, 32:D109-D111 (2004).
11. Bentwich et al., Nat Genet, 37:766-770 (2005).
12. Ota et al., Cancer Res, 64:3087-3095 (2004).
13. He et al., Nature, 435:828-833 (2005).
14. Akagi et al., Nucleic Acids Res, 32:D523-D527 (2004).
15. Collier et al., Nature, 436:272-276 (2005).
16. Dupuy et al., Nature, 436:221-226 (2005).
17. Suzuki et al., Nat Genet, 32:166-174 (2002).

18. Lund et al., Nat Genet, 32:160-165 (2002).
19. Hwang et al., Proc Natl Acad Sci USA, 99:11293-11298 (2002).
20. Mikkers et al., Nat Genet, 32:153-159 (2002).
21. Li et al., Nat Genet, 23:348-353 (1999).
22. Lovmand et al., J Virol, 72:5745-5756 (1998).
23. van Lohuizen et al., Cell, 65:737-752 (1991).
24. Nusse et al., Cell, 31:99-109 (1982).
25. Nusse et al., Nature, 307:131-136 (1984).
26. Berezikov et al., Cell, 120:21-24 (2005).
27. Tanzer et al., J Mol Biol, 339:327-335 (2004).
28. Lim et al., Science, 299:1540 (2003).
29. Xie et al., Nature, 434:338-345 (2005).
30. Weng et al., Science, 306:269-271 (2004).
31. Justice et al., Mamm Genome, 11:484-488 (2000).
32. Hallberg et al., J Virol, 65:4177-4181 (1991).
33. Nielsen et al., J Virol, 70:5893-5901 (1996).
34. Sorensen et al. J Virol, 70:4063-4070 (1996).
35. Kim et al., J Virol, 77:2056-2062 (2003).
36. Walther et al. Genome Res, 11:875-888 (2001).
37. Hofacker et al., Chemie, 125:167-148 (1994).
38. Zuker et al., Nucleic Acids Res, 9:133-148 (1981).
39. McCaskill, Biopolymers, 29:1105-1119 (1990).
40. Shtivelman et al., Proc Natl Acad Sci USA, 86: 3257-3260, (1989).
41. Arya et al., Expert Rev Mol Diagn: 5: 209-219 (2005).

II. BACKGROUND

[0002] MicroRNAs (miRNAs) are small, non peptide-coding RNAs that regulate gene expression in a variety of physiological and developmental processes [1, 2]. In the biogenesis of miRNAs, primary miRNA transcripts (pri-miRNAs) are first generated by RNA polymerase II [3, 4] and are then further processed like messenger RNA transcripts with the addition of a 5' cap structure and poly A tail. Because of this, the pri-miRNA transcripts can be found in standard cDNA libraries.

[0003] The primary transcript can be over 3 kb long and adopt one or several stem-loop structures which are subsequently processed by the enzymes Drosha [5] and/or Dicer [6] to generate mature miRNA. The mature miRNAs are generally 18 to 24 nucleotides long and are incorporated into the RNA-induced silencing complex (RISC), which inhibits translation by binding to similar, but not identical sequences, of the 3' untranslated region of mRNA. If the interaction is perfectly complementary, the miRNA may act as small inhibitory RNA (siRNA) leading to the degradation of the target mRNA. Often, a pri-miRNA transcript is polycistronic, i.e., one pri-miRNA transcript yields several different miRNAs. Further, miRNAs can be found within primary gene transcripts.

[0004] Dysregulated miRNA expression has been postulated to contribute to lymphoma formation in humans [7-9]. The miRNA registry [10] currently contains over 200 examples that are shared between humans and mice; another 89 miRNAs are found only in primates [11]. Of these, one miRNA cluster has been demonstrated to be overexpressed in human B cell lymphomas [12], and enforced overexpression of this cluster in hematopoetic stem cells from lymphoma-prone mice accelerated tumor development [13].

III. SUMMARY

[0005] The invention includes, in one aspect, a method for positively identifying a human miRNA sequence associated with a detectable disease state in humans, such as a cancer. The method includes the steps of (i) identifying, from each of at least two animals having a detectable disease state, such as a cancer, produced by insertional mutation, the sequence of a genomic segment that is common to both animals, and that contains an insertional mutation, (ii) identifying transcription units contained within the animal genome that are within about 200 Kbases, in either an upstream or downstream direction, of the sequenced genomic segment, (iii) identifying human genomic transcription units that are orthologous to the transcription units identified in step (ii), and (iv) for each human transcription unit identified in step (iii), employing a bioinformatics program capable of identifying putative miRNA sequences, to determine whether that transcription unit identified in step (iii) contains a putative miRNA sequence, in which case the putative miRNA sequence is positively identified as a human miRNA.

[0006] The detectable disease state may be a cancer, such as lymphoma, wherein step (i) of the method is carried out by isolating the genomic segment from each of at least two animals having a detectable cancer, such as lymphoma. The insertional mutation in step (i) may be a viral insertional mutation.

[0007] The sequence identified in step (iii) may be contained in a portion of a pri-miRNA that is outside the corresponding mature miRNA (fully processed miRNA), or it may contained completely within the mature miRNA, or it may be contained in both portions of pri-miRNA transcript.

[0008] In another aspect, the invention includes an assay kit for diagnosing the presence or risk of cancer in a human subject. The kit includes a first reagent designed to react specifically with a human pri-miRNA and/or mature miRNA sequence identified in accordance with the method of claim 2, to form a first detectable reaction product, and an indicator guide that indicates how the presence or amount of the reaction product correlates with the presence or risk of the disease state in a human subject.

[0009] The first reagent may be one of: (a) PCR reagents for detecting the presence or absence of the genomic sequence, or (b) oligonucleotide binding reagents for detecting the presence of absence of the genomic sequence. For use in diagnosing the presence of risk of a cancer in a human subject, step (i) in the method is carried out by isolating the genomic from each of at least two animals having a detectable cancer, such as a lymphoma. The kit's first reagent may be designed to react specifically with a mature human miRNA sequence identified in accordance with the method of claim 1.

[0010] Also disclosed is a method for treating a cancer in a human subject, by administering to the subject, a therapeutically effective amount of a compound capable of binding specifically to a mature human prim-miRNA and/or a mature miRNA sequence identified in accordance with the above method.

[0011] Further disclosed is an isolated mature human miRNA sequence selected from the group consisting of SEQ ID NOS: 1-55 .

[0012] In a more general aspect of the above method, the invention provides a method for identifying a human regulatory RNA (regRNA) sequence associated with a detectable disease state in humans. The method includes the steps of: (i) identifying, from each of at least two animals having a detectable disease state

produced by insertional mutation, the sequence of a genomic segment that is common to both animals, and that contains an insertional mutation, (ii) identifying transcription units contained within the animal genome that are within about 200 Kbases, in either an upstream or downstream direction, of the sequenced genomic segment, (iii) identifying human genomic transcription units that are orthologous to the transcription units identified in step (ii), (iv) for each human transcription unit identified in step (iii), using a bioinformatics program to determine whether that transcription unit is a non-coding RNA sequence, and (v) if the homologous human genomic sequence from step (iv) is a non-coding RNA sequence, classifying the sequence as a human regRNA sequence associated with the detectable disease state.

[0013] The insertional mutation in step (i) may be a viral insertional mutation. The detectable disease state may be a cancer, wherein step (i) is carried out by isolating the genomic segment from each of at least two animals having a detectable cancer.

[0014] The human regRNA sequence may be an miRNA, wherein step (iv) includes employing a bioinformatics program capable of identifying putative miRNA sequences to determine whether that transcription unit identified in step (iii) contains a putative miRNA sequence, in which case the putative miRNA sequence is positively identified as a human miRNA.

[0015] The method may further include utilizing the identified human regRNA sequence for diagnostic or therapeutic purposes.

[0016] Also disclosed is an assay kit for diagnosing the presence or risk of cancer in a human subject. The kit includes a first reagent designed to react specifically with a human regulatory RNA (regRNA) sequence identified in accordance with the method of claim 15, to form a first detectable reaction product, and an indicator guide that indicates how the presence or amount of the reaction product correlates with the presence or risk of the disease state in a human subject.

[0017] As above, the first reagent may be one of: (a) PCR reagents for detecting the presence or absence of the genomic sequence, or (ii) oligonucleotide binding reagents for detecting the presence of absence of the genomic sequence.

[0018] In still another aspect, the invention includes a novel regulatory RNA (regRNA), in addition to the novel miRNA identified above, which when

overexpressed or disrupted contribute to the formation of tumors. The human and mouse sequences for each regRNA in FASTA format are listed in Table 1B along with the identifying cluster ID. SEQ ID NO:1-55 are mature human miRNAs. SEQ ID NO: 56-110 are mature mouse miRNAs. SEQ ID NO: 111-165 are human pre-miRNAs. SEQ ID NO:166-220 are mouse pre-miRNAs. SEQ ID NO: 221-500 are human pri-miRNAs. SEQ ID NO: 501-822 are mouse pri-miRNAs.

[0019] The regRNA disclosed can regulate oncogenes and/or suppressors or actually be an oncogene and/or suppressor itself. The novel regRNA sequences may be used in diagnostic applications, for detecting the presence and/or risk of a given cancer type, or in therapeutics, e.g., for treating that cancer

[0020] These and other aspects, objects, advantages, and features of the invention will become apparent to those persons skilled in the art upon reading the details of the invention as more fully described below.

IV. BRIEF DESCRIPTION OF THE DRAWINGS

[0021] The invention is best understood from the following detailed description when read in conjunction with the accompanying drawings. It is emphasized that, according to common practice, the various features of the drawings are not to-scale. On the contrary, the dimensions of the various features are arbitrarily expanded or reduced for clarity. Included in the drawings are the following figures:

[0022] Figures 1A and 1B are customized screen prints of the UCSC genome web site browser (March 2005 version of the mm6 gene assembly), looking at the mir-17-20 locus (Fig. 1A); and at the mir-106a-92 locus (Fig. 1B). Mir-17-20 is the mouse cluster orthologous to the human mir-17-92 cluster. Mir-19b-1 only weakly maps to the mouse genome at the indicated location. Top, base position at chromosomes 14 and X, respectively. The handle bars below "Picobella_SL3" represent the retroviral insertions into the mir-17-20 locus (Fig. 1A) or the mir-106a-92 locus (Fig. 1B) in 29 or in 33 independent tumors, respectively. The bars below "miRNA", are miRNAs found in the miRNA registry 10 (//www.sanger.ac.uk/Software/Rfam/mirna/); the bars below "miRNA predicted" represent miRNAs predicted by use of the method herein. The exon/intron structure of mRNAs and ESTs of the mouse is shown below the predicted miRNA. Sequence

conservation between mouse and various other species (rat, human, dog, cow, opossum, chicken, tropicalis, zebrafish, and tetraodon) is also shown.

[0023] Figures 2A and 2B are each a customized screen print of the UCSC genome web site browser, looking at two loci with predicted miRNA located on chromosomes 8 and 12, (Figs. 2A and 2B, respectively). For Fig. 2A, the two handle bars below "Picobella_SL3" (1490S-206-1 and 1163S-137-14), represent retroviral insertions into the locus recovered in 2 independent tumors. Known miRNAs listed in the miRNAs registry 10 are not found in this locus; the 2 bars below "miRNA predicted" represent miRNAs predicted by use of the method herein. Two retroviral integrations (S3_306D and S5_030A1) represent independent tumors as listed in the RTCGD database [14] (Retrovirus Tagged Cancer Gene Database; //RTCGD.ncifcrf.gov). In Fig. 2B, the handle bars under "Picobella_SL3" represent retroviral insertions into the locus recovered in 8 independent tumors. The bars for "miRNA predicted" are miRNAs predicted by the method herein. Known miRNAs listed in the miRNAs registry 10 are not found in this locus. The AK019999, AI060616, BE848409, and BB634791 transcripts are thymus-specific. Sequence conservation between mouse and various other species is also shown.

[0024] Figs. 3A and 3B are each a customized screen print of the UCSC genome web site browser, looking at two loci with regulatory RNA. The top of the figures shows the base position at chromosomes 15 and 1 (Figs. 3A and 3B, respectively). The handle bars below "Picobella_SL3" represent the retroviral insertions recovered by the present method in 7 independent tumors (chr. 15, Fig. 3A); and 5 independent tumors (chr 1, Fig. 3B). Arrows within handle bars denote transcriptional direction. The exon/intron structure of mRNAs and ESTs of the mouse are shown below the predicted miRNAs. Transcripts AK040104 and AK041852 (Fig. 3A) and BY097680 (Fig. 3B) are thymus-specific. Sequence conservation between mouse and various other species is shown at the bottom.

[0025] Figure 4A is a table showing tumors assayed for the region containing mmu-mir-106a (Fig. 1B). Retroviral insertion site locations (August 2005 version of the mm7 genome assembly) are notated by the basepair located directly after the insertion. Orientation of the retrovirus is indicated by "+++" for directionality of left to right and by "---" for directionality of right to left on the chromosome.

[0026] Figure 4B is a graph of the relative expression of AY940616 as measured by quantitative PCR. Tumors with integrations located upstream of AY940616 (the predicted primary transcript for the mmu-mir-106a-92 locus) were assayed by qPCR using a dual labeled probe designed to AY940616. Integration sites assayed were located within (i) ~ 3 kb, (ii) ~14 kb, and (iii) ~18 kb upstream of AY940616. Tumors with no integrations in this region (iv) along with cDNA from a normal mouse spleen were run as controls. Beta-actin (ACTB) was used as the endogenous reference gene and 1735S, one of the tumor controls, was used as the calibrator sample in the calculation of $2^{-\Delta\Delta Ct}$ values. All $2^{-\Delta\Delta Ct}$ values were normalized such that the average of the tumor controls was set to 1.

[0027] Figure 4C is a graph of the relative expression levels of mmu-mir-106a by quantitative PCR. Tumors with integrations located upstream of the mmu-mir-106a -92 locus were assayed by qPCR using a reverse transcriptase primer/dual labeled probe system designed to mmu-mir-106a. Integration sites assayed were located within (i) ~ 3 kb, (ii) ~14 kb, and (iii) ~18 kb upstream of the miRNA cluster. Tumors with no integrations (iv) in this region were run as controls. Concentrations of mmu-mir-106a were determined using a standards curve generated with a synthetic mmu-mir-106a RNA oligo. Concentrations were then normalized by the average of the tumor controls to calculate relative expression levels.

[0028] Figure 5A is a map of the region containing AK030859. The genomic organization of retroviral insertion sites in the region containing AK030859 is shown by a screen capture of the UCSC genome website browser (August 2005 version of the mm7 genome assembly). Insertion sites are drawn as vertical handlebars below "PicoSL3".

[0029] Figure 5B is a table showing tumors assayed for the region containing AK030859. Tumor locations and orientations are notated as in Fig 4A.

[0030] Figure 5C is a graph showing the relative expression of AK030859 as measured by quantitative PCR. Tumors with integrations located in the region encompassing AK030859 were assayed by SYBR qPCR for the 5' end of AK030859. Integration sites assayed were located (i) up to 1.2 kb upstream, (ii) within, and (iii) up to 52 kb downstream of AK030859. Tumors with no integrations in this region (iv) were run as controls. Beta-actin (ACTB) was used as the endogenous reference gene and 1484S, one of the tumor controls, was used as the calibrator sample in the

calculation of $2^{-\Delta\Delta Ct}$ values. All $2^{-\Delta\Delta Ct}$ values were normalized such that the average of the tumor controls was set to 1.

