



(51) International Patent Classification:

A61K 48/00 (2006.01) C12N 15/86 (2006.01)

(21) International Application Number:

PCT/US2019/027672

(22) International Filing Date:

16 April 2019 (16.04.2019)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/658,358 16 April 2018 (16.04.2018) US

(71) Applicant: BAYLOR COLLEGE OF MEDICINE

[US/US]; One Baylor Plaza, Houston, TX 77030 (US).

(72) Inventors: YEN, Laising; 2507 Rose Bay Drive, Pearland,

TX 77584 (US). GUPTA, Sachin, Kumar; C/o Baylor Col-

lege of Medicine, One Baylor Plaza, Houston, TX 77030

(US). JEA, Jocelyn Duen-Ya; 8055 Cambridge Street, Apt.

67, Houston, Texas 77054 (US).

(74) Agent: SISTRUNK, Melissa, L. et al.; Norton Rose Ful-

bright US LLP, 1301 McKinney, Suite 5100, Houston, TX

77010-3095 (US).

(81) Designated States (unless otherwise indicated, for every

kind of national protection available): AE, AG, AL, AM,

AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ,

CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO,

DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,

HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP,

KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME,

MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ,

OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA,

SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, 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, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ,

UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ,

TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,

EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,

MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,

KM, ML, MR, NE, SN, TD, TG).

Published:

- without international search report and to be republished
upon receipt of that report (Rule 48.2(g))

(54) Title: CHIMERIC RNA-DRIVEN GENOMIC REARRANGEMENT IN MAMMALIAN CELLS

(57) Abstract: Embodiments of the disclosure include methods and compositions in which two inter- chromosomal or intra-chromo-
somal genomic DNA regions are rearranged upon exposure of the genomic DNA to a chimeric RNA that produces fusion of the two
regions. The chimeric RNA hybridizes to the two different parts of the genomic DNA and forms a DNA/RNA hybrid in a sequence-spe-
cific manner, thereby facilitating genomic rearrangement. In particular embodiments this mechanism is utilized to produce directed
gene fusion in targeted genomic DNA regions.



CHIMERIC RNA-DRIVEN GENOMIC REARRANGEMENT IN MAMMALIAN CELLS

[0001] This application claims priority to U.S. Provisional Patent Application Serial No. 62/658,358, filed April 16, 2018, which is incorporated by reference herein in its entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

[0002] This invention was made with government support under R01EB013584 awarded by National Institutes of Health. The government has certain rights in the invention.

TECHNICAL FIELD

[0003] Embodiments of the disclosure include at least the fields of cell biology, molecular biology, recombinant nucleic acid technology, gene editing, and medicine.

BACKGROUND

[0004] Fusion genes are among the most cancer-specific molecular signatures known. They are important for understanding cancer mechanisms and developing useful clinical biomarkers and anti-cancer therapies (Mitelman et al., 2007). Fusion gene formation as a result of chromosomal translocations is presumed to occur prior to fusion RNA expression. However, several studies have reported the presence of fusion transcripts in individuals without detectable fusion genes at the genomic DNA level. For instance, the *AML1-ETO* fusion transcript, associated with a subtype of acute myeloid leukemia, was present in some patients who were negative for chromosomal translocations (Langabeer et al., 1997). Other fusion RNAs, such as *BCR-ABL*, *MLL-AF4*, *TEL-AML1*, *PML-RAR α* , and *NPM-ALK*, were reported in healthy individuals (Janz et al., 2003). Although the discrepancy between the presence of fusion transcripts and the absence of fusion genes could result from detection limitations of the methodologies employed, fusion transcripts in normal cells could also arise from RNA trans-splicing in the absence of chromosomal translocations (Zaphiropoulos, 2011). Indeed, *JAZF1-JJAZ1* fusion transcripts are expressed in normal human endometrial tissue and an endometrial cell line in the absence of chromosomal translocation (Li et al., 2008). Furthermore, trans-splicing between *JAZF1* and *JJAZ1* was demonstrated to occur *in vitro* using cellular extracts, resulting in a fusion RNA similar to that transcribed from the *JAZF1-JJAZ1* fusion gene in endometrial stromal sarcomas (Li et al., 2008). These observations raise the possibility that cellular fusion RNAs created by trans-splicing act as guide RNAs to mediate genomic

rearrangements. A precedent for RNA-mediated genomic arrangements is found in lower organisms such as ciliates (Fang and Landweber, 2012; Nowacki et al., 2008). Rowley and Blumenthal (Rowley and Blumenthal, 2008) coined this as “the cart before the horse” hypothesis, in that “RNA before DNA” defies the normal order of the central dogma of biology: DNA → RNA → protein (Crick, 1970). Despite important implications in biology and human cancer, RNA-mediated genomic rearrangement in mammalian cells has not been directly demonstrated. In this report, we provide the first evidence that expression of a specific chimeric RNA can lead to specified gene fusion in mammalian cells.

BRIEF SUMMARY

[0005] The present invention is directed to methods and compositions that regard RNA-mediated gene fusion in mammalian cells. Methods of the disclosure utilize one or more composition(s) of exogenously provided chimeric RNA that guides genomic rearrangements that ultimately result in gene fusions. In methods encompassed herein, expression of a chimeric RNA drives formation of a specified gene fusion *via* genomic rearrangement in mammalian cells. In particular embodiments, the methods lack the involvement of immunogenic exogenous proteins. Included herein are chimeric RNA compositions for use in methods of the disclosure, and in specific embodiments the chimeric RNA is antisense. In particular embodiments, endogenous chimeric RNAs directed to gene fusion are both a clinical marker and therapeutic target for one or more medical conditions.

[0006] Embodiments of the disclosure include methods of producing genomic DNA rearrangements between two different regions of genomic DNA, comprising the step of: exposing to the genomic DNA an effective amount of an exogenously provided chimeric RNA, said chimeric RNA comprising fusion of (1) a first part that is reverse complement with respect to a first genomic region and (2) a second part that is reverse complement with respect to a second genomic region, wherein upon exposing of the chimeric RNA to the genomic DNA, the first and second genomic regions become fused through DNA recombination. In some cases, wherein upon exposing the chimeric RNA to the genomic DNA, the chimeric RNA hybridizes with the respective strands of the first and second genomic regions of the genomic DNA through the first and second parts of the chimeric RNA, thereby forming a DNA/RNA hybrid in a sequence-specific manner. In some cases, wherein upon hybridization of the chimeric RNA with the respective strands of the first and second genomic regions of the genomic DNA, a double

stranded DNA/DNA stem structure is produced between the first and second genomic regions adjacent to the site of hybridization with the chimeric RNA. In certain cases, wherein upon hybridizing of the chimeric RNA with the genomic DNA at the first and second genomic regions, a DNA spacer is produced in each of the respective strands of the first and second genomic regions between the DNA/DNA stem structure and the DNA/RNA hybrid structure. The two different regions of genomic DNA may be on the same chromosome or may be on different chromosomes. In some cases, at least one of the two different regions has active transcription. The chimeric RNA may be antisense with respect to the at least one region having active transcription. In certain embodiments, the two different regions lack active transcription. Whether or not a gene is being actively transcribed, for example in a cell, may be determined by northern to see if the RNA is being transcribed or it may be known based on the literature (for example, for certain genes and cell lines).

[0007] Regarding certain parameters for the chimeric RNA, in some cases the length of the first part of the chimeric RNA is in a range of 30-150, 30-125, 30-100, 30-75, 30-50, 50-150, 50-125, 50-100, 50-75, 75-150, 75-125, 75-100, 100-150, 100-125, 125-175, or 125-150 nucleotides in length. The length of the first part of the chimeric RNA may be in the range of 50-75 nucleotides. The length of the second part of the chimeric RNA may be in a range of 30-150, 30-125, 30-100, 30-75, 30-50, 50-150, 50-125, 50-100, 50-75, 75-150, 75-125, 75-100, 100-150, 100-125, 125-175, or 125-150 nucleotides in length. In some cases, the length of the second part of the chimeric RNA is in the range of 50-75 nucleotides. In particular embodiments, the length of the first part of the chimeric RNA and/or second part of the chimeric RNA is 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, or 75 nucleotides. The length of the chimeric RNA may be in the range of 60-300, 60-250, 60-200, 60-150, 60-100, 75-300, 75-250, 75-200, 75-150, 75-100, 100-300, 100-250, 100-200, 100-150, 150-300, 150-250, 150-200, 150-175, 175-300, 175-250, 175-200, 200-300, 200-275, or 200-250 nucleotides.

[0008] Regarding certain parameters for the DNA spacer that may or may not have a particular form, such as a bulge or loop, the DNA spacer may be single stranded or double stranded. In specific cases, the DNA spacer of the first genomic region is a length in a range of 15-50, 15-45, 15-30, 15-25, 15-20, 20-50, 20-45, 20-40, 20-35, 20-30, 20-25, 25-50, 25-45, 25-40, 25-35, 25-30, 30-50, 30-45, 30-40, 30-35, 40-50, 40-45, or 45-50 nucleotides. The DNA spacer of the first genomic region may be a length in a range of 25-45 nucleotides. The DNA

spacer of the first genomic region may be 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 nucleotides, for example. In specific cases, the DNA spacer of the first genomic region is a length of 35 nucleotides. The DNA spacer of the second genomic region may be a length in a range of 15-50, 15-45, 15-30, 15-25, 15-20, 20-50, 20-45, 20-40, 20-35, 20-30, 20-25, 25-50, 25-45, 25-40, 25-35, 25-30, 30-50, 30-45, 30-40, 30-35, 40-50, 40-45, or 45-50 nucleotides. In some cases, the DNA spacer of the second genomic region may be a length in a range of 25-45 nucleotides. In specific embodiments, the single stranded DNA region of the second genomic region is 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 nucleotides. The DNA spacer of the second genomic region may be a length of 35 nucleotides.

[0009] In particular embodiments, the genomic DNA is in a cell, such as an eukaryotic cell. The cell may be a mammalian cell, such as a human cell. The cell may be in a tissue or organ or body fluid. The cell may be in vivo. In some cases, the method occurs in vivo. The genomic DNA may be from a diseased cell or diseased cells. The genomic DNA may be from an individual in need of treatment. In specific embodiments, the method occurs at physiological hormone levels. In some cases, the chimeric RNA is produced in a cell. The cell may comprise the genomic DNA. In specific embodiments, the chimeric RNA is expressed from a vector, including a viral vector, such as a lentiviral vector, a retroviral vector, an adenoviral vector, or an adeno-associated viral vector, or or a non-viral vector, such as a plasmid.

[0010] In one embodiment, there is a method of inducing gene fusion between first and second target genes, comprising the steps of: designing or obtaining a chimeric RNA comprising (1) a first part that is antisense or sense with respect to the first target gene and (2) a second part that is antisense or sense with respect to a second target gene; and exposing an effective amount of the chimeric RNAs to nucleic acid comprising the first and second target genes, wherein upon exposing of the chimeric RNA to the nucleic acid, the first and second target genes become fused through DNA recombination.

[0011] In a particular embodiment, there is a method of identifying a cell, group of cells, or tissue at risk for becoming cancerous for an individual, comprising the step of identifying the presence of an endogenous chimeric RNA in the cell or one or more cells of the group of cells or tissue in a sample from the individual. The chimeric RNA may comprise sequence derived from

the parental genes of an oncogenic fusion gene or comprises sequence that is partially complimentary to that of the parental genes. In specific embodiments, the chimeric RNA is identified with reverse transcriptase-polymerase chain reaction, Northern gel analysis, and/or microarray chips designed for detecting endogenous chimeric RNAs.

[0012] An individual being treated may be provided an effective amount of a preventative (such as a vaccine) or therapeutic for the cancer. In specific embodiments, the agent that inhibits the chimeric RNA directly by binding, or inhibits the chimeric RNA indirectly by reducing its expression in cells. The agent that directly inhibits the chimeric RNA may be siRNA or antisense oligo DNA. In specific embodiments, the agent that indirectly inhibits the chimeric RNA is a compound that methylates the promotor that drives the endogenous chimeric RNA expression.

[0013] In some embodiments, there is a method of treating an individual for cancer, comprising the step of delivering an effective amount of an agent that inhibits a chimeric RNA in cancer cells of the individual. The agent may be a siRNA, for example one delivered to the individual in a vector, including a viral vector (lentiviral vector, a retroviral vector, an adenoviral vector, or an adeno-associated viral vector) or anti-viral vector.

[0014] The foregoing has outlined rather broadly the features and technical advantages of the present disclosure in order that the detailed description that follows may be better understood. Additional features and advantages will be described hereinafter which form the subject of the claims herein. It should be appreciated by those skilled in the art that the conception and specific embodiments disclosed may be readily utilized as a basis for modifying or designing other structures for carrying out the same purposes of the present designs. It should also be realized by those skilled in the art that such equivalent constructions do not depart from the spirit and scope as set forth in the appended claims. The novel features which are believed to be characteristic of the designs disclosed herein, both as to the organization and method of operation, together with further objects and advantages will be better understood from the following description when considered in connection with the accompanying figures. It is to be expressly understood, however, that each of the figures is provided for the purpose of illustration and description only and is not intended as a definition of the limits of the present disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

[0015] For a more complete understanding of the present disclosure, reference is now made to the following descriptions taken in conjunction with the accompanying drawing, in which:

[0016] **FIGS. 1A-1E. Exogenously expressed input chimeric RNAs induce the expression of endogenous fusion transcripts.** **FIG. 1A. Upper:** Schematics of the designed input RNAs containing complete *TPRSS2* exon-1 (78 nt, uc002yzj.3) and *ERG* exon-4 (218 nt, uc021wjd.1), expressed in the sense or antisense orientation from the CMV or U6 promoters. pA: poly-A signal. ts: transcriptional stop “TTTTTT” for U6 promoter. **Lower:** RT-PCR detection of induced fusion transcript (top gel) and input RNA (bottom gel). Parental plasmid vector containing mCherry sequence, DHT treatment without plasmid transfection, and PCR reaction without cDNA served as the controls (lane 5, 6, 8). **FIG. 1B.** Length and positional effect of antisense input RNAs. **Upper:** Antisense input RNAs with 75 nt (blue) targeting *ERG* exon-4 and varying lengths (82, 67, 52, and 33 nt, green) targeting *TPRSS2*. Dashed line links *ERG* and *TPRSS2* sequence within the input RNA and contains no sequence. A 20-nt mutation (black line) was introduced to the input RNAs to discern expressed input RNAs from the induced fusion transcript (see FIG. 8B). **Lower:** RT-PCR detection of induced fusion transcript (top gel), input RNA (middle gel), or detection of induced fusion transcript using a different primer pair (bottom gel). **FIG. 1C.** Corresponding sense input RNAs all failed to induce the fusion transcript. **FIG. 1D.** Induction by antisense-5 occurred at physiologically relevant DHT concentrations as low as 20 nM. Three-rounds of nested PCR were performed to reveal the lowest amount of DHT required. **FIG. 1E.** Antisense-5 led to clear induction while antisense-3 and 7 did not, indicating that it is not the length of input RNAs but targeted regions that is critical;

[0017] **FIGS. 2A-2G. Formation of a three-way junction may facilitate fusion induction.** **FIG. 2A. Left:** Schematics of three-way junction that could be formed between genomic DNA (black) and antisense-5 input RNA (green/blue). The sense genomic strands of both *TPRSS2* and *ERG* genes are on the minus strand of chromosome 21, separated by 3 Mb. Short lines in shaded regions represent base-pairings. Imperfect stem A includes a high energy G·T and A·C wobble pair known to have Watson-Crick-like geometry in a DNA double helix (Kimsey and Al-Hashimi, 2014; Watson and Crick, 1953). A spacer region of 36 nt and 47 nt

separate stem A from the regions targeted by antisense-5 input RNA. **Right:** Expressing antisense-5 as two separate halves that severed the link between *TMPRSS2* (52 nt) and *ERG* (75 nt) sequence within the input RNA. **FIG. 2B.** RT-PCR assays of fusion transcripts showed that the severed input RNAs resulted in the loss of induction. **FIG. 2C.** Locations of putative stems A to F identified by BLAST analyses. **FIG. 2D.** The putative three-way junction formed between the indicated genomic DNA stem B to G (Black) and designed antisense input RNA (green/blue). **FIG. 2E.** Targeting genomic DNA stem B, C, and D that exhibit higher DNA stem stability ($T_m = 40^\circ\text{C}$, 40°C , and 44°C , respectively) by the corresponding antisense input RNAs induced fusion transcripts (lanes 2 to 4). In contrast, targeting less stable stem E, F, and G ($T_m = 30^\circ\text{C}$, 24°C , and 16°C , respectively) failed to induce fusion transcripts (lanes 5 to 7). **FIG. 2F.** Antisense input RNAs designed to invade each of the respective genomic DNA stem B, C, and D resulted in the loss of induction. **FIG. 2G.** Corresponding sense input RNAs targeting stem B, C, and D failed to induce fusion transcripts (lanes 2 to 4).

[0018] FIGS. 3A-3F. Induced *TMPRSS2-ERG* fusion is the result of genomic rearrangements **FIG. 3A. Upper:** RT-PCR shows that LNCaP cells express *TMPRSS2* mRNA, and that the expression is upregulated by DHT. Primers used are specific to *TMPRSS2* exon-2 and exon-4. **Middle and lower:** *ERG* mRNA however was not detected in LNCaP cells under a wide range of DHT in the presence or absence of antisense-5 (lanes 1 to 11). Assays were performed using two independent primer pairs, one specific to exon-3 and -7 (middle panel), the other to exon-1 and -4 (lower panel), both capable of detecting *ERG* mRNA in VCaP cells (lane 12). These primer pairs were chosen because they selectively amplified endogenous *ERG* mRNA but not the induced *TMPRSS2-ERG* fusion transcript which has *ERG* exon-3 to exon-12. **FIG. 3B.** RT-PCR shows the transient nature of input RNA (lower panel) and the persistent nature of the induced fusion transcript (upper panel) in the enriched LNCaP population up to 52 days post initial treatment (see FIG. 18 for enrichment procedure). **FIG. 3C.** Schematics of identified genomic breakpoints and the primers used to amplify the breakpoints. The approximate location of genomic breakpoints were first determined using a primer set targeting *TMPRSS2* intron-1 multiplexed with a primer set targeting *ERG* intron-3 (see FIG. 19A). The region near the breakpoint was then further examined using primers A, B, C, and D. **FIG. 3D.** The un-rearranged wildtype *TMPRSS2* and *ERG* alleles were revealed by primer pair A/B (~1404 bp) and C/D (~1260 bp) respectively (lane 1, 2, 4, 5). The genomic fusion band of 862 bp amplified by fusion-specific primer pair A/D was present only in the enriched LNCaP population (lane 6)

and absent in untransfected LNCaP cells (lane 3). **FIG. 3E.** Sanger sequencing of the fusion band showed a 500bp segment of *TMPRSS2* intron-1 fused to 362 bp of *ERG* intron-3. The genomic breakpoint contains a 'CTG' microhomology (boxed). **FIG. 3F.** Prolonged expression of antisense-5 for 12 days induced the *TMPRSS2-ERG* fusion transcript in PNT1A cells as detected by three-round nested PCR. The results indicate that the induction of *TMPRSS2-ERG* fusion by antisense input RNA can occur in normal prostate epithelial cells prior to malignant transformation;

[0019] FIGS. 4A-4G. RNA-mediated inter-chromosomal gene fusion between *TMPRSS2* and *ETV1*. **FIG. 4A.** RT-PCR shows that only the antisense RNA designed to target stem *TMPRSS2-ETV1*-A, which has the highest stem stability ($T_m=72^\circ\text{C}$), led to induced fusion transcript (lane 1). Antisense RNAs targeting locations with lower genomic DNA stem stabilities (lanes 2 to 8), resulted in no detectable induction. **FIG. 4B.** The corresponding sense input RNA targeting the same *TMPRSS2-ETV1*-A stem failed to induce fusion transcript (lane 1 vs. 2). **FIG. 4C.** Gene fusion is specified by the sequence of input RNA used. The input RNA targeting *TMPRSS2* and *ETV1* induced *TMPRSS2-ETV1* fusion but not *TMPRSS2-ERG* fusion (lane 2). Conversely, antisense-5 targeting *TMPRSS2* and *ERG* induced *TMPRSS2-ERG* fusion but not *TMPRSS2-ETV1* fusion (lane 1). **FIG. 4D.** RT-PCR shows the transient nature of input RNA which was present at day 3 but not day 47 post initial treatment (lower panel, lane 1 vs. 2), and the persistent nature of the induced fusion transcript (upper panel) up to 47 days post initial treatment in the enriched LNCaP population. **FIG. 4E.** Schematics of three identified genomic breakpoints marked as x, y, and z. The approximate location of genomic breakpoints were first determined using a primer set targeting *TMPRSS2* intron-1 multiplexed with a primer set targeting *ETV1* intron-2. The un-rearranged wildtype *TMPRSS2* allele was then revealed by primer pair E/F (990 bp; lane 1 and 4) and G/H (956 bp; lane 7 and 10), and the un-rearranged wildtype *ETV1* allele by primer pair M/N (1024 bp; lane 2, 5, 8, 11). The genomic fusion band x (1150 bp) and y (1044 bp) amplified by fusion-specific primer pair E/N, and z (1043 bp) amplified by primer pair G/N, were present only in the induced and enriched LNCaP population but absent in untransfected LNCaP cells (lane 6 vs. 3, and lane 6 vs. 12). **FIG. 4F.** Sanger sequencing of the x, y, and z fusion band identified the exact genomic breakpoints. Region of microhomology at the breakpoints are boxed by solid lines, and indels by dash lines. **FIG. 4G.** FISH combined with 3D image reconstruction confirmed the gene fusion between *TMPRSS2* with *ETV1* in the enriched population. FISH probes against *TMPRSS2* gene (red) located on

chromosome 21 and against *ETV1* gene (green) located on chromosome 7 were used to reveal gene fusion based on the co-localized FISH signals in reconstructed 3D images. Examples of FISH signal in an untransfected cell (i) and a cell carrying induced *TMPRSS2-ETV1* gene fusion (ii) are shown. Arrow points to the co-localized FISH signals indicative of *TMPRSS2-ETV1* fusion, which is shown at a higher magnification in the inset. The table on the right shows the population analysis based on the 3D construction of FISH images. Approximately 0.9% of the enriched population (30 out of 3301 cells) was positive for *TMPRSS2-ETV1* fusion gene based on the co-localized FISH signals. In contrast, none of the cells from the untransfected population (0 out 620 cells) showed co-localized FISH signals.