[0031] Figure 6A is a map of region containing AK040062. The genomic organization of retroviral insertion sites in the region containing AK040062 is shown by a screen capture of the UCSC genome website browser (August 2005 version of the mm7 genome assembly). Insertion sites are drawn as vertical handlebars below "PicoSL3".

[0032] Figure 6B is a table showing the tumors assayed for the region containing AK040062. Tumor locations and orientations are notated as in Fig. 4A.

[0033] Figure 6C is a graph showing the relative expression of AK040062 exon 2 as measured by quantitative PCR. Tumors with integrations located in the region encompassing AK040062 were assayed by SYBR qPCR for AK040062 exon 2. Integration sites assayed were located (i) up to 6 kb upstream, (ii) within intron 1, (iii) within intron 2, and (iv) up to 16 kb downstream of AK040062. Tumors with no integrations in this region (v) along with normal mouse spleen samples (vi) were run as controls. Data was treated as previously mentioned for AK030859 except 3412S was used at the calibrator sample.

[0034] Figure 7A is a map of the region containing AK037419. The genomic organization of retroviral insertion sites in the region containing AK037419 is shown by a screen capture of the UCSC genome website browser (August 2005 version of the mm7 genome assembly). Insertion sites are drawn as vertical handlebars below "PicoSL3".

[0035] Figure 7B is a table showing the tumors assayed for the region containing AK037419. Tumor locations and orientations are notated as in Fig. 4A.

[0036] Figure 7C is a graph showing the relative expression of AK037419 exon3 as measured by quantitative PCR. Tumors with integrations located in the region encompassing AK037419 were assayed by SYBR qPCR for AK037419 exon 3. Integration sites assayed were located (i) up to 13 kb upstream, (ii) within intron 1, (iii) within intron 2, and (iv) within exon 3 of AK037419. Tumors with no integrations in this region (v) along with normal mouse spleen and thymus samples (vi) were run as controls. Data was treated as previously mentioned for AK030859 except 1438S was used as the calibrator sample.

[0037] Figure 8 is a graph showing relative expression of PVT1 exon 1 in matched human normal and tumor prostate RNA samples. Matched human normal and tumor prostate RNA samples were assayed by SYBR qPCR for PVT1 exon 1. Beta-actin (ACTB) was used as the endogenous reference gene and each normal RNA was used as a calibrator for its matched tumor RNA in calculating $2^{-\Delta\Delta Ct}$ values.

[0038] Table 1A includes a seven page list of regulatory RNA clusters. Tumors with proviral integrations, representative ESTs, and known and predicted miRNAs found at each loci are indicated. Chromosomal locations are from version mm6 of the mouse genome and the hg17 version of the human genome at the UCSC Genome Bioinformatics website (genome.ucsc.edu). "Known miRNAs" refers to miRNAs found in the miRNA registry (August 2005); "Predicted miRNAs" refers to miRNAs predicted as described in the text. Since the miRNA cluster mir-17-92 has been previously described as a possible oncogene [13], the mir-17-20 and mir-17-92 sequences are not included in Tables 1B. The human and mouse sequences for each regRNA in FASTA format are listed in Table 1B along with the identifying cluster ID. SEQ ID NO:1-55 are mature human miRNAs. SEQ ID NO: 56-110 are mature mouse miRNAs. SEQ ID NO: 111-165 are human pre-miRNAs. SEQ ID NO:166-220 are mouse pre-miRNAs. SEQ ID NO: 221-500 are human regRNAs. SEQ ID NO: 501-822 are mouse regRNAs. SEQ ID NO: 14, 26, 37-39, 41-43 are known human miRNAs that were not previously known to be associated with cancer.

[0039] Tables 2A and 2B are two and three page lists, respectively, of miRNAs, regRNA, ESTs, or genes, co-mutated with the mir-17-20 locus (Table 2A) or the mir-106a-92 locus (2B). The predicted miRNAs are in bold. Co-mutated regions in common between the mir-17-20 and the mir-106a-92 loci are indicated by asterisks (**). Chromosomal locations are from version mm6 of the mouse genome at the UCSC Genome Bioinformatics website (genome.ucsc.edu). "Indeterminate" refers to regions where the miRNA, EST, or gene could not be determined. "Desert" regions are those which appear to be void of miRNAs, ESTs, or genes.

V. DETAILED DESCRIPTION

A. Definitions

[0040] The following terms have the definitions given below, unless otherwise indicated in the specification.

[0041] "Regulatory RNA" or "regRNA" generally refers to non-protein encoding RNA molecules (including miRNA) that regulate the expression of genes.

[0042] "microRNA" or "miRNA" generally refer to ~18-24 -mer RNAs that regulate the expression of genes by binding to the 3'-untranslated regions (3'-UTR) of specific mRNAs. According to standard nomenclature, a pre-processed miRNA transcript prior is referred to an pri-miRNA. Enzymatic cleavage of pri-miRNA in the nuclear compartment by Drosha yields a pre-miRNA, which is further processed by Dicer in the cytoplasmic compartment in form mature miRNA. "miRNA" may be used herein to refer to pri-miRNA, pre-miRNA or mature miRNA, and the distinction, if any, will be understood from the context in which it is used.

[0043] "Stringent conditions" refers to a procedure including a stringent wash such as with 0.1% saline sodium citrate, and 0.1% sodium dodecyl sulfate (0.1% SSC, 0.1% SDS) at 65 °C. Appropriate stringent conditions are further described in Sambrook et al., Molecular Cloning, Cold Spring Harbor Laboratory Press, New York, 1989.

[0044] As used herein, a nucleotide or RNA sequence "specifically hybridizes" to a sequence under physiological conditions, with a Tm substantially greater than 37 °C, preferably at least 50 °C, and typically 60 °C, 80 °C or higher. Such hybridization preferably corresponds to stringent hybridization conditions, selected to be about 10 °C, and preferably about 50 °C lower than the thermal melting point (T[m]) for the specific sequence at a defined ionic strength and pH. At a given ionic strength and pH, the T[m] is the temperature at which 50% of a target sequence hybridizes to a complementary polynucleotide.

[0045] Polynucleotides are described as "complementary" to one another when hybridization occurs in an antiparallel configuration between two single-stranded sequences. Complementarity (the degree that one polynucleotide is complementary with another) is quantifiable in terms of the proportion of bases in opposing strands that are expected to form hydrogen bonds with each other, according to generally accepted base-pairing rules.

[0046] The term "overexpressed" refers to a range of expression of a protein which is greater than that generally observed for a given type of cells.

[0047] The term "insertional mutation" refers to a mutation that is introduced into a genome by insertion of an exogenous sequence or an endogenous sequence.

Such exogenous and endogenous sequences may be, for example, either viral or transposon-based. An insertional mutation may enhance the transcription of one or more coding or non-coding genes located within about 200 Kbases of the mutation.

[0048] The term "orthologous sequence" refers to a sequence having a direct evolutionary counterpart derived from a common ancestor by vertical descent; and, as a consequence, having conserved function to a high degree of likelihood.

[0049] A "bioinformatics program" refers to computer program designed to carry out one or more sequence analysis functions on database sequences. These functions may include sequence alignment, recognition of regions capable of forming secondary structure, recognition of various gene transcription and/or translation control sequences, and identification of one or many possible different classes of genomic sequences, including coding sequences in general, and coding sequences for particular types of proteins, non-coding gene sequences, transcription splice sites, secondary structure sites, identification of genes for various cellular RNAs, and recognition of orthologous genes from different organisms.

[0050] A "transcriptional unit" refers to a coding or non-coding gene, or the transcript produced thereby, and may be identified, for example, by the presence of a polyadenylation site on the corresponding processed transcript.

B. Methods of identifying regRNA

[0051] Regulatory RNA and miRNA sequences that contribute to tumor formation are described and disclosed. These regRNA and miRNA sequences were identified in mice, and were subsequently confirmed in humans. These sequences were identified by the following methods.

[0052] A retrovirus that induces tumors was used to identify 322 loci encoding regRNAs, many of which are expressed only in thymocytes. Of these loci, 29 are predicted by current algorithms to encode miRNA, and four are confirmed miRNA polycistrons listed in the miRNA registry. miRNA overexpression was confirmed for several tumors containing nearby integration sites predicted to activate transcription. These results (a) substantially increase the number of known miRNAs and (b) identify them as being oncogenic when dysregulated in T cells.

[0053] Because the expression of a large number of miRNAs is dysregulated in lymphomas [8, 9], it seemed likely that many more miRNAs than were previously

known act as oncogenes or tumor suppressor genes. The present method defines oncogenic miRNAs and other regRNAs in a high throughput manner using proviral tagging. Although viruses have not yet been implicated as a major cause of cancers in humans, research using tumor viruses has led to the discovery of many oncogenes and protooncogenes. In proviral tagging methods, mice are infected with a retrovirus that does not contain an oncogene (e.g., murine leukemia virus, MLV, or murine mammary tumor virus, MMTV). Recently, the host range of this approach has been broadened by the use of a transposon [15, 16].

[0054] During retroviral infection, the virus integrates into the cellular genome and inserts its DNA near or within genes, which leads to various outcomes:

[0055] (i) The insertion site is too far away from a protooncogene and thus does not activate it. In this case, there will be no selection for that cell.

[0056] (ii) The provirus inserts within 200 kb of a protooncogene, but not within the gene (type 1). Here, either the viral promoter or the viral enhancer increases the expression level of the protooncogene.

[0057] (iii) The provirus inserts within a gene, destroying or altering its function (type 2).

[0058] There will be no selection for a cell that contains either type 1 or type 2 insertion events in a gene that is not a protooncogene or tumor suppressor gene. If integration results in the formation of a tumor, genes adjacent to the integration site can be identified, and classified as either protooncogenes or tumor suppressor genes. This method has been used to identify many new protooncogenes as well as to confirm already known protooncogenes discovered by virtue of their homology to viral oncogenes [17-25]. A tumor suppressor may be scored if a retrovirus lands within a gene and truncates or destroys it. In these cases, the suppressor may be haplo-insufficient, or alternatively, the mutation on the other allele is provided spontaneously by the mouse. The integration event may also lead to more complex consequences, such as a dominant negative effect of the truncated gene product or the transcription of anti-sense or miRNA.

[0059] Because the mechanics of transcription of pri-miRNA and regular nuclear gene transcripts are the same, it was reasoned that retroviral insertions near or into these transcribed regions ought to have similar effects. Whereas to date, all mammalian miRNAs have been discovered by computational methods, the present

methods provide an extensive forward genetic approach to functionally identify novel oncogenic miRNAs in retrovirally generated tumors.

[0060] The present invention, in one non-limiting embodiment, provides a method of identifying novel human regulatory RNA (regRNA) sequences, including novel miRNA sequences, associated with a detectable disease state in humans. In practicing the method, an animal model, such as mouse or rat, having known disease states, and typically disease states that are similar to those found in humans, is subject to standard insertional mutagens, such as viral insertional mutagens, and then observed for development of one or more disease states, e.g., one or more cancer types, or hyperlipidemia, both diseases known to be associated with dysfunctions in regRNA. When a disease state is observed in a mutagenized animal, the genome, e.g., in a cancerous tissue or cell, is then analyzed for the presence and chromosomal locations of the one or more insertion mutations. This is done, for example, using PCR probes that overlap with the insertional mutagen sequence, to produce an amplified segment of the animal genome adjacent the mutation.

[0061] The sequence of this segment is then determined and used in a database search of the animal's genome, to find transcriptional units that are within a defined distance, typically less than 100 Kbases, but up to 200 Kbases, upstream and/or downstream of the insertional mutation site or sequenced segment containing that site. Transcriptional units are identified according to known procedures, e.g., by employing a bioinformatics program that stores information about transcription units that have been previously identified as such by the presence of polyadenylation in their transcripts.

[0062] For each transcriptional unit that is identified in this manner, the method now involves searching a human genomic database to identify human transcriptional units that are orthologous with the identified animal transcription units. This step is used in finding the human transcription unit corresponding to the one identified in the animal as possibly related to an identified disease state. Of course, since some animal transcription units are unique to that animal and/or do not overlap with human transcription units, not every animal transcription unit identified in the method will have a human ortholog.

[0063] Once the human transcription units corresponding to the disease-related animal transcription units have been identified, these are further analyzed using bioinformatics tools to (i) identify those non-coding units that will be classed as regRNAs, and (ii) among the regRNAs, those units that contain secondary structure and other sequence-related features associated with miRNAs. In the first case, a human transcription unit identified as above is compared against known coding sequences, or sequences with coding-gene sequence features, to determine whether the transcription unit is a coding or non-coding gene. If it is a non-coding gene, and not previously identified as a regRNA sequence, or not previously identified as having the newly identified disease association, the method identifies the transcription unit either as a novel regRNA, or a known regRNA having a newly-identified disease associated function.

[0064] If the regRNA is further determined to contain sequences characteristic of miRNAs, e.g., stem-loop regions characteristic of pre-miRNAs, then the method can further identify the regRNA as either a newly identified miRNA sequence (including the pri-miRNA, the pre-miRNA, and/or mature miRNA), or a previously known miRNA with a newly identified disease association (SEQ ID NOS: 13, 14, 26, 27, 37-39, and 41-43.)

[0065] It will be appreciated that the method just described, which combines a functional assay (disease association) with a bioinformatics analysis, allows confirmation or positive identification bioinformatics information, e.g., gene identification, and also allows for less stringent bioinformatics constraints, e.g., in the identification of novel miRNAs, as discussed below.

[0066] Also forming part of the invention are a comprehensive set of regRNAs including miRNAs that when overexpressed or deleted contribute to tumor formation. The miRNAs can regulate both oncogenes and suppressors, as well as represent both oncogenes and suppressors themselves. Although classic tumor suppressors require both alleles to be inactive, the present recovery of regRNA sequences used a modified retroviral tagging strategy. In this modified strategy, chemical mutagenesis was initially carried out on the paternal allele, followed by retroviral insertional mutagenesis (which can affect both the maternal and paternal alleles). Chemical mutagenesis was carried out using ENU (N-ethyl-N-nitrosourea; a potent germ line mutagen). If by chance the virus-disrupted (maternal) allele and the ENU-

inactivated (paternal) allele belong to the same locus, then the cell has no functional allele. Should this locus represent a tumor suppressor, the cell lacking it will have a growth advantage over other cells, which may result in tumor formation.

1. Viral tag recovery and locus identification

[0067] The viral integrations sites (tags) were determined in tumors generally by isolating and digesting genomic tumor DNA, followed by an anchored PCR technique [20]. This was performed by amplifying and sequencing a chimeric DNA fragment consisting of a short genomic sequence upstream of the viral 5' LTR and part of the viral 5' LTR itself. The tags were sequenced and mapped to the mouse genome sequence, and the affected transcription unit was determined. From 2373 tumors, 7300 tags were obtained, which mapped to 2,038 regions. Of these regions, 645 had two or more associated integration sites, with the largest region having 500 integrations.