[0020] FIGS. 5A-5B. The disparity between antisense versus sense input RNA is due to transcriptional conflict. FIG. 5A. RNA mediated gene fusion mechanism requires DNA:RNA hybrid formation. Induction of *TMPRSS2-ERG* fusion gene by antisense chimeric RNA was significantly reduced in the presence of wildtype RNaseH vs. the mutant RNaseH (**FIG. 5A**, lane 2 vs.1). Similarly, induction of *TMPRSS2-ETV1* fusion gene was also significantly reduced in the presence of wildtype vs. the mutant RNaseH (**FIG. 5A**, lane 6 vs. 5). These results indicate that the induction of gene fusions requires the formation of an RNA/DNA hybrid. Consistent with previous observations, sense input RNAs failed to induce fusion regardless of the expression of RNaseH (**FIG. 5A**, lane 3, 4, 7, 8). **FIG. 5B.** The input RNAs were expressed by U6 (a pol-III promoter) for one day, followed by α -amanitin-mediated inhibition of pol-II transcription for various time periods (0, 2, 6, 12 and 24 hrs) to shut down parental gene transcription. α -amanitin was then removed to resume cellular transcription and the induction by sense vs. antisense input RNA were compared. The corresponding sense input RNAs that previously failed to induce fusion, began to induce *TMPRSS2-ERG* (lane 9, 10) after 12 hours of α -amanitin treatment, and *TMPRSS2-ETV1* fusion (lane 20) after 24 hours of α -amanitin treatment, respectively. The results indicate that blocking the parental gene transcription by α -amanitin for a longer period (12-24 hrs for *TMPRSS2-ERG* and 24 hrs for *TMPRSS2-ETV1*) allows sense input RNA to induce gene fusion. Input RNAs used for the experiment: Antisense-5 vs. sense-5 for inducing *TMPRSS2-ERG* (upper panel, left two columns); antisense vs. sense *TMPRSS2-ETV1-A* for inducing *TMPRSS2-ETV1* (upper panel, right two columns). As controls, transfecting cells with a parental plasmid containing no input RNA sequences (lower panel, left column), cells without transfection (lower panel, center

column), and DHT and α -amanitin controls (lower panel, right column), also failed to induce fusion. In addition, GAPDH is used as internal control for the amount of RNA in each lane.

[0021] FIGS. 6A-6C. Specific endogenous RNA acts as the ‘initiator’ RNA to mediate gene fusion. FIG. 6A. Expression of full-length *AZII* mRNA for 3 days induced *TMPRSS2-ERG* fusion. The induction occurred at physiologically relevant concentrations of DHT as low as 40 nM (lane 4). Three-rounds of nested PCR were performed to reveal the lowest amount of DHT that permits *AZII*-mediated fusion induction. **FIG. 6B.** Expression of *AZII* exon 16-17 (lane 2), but not its antiparallel sequence (lane 3), led to *TMPRSS2-ERG* fusion. Inductions by full-length *AZII* (lane 1) and by antisense-5 (lane 4) were used as positive controls. **FIG. 6C.** Endogenous *AZII* RNA is expressed in LNCaP cells at low level and the expression is upregulated by DHT treatment.

[0022] FIGS. 7A-7C. A model of RNA-mediated gene fusion in mammalian cells. FIG. 7A. Our data support a model where the initiator RNA with chimeric sequence invades chromosomal DNA to stabilize a transient RNA/DNA duplex using DNA sequences located in two distant genes. Resolution of such an RNA/DNA duplex by DNA break/repair mechanisms might yield the final gene fusion through recombination in regions prone to DNA breaks. Such events however are infrequent (1 in 10^3 or 10^4 cells in the experimental cell population), and are dependent of the presence of hormone DHT in the case of LNCaP cells. **FIG. 7B** and **FIG. 7C.** The data also indicate that it is the antisense rather than sense chimeric RNAs that effectively drive gene fusion, and that this disparity can be explained by transcriptional conflict produced by the transcriptional activity of parental genes. The sense chimeric RNAs forming DNA/RNA hybrids with antisense strands of genomic DNA (the template strand used for transcription) are likely be frequently “bumped” off by RNA polymerase and unable to stabilize the structures required for initiating genomic rearrangements. The proposed RNA-driven model may provide a mechanism that can ‘specify’ gene fusion partners in early disease stages, and could have fundamental implications in the biology of mammalian genome stability, as well as gene editing technology via mechanisms native to mammalian cells.

[0023] FIGS. 8A-8B. Primer design used to distinguish induced endogenous fusion RNAs from exogenous input RNAs. FIG. 8A. The endogenous full-length *TMPRSS2-ERG* fusion RNA most commonly found in prostate cancer consisting of *TMPRSS2* exon 1 (78 nt, uc002yzj.3) spliced to *ERG* exons 4-12 (1289 nt, uc021wjd.1). In FIG. 1A, the exogenous input

RNAs expressed from plasmids consists of *TMPRSS2* exon-1 (78 nt) fused to *ERG* exon-4 (218 nt). The RT-PCR results shown in FIGS. 1A and 1B (middle panel) were performed using a forward primer on *TMPRSS2* exon-1 paired with a reverse primer on *ERG* exon-7, thereby specifically amplifying the endogenous fusion RNA transcribed from the *TMPRSS2-ERG* fusion gene, but not from transfected plasmids or input RNAs because both lack *ERG* exon-7. **FIG. 8B.** Because RNA-induced fusion is a low frequency event, a more efficient PCR method was developed to detect induced endogenous fusion RNA using a primer pair targeting *TMPRSS2* exon-1 and *ERG* exon-4. To distinguish induced endogenous fusion RNAs from exogenous input RNAs, mutations were introduced at nucleotide positions 16-35 of the input RNAs (gray box in *TMPRSS2* exon-1), allowing specific amplifications of induced endogenous fusion RNAs or exogenous input RNA using specific forward primers. The lower panel shows the wild type *TMPRSS2* exon-1 sequence in green. The mutated area in input RNA is shown in black and underlined. The primer used to specifically recognize induced fusion RNA is shown as the green arrow and was used in most of the figures. The primer for input RNA is denoted by the black arrow and was used in FIGS. 1B, 1C, 1D, 1E, FIG. 2B, and FIG. 3B.

[0024] FIGS. 9A-9B. Sense input RNAs that mimic trans-spliced products failed to induce fusion transcripts. FIG. 9A. Schematics of the designed input RNAs consisting of *TMPRSS2* (uc002yzj.3) and *ERG* (uc021wjd.1) exons. Upper panels: antisense-2 and sense-2 input RNA consisting of *TMPRSS2* exon-1 (78 nt) and *ERG* exon-4 (216 nt). Lower panel: sense-2-long RNA consisting of *TMPRSS2* exon-1 (78 nt) with *ERG* exons-4, -5, -6 and partial -7 (618 nt). ts: transcriptional stop “TTTTTT” for the U6 promoter. **FIG. 9B.** RT-PCR detection of induced fusion transcript (top gel) and input RNA (bottom gel). For induced fusion transcript, RT-PCR was performed using a forward primer on *TMPRSS2* exon-1 paired with a reverse primer on the 3' end of *ERG* exon-7 (see FIG. 8A), thereby specifically amplifying the endogenous fusion RNA transcribed from the *TMPRSS2-ERG* fusion gene, but not from transfected plasmids or input RNAs because both lack the 3' end of *ERG* exon-7. Control vector lacking input RNA sequence (lane 4), DHT treatment only (lane 5), and PCR reactions lacking cDNA (lane 7) were used as negative controls. Antisense RNA is capable of inducing the fusion transcript (lane 3), whereas sense RNAs failed to induce fusion transcripts regardless of their lengths (lanes 1 and 2).

[0025] FIG. 10. Antisense-5 but not its corresponding sense-5 induces *TMPRSS2-ERG* gene-fusion. Different amount of plasmids (1.000µg, 0.500µg, 0.250µg and 0.125µg)

expressing either sense-5 or antisense-5 were transfected to produce different amount of corresponding input RNAs in LNCaP cells. To maintain transfection efficiency, mCherry control plasmid was added to each transfection to make the final amount of plasmid to 1.0 μ g. Induction was done for three days in the presence of 0.9 μ M DHT. RT-PCR was performed to detect the levels of both induced *TMPRSS2-ERG* transcript and input RNA. RT-PCR results shows that even very low amount of antisense-5 input RNA is sufficient to induce *TMPRSS2-ERG* gene fusion (middle panel, lane 4) whereas very high level of sense-5 input RNA fails to induce *TMPRSS2-ERG* gene fusion (upper panel, lane 1). Thus, it is not the amount of the input RNA but the orientation of the input RNA (antisense vs. sense) that is important for fusion induction. Antisense-3, which is unable to induce fusion, was used as a negative control. mCherry alone was also transfected as a negative control. Additional controls (right panel): a vector expressing mCherry lacking input RNA sequence (lane 1), DHT treatment only (lane 2), and PCR reaction lacking cDNA (lane 4).

[0026] FIGS. 11A-11C. Expressed input RNA, but not the DNA plasmid, induces fusion transcript. FIG. 11A. Schematic of the expression plasmid containing a U6 promoter and antisense-5 sequence. Unique restriction sites are shown in red. The arrow indicates the transcriptional start. The plasmid was digested by SacII/KpnI, leaving U6 with antisense-5, or with SacII/PstI, separating U6 from antisense-5. **FIG. 11B.** Confirmation of complete digestion was visualized by agarose gel. Uncut plasmid vector was used as control. **FIG. 11C.** Induction of *TMPRSS2-ERG* fusion after transfection of digested plasmid fragments. The *TMPRSS2-ERG* fusion transcript was induced only when the U6 promoter was attached to antisense-5 sequence, indicating that it is the input RNA expressed by the U6 promoter, not the transfected DNA plasmid *per se*, that leads to fusion induction.

[0027] FIG. 12. DHT dosage effect in facilitating RNA-mediated gene fusion as assayed by one round RT-PCR. To determine the DHT concentrations that facilitate RNA-mediated gene fusion detectable by one round RT-PCR, we generated a dose-response curve using LNCaP cells transfected with antisense-5 and treated with different DHT concentrations for 3 days. Relative RT-PCR intensities of induced fusion transcript bands as generated by one round RT-PCR were plotted against DHT concentrations. In the presence of antisense-5, fusion induction is detectable at ~0.6 μ M DHT, reaches its maximum at 1.5 μ M DHT, with the effective concentration that yields 50% of maximal induction (EC_{50}) at ~0.9 μ M. In the absence of antisense-5, no induction was observed in the entire DHT range tested up to 2.0 μ M (upper

panel), indicating that DHT alone is ineffective in inducing *TMPRSS2-ERG* fusion. To streamline the experimental procedures, 0.9 μM DHT (the EC_{50} concentration) were used for 3 days followed by one round of RT-PCR. However, it is important to note that all key experiments were also performed under physiologically relevant DHT concentrations (<100 nM), and in those cases three-rounds of nested PCR were used to reveal the lowest amount of DHT required. As shown in FIG. 1D and FIG. 6A, gene fusion events induced by antisense-5 occurred at physiologically relevant DHT concentrations as low as 20 nM, and gene fusion events induced by endogenous *AZII* mRNA occurred at as low as 40 nM DHT.

[0028] FIG. 13. Estimation of the percentage of LNCaP cells transfected with antisense-5 that carry the induced *TMPRSS2-ERG* fusion. To estimate the percentage of LNCaP cells that underwent *TMPRSS2-ERG* gene fusion mediated by antisense-5 at 0.9 μM DHT (the EC_{50}), we mixed various amounts of total RNA from VCaP cells that express *TMPRSS2-ERG* (Mertz et al., 2007; Teles Alves et al., 2013) with the total RNA from untransfected LNCaP cells in ratios of 1:1 up to 1:10⁷. A standard dilution curve was determined using the relative RT-PCR intensity of the *TMPRSS2-ERG* fusion transcript. The relative intensity of induction by antisense-5 in LNCaP cells, as assayed under the same RT-PCR conditions, is shown as a red dot that corresponds to an equivalent dilution of 1:10³ to 1:10⁴, suggesting that approximately 1 in 10³ or 10⁴ LNCaP cells is positive of *TMPRSS2-ERG* fusion.

[0029] FIG. 14. Complementary base pairing of the genomic DNA sequences comprising the stems and their flanking sequences. The sense strand of genomic sequences from *TMPRSS2* and *ERG* that form the putative stems (in red boxes) with various degrees of stability as identified by BLAST analyses. Regions showing perfect complementary base pairing (A-T and G-C) are shaded in gray.

[0030] FIG. 15. Antisense input RNAs designed to disrupt the genomic stems B, C, and D. Stem B (upper row). Stem C (middle row). Stem D (lower row). Left column: Illustrations show the three-way junction formed by genomic DNA and corresponding antisense input RNA (antisense-B1, C1, and D1). Middle column: Disruption of stems via the *TMPRSS2* side using tailor-made input RNAs (antisense-B2, C2, and D2) that directly hybridize to the *TMPRSS2* stem sequence. Right column: Disruption of stems via the *ERG* side using tailor-made input RNAs (antisense-B3, C3, and D3) that directly hybridize to the *ERG* stem sequence.

Shaded regions indicate base pairing. The *TMPRSS2* portion of the input RNA is shown in green and the *ERG* portion is in blue.

[0031] FIGS. 16A-16B. Antisense input RNAs designed to disrupt the genomic stems

A. FIG. 16A. Disruption of the putative genomic stem using tailor-made input RNAs that directly hybridize to one side of the stem **A. FIG. 16B.** RT-PCR assays of fusion transcripts induced by the input RNAs illustrated in A. Interfering with three-way junction formation eliminated (lane 2) or significantly reduced the induction (lane 3). Controls include: vector lacking input RNA sequence (lane 4), DHT treatment only (lane 5), and PCR reaction lacking cDNA (lane 7).

[0032] FIGS. 17A-17D. Endogenous *ERG* mRNA is not detected in LNCaP cells by three-rounds of nested RT-PCR. FIG. 17A. Nested primer pairs used to detect low levels of endogenous *ERG* mRNA from LNCaP cells. Forward primers for *ERG* exon-3 are shown in green whereas reverse primers for *ERG* exon-7 are shown in blue (primer sequences are listed in the Materials and Methods). **FIG. 17B.** Endogenous *ERG* mRNA was not detected in LNCaP cells in the presence or absence of DHT or antisense-5 (lane 1 to 11) by three-rounds of nested RT-PCR. cDNA prepared from VCaP cells which express *ERG* mRNA was used as positive control (lane 13). **FIGS. 17C and 17D.** Similarly, using forward primers on *ERG* exon-1 and reverse primers on *ERG* exon-4 also failed to detect endogenous *ERG* mRNA by three-rounds of nested RT-PCR (lane 1 to 11, primer sequences are listed in Example 3). These primer sets were chosen because they specifically amplify endogenous *ERG* mRNA but not *TMPRSS2-ERG* fusion transcript that contains *ERG* exon-3 to exon-12.

[0033] FIG. 18. Procedure used to propagate and enrich the induced LNCaP population for 52 days in the absence of DHT. LNCaP cells were transfected on day 1 with antisense-5 and incubated for 3 days with DHT. Cells were then divided into sub-populations. Half of the cells from each sub-population were harvested for RT-PCR assays to detect the induced fusion transcript whereas the other half was used to propagate the population. The sub-population containing the highest intensity of induced fusion transcripts (circled in red) was selected for continuous propagation for the next round of division and RT-PCR assays. The process was continued for 52 days. RT-PCR was performed to detect both induced fusion transcripts and antisense input RNA during each division as shown in FIG. 3B.

[0034] FIGS. 19A-19B. Locations targeted by primers for mapping the genomic breakpoint and locations of identified *TMPRSS2-ERG* genomic breakpoints. FIG. 19A. The locations of 44 forward primers (green) spacing across *TMPRSS2* intron-1 (~10 kb) and 112 reverse primers (blue) spacing across *ERG* intron-3 (~130 kb). *ERG* intron-3 is significantly larger than that of *TMPRSS2* intron-1. Each vertical line represents a target location by a primer or a set of nested primers. The primers targeting *ERG* intron-3 are designed to concentrate in the region near exon-4 and in the hot spot regions previously identified from prostate cancer patients. The specific primers that amplify the genomic breakpoint shown in FIG. 3D are labeled as red. **FIG. 19B.** Locations of identified *TMPRSS2-ERG* genomic breakpoints. Locations of previously identified *TMPRSS2-ERG* genomic breakpoints from prostate cancer patients (Weier et al.) are shown in green in *TMPRSS2* intron-1 and blue in *ERG* intron-3. The number on top of each location is the ID number of a patient. The genomic breakpoint identified in our study is shown in red. Short arrows indicate sites targeted by antisense-5.

[0035] FIG. 20. Schematics of the putative three-way junctions formed between *TMPRSS2* and *ETVI* genomic DNA and the corresponding antisense input RNA. BLAST alignment was used to identify eight stems that could be formed with varying degrees of stability by the sense genomic *TMPRSS2* sequence paired with the sense genomic *ETVI* sequence. Matching antisense RNAs were then designed to forge the three-way junction with these putative intron stems. Shaded regions indicate base pairing between antisense input RNA (in green and purple) and genomic DNA (in black). *TMPRSS2* is on chromosome 21 whereas *ETVI* is on chromosome 7. For both, the genomic minus strand is the sense strand.

[0036] FIG. 21. Genome coordinates and T_m for the putative *TMPRSS2-ETVI* genomic stems described in FIG. 20. Upper panel: Locations of putative *TMPRSS2-ETVI* stems A, B, C, D, E, F, G, and H identified by BLAST alignment. All stem sequences are located in *TMPRSS2* intron-1 (chromosome 21) paired with sequences in *ETVI* intron-2 or -3 (chromosome 7). Genome coordinates are based on GRCh38/hg38 version. Coordinates of the exons (*TMPRSS2* exons-1 and -2; *ETVI* exons-2, -3, and -4) are also shown.