2. Calling regulatory RNA transcripts

[0068] At least one of the following, non-limiting, considerations should be taken into account to correctly identify the affected regRNA based on the retroviral screen. First, although vertebrates share extensive highly conserved non-coding sequences which might represent regRNA, not all non-translated (translatable) expressed sequence tags (ESTs) fragments represent true regulatory RNA. For example, a fraction represent small nuclear RNA of the spliceosome, another fraction results from DNA contamination, and yet another fraction may just be transcriptional noise not yet edited by evolution for energy efficiency. Second, viral integration into a potentially transcribed region does not necessarily mean this transcription unit is activated and contributes to tumorigenesis. There is the question whether or not the proviral enhancer/promoter can "leapfrog" the nearest gene and instead regulate the next one. In the past, it has been assumed that this is not (or only rarely) the case, and that proviruses can exert their function up to a distance of 200 kb from a gene. Such assumptions were reevaluated in light of more extensive genomic coverage and better annotation of non-coding transcripts. With the above potential complications in mind, the transcription unit nearest to a cluster of integration sites was identified. In the analysis, it was reasoned that if a gene is located, for example, 200 kb from an insertion site, then the other integration sites ought to be more or less evenly distributed over that distance. If, however, a cluster

of integration sites spans a few kilobases and is located within or next to a noncoding transcription unit, this unit was called rather than a far away gene.

3. miRNA identification

[0069] Early computational algorithms designed to predict miRNAs relied on sequence conservation between species, hairpin structure determination, and thermodynamic stability. A more recent prediction attempt has relaxed the species conservation requirement in an attempt to identify new primate-specific miRNAs [11]. Nonetheless, all computational approaches involve a trade off between maximizing sensitivity and minimizing false positives, and as such, may miss important classes of miRNAs. Since the retroviral screen provided complementary functional data, it was possible to modify the computational approach of Berezikov et al.[26] with relaxed input parameters and maintaining the sequence conservation between mouse and human as a necessary condition. This computational approach yielded 13,648 predicted miRNAs. Apart from non-translatability, ESTs terminating at the 3' or 5' end of the miRNA cluster were identified, which should be an indication for a site of Drosha processing activity. Based on these criteria, retroviral integrations at 322 loci with regRNAs were found, many of which are expressed only in thymocytes. These include integrations at: (1) mir-17-20, the mouse ortholog to the human miRNA cluster (mir-17-92) that has been demonstrated to be an oncogene in mouse and likely in humans [13]; (2) three other confirmed miRNAs in the registry; (3) 29 non-coding transcription units with predicted miRNA; and (4) 289 non-coding transcription units without miRNA predicted.

[0070] Table 1A is a list of the 322 mouse and 280 human regRNA and miRNA loci. For each cluster, the cluster ID, the chromosomal location, the tumors that contain the proviral integrations sites in that cluster, the ESTs within and adjacent to that cluster, the known and predicted miRNAs, and the genomic location of the corresponding human regRNA are listed. The chromosomal positions of the mouse regRNA and miRNAs are defined by the March 2005 UCSC genome assembly of the mouse genome (mm6) while the chromosomal positions of the human regRNA and miRNAs are defined by the hg17 UCSC human genome assembly. The sequences of the regRNA and miRNAs are listed in Table 1B in FASTA format, with the exception of the mir-17-20 and mir-17-92 loci. Examples of the groups are disclosed and described below.

4. mir-17-20 and mir-106a-92

[0071] The mir-17-20 polycistron contains four confirmed miRNAs, three of which are predicted by the bioinformatics approach of the present method (Fig. 1A; mir-19b-1 only weakly maps to this cluster). To date, this polycistron is the only one that has been shown to be an oncogene in the mouse [13]. Several of the ESTs terminate 3' of the cluster and all 5 miRNAs are contained in the intron of transcript AK053349. The 29 retroviral insertion sites fall into three groups, all contained in the 15 kb transcription unit. It is unclear why there are these three groups, but perhaps site specificity of Drosha or undetected novel miRNAs are the cause. Interestingly, all 11 integrations closest to the mir-17-20 polycistron have the same direction (left to right) of transcription as the miRNAs themselves (left to right; not shown). Conversely, 9 out of 10 of the integrations farthest from the polycistron have the opposite orientation (right to left) of the miRNAs. The orientation of the provirus is thought to be important in activation of protooncogenes. Either the viral promoter, in the same transcriptional orientation as the protooncogene, overrides the promoter of the protooncogene, or the enhancer, in either orientation, cooperates with the promoter to increase transcription of the cellular gene. In the classical insertions of type 2, i.e., within a gene, the result is either truncation or destruction. Because mir-17-92 polycistron acts as an oncogene [13], it ought to be the case that 3' to the integration sites there are transcripts generated at an increased level, and that these transcripts can be processed by Drosha and Dicer.

[0072] The mir-106a-92 polycistron is a cluster related by homology to mir-17-92 [27] and contains three previously identified miRNAs and one more predicted by us (Fig. 1B). The transcript AK084356 ends precisely where the miRNA cluster begins, and part of the intron is an exon of other transcripts. There are also several more near the miRNA cluster. The two leftmost proviral integrations (1505S, 1759S) have the same transcriptional orientation as the AK084356 transcript and thus may constitute "promoter insertions". Because of their distance to the transcription unit, the 3 rightmost retroviral insertions (558T, 569S, 2221S) ought to represent enhancer insertions. In these cases, the provirus has integrated 5' to a transcription unit, and the orientation of transcription of provirus and cellular transcription unit are opposite. This is because in the LTR of the provirus, the enhancer precedes the promoter and it is thought that the enhancer cooperates with

promoters without leapfrogging. The remaining integrations may be either promoter or enhancer insertions and thus may have either orientation. Transcript AY940616 and mature the mmu-mir-106a miRNA were both found to be overexpressed in mouse thymic tumors by quantitative PCR (Fig. 4A-C).

5. Oncogenic miRNAs not found in the miRNA registry

[0073] The number of existing miRNAs is growing monthly. In early 2005, the number in humans was roughly 200, and early estimates calculated 255 as the upper limit [28]. There are 321 human miRNAs in the most recent version of the miRNA registry (August 2005). Recent studies have suggested that the number of human miRNAs may be much greater, and as much as 800 [29].

[0074] As seen in Fig. 2A the predicted miRNAs are contained in transcript BC048951, and are close to other ESTs that may be processing products of Drosha and Dicer. Thus, part of transcripts AK045307 and AK087491 overlap with BC048951, and another part is contained in the intron of the much longer transcript AK050834. An additional transcript that covers part of the same intron is thymus specific AK079473. The retroviral insertion site 1490S is within the large intron of the AK050834 transcript, which presumably represents the largest piece of the pri-miRNA. The other insertion, 1163S, is 3' to the pri-miRNA, in the same transcriptional orientation, which allows the viral enhancer to cooperate with the promoter of the pri-miRNA.

[0075] Fig. 2B shows 8 insertions near two predicted mi-RNAs. Each miRNA is contained in a transcript that is found only in thymocytes (AI060616, BB634791). Interestingly, two other nearby transcripts are also found only in thymocytes.

[0076] The prediction program described herein was shown to find 81% of all registered miRNA in the mouse. There are other programs that compare regulatory motifs in promoters and 3' UTRs in several mammals [29]. The method also found many regions where no miRNA was predicted, but where the retroviral insertions were (1) within or nearby a transcript that was not translatable and (2) were often far away (>30kb) from any other gene.

6. Retroviral insertions into regRNA without miRNAs

[0077] The transcript AK040104 in Fig. 3A, for example, with eight proviral insertions sites, looks like a gene, except that it is not classifiable and is >300 kb away from the nearest known gene. There is a smaller transcript AK041852, which

covers two introns of the larger transcript, and both transcripts are expressed only in thymocytes.

[0078] Fig. 3B shows 5 integration sites upstream of transcript AK021325 which also lack predicted miRNAs and is ~40 kb away from the nearest authentic gene. All 5 integration sites have the same direction of transcription as the ESTs, suggesting that transcription of these ESTs in increased by the viral promoter. Thus, insertions into these types of regions were also surveyed, where there was a hint of Drosha processing activity and where thymocyte-specific expression is observed. These regions contain regulatory RNAs, resulting in identification of 289 new regions.

[0079] Figures 5-7 show three additional loci containing retroviral integrations near or within non-coding regRNAs. The expression levels of each regRNA were measured using quantitative methods; each of these regRNAs was found to be overexpressed in the majority mouse thymic tumors containing nearby integrations as compared control tumors that lacked such integrations.

7. Expression levels of regRNAs and miRNA in human tumors

[0080] The RNA expression level of a newly identified regRNA (PVT1) was measured in human tumors using quantitative methods (Fig. 8). In 3 out of 9 tumors, expression levels of the specific regRNA were elevated as compared to the level in matched normal tissue from the same patient. The change in expression levels may indicate how regRNAs and miRNAs can be used for diagnosis and therapy of the respective tumors for those skilled in the art.

8. Multistep tumorigenesis and co-mutation analysis

[0081] Co-mutation analysis may be a powerful way to find cooperating signaling pathways in tumorigenesis. Viral insertional mutagenesis, while perhaps not providing all the mutations necessary for a full-blown tumor, follows the multistep scenario of spontaneous tumorigenesis. Lymphocytic tumors that arise as a consequence of infection with MLV can contain up to 7 insertion sites. This fact can be used to differentiate between signaling pathways within a tumor: because multiple oncogenic hits along a signaling pathway may not be selected over a single hit, the genes actually recovered are likely not to be involved in the same pathway, but in complementary pathways that work together in tumorigenesis.

[0082] There are generally, however, two main caveats when considering co-mutation analysis. First, although in general, almost all viral insertions in a tumor are thought to be causative in its formation, the question arises whether there are any "passenger" insertions, i.e., insertional events that are not selected by tumorigenesis, but merely accompany other causative mutations. Passenger insertions do not seem to occur frequently due to the superinfection barrier and because secondary integration events are rare. These rare events, however, are responsible for the tumor formation by retroviruses. It is not clear whether the additional insertions are generated by re-infections or by retrotransposition. At any rate, even though passenger mutations have not yet been identified in previous studies, one needs to guard against interpreting such insertions as tumorigenic events—especially when the screen is large. The second confounding issue may be the potential oligoclonality of tumors. If the tumors are not clonal, then what is scored as a co-mutation may simply be a mutation in a different tumor.

[0083] With these caveats, co-mutation analysis provides valuable insight into the pathways that work together during tumorigenesis. The simplified reasoning can be summarized as follows: (i) genes that are co-mutated in a single cancer cell represent different pathways that cooperate during carcinogenesis; and, as a corollary, (ii) genes within the same pathway are never co-mutated.

9. Specific Co-mutations

[0084] Table 2 lists the co-mutations of the polycistrons mir-17-20 and mir-106a-92. From this table, at least three observations can be made:

(1) both polycistrons mir-17-20 and mir-106a-92 have recurrent co-mutations;

(2) they share 10 co-mutations between them; and

(3) both polycistrons cooperate with co-mutations in at least three other (predicted) miRNAs.

[0085] If a genomic region is hit with retroviral insertions only few times in the entire screen, the chance of scoring an accidental co-mutation is lower. While a low frequency may also indicate low importance in tumorigenesis, it may simply reflect the mechanistic restrictions of retroviral insertion at that locus. If a region is hit frequently, the chance of false co-mutations increases. However, careful analysis of the region can minimize false co-mutation assignments. For example, if one only considers known or predicted genes, then in the present screen, there are 500

insertions near or into the Evi5 locus. Not only is this locus an area of preferred integration, the nearby Gfi locus also has similar high integration frequencies. On the one hand, polycistron mir-17-20 seemingly has 11 co-mutations in the Evi5 locus, and polycistron mir-106a-92 has 5. But a closer inspection of these integration sites reveal that the two polycistrons share (five and two, respectively) co-mutations in the 429 nt transcript AK037419, which represents an EST from the neonate thymus. Thus, transcript AK037419 cooperates with polycistrons mir-17-20 and mir-106a-92, respectively. This otherwise nondescript transcript itself is an oncogene as well. On the other hand, there clearly are integrations into the Evi5 gene as well: polycistron mir-17-20 has four co-mutations in intron 17 of Evi5, and polycistron mir-106a-92 has one in intron 16 of Evi5.

[0086] Another frequently hit region in the present screen is Notch1, with 248 integrations. In human T acute lymphatic leukemia, the mutations in the Notch1 locus are not evenly distributed but they fall into two broad groups which affect heterodimerization of the receptor and stability of the cytoplasmic signaling portion of the molecule [30]. The mutations shown here fall into three broad groups, with 128 of the insertions into Notch1 in exon 34; these mutations presumably increase the stability of the cytoplasmic signaling portion [30]. Two of these mutations are each co-mutated with mir-17-20 and mir-106a-92, respectively. As mentioned above more confidence can be placed into a co-mutation, if only a few integration sites are scored in the entire screen, and most or all of these integrations are in the tumors with the first mutation. Thus, another group has 12 mutations in intron 2 of Notch1. Two of these co-mutate with mir-17-20 upon closer inspection, the insertions into intron 2 are only 531 nt apart, and they coincide with transcript BF720900. This is an indication that mir-17-20 is co-mutated with transcript BF720900.

[0087] The set of regulatory RNAs and miRNAs that cause tumors when overexpressed, deleted or otherwise mutated are of particular interest in the present methods. The invention dramatically increases the number of known oncogenic regRNAs and miRNAs and is useful for the diagnosis and therapeutic treatment of human cancers.

C. Diagnostic Methods and Reagents

[0088] In one aspect, the detection, identification, and quantitation of regRNA, including miRNA, and of mutations that affect the expression levels and/or function of these RNAs in tissue, body fluids, secretions and excretions are useful in cancer diagnostics are contemplate. Non-limiting examples include, but are not limited to (i) genotyping tumors for diagnosis, prognosis, and patient stratification in both therapy and clinical trials, and (ii) blood testing for early cancer detection of breast, ovary, colorectal and prostate cancer.

[0089] In one embodiment, an array (chip) containing complementary sequences of the regRNAs is used to score over- or under-expression of regRNAs in cancer tissue, which is linked to the cancer type and the precise diagnosis of it. This in turn allows better prognosis and therapy. In another embodiment of the invention, an oncogenic regRNA survey is carried out by the generally known methods of gel electrophoresis and detection by hybridization to complementary sequences. In yet another embodiment, DNA encoding regRNA is sequenced and mutations are recovered that may indicate non-physiological expression levels and/or function. When performed on bodily fluids, such as blood, these tests may be indicative of the presence of a tumor that escapes early detection by other means, or for which there are no early detection methods, or only detection methods that are more complicated and/or more expensive. Such tests may be carried out on material with or without prior amplification of nucleic acids.

D. Therapeutic Methods

[0090] Those skilled in the art will, upon reading this disclosure, further understand how the disclosure of oncogenic regRNA including miRNA sequences and their co-mutations are useful in therapy. When over-expressed in cancer, the expression of such sequences may be repressed and the physiological state of the tumor cell may be restored, which, in turn prevents further proliferation. When under-expressed in cancer, the expression of such sequences may be supplemented and the physiological state of the tumor cell may be restored, which, in turn prevents further proliferation. When mutated in a way that changes the function, the mutated sequence may be corrected or eliminated. The delivery of

drugs with these corrective effects may be accomplished by the known gene-therapy methods of transfection, infection and transduction.

[0091] In one general therapeutic method, molecules designed to bind specifically and with high affinity, e.g., by sequence-specific hybridization, may be may be employed to block overexpressed miRNA. For example, an oligonucleotide that targets a mature miRNA or its Drosha or Dicer cutting sites, may be employed for blocking levels or activity of a disease specific miRNA, as disclosed for example, in the Genetools website accessed at http://www.gene-tools.com/node/33.

VI. Materials and Methods

A. Generation of tumors by retroviral mutagenesis in mice with chemically mutagenized paternal haplotype

[0092] Cohorts of male BALB/c mice were injected three times with 0, 20, 50, 80 and 100 mg N-ethyl-N-nitrosourea (ENU) /kg body weight, with each injection one week apart [31]. The mice then became sterile, and the length of the sterility period was taken as a measure of the effectiveness of mutagenesis; only mice that had regained fertility after 11 weeks were used. After the sterility period the mice were mated with untreated BALB/cJ female mice to produce F1 pups. For each cohort infected with the SL3-3 virus [32-35], the experiment involved four groups of mice, experimental group (E1) as well as three control groups (C1-C3). For E1, 2500 newborn (less than 36 hours old) pups were injected i.p. with retrovirus (both male and female pups were used). Control group C1: 200 newborn (less than 36 hours old) pups, male or female, from BALB/cJ (ENU-treated) x BALB/cJ crosses were mock-injected i. p., with medium alone. C2: 200 newborn (less than 36 hours old) pups, male or female, from non-treated BALB/cJ x BALB/cJ crosses were injected i. p. with retrovirus. C3: 100 newborn (less than 36 hours old) pups, male or female, from non-treated BALB/cJ x BALB/cJ crosses were mock-injected i. p., with medium alone. In all groups, mice were individually labeled 3-4 weeks after birth. Then the mice were weaned and tumors were allowed to develop. The average latency period was 85±31 days for SL3-3 virus, for tumors in mice with or without ENU mutagenesis of one parent. Once they became moribund due to cancer development, the mice were euthanized, gross necropsy was performed and tumor tissues were prepared.

B. Viral tag recovery

[0093] To identify the integration sites of retroviral proviruses, the unknown flanking DNA was isolated using minor modifications of an anchored PCR method [20]. Genomic tumor DNA from spleen or thymus was digested with enzyme 1, and a splinkerette adapter was ligated. This was followed by digestion of enzyme 2, to remove the internal viral fragment. The ligated DNA was amplified by PCR with adapter and virus-specific primers, followed by two additional PCR amplification steps with nested primers. The PCR product was purified by gel electrophoresis and sequenced. The sequence chromatograms were then fed into the bioinformatics pipeline for gene identification.

C. Bioinformatics

[0094] The proviral inserts served as DNA tags for gene sequencing and identification. To extract and analyze genomic tags, a computational process was implemented. The input to this process was DNA sequencing chromatogram files, from which high-quality sequences were derived and matched to the mouse genome.

[0095] First, the sequence extraction step converted a chromatogram into a searchable tag sequence. The criteria for a searchable tag sequence include, but are not limited to, high-quality base-calls, non-vector sequence, non-repeat sequence and a length minimum. The base caller LifeTrace™ [36] was used to generate base calls from chromatograms and quality scores representing the accuracy of each base-call.

[0096] Second, using the base calls and quality scores from LifeTrace™, and a database of vector sequences, an algorithm was developed to automatically produce searchable sequences (i.e. sequences that can be matched to the mouse genome). In this algorithm, the region of high quality base calls was first determined by locating the longest stretch of base calls with a window-averaged quality score of 10. A window size of 11 was used to average the quality scores from five bases before to five bases after a central base-call. A quality score of ten indicated 90% accuracy of base calls.

[0097] Third, a database of vector sequences (entire retroviral genome sequences) was matched against the base calls to determine regions of viral sequence using the BLAST algorithm. Based on the sequencing construct, a stretch

of less than 50 bases of viral sequence is expected on the 5' end of the raw sequence; and read-through of short inserts can produce regions of 3' viral sequence starting with a specific restriction site. If a region of high-quality, non-vector sequence longer than 32 bases remains, it becomes a searchable tag sequence.

[0098] Finally, a searchable tag sequence is a stretch of high-quality base calls that should be derived from the mouse genome. The MegaBLAST algorithm was used to search the mouse genome with each searchable sequence. A version of the mouse genome that has been "masked" for repeat sequences (both low-information local repeats and dispersed repetitive elements are not allowed for matches) was used at this step so that non-informative matches are not pursued. For each significant match to a tag sequence (there is usually only one, but occasionally there are more), 2 kb of *unmasked* genomic sequence is retrieved and realigned to the tag sequence. This realignment produces a more complete match in cases where the global search was interrupted by masked repetitive regions. Lastly, the latest annotation files from the (March 2005) UC Santa Cruz build of the mouse genome (mm6 or mm7) were used to locate nearby known and predicted genes. The genomic region into which the provirus inserted is displayed in the UC Santa Cruz Genome Browser (//genome.ucsc.edu).

D. Algorithm to identify miRNAs

[0099] To identify miRNAs, a method that takes advantage of the characteristic form of conservation profiles observed for most known miRNAs [26] was used. This form consists of a drop in conservation immediately flanking pre-miRNA regions. Mouse-human (mm6-hgl7) whole genome alignments from the UCSC Genome Bioinformatics website (genome.ucsc.edu) were used. For every position in the alignments, the percentage conservation in a 15 nucleotide window was calculated and assigned a value of 0 to 9 for 0% to 90% identity and "o" for 100% identity. The resulting conservation strings were then searched for a match to the following Perl™ regular expression:

/([0-8] {50,60})([o98]{53,260})([0-8] {50,60 })/

[00100] Sequences that matched this were further analyzed with RNAfold [37-39] to compute optimal secondary structures. The secondary structure output is in bracket notation where parentheses represent base pairings and periods are

unpaired bases. A structure as an miRNA candidate was accepted if it matched the following Perl™ regular expression:

/((\((?:\.*\(){24,})(\.{2,17 }|\.*\({1,8}\.*\){1,8}\.*\({1,8}\.* \){8}\.*)(\)(?:\.*\)){150, }))/x

[00101] This method detected 81% of all known mouse miRNAs.

VII. EXAMPLES

[00102] The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the present invention, and are not intended to limit the scope of what the inventors regard as their invention nor are they intended to represent that the experiments below are all or the only experiments performed. Efforts have been made to ensure accuracy with respect to numbers used (e.g. amounts, temperature, etc.) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, molecular weight is weight average molecular weight, temperature is in degrees Centigrade, and pressure is at or near atmospheric.

Example 1

Viral tag recovery and locus identification

[00103] Viral integrations sites (tags) were determined from tumors that were isolated and digested genomic tumor DNA, by using an anchored PCR technique as described above. This was performed by amplifying and sequencing a chimeric DNA fragment consisting of a short genomic sequence upstream of the viral 5' LTR and part of the viral 5' LTR itself. The tags were sequenced and mapped to the mouse genome sequence, and the affected transcription unit was determined. From 2373 tumors, 7300 tags were obtained, which mapped to 2,038 regions. Of these regions, 645 had two or more associated integration sites, with the largest region having 500 integrations.

Example 2

Expression levels of regRNAs in mouse thymic tumors

[00104] The RNA expression levels of three regRNAs were measured in mouse thymic tumors using quantitative methods with the results shown in Figs 5-7. Mouse tumors with integrations located in regions containing the regRNAs and control tumors (which lack such integrations) were examined by quantitative PCR using

SYBR green. In all three regions, the majority of tumors have integrations which caused elevated expression of their respective noncoding RNAs.

[00105] The first region (R857:2) examined contains a group of noncoding transcripts located on chromosome 15, ~ 50 kb downstream of the Myc gene (Fig. 5A). A primer set was designed to the 5' end of AK030859 which is common to exon 1 of the other transcripts in the group. The sequence probed also falls within exon 1 of PVT1 (AK090048, plasmacytoma variant translocation 1), a region known for frequent chromosomal translocations [40]. Twenty seven tumors with integrations in this area were assayed for AK030859 expression levels (see Fig. 5B for tumor locations). In 11 of 19 tumors containing integrations located within and downstream of AK030859, expression of AK030859 was elevated 5 to 40 fold over tumors with no integrations in this region (Fig. 5C).

[00106] A second region (R894:1) with a high density of integration sites contains noncoding transcript AK040062 which is located on chromosome 2 (Fig. 6A). Primer sets were designed to AK040062 exon 2 and expression levels were measured for 24 tumors with integrations in this region (Fig. 6B). Elevated expression of AK040062 exon 2 was seen in tumors with integrations located upstream and within intron 1 of AK040062 (Fig. 6C). Of these 14 tumors, 10 had over 20 fold elevated expression of the noncoding RNA.

[00107] A third region (R217:3) examined for expression levels contains AK037419, a noncoding transcript located on chromosome 5, ~15 kb downstream of the Gfi1 gene (Fig. 7A). Expression levels of AK037419 exon 3 were measured by qPCR in 16 tumors containing integration sites in this region (Fig. 7B). Expression of AK037419 exon 3 was increased between 7 to 1000 fold in 11 of the 16 tumors tested as compared to control tumors with no integrations in this region (Fig. 7C).

Example 3

Expression levels of regRNAs in human tumors

[00108] The RNA expression levels of a newly identified regRNA (PVT1) was measured in human tumors using quantitative methods with the results shown in Fig. 8. The expression levels of PVT1 exon 1 were measured in matched human normal and cancer prostate RNA samples. Of nine matched tissue pairs, three tumor samples displayed 2 to 4 fold elevated expression of PVT1 exon1 as compared to

their matched normal sample. Expression levels of PVT1 were measured by SYBR Green qPCR [41] using primer sets designed to PVT1 exon 1.

[00109] The preceding merely illustrates the principles of the invention. It will be appreciated that those skilled in the art will be able to devise various arrangements which, although not explicitly described or shown herein, embody the principles of the invention and are included within its spirit and scope. Furthermore, all examples and conditional language recited herein are principally intended to aid the reader in understanding the principles of the invention and the concepts contributed by the inventors to furthering the art, and are to be construed as being without limitation to such specifically recited examples and conditions.

[00110] Moreover, all statements herein reciting principles, aspects, and embodiments of the invention as well as specific examples thereof, are intended to encompass both structural and functional equivalents thereof. Additionally, it is intended that such equivalents include both currently known equivalents and equivalents developed in the future, i.e., any elements developed that perform the same function, regardless of structure. The scope of the present invention, therefore, is not intended to be limited to the exemplary embodiments shown and described herein. Rather, the scope and spirit of present invention is embodied by the appended claims.

IT IS CLAIMED:

1. A method for positively identifying a human miRNA sequence associated with a detectable disease state in humans, comprising

(i) identifying, from each of at least two animals having a detectable disease state produced by insertional mutation, the sequence of a genomic segment that is common to both animals, and that contains an insertional mutation,

(ii) identifying transcription units contained within the animal genome that are within about 200 Kbases, in either an upstream or downstream direction, of the sequenced genomic segment,

(iii) identifying human genomic transcription units that are orthologous to the transcription units identified in step (ii), and

(iv) for each human transcription unit identified in step (iii), employing a bioinformatics program capable of identifying putative miRNA sequences, to determine whether that transcription unit identified in step (iii) contains a putative miRNA sequence, in which case the putative miRNA sequence is positively identified as a human miRNA.

2. The method of claim 1, wherein the detectable disease state is a cancer, and step (i) is carried out by isolating the genomic segment from each of at least two animals having a detectable cancer.

3. The method of claim 1, wherein the detectable cancer is a lymphoma, and step (i) is carried out by isolating the genomic segment from each of at least two animals having a lymphoma.

4. The method of claim 1, wherein the insertional mutation in step (i) is a viral insertional mutation.

5. The method of claim 1, wherein the sequence identified in step (iii) is contained in a pri-miRNA.

6. The method of claim 1, wherein the sequence identified in step (iii) is contained completely within the mature miRNA.

7. An assay kit for diagnosing the presence or risk of cancer in a human subject comprising

a first reagent designed to react specifically with a human pri-miRNA or mature miRNA sequence identified in accordance with the method of claim 2, to form a first detectable reaction product, and

an indicator guide that indicates how the presence or amount of the reaction product correlates with the presence or risk of the disease state in a human subject.

8. The kit of claim 7, wherein the first reagent includes one of: (a) PCR reagents for detecting the presence or absence of the genomic sequence, and (b) oligonucleotide binding reagents for detecting the presence or absence of the genomic sequence.

9. The kit of claim 7, for use in diagnosing the presence of risk of a cancer in a human subject, wherein step (i) in the method of claim 1 is carried out by isolating the genomic fragment from each of at least two animals having a detectable cancer.

10. The kit of claim 9, for use in diagnosing the presence of risk of a lymphoma in a human subject, wherein step (i) in the method of claim 1 is carried out by isolating the genomic segment from each of at least two animals having a detectable cancer.

11. The kit of claim 1, wherein the first reagent is designed to react specifically with a mature human miRNA sequence identified in accordance with the method of claim 1.

12. A method of treating a cancer in a human subject comprising administering to the subject, a therapeutically effective amount of a compound capable of binding specifically to a mature human miRNA sequence identified in accordance with the method of claim 2.

13. An isolated mature human miRNA sequence selected from the group consisting of SEQ ID NOS: 1-55.

14. A method for identifying a human regulatory RNA (regRNA) sequence associated with a detectable disease state in humans, comprising

(i) identifying, from each of at least two animals having a detectable disease state produced by insertional mutation, the sequence of a genomic segment that is common to both animals, and that contains an insertional mutation,

(ii) identifying transcription units contained within the animal genome that are within about 200 Kbases, in either an upstream or downstream direction, of the sequenced genomic segment,

(iii) identifying human genomic transcription units that are orthologous to the transcription units identified in step (ii),

(iv) for each human transcription unit identified in step (iii), using a bioinformatics program to determine whether that transcription unit is a non-coding RNA sequence, and

(v) if the orthologous homologous human genomic sequence from step (iv) is a non-coding RNA sequence, classifying the sequence as a human regRNA sequence associated with the detectable disease state.

15. The method of claim 14, wherein the detectable disease state is a cancer, and step (i) is carried out by isolating the genomic segment from each of at least two animals having a detectable cancer.

16. The method of claim 14, wherein the human regRNA sequence is an miRNA, and step (iv) includes employing a bioinformatics program capable of identifying putative miRNA sequences to determine whether that transcription unit identified in step (iii) contains a putative miRNA sequence, in which case the putative miRNA sequence is positively identified as a human miRNA.

17. The method of claim 14, wherein the insertional mutation in step (i) is a viral insertional mutation.

18. The method of claim 14, which further includes utilizing the identified human regRNA sequence for diagnostic or therapeutic purposes.

19. An assay kit for diagnosing the presence or risk of cancer in a human subject comprising

a first reagent designed to react specifically with a human regulatory RNA (regRNA) sequence identified in accordance with the method of claim 15, to form a first detectable reaction product, and

an indicator guide that indicates how the presence or amount of the reaction product correlates with the presence or risk of the disease state in a human subject.

20. The kit of claim 19, wherein the first reagent includes one of: (a) PCR reagents for detecting the presence or absence of the genomic sequence, and (b) oligonucleotide binding reagents for detecting the presence of absence of the genomic sequence.