[0037] FIG. 22. Locations targeted by primers for mapping the *TMPRSS2-ETVI* genomic breakpoints. The locations of 23 forward primers (green) spacing across *TMPRSS2* intron-1 (~10 kb) and 7 reverse primers (purple) spacing across *ETVI* intron-2 (~0.8 kb). Each

vertical line represents a target location by a primer or a set of nested primers. The specific primers that amplify the genomic breakpoint shown in FIG. 4E are labeled as red.

[0038] FIG. 23. Illustration of 3-way junction stem formation for TMPRSS2-ERG gene fusion having multiple different DNA/DNA genomic stems (Stem A-Stem D). Certain parameters for design of chimeric RNAs are characterized. Each illustration in this figure demonstrates 3-way junction stem formation between the chimeric RNA and two separate genomic regions of TMPRSS2 and ERG concomitantly. In each example, the 3-way junction produces regions of genomic DNA loop or spacer that in at least some cases may be described as a bulge at or near the base of the junction. Upper panel, Left: Schematics of three-way junction that could be formed between genomic DNA (black) and antisense-5 chimeric RNA (green/blue). The sense genomic strands of both *TMPRSS2* and *ERG* genes are on the minus strand of chromosome 21, separated by 3 Mb. Short lines in shaded regions represent base-pairings. Imperfect stem A includes a high energy G·T and A·C wobble pair known to have Watson-Crick-like geometry in a DNA double helix (Kimsey and Al-Hashimi, 2014; Watson and Crick, 1953). A DNA spacer region of 36 nt and 47 nt separate stem A from the regions targeted by antisense-5 chimeric RNA. Upper panel, Right: The putative three-way junction formed between the indicated genomic DNA stem B (black) and designed antisense chimeric RNA B1 (green/blue). Lower panel, Left: The putative three-way junction formed between the indicated genomic DNA stem C (black) and designed antisense chimeric RNA C1 (green/blue). Lower panel, Right: The putative three-way junction formed between the indicated genomic DNA stem D (black) and designed antisense chimeric RNA D1 (green/blue).

[0039] FIG. 24. Illustration of a generic 3-way junction stem formation for TMPRSS2-ERG gene fusion. The illustration shows the different regions of the 3-way junction stem formation that may be variable, including length of the bulge and lengths of the respective regions of the chimeric RNA, and the genomic DNA/DNA stem. The important variables important for gene fusion induction are highlighted in red.

[0040] FIG. 25 shows testing of different size of DNA spacers (bulge) for the Stem B 3-way junction stem formation for TMPRSS2-ERG gene fusion. In each test, the DNA bulge was set to the same size on both the left and the right side. For example, B-31 has bulge size of 31 nt on the TMPRSS2 side as well as the ERG side. RT-PCR was performed to detect the level of induced TMPRSS2-ERG fusion RNA. The optimum size of the bulge for fusion induction

was examined in FIGS. 25, 26, and 27 for the corresponding Stem B, C, and D of FIG. 23. As is shown, for fusion gene induction the optimum size of the bulge is around 35 nucleotides on each side, although other lengths also work.

[0041] FIG. 26 shows testing of different size of DNA spacers (bulge) for the Stem C 3-way junction stem formation for TMPRSS2-ERG gene fusion.

[0042] FIG. 27 shows testing of different size of DNA spacers (bulge) for the Stem D 3-way junction stem formation for TMPRSS2-ERG gene fusion.

[0043] FIG. 28 illustrates a schematic for testing particular lengths sizes for the different chimeric RNA designed to hybridize to TMPRSS2 and ERG. The length of chimeric RNA on each side was varied while the bulge is fixed at 35 nucleotides. The length variations were then tested for their ability to induce fusion gene. Length variations: 30nt - 30nt, 50nt-50nt, 75nt-75nt, 100nt-100nt, 150nt -150nt). The part of chimeric RNA targeting TMPRSS2 genomic locations are in blue whereas targeting ERG genomic locations are in red.

[0044] FIG. 29 shows testing of the different sizes for the different chimeric RNA parts corresponding to TMPRSS2 and ERG using the Stem A 3-way junction. For example, A(50) has 50 nt on each side of the chimeric RNA. RT-PCR was performed to detect the level of induced TMPRSS2-ERG fusion RNA. FIGS. 29-32 provide results of testing the various sizes of the chimeric RNA when using Stem A, Stem B, Stem C, or Stem D, respectively. For fusion gene induction, an optimum size for the chimeric RNAs is around 50-75 nucleotides on each side.

[0045] FIG. 30 shows testing of the different sizes for the different chimeric RNA parts corresponding to TMPRSS2 and ERG using the Stem B 3-way junction.

[0046] FIG. 31 shows testing of the different sizes for the different chimeric RNA parts corresponding to TMPRSS2 and ERG using the Stem C 3-way junction.

[0047] FIG. 32 shows testing of the different sizes for the different chimeric RNA parts corresponding to TMPRSS2 and ERG using the Stem D 3-way junction.

[0048] FIG. 33 illustrates a man-made designer chimeric RNA having a perfect match to the genes to be fused.

[0049] FIG. 34 illustrates an endogenous initiator chimeric RNA having a perfect match to the genes to be fused.

[0050] FIG. 35 shows an endogenous initiator chimeric RNA having a match to the genes to be fused that is not exact and may have a loop or secondary structure in the middle, such as at the point of fusion.

[0051] FIG. 36 demonstrates an endogenous initiator chimeric RNA wherein the endogenous RNA can have a non-perfect match adjacent to the point of fusion.

[0052] FIG. 37 shows an endogenous initiator chimeric RNA having mismatches with the genes to be fused.

[0053] FIG. 38 demonstrates an endogenous initiator chimeric RNA that is formed by trans-splicing of two different parental endogenous RNAs.

[0054] FIGS. 39A-39D. RNA-mediated inter-chromosomal JAZF1-SUZ12 fusion in 293T cells. **FIG. 39A.** The design of chimeric RNA is based on the three-way junction that could be formed between genomic DNA (in the stem) and chimeric RNA (perpendicular and below the stem). The three-way junction comprises an RNA/DNA hybrid and a genomic DNA stem between the parental genes (JAZF1 and SUZ12). The sense genomic strand of JAZF1 gene is on the negative strand of chromosome 7 while that of SUZ12 is on the positive strand of chromosome 17. Short vertical lines in gray regions between the chromosome and the chimeric RNA represent base pairings. **FIG. 39B.** An example of chimeric RNA sequences used to induce JAZF1-SUZ12 fusion. This chimeric RNA targets intron regions and has GC content of 27%. The fusion RNA transcribed from the induced fusion gene contains an annotated splice junction. In this particular case, two different fusion RNAs were confirmed based on the results derived from the Sanger sequencing data. **FIG. 39C.** Example of inter-chromosomal JAZF1-SUZ12 fusion transcript (arrow) induced by sequence-specific chimeric RNAs in 293T cells without additional hormone or other stimuli. 293T cells were transfected with chimeric RNA expression plasmid. Total RNA was then harvested at day 3. Nested RT-PCR using specific primer pairs were used to detect the induced fusion transcript. Bands were then excised and cloned in pGEM-T vector for Sanger sequencing which confirmed the identity of the chimeric RNA as JAZF1 exon 1 spliced to a cryptic splice site of SUZ12 exon 1. **FIG. 39D.** Another example of inter-chromosomal JAZF1-SUZ12 fusion transcript (arrow) induced by sequence-specific chimeric RNAs in 293T

cells without additional hormone or other stimuli. 293T cells were transfected with chimeric RNA expression plasmid. Total RNA was then harvested at day 3. Nested RT-PCR using specific primer pairs were used to detect the induced fusion transcript. Band marked by asterisk was then excised and cloned in pGEM-T vector for Sanger sequencing. The fusion transcript was from splicing JAZF1 exon 2 to SUZ12 exon 2.

[0055] FIGS. 40A-40C. RNA-mediated intra-chromosomal BCR-RANGAP1 fusion in 293T cells by designed chimeric RNA. FIG. 40A. The design of chimeric RNA is based on the three-way junction that could be formed between genomic DNA (in the stem) and chimeric RNA (perpendicular and below the stem). The three-way junction consists of an RNA/DNA hybrid and a genomic DNA stem between the parental genes (BCR and RANGAP1). The sense genomic strand of BCR gene is on the positive strand of chromosome 22 while that of RANGAP1 is on the negative strand of chromosome 22, separated by 18 Mb. Short lines in gray regions represent base pairings. **FIG. 40B.** An example of chimeric RNA sequence used to induce BCR-RANGAP1 fusion. This chimeric RNA target intron regions and has GC content of 48.4 %. The fusion RNA transcribed from the induced fusion gene contains an annotated splice junction. **FIG. 40C.** Intra-chromosomal BCR-RANGAP1 fusion transcript (arrow) induced by sequence-specific chimeric RNA in 293T cells without additional hormone or other stimuli. 293T cells were transfected with chimeric RNA expression plasmid. Total RNA was then harvested at day 3. Nested RT-PCR using specific primer pairs were used to detect the induced fusion transcript. The production of BCR-RANGAP1 fusion transcript was confirmed by Sanger sequencing to comprise BCR exon14 spliced to RANGAP1 exon3 (RANGAP1 isoform 1, NM_001278651.1). The RNA fusion junction is an annotated splice junction for both genes.

[0056] FIGS. 41A-41D. RNA-mediated inter-chromosomal BCR-KCNQ2 fusion in HL-60 cells by designed chimeric RNA. FIG. 41A. The design of chimeric RNA is based on the three-way junction that could be formed between genomic DNA (in the stem) and chimeric RNA (perpendicular and below the stem). The three-way junction consists of an RNA/DNA hybrid and a genomic DNA stem (black) between the parental genes (BCR and KCNQ2). The sense genomic strand of BCR gene is on the positive strand of chromosome 22 while that of KCNQ2 is on the negative strand of chromosome 20. Short lines in gray regions represent base pairings. **FIG. 41B.** An example of chimeric RNA sequence used to induce BCR-KCNQ2 fusion. This chimeric RNA target intron regions and has GC content of 55.4 %. The fusion RNA transcribed from the induced fusion gene contains an annotated splice junction. In this case, two

different fusion RNAs were confirmed based on the results derived from the Sanger sequencing data. **FIG. 41C.** Two BCR-KCNQ2 fusion transcript variants were induced in HL-60 cells by chimeric RNA targeting BCR and KCNQ2 genes in the genome without hormone or other stimuli. **FIG. 41D.** Schematic of induced BCR-KCNQ2 fusion transcripts derived from the Sanger sequencing data. HL-60 cells were transfected with chimeric RNA expression plasmid by nucleofection (Lonza) at day 1. Total RNA was then harvested at day 7. Nested RT-PCR using specific primer pairs were used to detect the induced fusion transcript.

[0057] FIGS. 42A-42E. BCR-KCNQ2 fusion induced in 293T cells by chimeric RNAs utilizing/excluding the SINE element for three-way junction formation. FIG. 42A. Examples of chimeric RNA design using Short Interspersed Nuclear Element (SINE) commonly present in the human genome. The designed chimeric RNA forge a three-way junction that could be formed between two SINEs in genomic DNA (in the stem) and chimeric RNA (perpendicular and below the stem). Genomic DNA stems formed by DNA repeats such as SINE/SINE repeats may constitute stems over 300 nts in length. Short lines in gray regions represent base pairings. SINEs are in crimson. **FIG. 42B.** An example of chimeric RNA sequence utilizing the SINE element for three-way junction formation to induce BCR-KCNQ2 fusion. This chimeric RNA has GC content of 52 %. **FIG. 42C.** Examples of chimeric RNA design targeting DNA sequences of BCR and KCNQ2 genes in the genome which form genomic stems using regions outside of known DNA repeats such as Short Interspersed Nuclear Element (SINE). SINEs are commonly present in the human genome are shown in solid boxes. The DNA stem could range from 10bp to >100bp with some mismatches. **FIG. 42D.** An example of chimeric RNA sequence used to induce BCR-KCNQ2 fusion in 293T cells. This chimeric RNA target intron regions and has GC content of 60.4 %. **FIG. 42E.** BCR-KCNQ2 fusion can be induced by chimeric RNAs that form 3-way junction with genomic stems using SINE repeats (Set1A) or outside of SINE repeats (Set1B). 293T cells were transfected with chimeric RNA expression plasmid for 3 days and then harvested for nested RT-PCR.

DETAILED DESCRIPTION

[0058] As used herein the specification, "a" or "an" may mean one or more. As used herein in the claim(s), when used in conjunction with the word "comprising", the words "a" or "an" may mean one or more than one. As used herein "another" may mean at least a second or more. In specific embodiments, aspects of the invention may "consist essentially of" or "consist

of' one or more sequences of the invention, for example. Some embodiments of the invention may consist of or consist essentially of one or more elements, method steps, and/or methods of the invention. It is contemplated that any method or composition described herein can be implemented with respect to any other method or composition described herein. The scope of the present application is not intended to be limited to the particular embodiments of the process, machine, manufacture, composition of matter, means, methods and steps described in the specification.

I. [0059] RNA-driven Genomic Rearrangement Methods

[0060] The present disclosure provides an example of a process that (1) utilizes three-way junction formation between genomic DNAs base-pairing with chimeric RNA, (2) is generally permissible regardless of intra-chromosomal or inter-chromosomal fusion, (3) in some cases may be facilitated by physiological hormone levels (androgen in the disclosed examples, although other hormone(s) or no hormones may be utilized, and (4) can occur in normal cells prior to malignant transformation. With respect to gene fusion of actively transcribed genes, it may be the antisense rather than sense chimeric RNAs that drives gene fusion, in specific embodiments, although the gene to be fused may dictate whether or not the chimeric RNA comprises sense sequence for both regions of the chimeric RNA, antisense sequence for both regions of the chimeric RNA or one of the regions of the chimeric RNA is sense while the other region of the chimeric RNA is antisense. While in some embodiments a three-way junction facilitates gene fusion induction, in specific cases RNA/DNA hybrid alone without a three-way junction induces gene fusion, for example under certain conditions.

[0061] Methods of the disclosure may or may not be practiced with exogenously provided hormone. In cases wherein hormone may be enhancing to the method or is known to be enhancing to the method, an effective amount of the hormone may be utilized. Whether or not hormone is utilized may depend on the gene(s) to be fused. For example, in cases wherein the cells in which the chimeric RNA is utilized are responsive to hormone, or the expression of one or both of the genes is sensitive to hormone, then an effective amount of hormone may be utilized in the method. As one example, if the cells in which the chimeric RNA is delivered are prostate cancer cells and at least one of the genes being fused is responsive to androgen or testosterone, then androgen, testosterone, or one of their derivatives may be employed in the method. In a situation where the cells are breast cancer cells, for example, and at least one of the genes being fused is responsive to estrogen, then estrogen may be employed in the method.

[0062] The formation of fusion genes (cancer fusion genes, merely as examples) demonstrated herein also represents a new and safer way of achieving gene editing in human cells, for example without involving immunogenic exogenous proteins such as Zinc finger or CRISPR/Cas9. Thus the RNA-driven genomic rearrangements encompassed by the disclosure are a useful technology for gene editing in mammalian cells in general, and overcomes certain difficulties in gene/cell therapy in particular.

[0063] Thus, provided herein are methods for rearranging regions of a genome. The method can produce genomic DNA rearrangements regardless of whether those genomic DNA segments are actively transcribed genes or untranscribed regions (or one of both). In the case of actively transcribed DNA segments (active genes), the chimeric RNA may comprise sequence that is antisense with respect to the gene sequences. In the case of untranscribed DNA segments (silent genes or intergenic space, for example), the chimeric RNA may comprise sequence that is either sense or antisense with respect to the genomic DNA segments. Thus, the method can target any segment of a genome.

[0064] The present disclosure concerns methods of generating fusion of two different regions of nucleic acid, including in particular two different regions of genomic DNA. Demonstrated herein is expression of an example of a synthetic chimeric RNA that drives formation of a specified gene fusion *via* genomic translocation in mammalian cells. In particular embodiments, a first region of a first gene in a genome is fused with a second region of a second gene in the genome, and the fusion is dictated by the sequence of one or more chimeric RNAs, and in some cases by the sequence of one or more antisense chimeric RNAs.

[0065] In particular embodiments, provided herein are methods of producing genomic DNA rearrangements between two different regions of genomic DNA by exposing to the genomic DNA an effective amount of an exogenously provided chimeric RNA. The particular chimeric RNA is not endogenous, in specific cases, and it may be synthetic and made by the hand of man. It may be derived from other chemically modified nucleic acids, from proteins, *etc.* In other cases the chimeric RNA is similar or identical to an endogenous chimeric RNA but is exogenously provided to the genomic DNA. In specific embodiments, the chimeric RNA comprises fusion of (1) a first part that is reverse complement with respect to a first genomic region and (2) a second part that is reverse complement with respect to a second genomic region.

Upon exposing of the chimeric RNA to the genomic DNA, the first and second genomic regions become fused through DNA recombination.

[0066] In particular embodiments when the chimeric RNA is exposed to the genomic DNA, the chimeric RNA then hybridizes with the respective strands of first and second genomic regions of the genomic DNA through the corresponding first and second parts of the chimeric RNA. Such hybridization thereby forms a DNA/RNA hybrid in a sequence-specific manner, in particular embodiments. In certain embodiments following or simultaneously with this hybridization of the chimeric RNA with the respective strands of the first and second genomic regions of the genomic DNA, a double stranded DNA/DNA stem structure is produced. The DNA/DNA stem structure is wholly genomic DNA and may be particularly produced between the first and second genomic regions. In specific embodiments the DNA/DNA stem structure is adjacent to the site of hybridization with the chimeric RNA. As used herein in this embodiment, adjacent to the site refers to within 15 and 50 nucleotides of the site of hybridization with the chimeric RNA. That is, in particular embodiments upon hybridizing of the chimeric RNA with the genomic DNA at the first and second genomic regions, a DNA spacer is produced in each of the respective strands of the first and second genomic regions between the DNA/DNA stem structure and the DNA/RNA hybrid structure, and this DNA spacer separates the DNA/DNA stem structure from the DNA/RNA hybrid structure. The DNA spacer may or may not be considered as a bulge or loop, in specific aspects, and the DNA spacer may be double stranded or single stranded, or have regions of both.

[0067] Turning to FIG. 24, illustrated therein is one example of targeting of a specifically designed chimeric RNA for the purpose of fusing the *TMPPRS2* and *ERG* genes, merely as examples. The particular structure produced upon this targeting generally approaches a 3-way stem formation. The chimeric RNA having particular sequence from the *TMPPRS2* and *ERG* genes may be designed specifically for the purpose of fusing the *TMPPRS2* and *ERG* genes. In this example, given that the genes are transcribed (as opposed to intergenic sequences, for example) the chimeric RNA is antisense with respect to the respective sense sequences of both *TMPPRS2* and *ERG* genes. Upon hybridization of the *TMPPRS2* and *ERG* genes with the chimeric RNA, a double-stranded DNA/DNA hybrid (which may be referred to as a stem) is generated as well as a spacer region (that may be referred to as a bulge) in the genomic DNA between the stem and the point of joint hybridization of the chimeric RNA with its corresponding sense sequences of *TMPPRS2* and *ERG* genes. Thus, in a 3-way junction one

“piece” comprises a DNA/RNA hybrid and another “piece” comprises a DNA spacer and DNA/DNA hybrid that in some cases may or may not be generally perpendicular with respect to the DNA/RNA hybrid. The DNA spacer may be single-stranded, double-stranded, or have regions of both single-strandedness and double-strandedness. Following this formation, endogenous machinery and/or reagents in the cell may then generate double-stranded breaks in the genomic regions and fusion. The double-stranded DNA breaks may occur outside or within the 3-way stem formation.

[0068] Also included herein are methods of inducing gene fusion between first and second target genes by designing or obtaining a chimeric RNA comprising (1) a first part that is antisense with respect to the first target gene and (2) a second part that is antisense with respect to a second target gene; and exposing an effective amount of the chimeric RNAs to nucleic acid comprising the first and second target genes, wherein upon exposing of the chimeric RNA to the nucleic acid, the first and second target genes become fused through DNA recombination.

[0069] In some cases, the fusion of two genomic DNA regions occurs in regions that are prone to DNA breaks, such as “hot spots” or at other DNA motifs, such as those with ALU elements and the repeated sequences.

[0070] In specific embodiments the genomic DNA is from or in a mammalian subject, although the genomic DNA may not be mammalian in some cases. The first and second parts of the chimeric RNA are non-identical and may be on the same chromosome or they may be on different chromosomes. The genomic DNA of the cell may be from any type of normal cell or any type of diseased cell. The methods may or may not be performed in vivo, such as ex vivo or in vitro. The cell or cells from which the genomic DNA is derived may be a human cell. The cell may be in a tissue or organ or body fluid. The genomic DNA may be from a diseased cell or diseased cells. In some cases, the genomic DNA is from an individual in need of treatment or suspected to be in need of treatment or prevention of a medical condition associated with one or more chimeric RNAs. In specific embodiments, the method occurs at physiological hormone levels.

[0071] In particular cases, recombination is induced by chimeric RNA that are designed such that the genomic DNA stem produced upon formation of a three-way junction with the chimeric RNA comprises sequence that is common in the genome, such as sequences that are repeats in the genome, including retrotransposons, such as short interspersed nuclear elements

(SINE) and long interspersed nuclear elements (LINE). In such cases, the SINEs or LINEs provide stability to the genomic stem of the three-way junction.

II. [0072] Chimeric RNA Compositions

[0073] Chimeric RNA compositions of the disclosure are utilized to induce genomic rearrangements, in particular cases. The chimeric RNAs in particular embodiments are synthetic and not from nature, although in specific embodiments they are isolated from nature. The synthetic chimeric RNAs may be used for a non-natural method of producing fusion of two different genomic DNA regions. The synthetic chimeric RNAs may be designed specifically for the purpose of generating fusion of two genomic DNA regions. The chimeric RNAs are the product of the hand of man and not natural. In such cases, two known genomic DNA regions are considered for targeting and the sequence of the chimeric RNA is determined based on the sequence of the two known genomic DNA regions. In cases wherein the two known genomic DNA regions are genes being transcribed (or in some cases, capable of being transcribed), the chimeric RNA is antisense with respect to one or both of the genomic DNA regions. In cases wherein the two known genomic DNA regions are genomic DNA regions that are not transcribed (for example, intergenic regions), the chimeric RNA may be designed to hybridize to either strand of the target genomic DNA non-transcribed region(s). In some cases, one of the targeted genomic DNA regions is not being transcribed and the corresponding region of the chimeric RNA may be either sense or antisense while the other targeted genomic region is transcribed and the corresponding region of the chimeric RNA is antisense.

[0074] In some cases, the synthetic chimeric RNA may have reverse complementary regions only at the ends, and in some cases reverse complementary for genomic region A at 5' end and reverse complementary for genomic region B at 3' separated by a loop-out stretch of unknown sequence of varying length.

[0075] In particular embodiments, the synthetic chimeric RNA comprises sequence and/or structure that is capable of producing a 3-way junction stem formation in the genomic DNA between the two regions (see, for example, FIG. 24). In producing the 3-way junction stem formation in the genomic DNA between the two regions, there is a DNA spacer (that may or may not structurally bulge from the junction) on both sides of the base of the 3-way junction. In at least some cases, there is an optimum size of the DNA spacer for fusion induction. In specific embodiments the size of the DNA spacer of either genomic DNA strand is around 15,

16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 nucleotides on each side. In particular embodiments the size of the DNA spacer of either genomic DNA strand is around 35 nucleotides in size. In some cases, the length of the DNA spacer in a first genomic DNA strand is the same or different compared to the length of the DNA spacer in a second genomic DNA strand. For example, the DNA spacer of first and second genomic DNA strand may be in a length in a range of 15-50, 15-45, 15-30, 15-25, 15-20, 20-50, 20-45, 20-40, 20-35, 20-30, 20-25, 25-50, 25-45, 25-40, 25-35, 25-30, 30-50, 30-45, 30-40, 30-35, 40-50, 40-45, or 45-50 nucleotides at either side. In some cases, the length of the DNA spacer in a first genomic DNA strand is the same or different compared to the length of the DNA spacer in a second genomic DNA strand.

[0076] In particular cases, there is an optimum synthetic chimeric RNA size for fusion induction. In specific embodiments, the chimeric RNA is a fusion of RNA corresponding to two different genes and each region of the chimeric RNA may be of a particular length. In specific embodiments, the two different regions of the chimeric RNA are substantially the same in length, although in other embodiments the two different regions of the chimeric RNA are not the same or are not substantially the same in length. In specific embodiments, the length of each side of the chimeric RNA is about 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, and so on. In some embodiments, the length of two different regions of the chimeric RNA may be in a range of 30-150, 30-125, 30-100, 30-75, 30-50, 50-150, 50-125, 50-100, 50-75, 75-150, 75-125, 75-100, 100-150, 100-125, 125-175, or 125-150. In a particular embodiment, the length of each side of the chimeric RNA is in a range of 50-75 nucleotides. In specific embodiments, the length of 50-75 nucleotides on each side of the chimeric RNA may be utilized when the DNA spacer bulge is 35 nt. In some embodiments, any length of each side of the chimeric RNA may be utilized when the DNA spacer bulge is in a range of 15-50 nt.

[0077] The length of the synthetic chimeric RNA may be of any suitable length but in specific embodiments it is in the range of 60-300, 60-250, 60-200, 60-150, 60-100, 75-300, 75-250, 75-200, 75-150, 75-100, 100-300, 100-250, 100-200, 100-150, 150-300, 150-250, 150-200,

150-175, 175-300, 175-250, 175-200, 200-300, 200-275, or 200-250 nucleotides, for example. The length of the chimeric RNA may be 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, or 155 nucleotides, or shorter or longer.

[0078] In specific embodiments for fusion gene induction, an optimum size of chimeric RNA is around 50nt on each side to 75nt on each side.

[0079] In some embodiments, chimeric RNAs are designed to have a particular percentage of GC content. In specific embodiments, the GC content is from about 27-65, 27-60, 27-55, 27-50, 27-45, 27-40, 27-35, 27-30, 35-65, 35-60, 35-55, 35-50, 35-45, 35-40, 40-65, 40-60, 40-55, 40-50, 40-45, 50-65, 50-60, 50-55, 55-65, 55-60, or 60-65%GC. The GC content may be about or may be 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, or 65%GC.

[0080] The sequence of endogenous or exogenously provided chimeric RNA need not be identical or 100% match to that of the parental genes for induction of fusion. For example, where FIG. 33 and FIG. 34 illustrate examples where a man-made chimeric RNA and an endogenous initiator chimeric RNA, respectively have a perfect match to the genes to be fused, there are examples where the match may or may not be perfect (FIGS. 35-38).

[0081] In specific embodiments, the chimeric RNA is designed such that upon generation of a three-way junction with a genomic DNA as the stem, the genomic DNA may comprise SINE or LINE sequence, or the chimeric RNA may be designed to exclude SINE or LINE sequence. In particular cases, the design of the chimeric RNA is such that the genomic DNA stem comprises SINE or LINE and the length of such elements (for example, 200-300 bp) provides stability to the stem of the junction.

[0082] Delivery of the chimeric RNA to cells of interest (whether cell lines or other cells) may be by any suitable method, including in a vector or without a vector. In cases wherein the chimeric RNA is delivered without a vector, the delivery may comprise electroporation. In cases wherein the RNA is delivered with or in a vector, the delivery may be *via* viral vectors or non-viral vectors. Examples of viral vectors include adenoviral-associated, lentiviral, adenoviral, or

retroviral, and examples of non-viral vectors include plasmids, nanoparticles, liposomes, and so forth.

III. [0083] Methods of Detecting and Targeting Endogenous Chimeric RNAs

[0084] Endogenous chimeric RNAs, as opposed to synthetically produced, man-made chimeric RNAs for other embodiments of the disclosure, may be a component of methods for detecting and targeting. The detecting and targeting methods may be related to clinical aspects of the disclosure, including diagnosis, treatment, and monitoring of therapy, for example.

A. Detecting Endogenous Chimeric RNAs

[0085] Embodiments of the disclosure include diagnostic, prognostic, preventive, and therapeutic strategies relating to endogenous chimeric RNAs present *in vivo*. In particular embodiments, prior to the occurrence of a disease-causing fusion gene, the presence of an endogenous chimeric RNA that is directed to two different genomic regions in a cell is a clinical marker that the cell is cancerous or is at elevated risk for becoming cancerous. The cancer may be of any kind, including any solid tumor or blood cancers.

[0086] Identification of an endogenous chimeric RNA, prior to the occurrence of a disease-causing fusion gene, in one or more cells from a tissue from an individual may indicate that the cell(s) (and by extrapolation, the tissue) is cancerous or at risk for becoming cancerous. Therefore, the present disclosure encompasses methods of identifying endogenous chimeric RNAs from samples from individuals for the purpose of diagnosing cancer or determining a risk for cancer for the individual. In particular embodiments, the chimeric RNA corresponds to a certain cancer fusion gene, and many cancer fusion genes are known in the art (see, for example, the TCGA Fusion Gene Database and ChimerDB 3.0 database on the world wide web). Therefore, in particular embodiments, from samples from an individual suspected of having cancer or at risk for cancer, the presence of one or more particular chimeric RNAs are searched for based on known sequences of known cancer fusion genes.

[0087] For example, the sequence of the chimeric RNA may be predicted based on the known sequences of cancer fusion genes, and the methods of identifying whether or not there are chimeric RNAs in a sample utilize such sequences. For example, a method of identifying the presence of chimeric RNAs in a sample may be based on identifying the nucleic acid sequence of an endogenous chimeric RNA based on known cancer fusion genes that are associated with the cells and tissue of interest. For example, an endogenous chimeric RNA may have antisense

sequences that are corresponding to the intronic or exonic genomic regions of actively transcribing genes. For example, an endogenous chimeric RNA may have sense or antisense sequences that are corresponding to the intronic or exonic genomic regions of the untranscribed genes. For example, the sequence of an endogenous chimeric RNA may be fully complementary or partially complementarity to the sequence of genes of interest. or comprises sequence that is partially complimentary to that of the parental genes.

[0088] An example of such a method includes reverse transcriptase-polymerase chain reaction (RT-PCR), Northern RNA gel analysis, and microarray chips designed for detecting endogenous chimeric RNAs. In such cases for RT-PCR as an example, primers for the PCR may be designed to determine the presence of the chimeric RNA based on the sequence of the cancer gene fusion. For example, the sequence of the chimeric RNA may be determined based on sequence from the cancer gene fusion and in specific cases may include the fusion point between the two genes of the cancer gene fusion.

[0089] The endogenous chimeric RNAs may be considered to be initiatory RNAs to induce gene fusion. In some cases, their presence in cells from a sample indicates that gene fusion to produce cancer fusion gene(s) may or may not have already occurred or will occur.

[0090] A particular type of cancer may have multiple cancer gene fusions, and one or more chimeric RNAs directed to one or more of the multiple cancer gene fusions may be assayed for in a sample.

[0091] In some embodiments, the presence of one or more endogenous chimeric RNAs may be assayed for as a means to monitor therapy for a medical condition, such as cancer or another medical condition. The monitoring of therapy may include analysis of samples for one or more endogenous chimeric RNAs before and after at least one round of therapy for the individual. In some cases, analysis of a sample for endogenous chimeric RNAs occurs after a therapy but not before. When a treatment is effective, there is a detectable reduction in the amount of chimeric RNAs following the treatment.

[0092] In alternative embodiments, the presence of one or more endogenous chimeric RNAs is an indicator of presence or risk for a medical condition other than cancer, such as heart disease, neurodegenerative disease, or diabetes.

[0093] Following identification of endogenous chimeric RNAs, an individual may be given one or more treatments because of the chimeric RNA presence. The treatment may or may not be directed against the chimeric RNA itself. In cases wherein a cancer treatment does not directly target the endogenous chimeric RNA, the treatment may include chemotherapy, immunotherapy, radiation, hormone therapy, surgery, or a combination thereof, as examples.

[0094] Samples from which the presence or absence of endogenous chimeric RNA is analyzed generally comprise one or more cells. The samples may be of any kind, including blood, urine, biopsy, hair, sputum, semen, saliva, cheek scrapings, cerebrospinal fluid, nipple aspirate, gastric fluid, Pap smears, or a combination thereof.

B. Targeting Endogenous Chimeric RNAs

[0095] In particular embodiments of the disclosure, there are methods of targeting endogenous chimeric RNAs to treat a medical condition associated with the presence of the chimeric RNAs. In specific embodiments, the targeting methods neutralize the disease-causing endogenous chimeric RNA. Thus, the chimeric RNA may be a therapeutic target for treatment methods and may be targeted with a structural and/or functional inhibitor of the chimeric RNA. In specific embodiments, an agent that directly binds the chimeric RNA is utilized, such as a siRNA or other RNA interference molecule or antisense oligo DNA. In specific embodiments, an agent that indirectly inhibits the chimeric RNA is utilized, such as a compound that methylates the promotor that drives the expression of endogeneous chimeric RNA.

[0096] Knowing the sequence of an endogenous chimeric RNA, one can develop siRNA to target the chimeric RNA as a therapy. In particular embodiments, the siRNA is delivered *via* a vector that expresses the siRNA, such as a viral vector or non-viral vector (such as a plasmid). Viral vectors include lentiviral, retroviral, adenoviral, or adeno-associated viral vectors. Promoters suitable for effective expression of the siRNA are known in the art. The siRNA may be provided to the individual locally or systemically.

[0097] In some cases, an individual is given an effective amount of an inhibitor of endogenous chimeric RNAs associated with cancer or other deleterious gene fusions when the individual has not been proven to have the presence of the cancer. That is, a medical practitioner finding cancer in an individual can presume that chimeric RNAs are present and provide an effective amount of one or more agents that inhibit the chimeric RNA.

[0098] A particular type of cancer may have multiple cancer gene fusions, and one or more agents directed to one or more endogenous chimeric RNAs may be utilized for a particular therapy.

EXAMPLES

[0099] The following examples are presented in order to more fully illustrate the preferred embodiments of the invention. They should in no way, however, be construed as limiting the broad scope of the invention.

EXAMPLE 1

RNA-MEDIATED GENE FUSION IN MAMMALIAN CELLS

[0100] One of the hallmarks of cancer is the formation of oncogenic fusion genes as a result of chromosomal translocations. Fusion genes are presumed to occur prior to fusion RNA expression. However, studies have reported the presence of fusion RNAs in individuals who were negative for chromosomal translocations. These observations give rise to “the cart before the horse” hypothesis, in which fusion RNA precedes the fusion gene and guides the genomic rearrangements that ultimately result in gene fusions. Yet RNA-mediated genomic rearrangement in mammalian cells has never been demonstrated. Here the inventors provide evidence that expression of a chimeric RNA drives formation of a specified gene fusion via genomic rearrangement in mammalian cells. The process is (1) specified by the sequence of chimeric RNA involved, (2) facilitated by physiological hormone levels, (3) permissible regardless of intra-chromosomal (*TMPRSS2-ERG*) or inter-chromosomal (*TMPRSS2-ETV1*) fusion, and (4) can occur in normal cells prior to malignant transformation. It is demonstrated that in at least some cases, contrary to “the cart before the horse” model, it is the antisense rather than sense chimeric RNAs that effectively drive gene fusion, and that this disparity can be explained by transcriptional conflict. Furthermore, an endogenous RNA *AZII* is identified that acts as the ‘initiator’ RNA to induce *TMPRSS2-ERG* fusion. RNA-driven gene fusion demonstrated in this disclosure provides important insight in early disease mechanism, and allows for applications in the biology of mammalian genome stability, as well as gene editing technology *via* mechanisms native to mammalian cells.

EXAMPLE 2

RNA-MEDIATED GENE FUSION IN MAMMALIAN CELLS

[0101] To test whether the expression of a fusion RNA in mammalian cells can lead to a specific gene fusion, the *TMPRSS2-ERG* fusion (Perner et al., 2006; Tomlins et al., 2005), found in ~50% of prostate cancers, was selected as a model. Both the *TMPRSS2* and *ERG* genes are located on chromosome 21, an intra-chromosomal configuration prone to rearrangements. To recapitulate *TMPRSS2-ERG* fusion gene formation, the LNCaP prostate cancer cell line that lacks the *TMPRSS2-ERG* fusion (Horoszewicz et al., 1980; Tomlins et al., 2005) was used. Furthermore, treating LNCaP cells with androgen increases the chromosomal proximity between the *TMPRSS2* and *ERG* genes (Bastus et al., 2010; Lin et al., 2009; Mani et al., 2009), which was thought to increase the possibility of gene fusion. To test “the cart before the horse” hypothesis (Rowley and Blumenthal, 2008; Zaphiropoulos, 2011), a short fusion RNA was transiently expressed consisting of two exons, *TMPRSS2* exon-1 joined to *ERG* exon-4, which is a short fragment of a full-length *TMPRSS2-ERG* fusion RNA that is most common in prostate cancer (FIG. 1A, upper panel). This short fusion RNA mirrors the presumptive trans-spliced fusion RNA product that is generated only in the sense orientation because the correct splice sites are absent in the antisense orientation. However, because the ‘antisense’ sequence should, in theory, contain the same template information for guiding genomic rearrangements, both the sense and antisense short fusion RNA were tested. Each was individually expressed using either a CMV or a U6 promoter (FIG. 1A, upper panel) and designated as ‘input RNA’ to distinguish them from the ‘endogenous’ full-length fusion RNA transcribed from the genome.

[0102] LNCaP cells were transiently transfected with either plasmid and treated the cells with dihydrotestosterone (DHT, a metabolite of testosterone) for 3 days. If the expression of an input RNA leads to a *TMPRSS2-ERG* gene fusion, it is expected that the endogenous full-length fusion RNAs would be transcribed from the newly induced fusion gene. Specific RT-PCR assays were designed to distinguish between endogenous full-length fusion RNAs and the input RNAs exogenously expressed from the plasmids (see FIG. 8A for primer designs). As shown in FIG. 1A, expression of the sense short fusion RNA resembling the trans-spliced product, either by the CMV or U6 promoter (FIG. 1A lower panel, lane 1 and 2 respectively), led to no detection of an induced endogenous fusion transcript. Expression of a longer version of sense fusion RNA consisting of four exons (*TMPRSS2* exon-1 joined to *ERG* exon-4/5/6) also failed to induce the

endogenous fusion transcript (FIGS. 9A-9B). In contrast, expression of antisense short fusion RNAs induced a clear band of 721 bp (FIG. 1A lower panel, lane 3 and 4). Sanger sequencing revealed that the induced band contains *TMPRSS2* exon-1 fused to *ERG* exons-4/5/6/7, and that the exons are joined by annotated splice sites, which would be expected of mature endogenous fusion mRNA derived from the *TMPRSS2-ERG* fusion gene. This induced fusion transcripts cannot possibly arise from the sequence of input RNAs as the expression plasmids used contain only *TMPRSS2* exon-1 and *ERG* exon-4 without the *ERG* exon-5/6/7 sequence. Notably, the induction was more pronounced when the antisense input RNA was driven by the U6 promoter (FIG. 1A lower panel, antisense-2, lane 4) compared to the CMV promoter (antisense-1, lane 3). These differences (antisense vs. sense, U6 vs. CMV) are not caused by differing amounts of input RNA because all input RNAs were expressed at relatively equal levels (FIG. 1A, lower panel). Transfection with a parental plasmid containing mCherry sequence (FIG. 1A, lane 5), DHT treatment without plasmid transfection (FIG. 1A, lane 6), and PCR reaction without cDNA served as RT-PCR controls (FIG. 1A, lane 8), all resulted in no endogenous fusion transcript. In addition, all experiments were performed independently at least four times and results were identical. Together, the data suggest that expression of an input RNA with chimeric sequence can lead to the induction of a specified endogenous fusion transcript in human cells. Surprisingly, the antisense, rather than the sense version of input RNA, exhibits the capacity of induction.