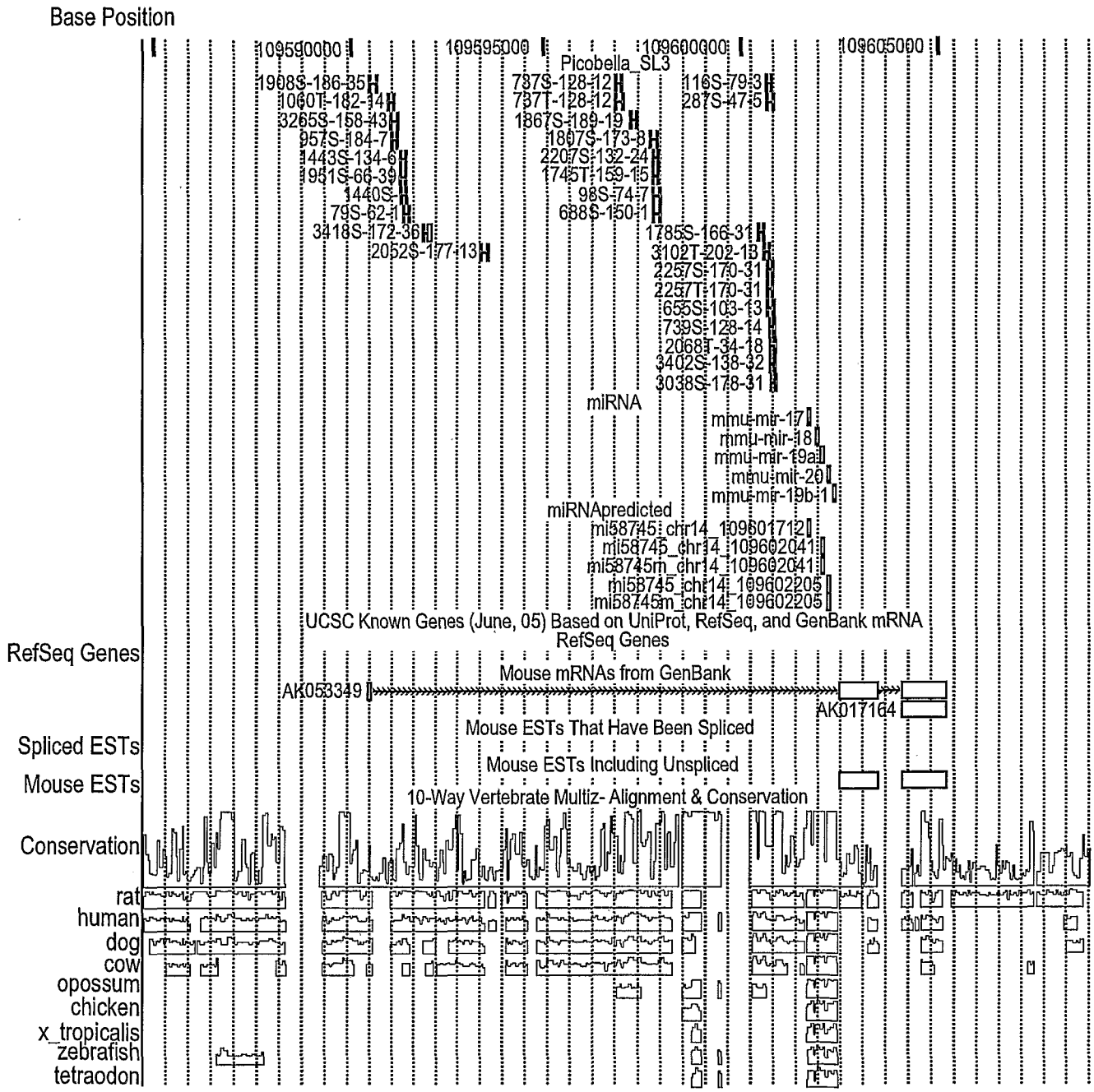
Base Position
109590000
109595000
109600000
109605000
Picobella_SL3
1908S-186-35
1060T-182-14
3265S-158-43
957S-184-7
1443S-134-6
1951S-66-39
1440S-
79S-62-1
3418S-172-36
2052S-177-13
787S-128-12
787T-128-12
1867S-189-19
1807S-173-8
2207S-132-24
1745T-159-15
98S-74-7
688S-150-1
116S-79-3
287S-47-5
1785S-166-31
3102T-202-13
2257S-170-31
2257T-170-31
655S-103-13
739S-128-14
2068T-34-18
3402S-138-32
3038S-178-31
miRNA
mmu-mir-17
mmu-mir-18
mmu-mir-19a
mmu-mir-20
mmu-mir-19b-1
miRNApredicted
mi58745_chr14_109601712
mi58745_chr14_109602041
mi58745m_chr14_109602041
mi58745_chr14_109602205
mi58745m_chr14_109602205
UCSC Known Genes (June, 05) Based on UniProt, RefSeq, and GenBank mRNA
RefSeq Genes
RefSeq Genes
Mouse mRNAs from GenBank
AK053349
AK017164
Spliced ESTs
Mouse ESTs That Have Been Spliced
Mouse ESTs
Mouse ESTs Including Unspliced
10-Way Vertebrate Multiz- Alignment & Conservation
Conservation
rat
human
dog
cow
opossum
chicken
x_tropicalis
zebrafish
tetraodon

Fig. 1A

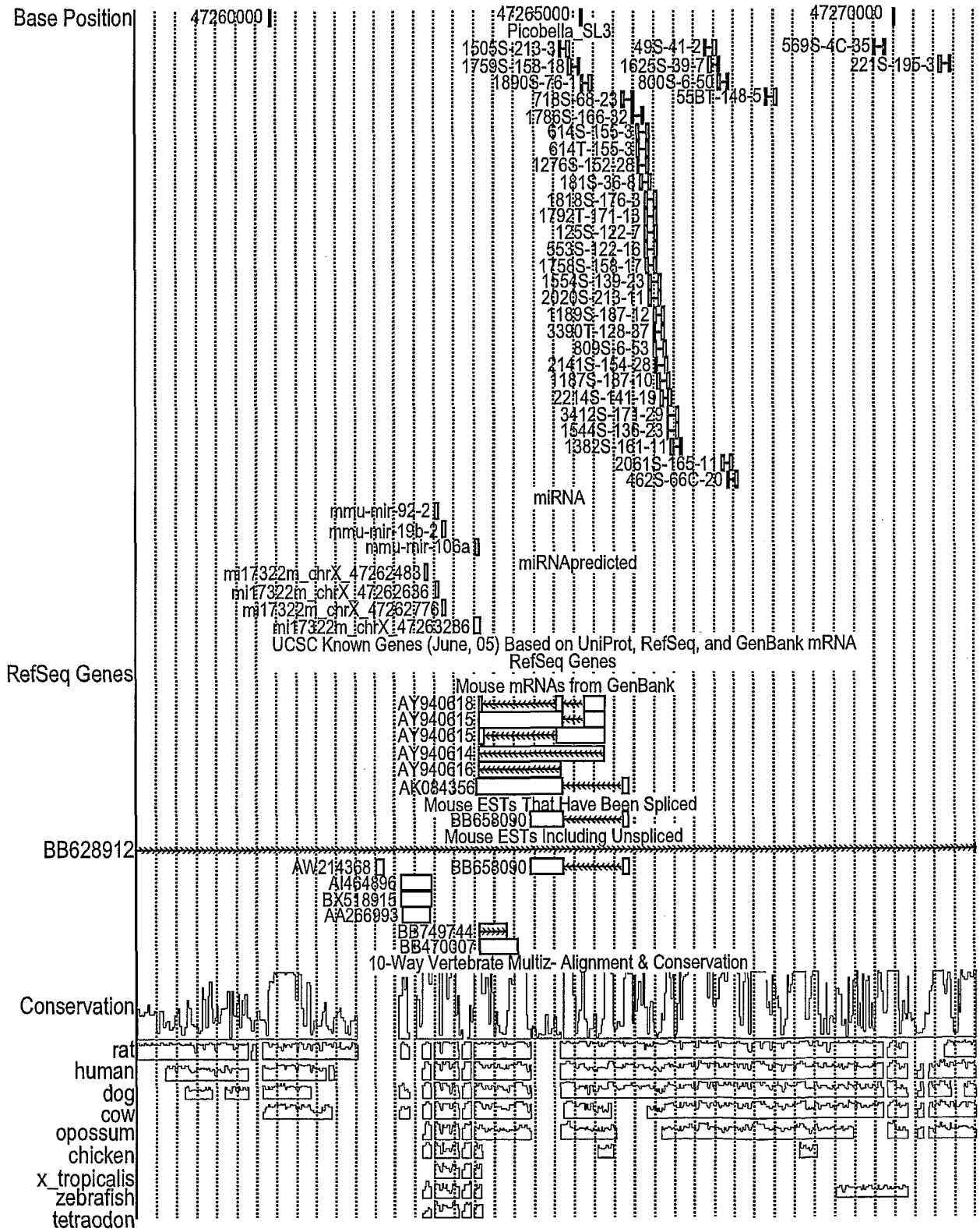
Base Position
47260000
47265000
47270000
Picobella_SL3
miRNA
miRNApredicted
UCSC Known Genes (June, 05) Based on UniProt, RefSeq, and GenBank mRNA
RefSeq Genes
RefSeq Genes
Mouse mRNAs from GenBank
Mouse ESTs That Have Been Spliced
Mouse ESTs Including Unspliced
BB628912
10-Way Vertebrate Multiz- Alignment & Conservation
Conservation
rat
human
dog
cow
opossum
chicken
x_tropicalis
zebrafish
tetraodon

Fig. 1B

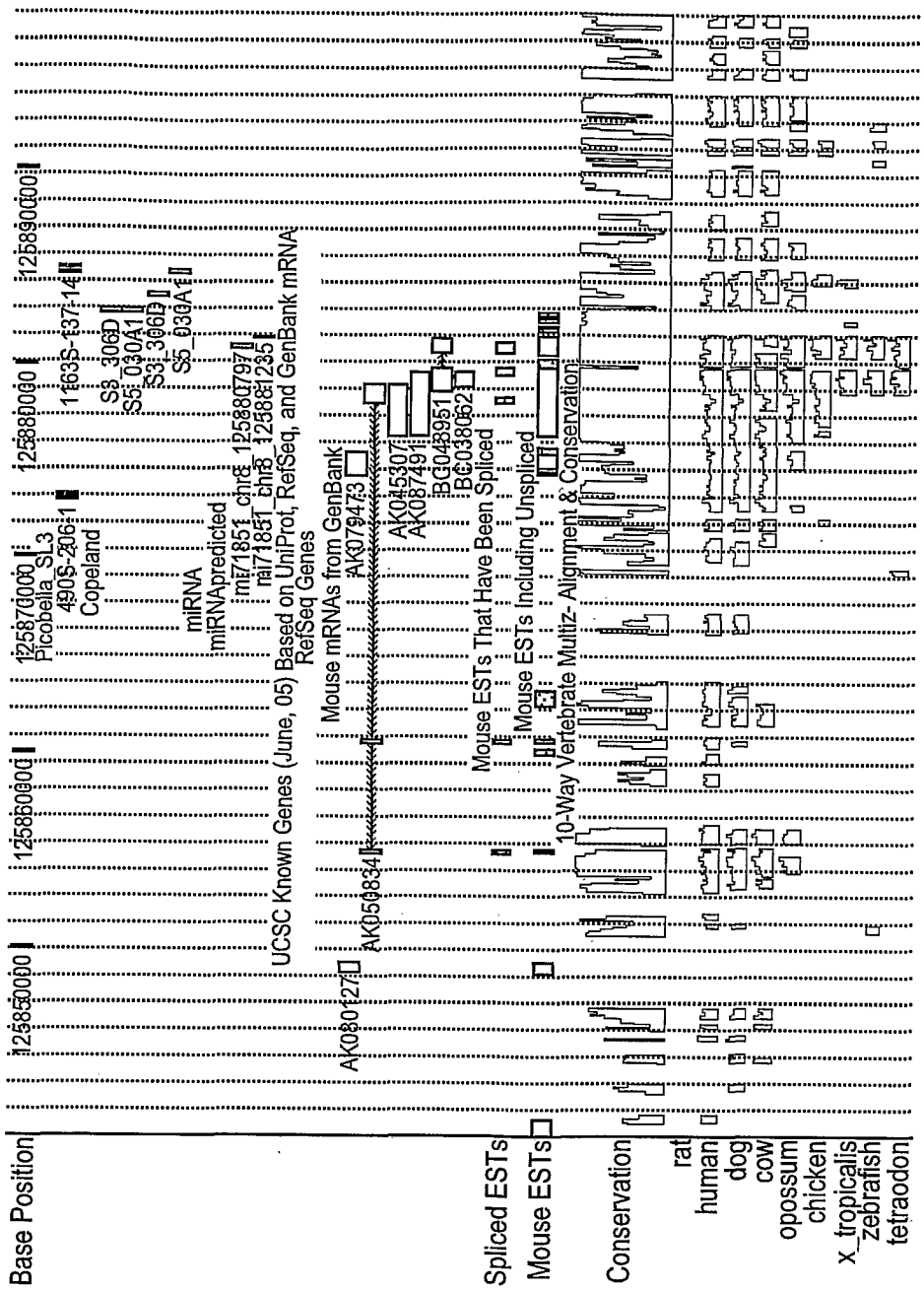
Base Position
Spliced ESTs
Mouse ESTs
Conservation
rat
human
dog
cow
opossum
chicken
x_tropicalis
zebrafish
tetraodon
Picobella_SL3
Copeland
miRNA
miRNApredicted
UCSC Known Genes (June, 05) Based on UniProt, RefSeq, and GenBank mRNA
RefSeq Genes
Mouse mRNAs from GenBank
Mouse ESTs That Have Been Spliced
Mouse ESTs Including Unspliced
10-Way Vertebrate Multiz- Alignment & Conservation