[0103] Antisense input RNAs described above contain 218 nt against the entire *ERG* exon-4 and 78 nt against the entire *TMPRSS2* exon-1 (FIG. 1A), suggesting that 78 nt is sufficient to specify a parental gene for gene fusion. Furthermore, because the effective input RNAs are of the ‘antisense’ orientation, the data imply that the input RNAs may not require an RNA junction resembling that of the *TMPRSS2-ERG* fusion transcript generated by splicing in the sense orientation. To further analyze the sequence requirement, the U6 promoter was used to express a series of antisense input RNAs with 75 nt complementary to *ERG* exon-4 joined to various segments (33, 52, 67, 82 nt) that are complementary to *TMPRSS2* near the exon-1/intron-1 boundary (FIG. 1B). A parallel set of sense input RNAs were also tested as controls (FIG. 1C). As shown in FIG. 1B (lower panel), all antisense RNAs, with the exception of antisense-3, induced fusion transcripts even though their target regions span the exon/intron boundary. The level of induction peaked for antisense-5, which contains 52 nt designed to anneal with *TMPRSS2*, suggesting that this number of nucleotides might be optimal to specify a parental gene for induction. The results were confirmed using a different, but more efficient, primer pair

(FIG. 1B, lower panel; primer design in FIG. 8B) followed by Sanger sequencing of the induced band. In contrast to the antisense input RNA, all corresponding sense input RNAs failed to induce endogenous fusion transcripts (FIG. 1C, lower panel). This was true even when the sense input RNA was intentionally expressed at a much higher level than the antisense RNA (FIG. 10). Additional experiments using plasmids with a severed U6 promoter (FIGS. 11A-11C), to eliminate input RNA expression, also confirmed that it is the antisense input RNAs expressed from plasmids, not the DNA sequence of plasmids, that induce the observed *TMPRSS2-ERG* fusion transcripts.

[0104] As shown in FIG. 1D, the amount of endogenous fusion transcript induced by antisense-5 (the most effective antisense input RNA) appears to correlate with the concentration of DHT used, presumably because the hormone increases the chromosomal proximity between the *TMPRSS2* and *ERG* genes (Bastus et al., 2010; Lin et al., 2009; Mani et al., 2009). Antisense-5 was effective at DHT concentrations as low as 20 nM as revealed by sensitive nested PCR (FIG. 1D lane 3), indicating that fusion events induced by input RNA can occur under physiologically relevant androgen conditions (Boyce et al., 2004). As a control, DHT treatment alone up to 2 μ M failed to induce fusion (FIG. 12). Titration of DHT showed that the induction by antisense-5 reaches the 50% of maximal level (EC_{50}) at 0.9 μ M DHT (FIG. 12). Under this standard EC_{50} condition, we estimated that the percentage of LNCaP cells induced by antisense-5 to express the *TMPRSS2-ERG* fusion transcript is approximately 1 in 10^3 or 10^4 cells (see assay in FIG. 13). Together, these results demonstrated that the induction by input RNA can occur at physiologically relevant hormone levels, but does not represent a high frequency event.

[0105] Although induced fusion is infrequent, all antisense RNAs described in FIG. 1B successfully induced endogenous fusion RNA except antisense-3, which is only 30 nt longer than antisense-5 in the arm targeting *TMPRSS2* intron-1 (FIG. 1B, upper panel). To test whether its inability to induce was due to input RNA length or the specific target sequence in *TMPRSS2* intron-1, we constructed a hybrid antisense (antisense-7) that shifted the 52 nt recognition window of antisense-5 to target the *TMPRSS2* intron-1 region covered by antisense-3 (FIG. 1E). This alteration resulted in the loss of induction (FIG. 1E, lane 1 vs. 3), implying that the inability of antisense-3 to induce is not reflective of input RNA length. Rather its targeting arm may interfere with a motif important for the fusion process. BLAST alignment of the genomic DNA sequence revealed an imperfect stem (named stem A) potentially formed by the sense genomic *TMPRSS2* sequence complementary to the sense genomic *ERG* sequence (FIG. 2A, left). It was

considered that this genomic DNA stem ($T_m = \sim 44^\circ\text{C}$) could potentially stabilize a three-way junction that involves an RNA/DNA duplex formed by the antisense-5 RNA and its targeted genomic DNA in a sequence-specific manner (FIG. 2A, left). If correct, then the formation of this putative three-way junction would be disrupted by antisense-3 because its recognition sequence invades the genomic DNA stem (FIG. 2A, left). Consistent with the idea that induction requires bringing *TMPRSS2* and *ERG* gene in close proximity, expression of antisense-5 as two separate halves (FIG. 2A right panel, antisense-5A and -5B) severed the link between *TMPRSS2* (52 nt) and *ERG* (75 nt) sequence within the input RNA, resulting in the loss of induction (FIG. 2B, lane 1 to 3).

[0106] To test whether the proposed three-way junction formation could facilitate induction, BLAST alignment was used to identify several intron locations where genomic DNA stems could be formed by the sense genomic *TMPRSS2* sequence paired with the sense genomic *ERG* sequence (stems B to G in FIGS. 2C and 2D; sequences flanking the stems in FIG. 14). Matching antisense input RNAs (termed antisense-B1 to G1) were then designed to facilitate the formation of a three-way junction with the possible intron stems (FIG. 2D) that would mirror the three-way junction formed by antisense-5 on stem A as postulated in FIG. 2A. Because these input RNAs target the introns (FIG. 2C) and contain no exon sequence, any observed induction of endogenous fusion transcripts composed of exons cannot arise from the sequence of input RNAs or plasmids used to express them. As shown in FIG. 2E, targeting genomic DNA stem B, C, and D that exhibit higher DNA stem stability ($T_m = 40^\circ\text{C}$, 40°C , and 44°C , respectively) by the corresponding antisense input RNAs clearly induced fusion transcripts (FIG. 2E, lanes 2 to 4). In contrast, targeting less a stable stem E, F, and G ($T_m = 30^\circ\text{C}$, 24°C , and 16°C , respectively) failed to induce fusion transcripts (FIG. 2E, lanes 5 to 7). To disrupt the three-way junction involving stem B, C, and D, six additional antisense RNAs (antisense-B2, B3, C2, C3, D2, and D3) were designed with one side of their recognition sequence altered to invade each of the respective genomic DNA stems on the *TMPRSS2* side or the *ERG* side (FIG. 15). These modifications were chosen to mirror the interference on stem A by antisense-3. Similar antisense RNAs were also designed to invade stem A (FIGS. 16A-16B). In all cases, invasion of the genomic DNA stems by the modified input RNAs resulted in the complete loss of induction (FIG. 2F). While these results by no means necessitate that a three-way junction is required for fusion transcript induction, they nevertheless suggest that such transiently stabilized structures may ‘facilitate’ the process and could have important implications in developing gene editing

technologies via mechanisms native to mammalian cells. Consistent with earlier observations, the corresponding sense version of the effective antisense input RNAs (sense-B1, C1, D1) all failed to induce fusion transcripts (FIG. 2G, lanes 2 to 4).

[0107] The fact that antisense input RNAs, but not their sense counterparts, induce fusion transcripts, raises the possibility that the former act as a docking station to mediate trans-splicing between endogenous sense *TMPRSS2* and *ERG* pre-mRNAs. Because that the antisense, but not the sense input RNAs, are complimentary to both sense *TMPRSS2* and *ERG* pre-mRNAs, they can base-pair with both parental pre-mRNAs, thus resulting in spliced fusion transcripts without the requirement of genomic rearrangement. This mechanism, however, is unlikely as the major contributor to the observed induction for the following reasons. First, although *TMPRSS2* is expressed in LNCaP cells (FIG. 3A, upper panel), endogenous *ERG* mRNA is not detected in LNCaP cells (Tomlins et al., 2005) in the presence or absence of DHT or before and after transfection of antisense-5 (FIG. 3A, middle and lower panel with different primer pairs). In fact, parental *ERG* mRNA was not detected even using three rounds of nested RT-PCR using various primer sets (FIGS. 17A-17D). Therefore prior to and during induction, no or an insufficient number of parental *ERG* mRNAs are available in LNCaP cells as raw material for trans-splicing to account for the level of induced fusion transcript. Second, after initial transient transfection and DHT treatment for 3 days, we continued to propagate and enrich the induced LNCaP population for 52 days in the absence of DHT (experimental procedures described in FIG. 18). As shown in the lower panel of FIG. 3B, antisense-5 RNA transiently expressed by plasmids was degraded and completely absent beyond day 17. In contrast, the induced fusion transcript was continuously expressed and enriched up to day 52 in the absence of antisense input RNA and DHT (FIG. 3B upper panel), indicating the persistent nature of the induced fusion product. Taken together, these results strongly suggest that the induced expression of the *TMPRSS2-ERG* fusion transcript is the consequence of gene fusion at the DNA level, which has a permanent nature. This is in contrast to the result of induced trans-splicing at the RNA level mediated by antisense input RNA, which is transient and requires the continuous presence of input RNAs.

[0108] To provide definite evidence of gene fusion *via* genomic rearrangement, genomic PCR was used to identify the genomic breakpoint induced by antisense-5 in the enriched LNCaP population (primer designs in FIG. 19A and FIG. 3C). As shown in FIG. 3D, the un-rearranged wildtype *TMPRSS2* and *ERG* alleles were amplified by gene-specific primer pair A/B and C/D both in untransfected cells (FIG. 3D, lane 1 and 2) and enriched LNCaP cells (FIG. 3D, lane 4

and 5). In contrast, a genomic fusion band of ~862 bp amplified by fusion-specific primer pair A/D was present only in the enriched LNCaP population (FIG. 3D, lane 6) and absent in untransfected LNCaP cells (lane 3). Sanger sequencing of the excised fusion band (FIG. 3D, lane 6) revealed the exact genomic breakpoint located within *TMPRSS2* intron-1 (chr21:41502038, GRCh38/hg38) and *ERG* intron-3 (chr21:38501207, GRCh38/hg38) (FIG. 3E). Intriguingly, within *TMPRSS2* intron-1 the induced breakpoint lies within an Alu, a transposable element known to contribute to genomic rearrangements (Rudiger et al., 1995). In *ERG* intron-3, the breakpoint resides in a hot spot clustered with genomic breakpoints previously identified in prostate cancer patients (FIG. 19B) (Weier et al., 2013). There is no obvious sequence homology between *TMPRSS2* and *ERG* at the genomic breakpoint except for a three nucleotide 'CTG' microhomology (FIG. 3E), suggesting that this gene fusion may be mediated by non-homologous break repair mechanisms (Lieber, 2010; Zhang et al., 2009).

[0109] To test whether antisense input RNA can cause *TMPRSS2-ERG* fusion in non-malignant cells prior to cancerous transformation, experiments were performed using immortalized normal prostate epithelium cells (PNT1A), that express very low levels of androgen receptors (Coll-Bastus et al., 2015). As shown in the lower panel of FIG. 3F, prolonged expression of antisense-5 for 12 days induced fusion transcripts. This induction was not due to prolonged exposure to DHT because continuous treatment of 0.9 μ M DHT alone for up to 2 months resulted in no detectable fusion transcripts in PNT1A cells (FIG. 3F, lane 8). Thus, the results indicate that the induction of *TMPRSS2-ERG* fusion by antisense input RNA can occur in normal prostate epithelial cells prior to malignant transformation and is not restricted to the pathological cellular context of malignant cells.

[0110] To test whether an input RNA can specify a pair of genes to undergo fusion other than *TMPRSS2-ERG* in a sequence-specific manner, a series of input RNAs were designed to induce *TMPRSS2-ETV1*, an inter-chromosomal fusion gene found in approximately 1% of prostate cancers (Rubin et al., 2011; Tomlins et al., 2005). Eight antisense RNAs (FIG. 20) were designed to target different chosen regions in the introns where three-way junctions potentially can be forged between the genomic DNA and input RNAs (FIGS. 20 and 21). Again, because these input RNAs target introns and contain no exon sequence, it rules out the possibility that induced endogenous fusion transcripts composed of exons arise from the sequence of input RNAs or the plasmids. As shown in FIG. 4A, targeting stem *TMPRSS2-ETV1-A*, which has the highest genomic DNA stem stability ($T_m = 72^\circ\text{C}$) among this group, led to clear induction of the

TMPRSS2-ETV1 fusion transcript (FIG. 4A, lane 1). Sanger sequencing validated that the induced transcript contains *TMPRSS2* exon-1 joined with *ETV1* exon-3 (uc003ssw.4) by annotated splice sites. Similar to earlier observations, targeting with sense versions of input RNAs (FIG. 4B, lane 1 vs. 2), or using antisense input RNAs designed to form three-way junctions with lower genomic DNA stem stabilities (FIG. 4A, lanes 2 to 8; FIG. 20), resulted in no detectable induction. Furthermore, the input RNA designed to target *TMPRSS2* and *ETV1* induced *TMPRSS2-ETV1* fusion but not *TMPRSS2-ERG* fusion (FIG. 4C, lane 2). Conversely, antisense-5 targeting *TMPRSS2* and *ERG* induced *TMPRSS2-ERG* fusion but not *TMPRSS2-ETV1* fusion (FIG. 4C, lane 1), indicating that fusion formation is specified by the sequence of input RNA and not secondary effects such as global genomic stability.

[0111] To verify that *TMPRSS2-ETV1* as a second example of induced fusion that is indeed the consequence of genomic translocation, the inventors propagated and enriched the induced LNCaP population for 47 days after the initial transfection of input RNA and DHT treatment (experimental procedures same as described for *TMPRSS2-ERG* enrichment in FIG. 18). The transiently expressed antisense input RNA had been degraded and was absent by day 47 (FIG. 4D, lane 1 vs. 2). The induced *TMPRSS2-ETV1* fusion transcript, however, was continuously expressed beyond day 47 (FIG. 4D, lane 2). Once again this observation indicated that the sustained expression of an induced fusion does not require the continuous presence of input RNA. Moreover, genomic PCR assays identified three distinct genomic breakpoints between *TMPRSS2* and *ETV1* gene (labeled as x, y, z in FIG. 4E) that were present only in the enriched LNCaP population but absent in untransfected LNCaP cells (FIG. 4E, lane 3 vs. 6, lane 9 vs. 12). Similar to earlier observations, no obvious sequence homology between *TMPRSS2* and *ETV1* was observed at the genomic breakpoints except for a few nt of microhomology (FIG. 4F), indicating that the gene fusion is mediated by non-homologous break repair mechanisms (Lieber, 2010; Zhang et al., 2009).

[0112] Unlike *TMPRSS2* and *ERG* that are located near each other on the same chromosome, *TMPRSS2* and *ETV1* are located on different chromosomes. Thus, gene fusion as a result of chromosomal translocation could be confirmed unequivocally by evidence of chromosomal co-localization of the latter pair. Using probes specific to *TMPRSS2* and *ETV1*, fluorescence in situ hybridization (FISH) was performed, followed by deconvolution microscopic imaging, of 3301 cells from the enriched LNCaP cell population and 620 cells from the control untransfected LNCaP population. Analyses of constructed 3D images showed that

approximately 0.9% of the enriched population (30 out of 3301 cells) were positive for co-localization of *TMPRSS2* and *ETV1* gene in the cellular nucleus (FIG. 4G). In contrast, none of the cells from the untransfected population showed co-localized FISH signals as determined by the same 3D image criteria. These results, together with the genomic breakpoints identified by genomic PCR at single base resolution (FIGS. 4E and 4F), provide strong evidence that the induced expression of the *TMPRSS2-ETV1* fusion transcript represents the consequence of gene fusion caused by chromosomal translocation.

[0113] RNA mediated gene fusion mechanism requires DNA:RNA hybrid formation. The mechanism central to at least certain embodiments of the disclosure is that the input chimeric RNA acts as a guide RNA to mediate genome rearrangement by annealing to *TMPRSS2* or *ERG* genes. Resolving such an RNA/DNA duplex by DNA break/repair mechanisms yield the final gene fusion through recombination in regions prone to DNA breaks. Accordingly, over-expression of RNaseH in cells, which degrades the RNA in an RNA/DNA duplex, should reduce the probability of fusion gene formation. To test whether RNA/DNA duplex is indeed required for RNA-mediated fusion gene, the inventors co-transfected input chimeric RNA expression plasmid together with a second plasmid that expresses wildtype RNaseH (Britton, Dernoncourt et al. 2014) which degrades the RNA in the RNA/DNA duplex. As a control, an inactive mutant RNaseH (D10R E48R mutant) (Britton, Dernoncourt *et al.* 2014) that lacks the ability to degrade RNA was used for head-to-head comparisons. As shown in FIG. 5A, induction of *TMPRSS2-ERG* fusion gene by antisense chimeric RNA was significantly reduced in the presence of wildtype RNaseH vs. the mutant RNaseH (FIG. 5A, lane 2 vs.1). Similarly, induction of *TMPRSS2-ETV1* fusion gene was also significantly reduced in the presence of wildtype vs. the mutant RNaseH (FIG. 5A, lane 6 vs. 5). These results indicate that the induction of gene fusions requires the formation of an RNA/DNA hybrid. Consistent with previous observations, sense input RNAs failed to induce fusion regardless of the expression of RNaseH (FIG. 5A, lane 3, 4, 7, 8).

[0114] One important observation emerging from our study is that all sense input RNAs failed to induce gene fusion. In particular, of the ten antisense input RNAs that were demonstrated to induce fusion (FIGS. 1A, 1B, 2E, 4A), all of their corresponding sense counterparts failed to induce fusion (FIGS. 1A, 1C, 2G, 4B). This specificity was observed despite the fact that sense input RNAs in theory can anneal to the same genomic sites targeted by their antisense counterparts and form similarly stable DNA/RNA hybrids. To test whether the

disparity between antisense and sense is due to transcriptional activity of parental genes, the input RNAs were expressed by U6 (a pol-III promoter) for one day, followed by α -amanitin-mediated inhibition of pol-II transcription for various time periods to shut down parental gene transcription. α -amanitin was then removed to resume cellular transcription and the induction by sense vs. antisense input RNA were compared. As shown in FIG. 5B upper panel, the corresponding sense input RNAs that previously failed to induce fusion began to induce *TMPRSS2-ERG* after 12 hours of α -amanitin treatment (lane 9 and 10), and *TMPRSS2-ETV1* fusion (lane 20) after 24 hours of α -amanitin treatment. This latent induction is not a property of general cellular toxicity of α -amanitin because the toxicity caused by the same treatment actually reduced the induction by antisense input RNAs (lane 1 vs. 5 for *TMPRSS2-ERG*, lane 11 vs. 15 for *TMPRSS2-ETV1*). Furthermore, the input RNAs designed to target *TMPRSS2* and *ERG*, regardless of their sense or antisense nature, induced *TMPRSS2-ERG* fusion but not that of *TMPRSS2-ETV1* (lane 1 to 10). Conversely, input RNAs targeting *TMPRSS2* and *ETV1*, regardless of their directional orientation, induced *TMPRSS2-ETV1* fusion but not *TMPRSS2-ERG* fusion (lane 11 to 20). Additional control experiments using a parental plasmid vector lacking the input RNA sequences, DHT treatment without plasmid transfection, and PCR reactions without cDNA, all induced no endogenous fusion transcript under the same α -amanitin treatment (FIG. 5B, lower panel). The specificity exhibited by these experiments argues against general toxicity effects. The results suggest that the antisense versus sense disparity is largely due to transcriptional conflict.