Fig. 2A

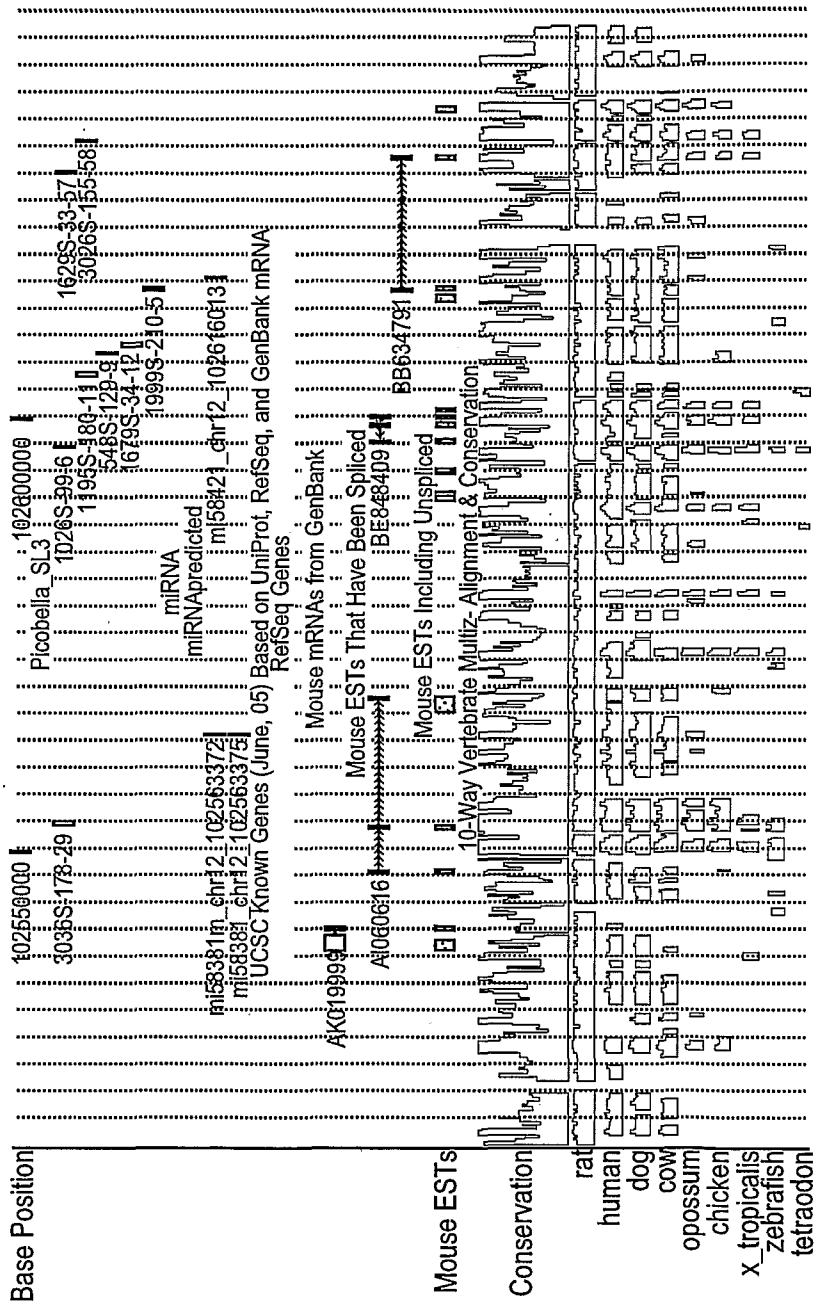
Base Position
Picobella_SL3
miRNA
miRNApredicted
UCSC Known Genes (June, 05) Based on UniProt, RefSeq, and GenBank mRNA
RefSeq Genes
Mouse mRNAs from GenBank
Mouse ESTs That Have Been Spliced
Mouse ESTs Including Unspliced
Mouse ESTs
10-Way Vertebrate Multiz Alignment & Conservation
Conservation
rat
human
dog
cow
opossum
chicken
x_tropicalis
zebrafish
tetraodon

Fig. 2B

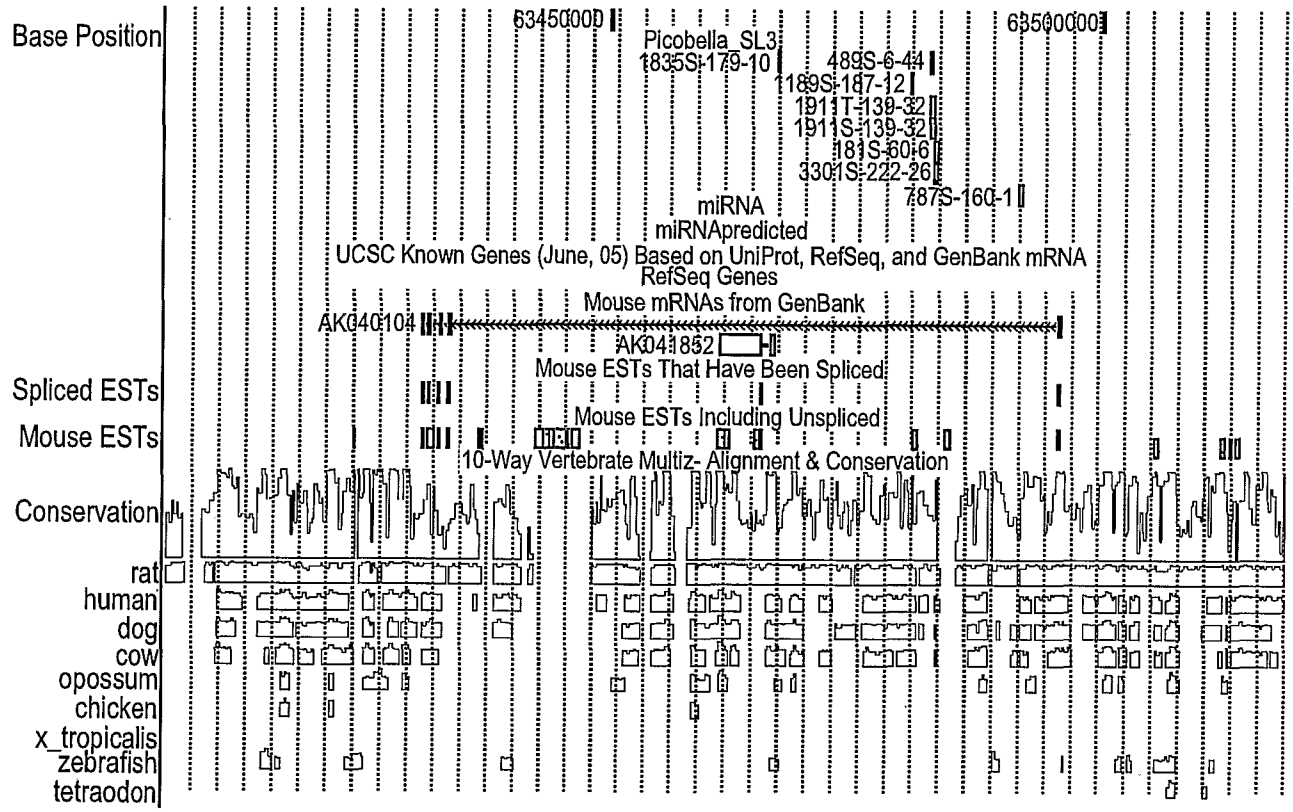
Base Position
63450000
63500000
Picobella_SL3
1835S-179-10
489S-6-44
1189S-187-12
1911T-139-32
1911S-139-32
181S-60-6
330-1S-222-26
787S-160-1
miRNA
miRNApredicted
UCSC Known Genes (June, 05) Based on UniProt, RefSeq, and GenBank mRNA
RefSeq Genes
Mouse mRNAs from GenBank
AK040104
AK041852
Spliced ESTs
Mouse ESTs That Have Been Spliced
Mouse ESTs
Mouse ESTs Including Unspliced
10-Way Vertebrate Multiz- Alignment & Conservation
Conservation
rat
human
dog
cow
opossum
chicken
x_tropicalis
zebrafish
tetraodon

Fig. 3A

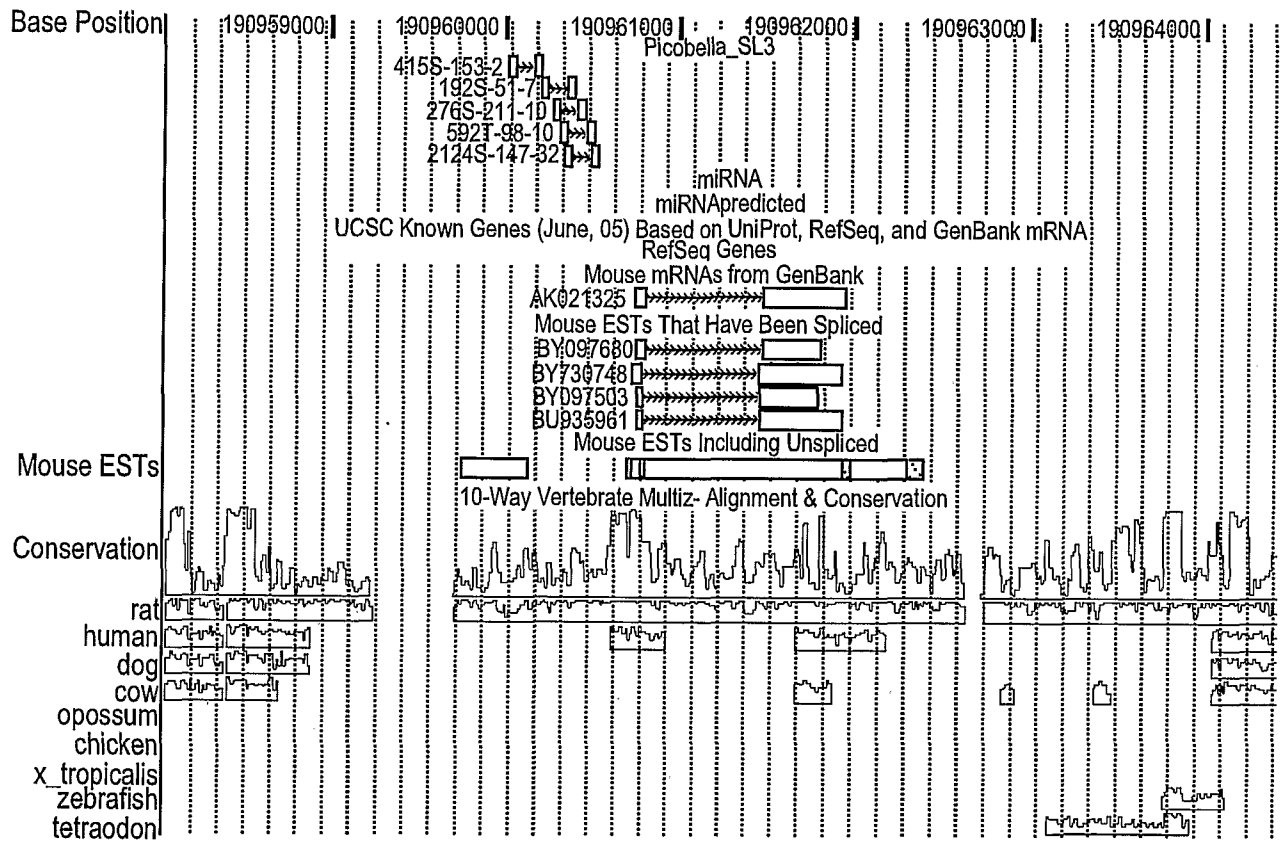
Base Position
190959000
190960000
190961000
190962000
190963000
190964000
Picobella_SL3
4158-153-2
192S-51-7
276S-211-10
592T-98-10
2124S-147-32
miRNA
miRNApredicted
UCSC Known Genes (June, 05) Based on UniProt, RefSeq, and GenBank mRNA
RefSeq Genes
Mouse mRNAs from GenBank
AK021325
Mouse ESTs That Have Been Spliced
BY097680
BY730748
BY097503
BU935961
Mouse ESTs Including Unspliced
Mouse ESTs
10-Way Vertebrate Multiz- Alignment & Conservation
Conservation
rat
human
dog
cow
opossum
chicken
x_tropicalis
zebrafish
tetraodon

Fig. 3B

Tumor	Orientation	Location
1759S	---	chrX:47836458
1890S	+++	chrX:47836460
718S	+++	chrX:47837112
1786S	---	chrX:47837483
818S	+++	chrX:47837396
553S	---	chrX:47837694
1758S	+++	chrX:47837503
1189S	+++	chrX:47837606
3390T	---	chrX:47837806
3412S	+++	chrX:47837818
1544S	---	chrX:47838017
49S	---	chrX:47838631
800S	+++	chrX:47838650
2061S	---	chrX:47838885
462S	+++	chrX:47838785
494S	+++	chrX:47844147
1469S	+++	chrX:47847472
195S	+++	chrX:47848368
463S	+++	chrX:47849934
1415S	+++	chrX:47851741
2057S	---	chrX:47852851

Transcript:	mmu-mir-106a	---	chrX:47834755-47834819

Fig. 4A

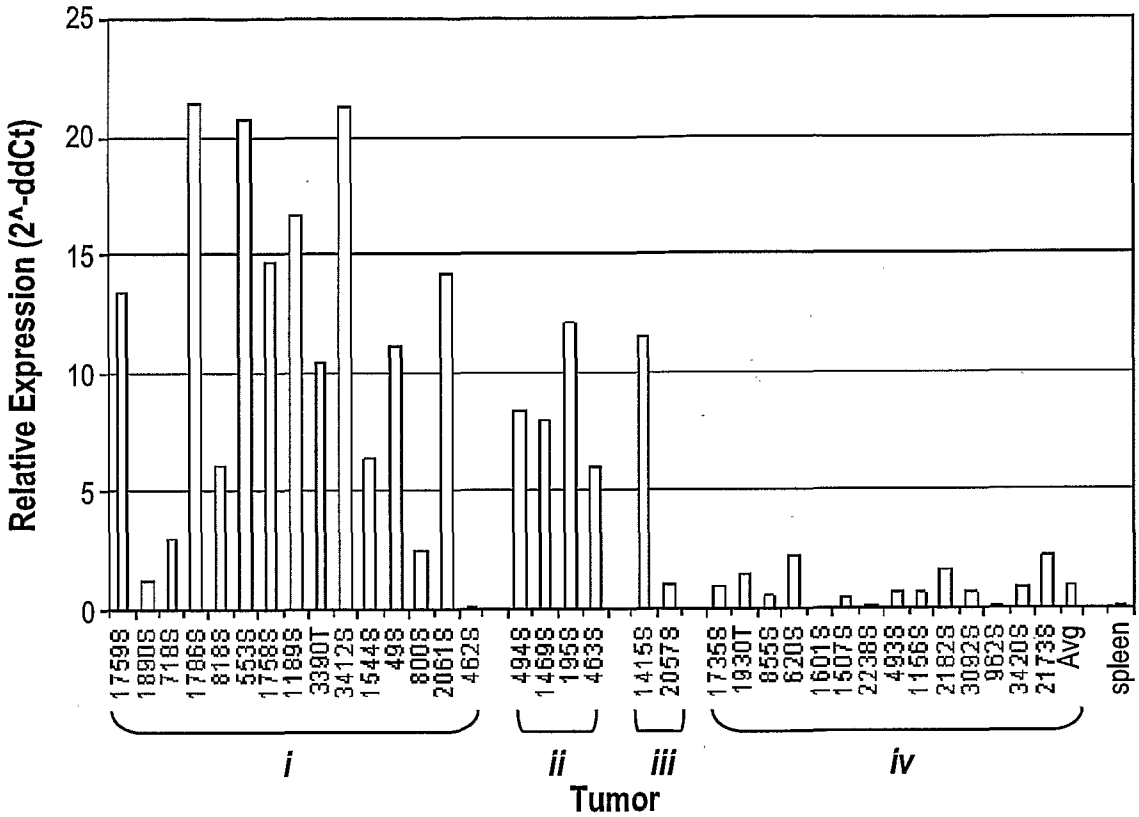
Relative Expression (2^-ddCt)
25
20
15
10
5
0
1759S
1890S
718S
1786S
818S
553S
1758S
1189S
3390T
3412S
1544S
49S
800S
2061S
462S
494S
1469S
195S
463S
1415S
2057S
1735S
1930T
855S
620S
1601S
1507S
2238S
493S
1156S
2182S
3092S
962S
3420S
2173S
Avg
spleen
i
ii
iii
iv
Tumor