[0115] With the plausibility of RNA-mediated gene fusion established, it was considered that specific endogenous cellular RNAs can act as the ‘initiator’ to induce *TMPRSS2-ERG* fusion, which is found in ~50% of prostate cancers. To identify candidate cellular initiator RNAs, an available mRNA-seq database was analyzed consisting of prostate tumors and matched benign tissues (Kannan et al., 2011). However, there was no evidence of perfect endogenous antisense chimeric RNAs in which the *TMPRSS2* sequence was joined to any *ERG* sequence by discernable 5’ and 3’ splice sites in the antisense orientation. This suggests that if endogenous initiator RNAs do exist, they might arise from unrelated genomic sources that coincidentally resemble an imperfect chimeric RNA antisense to both *TMPRSS2* and *ERG*. Thermodynamic calculations of RNA/DNA hybrids are performed to identify cellular RNAs with partial sequence complementarity to the *TMPRSS2* and *ERG* genes. *AZII* mRNA (also known as *CEP131*) (Aoto et al., 1997; Aoto et al., 1995) could form high affinity RNA/DNA

hybrids with *TMPRSS2* and *ERG* genomic sequences. As shown in FIG. 6A, overexpressing full-length *AZII* mRNA (3619 nt, uc002jzn.1) induced the *TMPRSS2-ERG* fusion transcript in LNCaP cells. The induction was observed at a physiologically relevant concentration (40 nM) of DHT (FIG. 6A, lane 4). Furthermore, expression of exon16-17 of *AZII*, a short 220 nt segment containing an imperfect sequence antisense to *TMPRSS2* and *ERG*, was sufficient to induce *TMPRSS2-ERG* fusion (FIG. 6B, lane 2). This result indicates that the induction of gene fusion is mediated by an RNA sequence that resides in exon16-17 and requires no *AZII* protein. Consistent with previous observations that sense input RNAs are ineffective for the fusion process, the expression of exon16-17 in the antiparallel orientation also failed to induce *TMPRSS2-ERG* fusion (Fig. 6B, lane 3).

[0116] Significance of Certain Embodiments

[0117] In summary, this disclosure provides the first evidence that expression of a chimeric RNA can drive the formation of gene fusions in mammalian cells. The data support an embodiment (shown in FIG. 7) where the initiator RNA with chimeric sequence invades chromosomal DNA to stabilize a transient RNA/DNA duplex using DNA sequences located in two distant genes. Resolution of such an RNA/DNA duplex by DNA repair mechanisms would yield the final gene fusion through recombination in regions prone to DNA breaks (FIG. 7A). Such events were rare in the initial population of transfected cells (1 in 10^3 or 10^4 cells occurred within 3 days). However, the necessary machinery is clearly present in normal prostate epithelial cells prior to malignant transformation. If the resulting gene fusion (such as *TMPRSS2-ERG*) provides a growth advantage, a single affected cell among billions of cells in a normal prostate tissue may proliferate abnormally and eventually contribute to cancer formation. Identifying such initiator RNAs might provide novel insights into early disease mechanisms, as well as the discovery of new preventive and therapeutic strategies to combat diseases such as cancer.

[0118] Contrary to the previous “cart before the horse” model (Rowley and Blumenthal, 2008; Zaphiropoulos, 2011), the results do not support the postulation that a sense fusion mRNA derived from trans-splicing between two pre-mRNAs effectively directs gene fusion. Expressing sense input RNAs mirroring the trans-spliced mRNA failed to induce fusion in LNCaP cells (FIG. 1A and 9). Of ten antisense RNAs that were demonstrated to be capable of inducing fusion (FIGS. 1A, 1B, 2E, 4A, 6B), all of their corresponding sense RNAs failed to induce fusion (FIGS. 1A, 1C, 2G, 4B, 6B). This occurred even though the sense RNAs could, in theory, anneal

to the same genomic sites targeted by their antisense counterparts and form similar DNA/RNA hybrids when paired with the antisense strand of genomic DNA. As demonstrated in FIG. 5, this antisense versus sense disparity can be explained by transcriptional conflict (FIGS. 7B and 7C). Because the *TMPRSS2* promoter is highly active in LNCaP cells, sense chimeric RNAs forming DNA/RNA hybrids with antisense strands of genomic DNA (the template strand used for transcription) would be frequently “bumped” off and unable to stabilize the structures required for initiating genomic rearrangements. The mechanistic basis for the antisense versus sense RNA, and whether these phenomena can be generalized, may be characterized.

[0119] The results also do not support the hypotheses that antisense input RNAs, acting as a docking station, mediate trans-splicing by base-pairing with both endogenous sense parental pre-mRNAs, or by bringing the parental genes in close proximity thus facilitating trans-splicing of parental pre-mRNAs transcribed from two genomic loci. Both mechanisms would require the continuous presence of antisense input RNAs to sustain the expression of induced fusion transcripts. Yet it was shown that the induced fusion expression has a permanent nature and requires no continuous presence of input RNAs (FIG. 3B and FIG. 4D). Furthermore, in the case of *TMPRSS2-ERG* there is no detectable *ERG* parental RNA as raw material (FIG. 3A) to account for the trans-splicing models. Moreover, sense input RNAs, which are not complimentary to the sense parental pre-mRNAs thus cannot act as their docking station, are able to induce the fusion transcripts after a brief period of transcriptional inhibition, again arguing against the docking model. On the contrary, the genomic breakpoints identified by genomic PCR and chromosomal co-localization provided by FISH, strongly support that the induced expression of fusion transcript is largely the consequence of gene fusion resulting from chromosomal translocation. While prior works have shown that infrequent *TMPRSS2-ERG* fusions can be induced through genotoxic stress such as gamma radiation in the presence of DHT that increases double-stranded DNA breaks (Lin et al., 2009; Mani et al., 2009), such mechanisms of general genotoxicity fail to account for the specificity of gene fusion partners found in cancer. This report is the *first* to demonstrate RNA-mediated gene fusion in mammalian cells, and provides an RNA-driven mechanism that can account for the ‘specificity’ of gene fusion partners that were selected to undergo gene fusion in early disease stages. In specific embodiments the results represent a pathological example in a broad spectrum of potential RNA-mediated genome rearrangements and could have fundamental implications in the biology of

mammalian genome stability, as well as gene editing technology *via* mechanisms native to mammalian cells.

EXAMPLE 3

EXAMPLES OF MATERIALS AND METHODS

[0120] LNCaP cell culture

[0121] LNCaP cells were routinely cultured in RPMI 1640 medium (RPMI 1640, 1X, with L-glutamine, #10-040-CV, CORNING cellgro) containing 10% fetal bovine serum (premium grade FBS, #1500-500, Seradigm) and 1% penicillin/streptomycin (#15140-122, Gibco) in a 5% CO₂ humidified incubator. For experiments involving the induction of fusion gene by input RNA, regular fetal bovine serum in the culture medium was replaced by Charcoal:Dextran stripped fetal bovine serum (catalog #100-119, Gemini Bioproducts) to remove hormones present in serum. LNCaP cells were cultured in this special medium for 24 hrs prior to plasmid transfection.

[0122] PNT1A cell culture

[0123] PNT1A cells were routinely cultured in RPMI 1640 medium containing 10% fetal bovine serum (premium grade FBS, #1500-500, Seradigm) and 1% penicillin/streptomycin (#15140-122, Gibco) in a 5% CO₂ humidified incubator.

[0124] Transient transfection of plasmids for expressing the chimeric RNAs

[0125] Twenty hours prior to transfection, LNCaP cells were seeded in 12-wells plate (BioLite 12 Well Multidish, #130185, Thermo Fisher Scientific) with a density of 5×10^5 cells/well and 1 ml/well of culture medium containing Charcoal:Dextran stripped fetal bovine as described above. Transfection was performed using Turbofect transfection reagent (Thermo Scientific, #R0531) according to manufacturer's protocol. Briefly, 1 µg of a particular plasmid was first diluted in 100 µl of the serum-free DMEM followed by immediate mixing by pipetting. 4 µl of the transfection reagent was then added to the diluted DNA followed by mixing and incubation for 20 min. The DNA/transfection reagent mixture was then added drop wise to a well containing LNCaP cells in 1ml medium.

[0126] For transfection in PNT1A cells, 5×10^5 cells/well were plated in 12-wells plate in 1 ml/well of cultured medium 24 hrs prior to transfection. Transfection was performed using the same formula described for LNCaP cells. For repetitive transfections, initially transfected PNT1A cell population were split every three days, half was processed for RT-PCR assay and half was seeded again in a new well for the next transfection.

[0127] DHT preparation and treatment

[0128] DHT (Dihydrotestosterone) was purchased from Sigma Adrich (5α -Androstan- 17β -ol-3-one, #A8380). Concentrated stock of $1500\mu\text{M}$ was prepared by dissolving 4.3566 mg of DHT powder in 10 ml of 100% ethanol (200 proof ethanol, Koptec, #V1016) and then aliquoted in 1ml tubes and stored at -80°C .

[0129] For treating cultured cells, concentrated DHT stock was diluted as 10x working solutions (for example, for $0.9\mu\text{M}$ final concentration, 10x is prepared as $9.0\mu\text{M}$) with the appropriate complete culture medium and used immediately. Complete media for LNCaP cells: RPMI 1640 + 10% Charcoal:Dextran stripped fetal bovine serum + 1% penicillin/streptomycin. Complete media for PNT1A: RPMI 1640 + 10% fetal bovine serum + 1% penicillin/streptomycin. Six hrs post transfection, 111 μl of fresh 10x DHT working solutions was added to each well of 12-wells plate containing 1ml medium and transfected cells.

[0130] For long-term treatment, medium was changed with fresh DHT every three days.

[0131] RNA isolation

[0132] Total RNA from cultured cells was extracted using High Pure RNA isolation Kit according to manufacturer's instructions (#11828665001, Roche). Briefly, cells were suspended in 200 μl of PBS buffer and were then lysed with 400 μl of lysis buffer. The sample was then passed through the filter assembly resulting in the binding of the nucleic acids to the filter. The filter containing nucleic acids was then incubated with DNase I dissolved in DNase incubation buffer to degrade genomic and plasmid DNAs. The column was then rinsed with wash buffer and total RNA then eluted in a new tube for further analysis.

[0133] For detection of residual genomic and plasmid DNA, eluted RNA was subject to PCR reaction with primers specific to intron regions of house-keeping gene GAPDH, and with

primers specific to plasmid transfected. Total RNA was converted to cDNA only if it is validated as free of DNA contamination.

[0134] Reverse transcription reaction

[0135] 1 µg of total RNA was used for each reverse transcription reaction according to manufacturer instruction (superscript III RT, # 18080-051, Invitrogen). RNA was converted to cDNA either with Oligo dT primer (for induced fusion transcripts) or with random hexamers (for input RNAs expressed by U6 promoter). After the addition of dNTPs, the mixture was denatured at 65°C for 5 minutes. This was followed by the addition of a master-mix containing 1× superscript buffer, 10 mM DTT, 5 mM Magnesium chloride, RNaseOUT and Superscript III reverse transcriptase. Reactions were carried out at 50°C for 50 minutes and then terminated by incubation at 85°C for 5 minutes. cDNA was then treated with RNase-H for 20 minutes at 37°C to degrade RNA in DNA/RNA hybrid. 1 µl of cDNA was used as template for each subsequent PCR reaction.

[0136] RT-PCR for detecting induced fusion transcripts

[0137] The majority of induced fusion RNAs in this manuscript were detected using one-round RT-PCR. The following cases were assayed using three-round nested PCR: (1) the results of DHT treatment at physiological concentrations as shown in FIG. 1D and 6A, (2) the induction of *TMPRSS2-ERG* fusion transcript in non-malignant PNT1A cells as shown in FIG. 3F, (3) the specificity of input RNAs assayed in FIG. 4C, and (4) endogenous *ERG* level detection in LNCaP cells in FIGS. 17A-17D. The following cases were assayed using two-round nested PCR: the induction of *TMPRSS2-ETVI* fusion RNA as shown in FIGS. 4A and 4B, and the detection of *TMPRSS2-ETVI* fusion in the enriched population in FIG. 4D.

[0138] PCR was done with a standard three-step protocol using REDTaq DNA polymerase (#D5684-1KU, Sigma) according to manufacturer instruction.

[0139] Reaction was set as follows:

[0140] PCR reaction:

[0141] Forward primer : 1.0µl (from 10µM stock, Sigma)

[0142] Reverse primer : 1.0µl (from 10µM stock, Sigma)

- [0143] 10x reaction buffer: 5.0 μ l (comes with REDTaq, Sigma)
- [0144] dNTPs: 1.0 μ l (from 10mM stock, #11969064001, Roche)
- [0145] DMSO: 1.5 μ l (#154938, Sigma-Aldrich)
- [0146] cDNA: 1.0 μ l (from 20 μ l stock prepared from 1 μ g RNA)
- [0147] Autoclaved Milli-Q water: 38.5 μ l
- [0148] REDTaq DNA polymerase: 1.0 μ l (#D5684-1KU, Sigma)
- [0149] Total volume: 50 μ l

[0150] Standard one-round PCR conditions for *TMPRSS2-ERG*:

- [0151] Pre-denaturation 94⁰C, 4 min
- [0152] Denaturation 94⁰C, 30 sec
- [0153] Annealing 58⁰C, 45 sec
- [0154] Extension 72⁰C, 60 sec
- [0155] Final Extension 72⁰C, 5 min

[0156] The Denaturation, Annealing and Extension steps were performed 32 cycles for induced fusion RNA and 27 cycles for input RNA.

[0157] PCR conditions for three-round nested PCR for *TMPRSS2-ERG*:

[0158] **1st round** : PCR with *TMPRSS2* ex-1 F1 and *ERG* ex-4 R1 on 1 μ l of cDNA

- [0159] Pre-denaturation 94⁰C, 4 min
- [0160] Denaturation 94⁰C, 30 sec
- [0161] Annealing 58⁰C, 45 sec
- [0162] Extension 72⁰C, 60 sec

[0163] Final Extension 72⁰C, 5 min

[0164] The Denaturation, Annealing and Extension steps were performed 32 cycles for induced fusion RNA.

[0165] **2nd round** : PCR with *TMPRSS2* ex-1 F2 and *ERG* ex-4 R2 on 1μl of 1st round product, PCR conditions same as 1st round.

[0166] **3rd round** : PCR with *TMPRSS2* ex-1 F3 and *ERG* ex-4 R3 on 1μl of 2nd round product, PCR conditions same as 1st round.

[0167] PCR conditions for two-round nested PCR for *TMPRSS2-ETVI*:

[0168] **1st Round**: Top down PCR with *TMPRSS2* ex-1 F1 and *ETVI* ex-6 R1

[0169] Denaturation 94⁰C, 30 sec

[0170] Annealing * , 45 sec

[0171] Extension 72⁰C, 60 sec

[0172] The Denaturation, Annealing and Extension steps were performed 2 cycles.

[0173] *2 cycles at each temperature: 62⁰C, 61⁰C, 60⁰C.....49⁰C followed by

[0174] Denaturation 94⁰C, 30 sec

[0175] Annealing 48 ⁰C, 45 sec

[0176] Extension 72⁰C, 60 sec

[0177] The Denaturation, Annealing and Extension steps were performed 20 cycles

[0178] **2nd Round**: PCR with *TMPRSS2* ex-1 F2 and *ETVI* ex-5 R1 on 1μl of 1st round.

[0179] Pre-denaturation 94⁰C, 4 min

[0180] Denaturation 94⁰C, 30 sec

[0181] Annealing 57⁰C, 45 sec

[0182] Extension 72⁰C, 60 sec

[0183] Final Extension 72⁰C, 5 min

[0184] The Denaturation, Annealing and Extension steps were performed 32 cycles for Induced fusion RNA.

[0185] **3rd round** (for FIG. 4D): PCR with *TMPRSS2* ex-1 F3 and *ETV1* ex-5 R2 on 1μl of 2nd round, PCR conditions same as 2nd round.

[0186] Long range PCR for detecting genomic DNA fusion junction

[0187] Nested long-range PCRs according to the manufacturer's protocols using LA PCR kit (Takara, # RR002M). 200 ng of genomic DNA was used in each reaction and PCR was performed with annealing and extension at 68°C for 20 minutes. 1μl from the above reaction (1st round PCR) was used as template for the 2nd round PCR.

[0188] For the genomic breakpoint identified in this manuscript, 1st round long range PCR was done using primers *TMPRSS2* genomic bk-F1 and *ERG* genomic bk-R1 shown in primer list below. 1μl from the above reaction (1st round PCR) was used as template for the 2nd round PCR using inner primers *TMPRSS2* genomic bk-F2 and *ERG* genomic bk-R2.

[0189] Cloning and Sanger sequencing of induced fusion transcripts

[0190] PCR amplified cDNA bands were excised from the gel and eluted using QIAquick Gel Extraction Kit (#28706, Qiagen). The eluted bands were then cloned to pGEM-T vector (pGEM-T vector system I, # A3600) following manufacturer instruction. Sanger sequencing was performed using the service of Beckman Coulter Genomics.

[0191] Tm calculations

[0192] Melting temperature (T_m) of putative genomic DNA stems were calculated using the following formula (Rychlik and Rhoads, 1989):

[0193] $T_m (^{\circ}\text{C}) = 4 \times (\text{number of G's and C's}) + 2 \times (\text{number of A's and T's})$

[0194] A high energy G·T and A·C wobble pair known to have Watson-Crick like geometry in DNA double helix (Kimsey and Al-Hashimi, 2014; Watson and Crick, 1953) are considered as having the same stability as an A·T pair.

[0195] Fluorescent in situ hybridization (FISH)

[0196] Enriched population carrying *TMPPRSS2-ETVI* fusion events were first grown on 18 mm round #1 coverglass in a 12-well cell culture plate at the initial density of 200-400k/well. Cells were then fixed with 4% (vol./vol.) formaldehyde followed by denaturation of DNA with 0.1 N HCl for 5 min and with 70% formamide at 85°C for 7 min. Hybridization of target DNA with probes were done at 37°C for 16hr in a humidified chamber. Cells were then washed, stained with DAPI and imaged with microscope. FISH probes for *TMPPRSS2* (RP11-35C4, red) and *ETVI* (RP11-769K2, green) were purchased from Empire Genomics.

[0197] α -Amanitin assay

[0198] Twenty hours prior to transfection, LNCaP cells were seeded in 12-wells plate (BioLite 12 Well Multidish, #130185, Thermo Fisher Scientific) with a density of 5×10^5 cells/well and transfection was performed using Turbofect transfection reagent (Thermo Scientific, #R0531) as described earlier. DHT was added at the final concentration of $0.9 \mu\text{M}$ six hours post transfection. Following overnight incubation, cells were then treated with $4 \mu\text{g/ml}$ α -amanitin for various time periods (0, 2, 6, 12 and 24 hours). Cells were then revived in fresh medium containing $0.9 \mu\text{M}$ DHT without α -amanitin and RT-PCR was performed for either *TMPRSS2-ERG* or *TMPRSS2-ETV1* fusion.