Fig. 4B

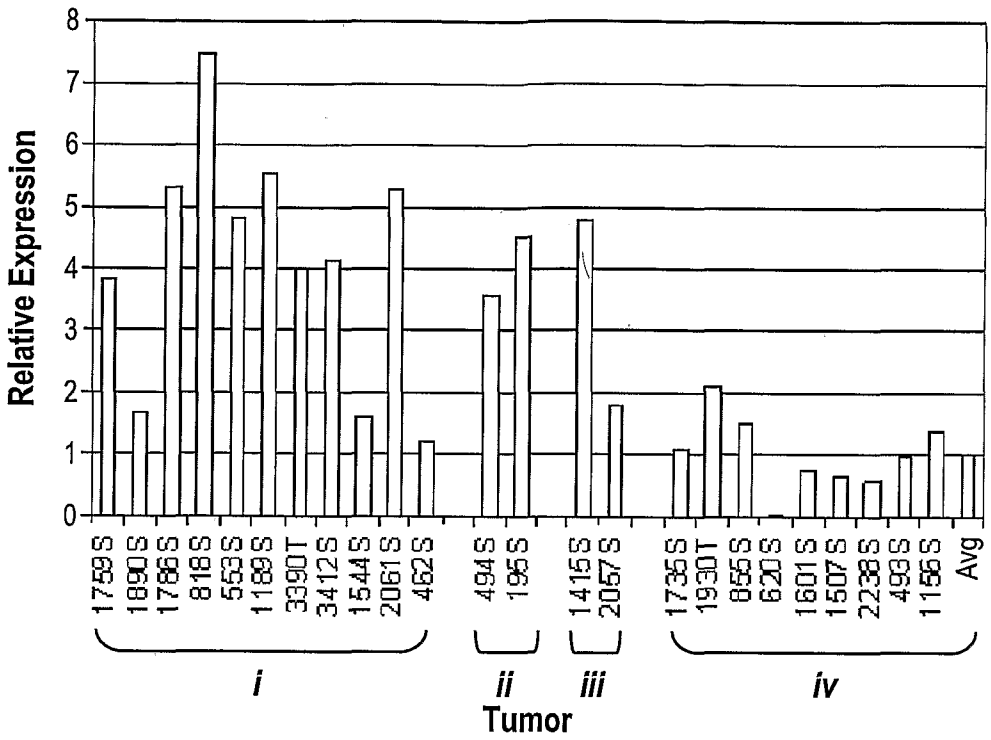
Relative Expression
8
7
6
5
4
3
2
1
0
1759S
1890S
1786S
818S
553S
1189S
3390T
3412S
1544S
2061S
462S
494S
195S
1415S
2057S
1735S
1930T
855S
620S
1601S
1507S
2238S
493S
1156S
Avg
i
ii
iii
iv
Tumor

Fig. 4C

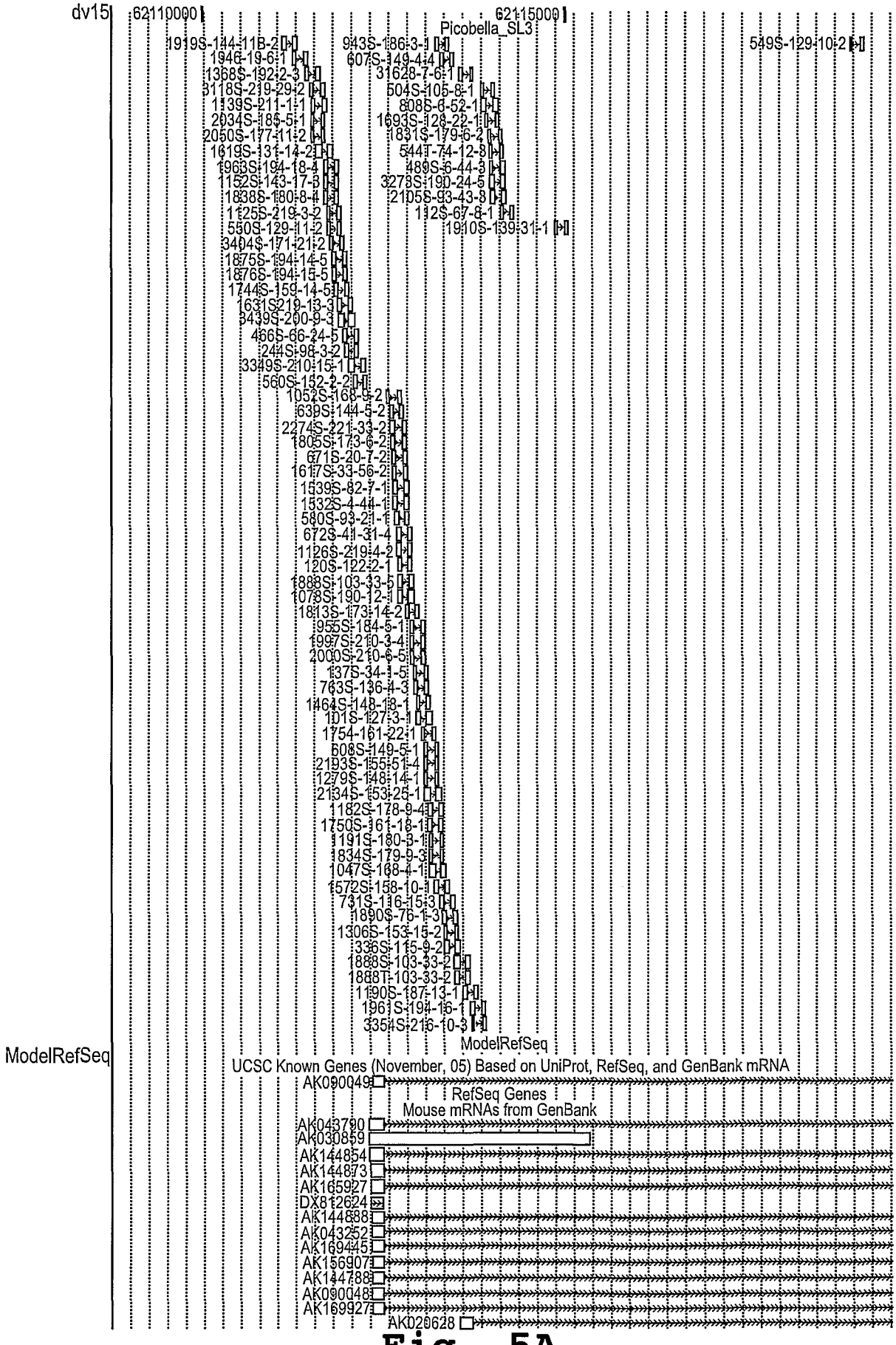
dv15
62110000
62115000
Picobella_SL3
ModelRefSeq
ModelRefSeq
UCSC Known Genes (November, 05) Based on UniProt, RefSeq, and GenBank mRNA
RefSeq Genes
Mouse mRNAs from GenBank
AK090049
AK043790
AK030859
AK144854
AK144873
AK165927
DX812624
AK144888
AK043252
AK169445
AK156907
AK144788
AK090048
AK169927
AK026628

Fig. 5A

Tumor	Orientation	Location
1919S	+++	chr15:62111106
194S	+++	chr15:62111238
1139S	+++	chr15:62111498
1963S	+++	chr15:62111686
3404S	+++	chr15:62111761
1744S	+++	chr15:62111835
244S	+++	chr15:62111963
560S	+++	chr15:62112094
1052S	+++	chr15:62112545
672S	+++	chr15:62112681
1813S	+++	chr15:62112803
955S	---	chr15:62113068
763S	+++	chr15:62112926
1754S	+++	chr15:62113009
1182S	+++	chr15:62113130
1572S	+++	chr15:62113215
1890S	+++	chr15:62113324
1888S	---	chr15:62113689
1190S	+++	chr15:62113599
2105S	+++	chr15:62113987
1910S	+++	chr15:62114870
504S	+++	chr15:62113841
549S	+++	chr15:62118947
3005S	+++	chr15:62129190
1437S	+++	chr15:62138613
455S	+++	chr15:62154036
2262S	+++	chr15:62167000

Transcript:	AK030859	+++	chr15:62112297-62115364

Fig. 5B

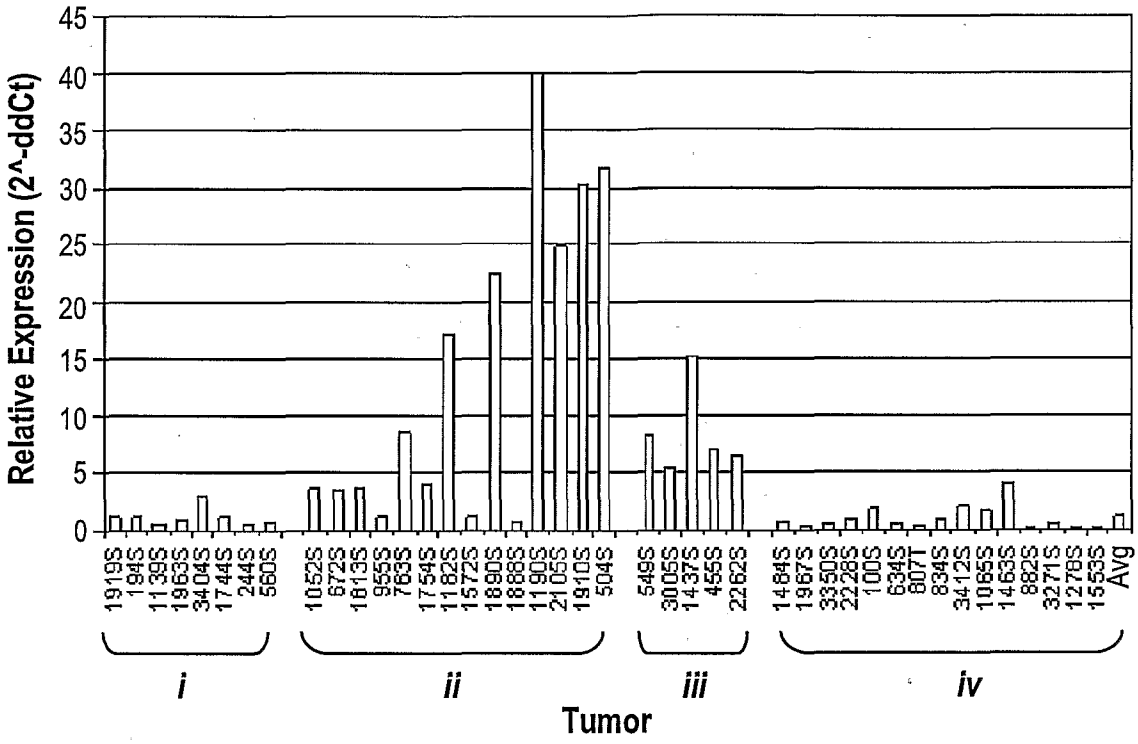
Relative Expression (2^-ddCt)
45
40
35
30
25
20
15
10
5
0
1919S
194S
1139S
1963S
3404S
1744S
244S
560S
1052S
672S
1613S
955S
763S
1754S
1182S
1572S
1890S
1888S
1190S
2105S
1910S
504S
549S
3005S
1437S
455S
2262S
1484S
1967S
3350S
2228S
100S
634S
807T
834S
3412S
1065S
1463S
882S
3271S
1278S
1553S
Avg
i
ii
iii
iv
Tumor

Fig. 5C

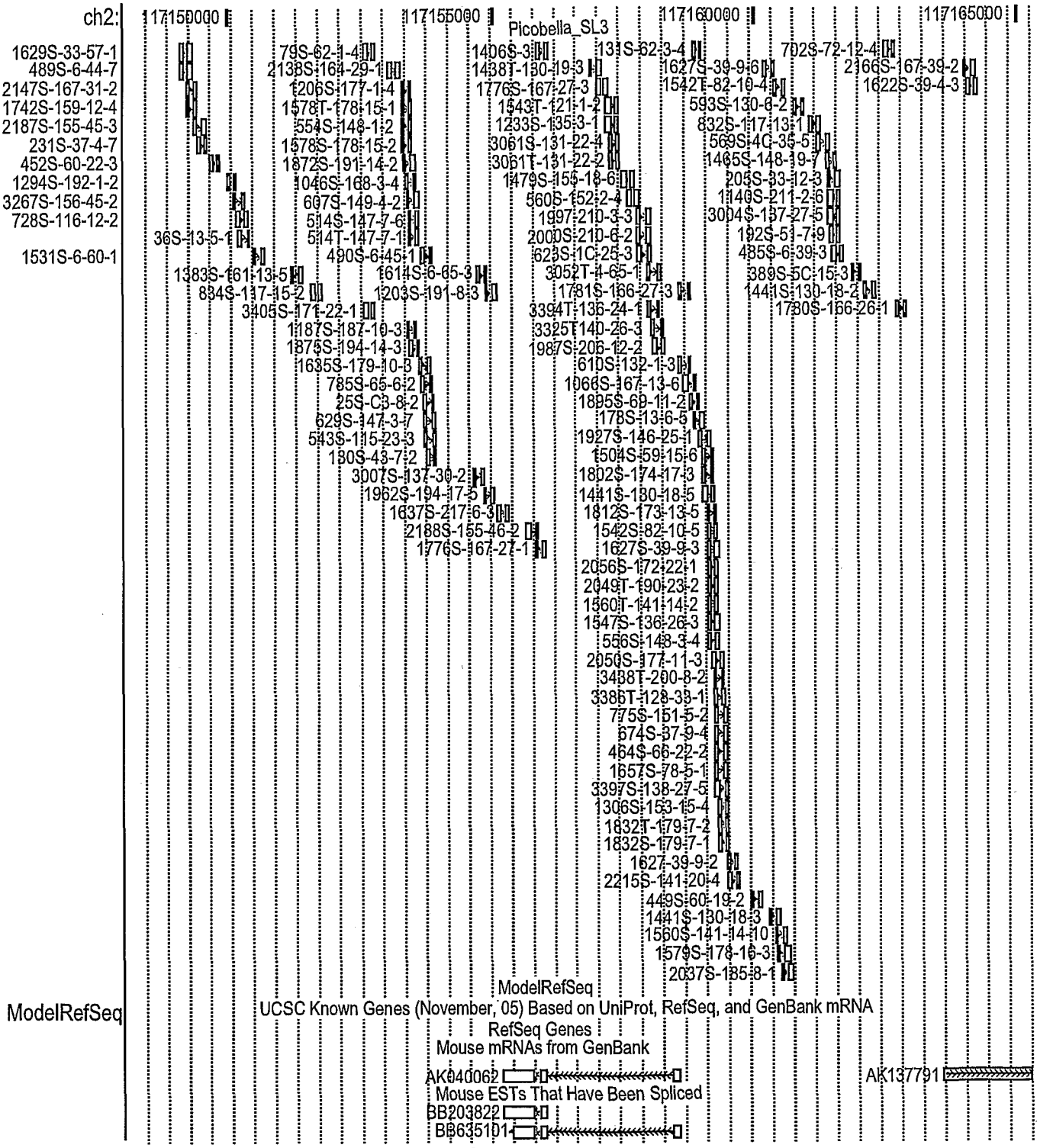
ch2:
117150000
117155000
117160000
117165000
Picobella_SL3
1629S-33-57-1
489S-6-44-7
2147S-167-31-2
1742S-159-12-4
2187S-155-45-3
231S-37-4-7
452S-60-22-3
1294S-192-1-2
3267S-156-45-2
728S-116-12-2
1531S-6-60-1
79S-62-1-4
2138S-164-29-1
1206S-177-1-4
1578T-178-15-1
554S-148-1-2
1578S-178-15-2
1872S-191-14-2
1046S-168-3-4
607S-149-4-2
514S-147-7-6
514T-147-7-1
490S-6-45-1
1383S-161-13-5
834S-117-15-2
3405S-171-22-1
1187S-187-10-3
1875S-194-14-3
1635S-179-10-3
785S-65-6-2
25S-C3-8-2
629S-147-3-7
543S-115-23-3
130S-43-7-2
3007S-137-30-2
1962S-194-17-5
1637S-217-6-3
2188S-155-46-2
1776S-167-27-1
1614S-6-65-3
1203S-191-8-3
1406S-3
1438T-130-19-3
1776S-167-27-3
1543T-121-1-2
1233S-135-3-1
3061S-131-22-4
3061T-131-22-2
1479S-155-18-6
560S-152-2-4
1997-210-3-3
2000S-210-6-2
623S-1C-25-3
3052T-4-65-1
1781S-166-27-3
3394T-136-24-1
3325T140-26-3
1987S-206-12-2
610S-132-1-3
1066S-167-13-6
1895S-69-1-2
178S-13-6-5
1927S-146-25-1
1504S-59-15-6
1802S-174-17-3
1441S-130-18-5
1812S-173-13-5
1542S-82-10-5
1627S-39-9-3
2056S-172-22-1
2049T-190-23-2
1560T-141-14-2
1547S-136-26-3
556S-148-3-4
2050S-177-11-3
3438T-200-8-2
3386T-128-33-1
775S-151-5-2
674S-37-9-4
464S-66-22-2
1657S-78-5-1
3397S-138-27-5
1306S-153-15-4
1832T-179-7-2
1832S-179-7-1
1627-39-9-2
2215S-141-20-4
449S-60-19-2
1441S-130-18-3
1560S-141-14-10
1579S-178-16-3
2037S-135-8-1
1627S-39-9-6
1542T-82-10-4
593S-130-6-2
832S-117-13-1
569S-4C-35-5
1465S-148-19-7
205S-33-12-3
1140S-211-2-6
3004S-137-27-5
192S-51-7-9
485S-6-39-3
389S-5C-15-3
1441S-130-18-2
1780S-166-26-1
702S-72-12-4
2166S-167-39-2
1622S-39-4-3
ModelRefSeq
ModelRefSeq
UCSC Known Genes (November, 05) Based on UniProt, RefSeq, and GenBank mRNA
RefSeq Genes
Mouse mRNAs from GenBank
AK040062
AK137791
Mouse ESTs That Have Been Spliced
BB203822
BB635101