[0199] Examples of Input RNA sequences

[0200] Sense-1

[0201]

**+1tcagatcgccctggagacgccatccacgctgttttgacctccatagaagacaccgggaccgatccagcctcccctcgaagct
gatactgagaacttcaggctcctgggcaacgtgctggtctgtgtgctggcccatcactttggcaagaattcGAGTAGGCGCGA
GCTAAGCAGGAGGCGGAGGCGGAGGCGGAGGGCGAGGGGCGGGGAGCGCCG
CCTGGAGCGCGGCAGGAAGCCTTATCAGTTGTGAGTGAGGACCAGTCGTTGTTTGA
GTGTGCCTACGGAACGCCACACCTGGCTAAGACAGAGATGACCGCGTCCCTCCTCCA
GCGACTATGGACAGACTTCCAAGATGAGCCCACGCGTCCCTCAGCAGGATTGGCTG**

TCTCAACCCCCAGCCAGGGTCACCATCAAAATGGAATGTAACCCTAGCCAGGTGAA
TGGCTCAAGaagcttatcgataccgtcgacctcgagggccagatctaattcaccaccagtcaggctgcctatcagaaagtgg
 tggctggtgtggctaatagccttgccccacaagtatcactaagctcgctttcttctgtgtccaatttctattaaaggttcctttgtccctaagtccaa
 ctactaaactgggggatattatgaagggccttccggagcatctggattctgcctaataaaaaacatttatttcattgcaaaaaa.....a
 (SEQ ID NO:1)

[0202] In **BOLD CAPS**: *TMPRSS2* double underline CAPS: *ERG* **+1**: Transcription
 start aataaa: poly(A) signal aaaaaa.....a: poly(A) tail gaattc: EcoRI
aagctt: HindIII

[0203] Sense-2

[0204]

+1tgtgctcgcttcggcagcacatataactgaacgatcctgcag**GAGTAGGCGCGAGCTAAGCA**
GGAGGCGGAGGCGGAGGCGGAGGGCGAGGGGCGGGGAGCGCCGCCTGGAGC
GCGGCAGGAAGCCTTATCAGTTGTGAGTGAGGACCAGTCGTTGTTTGAGTGTGCCT
ACGGAACGCCACACCTGGCTAAGACAGAGATGACCGCGTCCTCCTCCAGCGACTAT
GGACAGACTTCCAAGATGAGCCACGCGTCCCTCAGCAGGATTGGCTGTCTCAACCC
CCAGCCAGGGTCACCATCAAAATGGAATGTAACCCTAGCCAGGTGAATGGCTCAAG
 (SEQ ID NO:2)

[0205] In **BOLD CAPS**: *TMPRSS2* double underline CAPS: *ERG* **+1**:
 Transcription start ctgcag: PstI

[0206] Sense-2 long

[0207]

+1tgtgctcgcttcggcagcacatataactgaacgatcctgcag**GAGTAGGCGCGAGCTAAGCA**
GGAGGCGGAGGCGGAGGCGGAGGGCGAGGGGCGGGGAGCGCCGCCTGGAGC
GCGGCAGGAAGCCTTATCAGTTGTGAGTGAGGACCAGTCGTTGTTTGAGTGTGCCT
ACGGAACGCCACACCTGGCTAAGACAGAGATGACCGCGTCCTCCTCCAGCGACTAT
GGACAGACTTCCAAGATGAGCCACGCGTCCCTCAGCAGGATTGGCTGTCTCAACCC
CCAGCCAGGGTCACCATCAAAATGGAATGTAACCCTAGCCAGGTGAATGGCTCAAG
GAAGTCTCCTGATGAATGCAGTGTGGCCAAAGGCGGGAAGATGGTGGGCAGCCCAG
ACACCGTTGGGATGAACTACGGCAGCTACATGGAGGAGAAGCACATGCCACCCCCA
AACATGACCACGAACGAGCGCAGAGTTATCGTGCCAGCAGATCCTACGCTATGGAG

TACAGACCATGTGCGGCAGTGGCTGGAGTGGGCGGTGAAAGAATATGGCCTTCCAG
ACGTCAACATCTTGTTATTCCAGAACATCGATGGGAAGGAACTGTGCAAGATGACC
AAGGACGACTTCCAGAGGCTCACCCCCAGCTACAACGCCGACATCCTTCTCTCACAT
CTCCACTACCTCAGAGAGACTCCTCTTCCACATTTGACTTCAGATGATGTTGATAAA
GCCTTA (SEQ ID NO:3)

[0208] **BOLD CAPS:** *TMPRSS2* double underline CAPS: *ERG* +1: Transcription
 start ctgcag: PstI

[0209] Antisense-1

[0210]

+1tcagatgcctggagacgccatccacgtgtttgacctcatagaagacaccgggaccgatccagcctccccctgaagct
gacctgagaactcaggctcctgggcaacgtgctggtctgtgtgctggccatcactttggcaaagaattcCTTGAGCCATTCA
CCTGGCTAGGGTTACATTCCATTTTGGATGGTGACCCTGGCTGGGGGTTGAGACAGCC
AATCCTGCTGAGGGACGCGTGGGCTCATCTTGGAAGTCTGTCCATAGTCGCTGGAGG
AGGACGCGGTCATCTCTGTCTTAGCCAGGTGTGGCGTTCCGTAGGCACACTCAAACA
ACGACTGGTCCTCACTCACAACTGATAAGGCTTCCTGCCGCGCTCCAGGCGGCGC
TCCCCGCCCTCGCCCTCCGCCTCCGCCTCCGCCTCCTGCTTAGCTCGCGCCTA
CTCaagcttatcgataccgtcgacctcgagggccagatctaattcacccaccagtgaggtgcctatcagaaagtgggtggtggtg
 ggctaatgcctggccacaagtatacctaagctcgcttcttctgtccaatttctattaaaggttcctttgtccctaagtccaactactaaact
 ggggatattatgaagggcctccggagcatctggattctgcctaataaaaaaacatttattttcattgcaaaaaa.....a (SEQ ID
 NO:4)

[0211] **BOLD CAPS :** *TMPRSS2* double underline CAPS : *ERG* +1: Transcription
 start aataaa: poly(A) signal aaaaaa.....a: poly(A) tail gaattc: EcoRI aagctt:
 HindIII

[0212] Antisense-2

[0213]

+1tgtgtcgtctcggcagcacatataactaacattggaacgatcctgcagCTTGAGCCATTACCTGGCTAG
GGTTACATTCCATATTGATGGTGACCCTGGCTGGGGGTTGAGACAGCCAATCCTGCT
GAGGGACGCGTGGGCTCATCTTGGAAGTCTGTCCATAGTCGCTGGAGGAGGACGCG
GTCATCTCTGTCTTAGCCAGGTGTGGCGTTCCGTAGGCACACTCAAACAACGACTGG
TCCTCACTCACAACTGATAAGGCTTCCTGCCGCGCTCCAGGCGGCGCTCCCCGCC

CCTCGCCCTCCGCCTCCGCCTCCGCCTCCTGCTTAGCTCGCGCCTACTC (SEQ ID NO:5)

[0214] **BOLD CAPS** : *TMPRSS2* double underline CAPS : *ERG* **+1**:

Transcription start

[0215] **BOLD CAP** and double underline A: T to A change to inactivate the cryptic transcription termination by U6 promoter ctgcag: PstI

[0216] Antisense-3

[0217]

+1tggtctgcttcggcagcacatataactaacattggaacgatcctgcagCTTGAGCCATTACCTGGCTAG
GGTTACATTCCATATTGATGGTGACCCTGGCTGGGGGTTGAGACAGCCAATCCCGG
ACCCCGAGCCGGGACCCTGGTACCGGCGCCGCTCACCTGCCGCGCTCCAGGGCG
GCGCTCCCCGCCCCCTCGCCCTCCGCCagacaggagtgagagatggaagctcgcgctactc (SEQ ID NO:6)

[0218] **BOLD CAPS** : *TMPRSS2* double underline CAPS: *ERG* G : Mutated region of *TMPRSS2* **+1**: Transcription start A: T to A change to inactivate the cryptic transcription termination by U6 promoter ctgcag: PstI

[0219] Antisense-4

[0220]

+1tggtctgcttcggcagcacatataactaacattggaacgatcctgcagCTTGAGCCATTACCTGGCTAG
GGTTACATTCCATATTGATGGTGACCCTGGCTGGGGGTTGAGACAGCCAATCCGACC
CTGGTACCGGCGCCGCTCACCTGCCGCGCTCCAGGCGGCGCTCCCCGCCCCCTC
GCCCTCCGCCagacaggagtgagagatggaagctcgcgctactc (SEQ ID NO:7)

[0221] **BOLD CAPS**: *TMPRSS2* double underline CAPS: *ERG* G: Mutated region of *TMPRSS2* **+1**: Transcription start **BOLD CAP** and double underline A: T to A change to inactivate the cryptic transcription termination by U6 promoter ctgcag: PstI

[0222] Antisense-5

[0223]

+1tggtctcgcttcggcagcacatataactaacattggaacgatcctgcagCTTGAGCCATTACCTGGCTAG
 GGTTACATTCCATATTGATGGTGACCCTGGCTGGGGGTTGAGACAGCCAATCCGCC
 GCTCACCTGCCGCGCTCCAGGCGGCGCTCCCCGCCCCCTCGCCCTCCGCCagacag
 gagtgagagatggaagctcgcgcctactc (SEQ ID NO:8)

[0224] **BOLD CAPS:** *TMPRSS2* double underline CAPS: *ERG* lower case bold:
 Mutated region of *TMPRSS2* +1: Transcription start BOLD CAP and double underline A: T
 to A change to inactivate the cryptic transcription termination by U6 promoter

[0225] ctgcag: PstI

[0226] Antisense-6

[0227]

+1tggtctcgcttcggcagcacatataactaacattggaacgatcctgcagCTTGAGCCATTACCTGGCTAG
 GGTTACATTCCATATTGATGGTGACCCTGGCTGGGGGTTGAGACAGCCAATCCCCAG
 GCGGCGCTCCCCGCCCCCTCGCCCTCCGCCagacaggagtgagagatggaagctcgcgcctactc
 (SEQ ID NO:9)

[0228] **BOLD CAPS:** *TMPRSS2* double underline CAPS: *ERG* lower case bold:
 Mutated region of *TMPRSS2* +1: Transcription start BOLD CAP and double underline A: T
 to A change to inactivate the cryptic transcription termination by U6 promoter ctgcag:
 PstI

[0229] Antisense-7

[0230]

+1tggtctcgcttcggcagcacatataactaacattggaacgatcctgcagCTTGAGCCATTACCTGGCTAG
 GGTTACATTCCATATTGATGGTGACCCTGGCTGGGGGTTGAGACAGCCAATCCCGG
 ACCCGAGCCGGGACCCTGGTACCGGCGCCGCTCACCTGCCGCGCTCCAagacag
 gagtgagagatggaagctcgcgcctactc (SEQ ID NO:10)

[0231] **BOLD CAPS:** *TMPRSS2* double underline CAPS: *ERG* lower case bold:
 Mutated region of *TMPRSS2* +1: Transcription start BOLD CAP and double underline A: T

to A change to inactivate the cryptic transcription termination by U6 promoter ctgcag:
PstI

[0232] Antisense-8

[0233]

+1gtgctcgcttcggcagcacatataactaacattggaacgatactgcagCTTGAGCCATTACCTGGCTAG
GGTTACATTCCATATTGATGGTGACCCTGGCTGGGGGTTGAGACAGCCAATCCCCAG
GTTCCCCTCCCCAGCCCCGACCCCCGAGCCGGGACCCTGGTACCGGCGCagacagg
agtgagagatggaagctcgcgcctactc (SEQ ID NO:11)

[0234] BOLD CAPS: *TMPRSS2* double underline CAPS: *ERG* lower case bold:
Mutated region of *TMPRSS2* +1: Transcription start BOLD CAP and double underline A: T
to A change to inactivate the cryptic transcription termination by U6 promoter ctgcag:
PstI

[0235] Antisense-9

[0236]

+1gtgctcgcttcggcagcacatataactaacattggaacgatactgcagGAGACAGCCAATCCTGCTGAG
GGACGCGTGGGCTCATCTTGGAAGTCTGTCCATAGTCGCTGGAGGAGGACGCGGGC
CGCTCACCTGCCGCGCTCCAGGCGGCGCTCCCCGCCCCCTCGCCCTCCGCCagaca
ggagtgagagatggaagctcgcgcctactc (SEQ ID NO:12)

[0237] BOLD CAPS: *TMPRSS2* double underline CAPS: *ERG* lower case bold:
Mutated region of *TMPRSS2* +1: Transcription start ctgcag: PstI

[0238] Antisense-5A

[0239]

+1gtgctcgcttcggcagcacatataactaacattggaacgatactgcagGCCGCTCACCTGCCGCGCTC
CAGGCGGCGCTCCCCGCCCCCTCGCCCTCCGCCagacaggagtgagagatggaagctcgcgcctactc
(SEQ ID NO:13)

[0240] BOLD CAPS: *TMPRSS2* G: Mutated region of *TMPRSS2* +1: Transcription
start ctgcag: PstI

[0241] Antisense-5B**[0242]**

+1tggtctgcttcggcagcacatataactaacattggaacgatcctgcagCTTGAGCCATTCACCTGGCTAG
GGTTACATTCCATATTGATGGTGACCCTGGCTGGGGGTTGAGACAGCCAATCC (SEQ
ID NO:14)

[0243] double underline CAPS: *ERG* **+1**: Transcription start ctgcag: PstI
BOLD CAP and double underline **A**: T to A change to inactivate the cryptic transcription
termination by U6 promoter

[0244] Antisense-B1**[0245]**

+1tggtctgcttcggcagcacatataactaacattggaacgatcctgcagTGGCCTGAGCCTTGAAGAATG
GGGTGTACTGGGTAAATCAAAATGGTGGGGAGCATTTCAGATGGAGAAACTGCCT
CACCTGCCGCGCTCCAGGCGGCGCTCCCCGCCCTCGCCCTCCGCTCCG (SEQ
ID NO:15)

[0246] BOLD CAPS: *TPRSS2* double underline CAPS: *ERG* **+1**: Transcription
start ctgcag: PstI

[0247] Antisense-B2**[0248]**

+1tggtctgcttcggcagcacatataactaacattggaacgatcctgcagTGGCCTGAGCCTTGAAGAATG
GGGTGTACTGGGTAAATCAAAATGGTGGGGAGCATTTCAGATGGAGAAACTGCC
CTCCCCAGCCCGGACCCCGAGCCGGGACCCTGGTACCGGCGCCGCTCACC
(SEQ ID NO:16)

[0249] BOLD CAPS: *TPRSS2* double underline CAPS: *ERG* **+1**: Transcription
start ctgcag: PstI

[0250] Antisense-B3**[0251]**

+1tggtctgcttcggcagcacatataactaacattggaacgatcctgcagTGGGGAGCATTTCAGATGGA

GAAACTGCAAGGAAAAGCATAGAAGTGGGGCCACCCCTCGTGAGCTGGGGAGGGC
TCACCTGCCGCGCTCCAGGCGGCGCTCCCCGCCCTCGCCCTCCGCCTCCG
 (SEQ ID NO:17)

[0252] **BOLD CAPS:** *TMPRSS2* double underline CAPS: *ERG* **+1**: Transcription
 start ctgcag: PstI

[0253] Antisense-C1

[0254]

+1tggtctcgcttcggcagcacatataactaacattggaacgatcctgcagTGAGCTCATGCTATTCCTATGA
CATAGATGAGCACTGGGTAGACCCCGTCCTGGTAACACTATTCATGCACTAACCCCA
GGCGGGGGCCGTGGAGGGCAGGCGGACTAGGAGCCAGCTTTGGGGACC (SEQ
 ID NO:18)

[0255] **BOLD CAPS:** *TMPRSS2* double underline CAPS: *ERG* **+1**: Transcription
 start BOLD CAP and double underline A: T to A change to inactivate the cryptic
 transcription termination by U6 promoter ctgcag: PstI

[0256] Antisense-C2

[0257]

+1tggtctcgcttcggcagcacatataactaacattggaacgatcctgcagTGAGCTCATGCTATTCCTATGA
CATAGATGAGCACTGGGTAGACCCCGTCCTGGTAACACTATTCATGCACTAACCCAG
CACTCCCAGTCCTCCTCCCCAAAGAGAAAAGGCGCACCGGTGCTCCCAG (SEQ
 ID NO:19)

[0258] **BOLD CAPS:** *TMPRSS2* double underline CAPS: *ERG* **+1**: Transcription
 start BOLD CAP and double underline A: T to A change to inactivate the cryptic
 transcription termination by U6 promoter ctgcag: PstI

[0259] Antisense-C3

[0260]

+1tggtctcgcttcggcagcacatataactaacattggaacgatcctgcagGTCCTGGTAACACTATTCATGC
ACTAACAAGTTGGTTGCCAGTGAGACTTGATTATTATGACTCTGGGAGTGCTGCCCA

GGCGGGGGCCGTGGAGGGCAGGCGGACTAGGAGCCAGCTTTGGGGACC (SEQ ID NO:20)

[0261] BOLD CAPS: *TMPRSS2* double underline CAPS: *ERG* **+1**: Transcription start
BOLD CAP and double underline **A**: T to A change to inactivate the cryptic transcription termination by U6 promoter ctgcag: PstI

[0262] Antisense-D1

[0263]

+1tggtctcgttcggcagcacatataactaacattggaacgatcctgcagGAGAGACAGAGAGAGAGAGG
CTGGTAGAGGGAAGAGACAGAAAGAAAGATGAAGGGATAAGTGTCCAGAATCCCTG
AGCGCTCGACCCTCGGGCGCACTCACCTGCCGCGCCGCGCTCCTCACACCCG
(SEQ ID NO:21)

[0264] BOLD CAPS: *TMPRSS2* double underline CAPS: *ERG* **+1**: Transcription start
ctgcag: PstI

[0265] Antisense-D2

[0266]

+1tggtctcgttcggcagcacatataactaacattggaacgatcctgcagGAGAGACAGAGAGAGAGAGG
CTGGTAGAGGGAAGAGACAGAAAGAAAGATGAAGGGATAAGTGTCCAGAATCCCTG
CCCAGCACTCTCCCAGCACCCCGGGAGGCGCCCTGCCCGGCTGGCCCCAGCG
(SEQ ID NO:22)

[0267] BOLD CAPS: *TMPRSS2* double underline CAPS: *ERG* **+1**: Transcription start
ctgcag: PstI

[0268] Antisense-D3

[0269]

+1tggtctcgttcggcagcacatataactaacattggaacgatcctgcagTGAAGGGATAAGTGTCCAGAA
TCCCTGGATCTGGGATGGAATAAAGGATCTGGATGGTAAACGGAGAGTGCTGGGAG
CGCTCGACCCTCGGGCGCACTCACCTGCCGCGCCGCGCTCCTCACACCCG (SEQ ID NO:23)

[0270] **BOLD CAPS:** *TMPRSS2* double underline CAPS: *ERG* **+1**: Transcription start ctgcag: PstI

[0271] Antisense-E1

[0272]

+1tggtctgcttcggcagcacatataactaacattggaacgatactgcagAGGGTATTCAGTATTACTATTTGGCTTAGATAAGCTGGTAGTTACTTGCTAAAATTAATCTTTATTATAAAGCAGAAAGCCGCATTCTGACATCACTCTCCATGGACAAAGATTCTTCGCTTGATCA (SEQ ID NO:24)

[0273] **BOLD CAPS:** *TMPRSS2* double underline CAPS: *ERG* **+1**: Transcription start BOLD CAP and double underline A: T to A change to inactivate the cryptic transcription termination by U6 promoter ctgcag: PstI

[0274] Antisense-F1

[0275]

+1tggtctgcttcggcagcacatataactaacattggaacgatactgcagAATGGATGAATACATAAAATAAATTGTGGTGGATATATACAACGGAATATCATTTAGCCTTTATTATTATTGAGAACTGTGCCGAGCCGGGCAGGACAGGATGAGGTGGACCGAAGCGCCCAGGTGC (SEQ ID NO:25)

[0276] **BOLD CAPS:** *TMPRSS2* double underline CAPS: *ERG* **+1**: Transcription start BOLD CAP and double underline A: T to A change to inactivate the cryptic transcription termination by U6 promoter ctgcag: PstI

[0277] Antisense-G1

[0278]

+1tggtctgcttcggcagcacatataactaacattggaacgatactgcagTTTATTCATGAGAAAATTCTGGCACAATGGAAAACCCTGGCAAGCAAAAGATAGGGGCAGCAGATGTCCTGGCCTACAAGAAGTCCAAGCCCATCGTCCCTAGAAAGCATGGTCTCCCATGACCCC (SEQ ID NO:26)

[0279] **BOLD CAPS:** *TMPRSS2* double underline CAPS: *ERG* **+1**: Transcription start ctgcag: PstI

[0280] Sense-3**[0281]**

+1tggtctgcttcggcagcacatataactaacattggaacgatcctgcaggagtaggcgcgagcttccatctctcactcctgtct
GGCGGAGGGCGAGGGGCGGGGAGCGCCGCCTGGAGCGCGGCAGGTGAGCGG
CGCCGGTACCAGGGTCCCGGCTCGGGGTCCGGGATTGGCTGTCTCAACCCCCAG
CCAGGGTCACCATCAAAATGGAATGTAACCCTAGCCAGGTGAATGGCTCAAG (SEQ
 ID NO:27)

[0282] BOLD CAPS: *TMPRSS2* double underline CAPS: *ERG* Bold lower case:
 Mutated region of *TMPRSS2* **+1**: Transcription start ctgcag: PstI

[0283] Sense-4**[0284]**

+1tggtctgcttcggcagcacatataactaacattggaacgatcctgcaggagtaggcgcgagcttccatctctcactcctgtct
GGCGGAGGGCGAGGGGCGGGGAGCGCCGCCTGGAGCGCGGCAGGTGAGCGG
CGCCGGTACCAGGGTCCGGATTGGCTGTCTCAACCCCCAGCCAGGGTCACCATCAA
TATGGAATGTAACCCTAGCCAGGTGAATGGCTCAAG (SEQ ID NO:28)

[0285] BOLD CAPS: *TMPRSS2* double underline CAPS: *ERG* Bold lower case:
 Mutated region of *TMPRSS2* **+1**: Transcription start ctgcag: PstI