Fig. 6A

Tumor	Orientation	Location
1622S	+++	chr2:117164076
2166S	+++	chr2:117164028
702S	---	chr2:117162701
205S	+++	chr2:117161427
593S	+++	chr2:117160770
1560S	---	chr2:117160586
1441S	---	chr2:117159236
1627S	+++	chr2:117159458
1812S	+++	chr2:117159081
1987S	+++	chr2:117158114
2000S	+++	chr2:117157821
560S	+++	chr2:117157627
1479S	+++	chr2:117157538
3061S	+++	chr2:117157283
1406S	+++	chr2:117155891
1776S	+++	chr2:117157046
1203S	+++	chr2:117154878
785S	+++	chr2:117153665
1206S	+++	chr2:117153260
1383S	+++	chr2:117151179
1742S	+++	chr2:117149169
652T	+++	chr2:117146294
890S	+++	chr2:117144963
670S	+++	chr2:117139386

Transcript:	AK040062	---	chr2:117155240-117158545

Fig. 6B

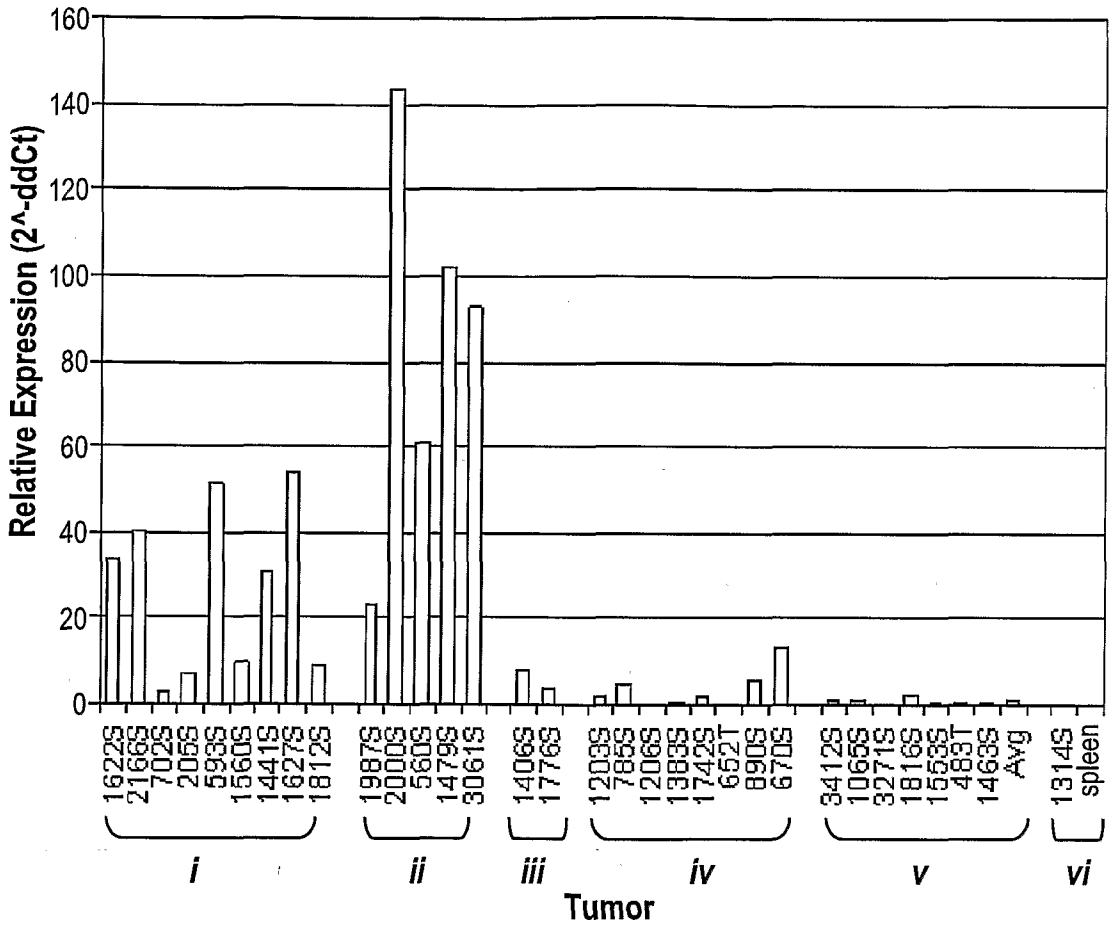
Relative Expression (2^-ddCt)
160
140
120
100
80
60
40
20
0
1622S
2166S
702S
205S
593S
1560S
1441S
1627S
1812S
1987S
2000S
560S
1479S
3061S
1406S
1776S
1203S
785S
1206S
1383S
1742S
652T
890S
670S
3412S
1065S
3271S
1816S
1553S
488T
1463S
Avg
1314S
spleen
i
ii
iii
iv
v
vi
Tumor

Fig. 6C

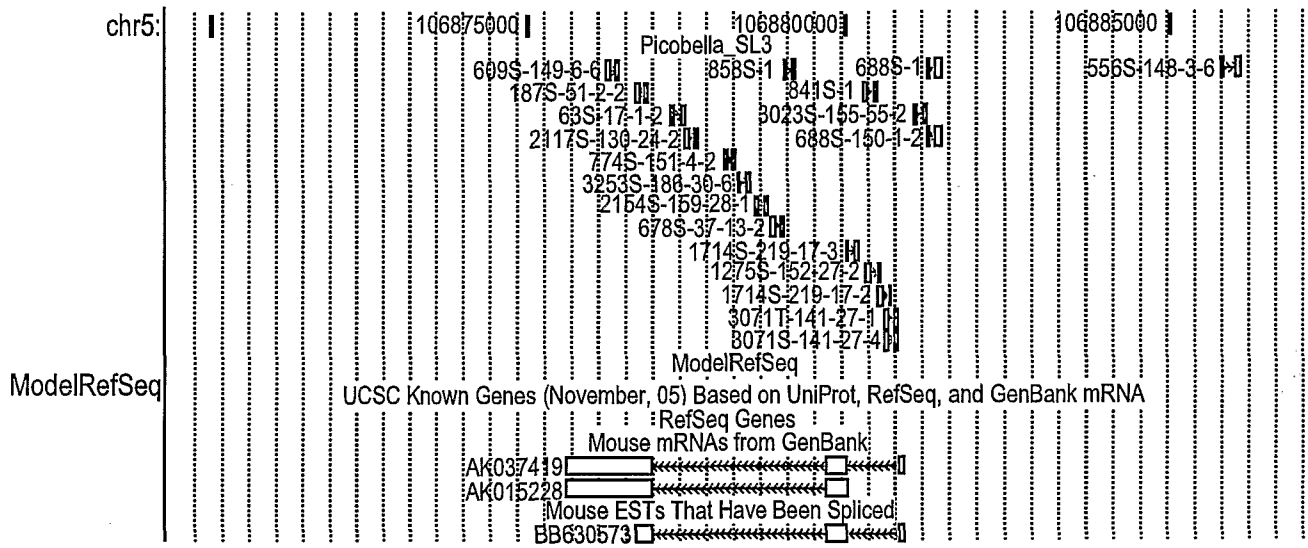
chr5:
106875000
106880000
106885000
Picobella_SL3
609S-149-6-6
858S-1
688S-1
556S-148-3-6
187S-51-2-2
841S-1
63S-17-1-2
3023S-155-55-2
2117S-130-24-2
688S-150-1-2
774S-151-4-2
3253S-186-30-6
2154S-159-28-1
678S-37-13-2
1714S-219-17-3
1275S-152-27-2
1714S-219-17-2
3071T-141-27-3
3071S-141-27-4
ModelRefSeq
ModelRefSeq
UCSC Known Genes (November, 05) Based on UniProt, RefSeq, and GenBank mRNA
RefSeq Genes
Mouse mRNAs from GenBank
AK037419
AK015228
Mouse ESTs That Have Been Spliced
BB630573

Fig. 7A

Tumor	Orientation	Location
1063S	---	chr5:106893798
759S	---	chr5:106894698
1967S	---	chr5:106894785
1206S	---	chr5:106895011
688S	---	chr5:106881416
3023S	---	chr5:106881190
3071S	+++	chr5:106880529
1714S	---	chr5:106880620
1275S	---	chr5:106880447
841S	---	chr5:106880410
858S	---	chr5:106879114
2154S	---	chr5:106878686
3253S	+++	chr5:106878208
2117S	+++	chr5:106877396
63S	---	chr5:106877369
609S	---	chr5:106876360

Transcript:	AK037419	---	chr5:106875550-106880818

Fig. 7B

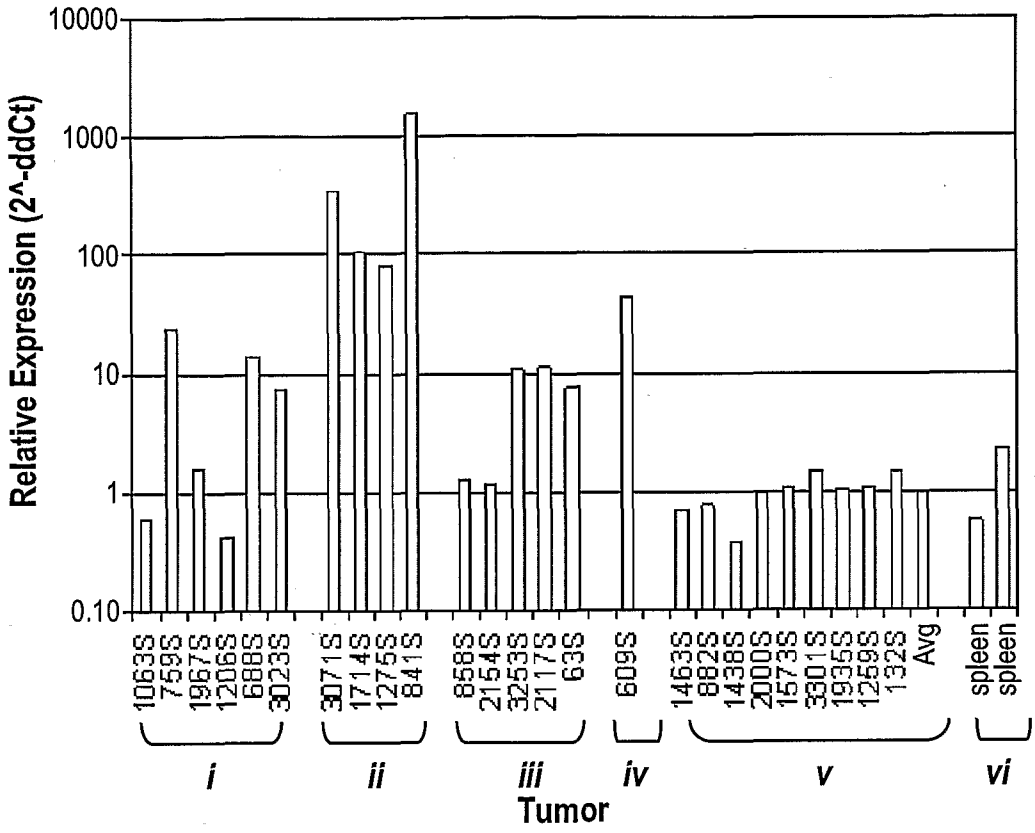
Relative Expression (2^-ddCt)
10000
1000
100
10
1
0.10
1063S
759S
1967S
1206S
688S
3023S
3071S
1714S
1275S
841S
858S
2154S
3253S
2117S
63S
609S
1463S
882S
1438S
2000S
1573S
3301S
1935S
1259S
132S
Avg
spleen
spleen
i
ii
iii
iv
v
vi
Tumor

Fig. 7C

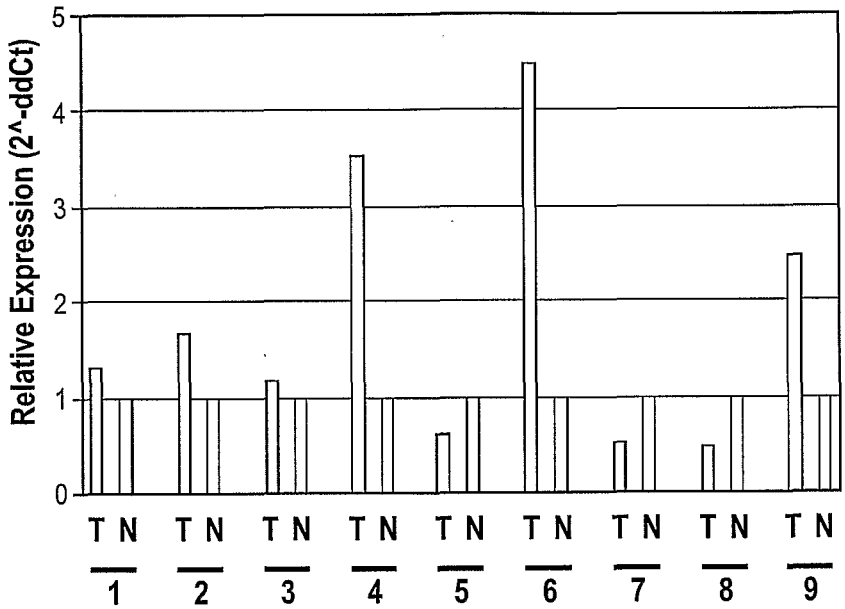
Relative Expression (2^-ddCt)
5
4
3
2
1
0
T N T N T N T N T N T N T N T N T N
1 2 3 4 5 6 7 8 9

Fig. 8