[0286] Sense-5**[0287]**

+1tggtctgcttcggcagcacatataactaacattggaacgatcctgcaggagtaggcgcgagcttccatctctcactcctgtct
GGCGGAGGGCGAGGGGCGGGGAGCGCCGCCTGGAGCGCGGCAGGTGAGCGG
CGGATTGGCTGTCTCAACCCCCAGCCAGGGTCACCATCAAAATGGAATGTAACCCT
AGCCAGGTGAATGGCTCAAG (SEQ ID NO:29)

[0288] BOLD CAPS: *TMPRSS2* double underline CAPS: *ERG* Bold lower case:
 Mutated region of *TMPRSS2* **+1**: Transcription start ctgcag: PstI

[0289] Sense-6

[0290]

+1tgctgcgttcggcagcacatataactaacattggaacgatcctgcaggagtaggcgcgagcttccatctctcactcctgtct
GGCGGAGGGCGAGGGGCGGGGAGCGCCGCCTGGGGATTGGCTGTCTCAACCCC
CAGCCAGGGTCACCATCAAAATGGAATGTAACCCTAGCCAGGTGAATGGCTCAAG
 (SEQ ID NO:30)

[0291] **BOLD CAPS:** *TPRSS2* double underline CAPS: *ERG* bold lower case:
 Mutated region of *TPRSS2* +1: Transcription start ctgcag: PstI

[0292] Sense-B1

[0293]

+1tgctgcgttcggcagcacatataactaacattggaacgatcctgcag**CGGAGGCGGAGGGGCGAGGG**
GCGGGGAGCGCCGCCTGGAGCGCGGCAGGTGAGGCAGTTTCTCCATCTGGAAAT
GCTCCCCACCATATTGATTTACCCAGTACACCCCATTTCTTCAAGGCTCAGGCCA (SEQ
 ID NO:31)

[0294] **BOLD CAPS:** *TPRSS2* double underline CAPS: *ERG* +1: Transcription
 start BOLD CAP and double underline A: T to A change to inactivate the cryptic
 transcription termination by U6 promoter ctgcag: PstI

[0295] Sense-C1

[0296]

+1tgctgcgttcggcagcacatataactaacattggaacgatcctgcag**GGTCCCCAAAGCTGGCTCCT**
AGTCCGCCTGCCCTCCACGGCCCCCGCCTGGGGTTAGTGCATGAAGAGTGTTAC
CAGGACGGGGTCTACCCAGTGCTCATCTATGTCATAGGAAGAGCATGAGCTCA (SEQ
 ID NO:32)

[0297] **BOLD CAPS:** *TPRSS2* double underline CAPS: *ERG* +1: Transcription
 start ctgcag: PstI

[0298] Sense-D1

[0299]

+1tgctgcgttcggcagcacatataactaacattggaacgatcctgcag**CGGGTGTGAGGAGCGCGGC**
GCGGCAGGTGAGTGCGCCCGAGGGTCGAGCGCTCAGGGATTCTGGACACTTATC

CCTTCATCTTTCTTCTGTCTCTTCCCTCTACCAGCCTCTCTCTCTGTCTCTC (SEQ ID NO:33)

[0300] **BOLD CAPS:** *TMPRSS2* double underline CAPS: *ERG* **+1**: Transcription start promoter ctgcag: PstI

[0301] Antisense-*TMPRSS2-ETV1-A1*

[0302]

+1tggtctcgcttcggcagcacatataactaacattggaacgatcctgcagTGCCGCATTATGTAAATCGTTC
CAAGTTAAAGTCTTAGTTAGATTCACTAGACTAGGAGCCAGCTTTGGGGACCCCG
GGGGACTCTCTTCCACCAACTGG (SEQ ID NO:34)

[0303] **BOLD CAPS:** *TMPRSS2* double underline CAPS: *ETV1* **+1**: Transcription start ctgcag: PstI

[0304] Antisense-*TMPRSS2-ETV1-B1*

[0305]

+1tggtctcgcttcggcagcacatataactaacattggaacgatcctgcagGATTCAGTAATTTCAAAGTTTA
TTATATTTAAGATAAGACTGAAGTGCTCAACAACTTAGTCTCACTTTAGGTATTC
CAAATGCCTTGTAAGTGGGCTG (SEQ ID NO:35)

[0306] **BOLD CAPS:** *TMPRSS2* double underline CAPS: *ETV1* **+1**: Transcription start BOLD CAP and double underline A: T to A change to inactivate the cryptic transcription termination by U6 promoter ctgcag: PstI

[0307] Antisense-*TMPRSS2-ETV1-C1*

[0308]

+1tggtctcgcttcggcagcacatataactaacattggaacgatcctgcagCTCTGAATAGAAAAATAGAAG
TCCATAGTATCAACTCTAATATTCATATTTGGCTGCATCCCCACTTCCTGGAGTAC
CTTCCCAGATCTCCTGGGACAGG (SEQ ID NO:36)

[0309] **BOLD CAPS:** *TMPRSS2* double underline CAPS: *ETV1* **+1**: Transcription start BOLD CAP and double underline A: T to A change to inactivate the cryptic transcription termination by U6 promoter ctgcag: PstI

[0310] Antisense-TMPRSS2-ETV1-D1**[0311]**

+1tggtcgtcgttcggcagcacatataactaacattggaacgatcctgcagATATAAACAAAAAGTGTCAGC
ATTGTCTCAACTTCATTCTATTCAATGTAAGGCCCTTTGCGCTGGTAAACTCTCC
 CTGCCACACTCCCAACCCCCATC (SEQ ID NO:37)

[0312] BOLD CAPS: *TMPRSS2* double underline CAPS: *ETV1* **+1:** Transcription
 start ctgcag: PstI

[0313] Antisense-TMPRSS2-ETV1-E1**[0314]**

+1tggtcgtcgttcggcagcacatataactaacattggaacgatcctgcagCTCTGAATAGAAAAATAGAAG
TCCATAGTATCAACTCTAATATTCATATTCTTCAGCAACCAAAACTGAACAAGCAC
 TCCATTGACCATTACCTTTCCT (SEQ ID NO:38)

[0315] BOLD CAPS: *TMPRSS2* double underline CAPS: *ETV1* **+1:** Transcription
 start BOLD CAP and double underline A: T to A change to inactivate the cryptic transcription
 termination by U6 promoter ctgcag: PstI

[0316] Antisense-TMPRSS2-ETV1-F1**[0317]**

+1tggtcgtcgttcggcagcacatataactaacattggaacgatcctgcagAGGAGGCCTTAAGTATGATTC
AGTGAACACATTATGGTCGATAAACAAGGTGGGCCCCATTCTCAGAGTCTGATGT
 AATAATTGGGACCAAGGCAATGA (SEQ ID NO:39)

[0318] BOLD CAPS: *TMPRSS2* double underline CAPS: *ETV1* **+1:** Transcription
 start ctgcag: PstI

[0319] Antisense-TMPRSS2-ETV1-G1**[0320]**

+1tggtcgtcgttcggcagcacatataactaacattggaacgatcctgcagTGTAATTGACTTAGATCTTGAA
AGAGTTCTAAAAAACAAGTCAAAGACATCTAGAAGAATCTCTAGATGAAGGTTAC
 CTACAACAAAGACCAGTGTTGCC (SEQ ID NO:40)

[0321] **BOLD CAPS:** *TMPRSS2* double underline CAPS: *ETV1* **+1**: Transcription
start ctgcag: PstI

[0322] Antisense-TMPRSS2-ETV1-H1

[0323]

+1tgtgctcgttcggcagcacatataactaacattggaacgacctgcagTTAACAAACAGCTTAATAAAT
AAGCTCAGGGATACCAGAATTCACAAAAAGAAAGCACTCTCCTCTGGGATCAGAG
TGGGTAGGAGGATGGGGTGCAATTG (SEQ ID NO:41)

[0324] **BOLD CAPS:** *TMPRSS2* double underline CAPS: *ETV1* **+1**: Transcription
start ctgcag: PstI

[0325] Sense-TMPRSS2-ETV1-A1

[0326]

+1tgtgctcgttcggcagcacatataactaacattggaacgacctgcagCCAGTTGGTGGAAGAGAGTC
CCCCGGGGTCCCCAAAGCTGGCTCCTAGTCTACTGAATCTAACTAAGACTTTAAC
TTGGAACGATTACATAATGCGGCA (SEQ ID NO:42)

[0327] **BOLD CAPS:** *TMPRSS2* double underline CAPS: *ETV1* **+1**: Transcription
start ctgcag: PstI

[0328] List of primers used

[0329] RT-PCR primers for amplifying induced fusion RNAs:

Induced fusion RNA (<i>TMPRSS2</i> ex1- <i>ERG</i> ex7)	
<i>TMPRSS2</i> ex-1 F1	5'- TAGGCGCGAGCTAAGCAGGAG-3' (SEQ ID NO:43)
<i>ERG</i> ex-7 R1	5'- TAGCATGCATTAACCGTGGAGA-3' (SEQ ID NO:44)
Induced fusion RNA (<i>TMPRSS2</i> ex1- <i>ERG</i> ex4)	
<i>TMPRSS2</i> ex-1 F1	5'- TAGGCGCGAGCTAAGCAGGAG -3' (SEQ ID NO:45)
<i>ERG</i> ex-4 R1	5'- CTTGAGCCATTACCTGGCTAG-3' (SEQ ID NO:46)
Induced fusion RNA (<i>TMPRSS2</i> ex1- <i>ETV</i> ex5)	
<i>TMPRSS2</i> ex-1 F1	5'- TAGGCGCGAGCTAAGCAGGAG-3' (SEQ ID NO:47)
<i>ETV1</i> ex-6 R1	5'- TTCTTGACTGCAGGCAGAGCT -3' (SEQ ID NO:48)

<i>TMPRSS2</i> ex-1 F2	5'- CAGGAGGCGGAGGCGGA-3' (SEQ ID NO:49)
<i>ETV1</i> ex-5 R1	5'- CTTTCAGCCTGATAGTCTGGT-3' (SEQ ID NO:50)
<i>TMPRSS2</i> ex-1 F3	5'- CGGAGGGCGAGGGGCGGGGA-3' (SEQ ID NO:51)
<i>ETV1</i> ex-5 R2	5'- AACTGCTCATCATTGTCAGGT-3' (SEQ ID NO:52)

[0330] Primers used in three-round PCR for amplifying induced fusion RNAs:

1 st round PCR	
<i>TMPRSS2</i> ex-1 F1	5'- TAGGCGCGAGCTAAGCAGGAG -3' (SEQ ID NO:53)
<i>ERG</i> ex-4 R1	5'- CTTGAGCCATTCACCTGGCTAG -3' (SEQ ID NO:54)
2 nd round PCR	
<i>TMPRSS2</i> ex-1 F2	5'- CAGGAGGCGGAGGCGGA -3' (SEQ ID NO:55)
<i>ERG</i> ex-4 R2	5'- TGACCCTGGCTGGGGGTTGAGA -3' (SEQ ID NO:56)
3 rd round PCR	
<i>TMPRSS2</i> ex-1 F3	5'- CGGAGGGCGAGGGGCGGGGA -3' (SEQ ID NO:57)
<i>ERG</i> ex-4 R3	5'- TCCTGCTGAGGGACGCGTGG -3' (SEQ ID NO:58)

[0331] RT-PCR primers for amplifying endogenous parental mRNAs:

<i>TMPRSS2</i> parental mRNA (<i>TMPRSS2</i> ex2- <i>TMPRSS2</i> ex4)	
<i>TMPRSS2</i> ex-2 F1	5'- GTCATATTGAACATTCCAGA-3' (SEQ ID NO:59)
<i>TMPRSS2</i> ex-4 R1	5'- GCGCAGCTCCCACGAGGAAGGT-3' (SEQ ID NO:60)
<i>ERG</i> parental mRNA (<i>ERG</i> ex3- <i>ERG</i> ex7): one-round PCR	
<i>ERG</i> ex-3 F1	5'- CAGGTTCTGAACAGCTGGTA-3' (SEQ ID NO:61)
<i>ERG</i> ex-7 R1	5'- TAGCATGCATTAACCGTGGAGA-3' (SEQ ID NO:62)
<i>ERG</i> ex-1 F1	5'- CCCCCGAGGGACATGAGAGAA-3' (SEQ ID NO:63)
<i>ERG</i> ex-4 R1	5'- TGGGGGTTGAGACAGCCAAT-3' (SEQ ID NO:64)
<i>ERG</i> parental mRNA (<i>ERG</i> ex3- <i>ERG</i> ex7): three-round PCR	
<i>ERG</i> ex-3 F1	5'- CAGGTTCTGAACAGCTGGTA-3' (SEQ ID NO:65)
<i>ERG</i> ex-7 R1	5'- TAGCATGCATTAACCGTGGAGA-3' (SEQ ID NO:66)
<i>ERG</i> ex-3 F2	5'- TGGGCTGGCTTACTGAAGGA-3' (SEQ ID NO:67)
<i>ERG</i> ex-7 R2	5'- TTGTAAGGCTTTATCAACAT-3' (SEQ ID NO:68)

<i>ERG</i> ex-3 F3	5'- ATGATTCAGACTGTCCCGGA-3' (SEQ ID NO:69)
<i>ERG</i> ex-7 R3	5'- CATCTGAAGTCAAATGTGGA-3' (SEQ ID NO:70)
<i>ERG</i> parental mRNA (<i>ERG</i> ex1- <i>ERG</i> ex4): three-round PCR	
<i>ERG</i> ex-1 F1	5'- CCCCCGAGGGACATGAGAGAA-3' (SEQ ID NO:71)
<i>ERG</i> ex-4 R1	5'- TGGGGGTTGAGACAGCCAAT-3' (SEQ ID NO:72)
<i>ERG</i> ex-1 F2	5'- AGGGACATGAGAGAAGAGGA-3' (SEQ ID NO:73)
<i>ERG</i> ex-4 R2	5'- TGAGGGACGCGTGGGCTCAT-3' (SEQ ID NO:74)
<i>ERG</i> ex-1 F3	5'- GAGAGAAGAGGAGCGGCGCT-3' (SEQ ID NO:75)
<i>ERG</i> ex-4 R3	5'- CTTGGAAGTCTGTCCATAGT-3' (SEQ ID NO:76)

[0332] RT-PCR primers for amplifying BCAM-AKT2 chimeric RNA:

1 st round PCR	
BCAM-AKT2 F1	5'- GTGAGCAGCTCTCTGACCCTGA -3' (SEQ ID NO:77)
BCAM-AKT2 R1	5'- CGCACCAGGATGACTTTGCCA-3' (SEQ ID NO:78)
2 nd round PCR	
BCAM-AKT2 F2	5'- TGTCTTCTACTGCGTGAGACGCA-3' (SEQ ID NO:79)
BCAM-AKT2 R2	5'- TCTCCTCAGTCGTGGAGGAGT-3' (SEQ ID NO:80)

[0333] RT-PCR primers for amplifying input RNAs:

Sense-1	
Sense 1 F1	5'- ACGGAATTCGAGTAGGCGCGAGCTAAGCA-3' (SEQ ID NO:81)
Sense 1 R1	5'- TAGAAGCTTCTTGAGCCATTACCTGGCT-3' (SEQ ID NO:82)
Sense-2	
Sense 2 F1	5'- ACGCTGCAGGAGTAGGCGCGAGCTAAGCA-3' (SEQ ID NO:83)
Sense 2 R1	5'- TAGAAGCTTAAAAAACTTGAGCCATTACCTGGCT-3' (SEQ ID NO:84)
Sense-2 long	
Sense 2 long F1	5'- ACGCTGCAGGAGTAGGCGCGAGCTAAGCA-3' (SEQ ID NO:85)

Sense 2 long R1	5'- TAGAAGCTTAAAAAATAAGGCTTTATCAACATCAT-3' (SEQ ID NO:86)
Antisense-1	
Antisense 1 F1	5'- ACGGAATTCCTTGAGCCATTACCTGGCT-3' (SEQ ID NO:87)
Antisense 1 R1	5'- TAGAAGCTTGAGTAGGCGCGAGCTAAGCA-3' (SEQ ID NO:88)
Antisense-2	
Antisense 2 F1	5'- ACGCTGCAGCTTGAGCCATTACCTGGCTAGGGTT-3' (SEQ ID NO:89)
Antisense 2 R1	5'- TAGAAGCTTAAAAAAGAGTAGGCGCGAGCTAAGCA-3' (SEQ ID NO:90)
Antisense-3	
Antisense 3 F1	5'- ACGCTGCAGCTTGAGCCATTACCTGGCTAGGGTT-3' (SEQ ID NO:91)
Antisense 3 R1	5'- TAGAAGCTTAAAAAAGAGTAGGCGCGAGCTTCCAT-3' (SEQ ID NO:92)
Antisense-4	
Antisense 4 F1	5'- ACGCTGCAGCTTGAGCCATTACCTGGCTAGGGTT-3' (SEQ ID NO:93)
Antisense 4 R1	5'- TAGAAGCTTAAAAAAGAGTAGGCGCGAGCTTCCAT-3' (SEQ ID NO:94)
Antisense-5	
Antisense 5 F1	5'- ACGCTGCAGCTTGAGCCATTACCTGGCTAGGGTT-3' (SEQ ID NO:95)
Antisense 5 R1	5'- TAGAAGCTTAAAAAAGAGTAGGCGCGAGCTTCCAT-3' (SEQ ID NO:96)
Antisense-6	
Antisense 6 F1	5'- ACGCTGCAGCTTGAGCCATTACCTGGCTAGGGTT-3' (SEQ ID NO:97)
Antisense 6 R1	5'- TAGAAGCTTAAAAAAGAGTAGGCGCGAGCTTCCAT-3' (SEQ ID NO:98)

Antisense-7	
Antisense 7 F1	5'- ACGCTGCAGCTTGAGCCATTACCTGGCTAGGGTT-3' (SEQ ID NO:99)
Antisense 7 R1	5'- TAGAAGCTTAAAAAAGAGTAGGCGCGAGCTTCCAT-3' (SEQ ID NO:100)
Antisense-8	
Antisense 8 F1	5'- ACGCTGCAGCTTGAGCCATTACCTGGCTAGGGTT-3' (SEQ ID NO:101)
Antisense 8 R1	5'- TAGAAGCTTAAAAAAGAGTAGGCGCGAGCTTCCAT-3' (SEQ ID NO:102)
Antisense-9	
Antisense 9 F1	5'- ACGCTGCAGGAGACAGCCAATCCTGCTGAGGGACGCGTGGGC- 3' (SEQ ID NO:103)
Antisense 9 R1	5'- TAGAAGCTTAAAAAAGAGTAGGCGCGAGCTTCCAT-3' (SEQ ID NO:104)
Antisense-5A	
Antisense 5A F1	5'- TAGAAGCTTAAAAAAGAGTAGGCGCGAGCTTCCAT-3' (SEQ ID NO:105)
Antisense 5A R1	5'- TAGCTGCAGGCCGCTCACCTGCCGCGCTCC-3' (SEQ ID NO:106)
Antisense-5B	
Antisense 5B F1	5'- ACGAAGCTTAAAAAAGGATTGGCTGTCTCAACCCCCA-3' (SEQ ID NO:107)
Antisense 5B R1	5'- ACGCTGCAGCTTGAGCCATTACCTGGCTAGGGTT-3' (SEQ ID NO:108)
Antisense-B1	
Antisense B1 F1	5'-TAGCTGCAGTGGCCTGAGCCTTGAAGAAT-3' (SEQ ID NO:109)
Antisense B1 R1	5'-ACGAAGCTTAAAAAACGGAGGCGGAGGGCGAGGGG-3' (SEQ ID NO:110)
Antisense-B2	

Antisense B2 F1	5'- TAGCTGCAGTGGCCTGAGCCTTGAAGAAT-3' (SEQ ID NO:111)
Antisense B2 R1	5'- ACGAAGCTTAAAAAAGGTGAGCGGCGCCGGTAC-3' (SEQ ID NO:112)
Antisense-B3	
Antisense B3 F1	5'- TAGCTGCAGTGGGGAGCATTTCCAGATGGAGAACTGCAAGG- 3' (SEQ ID NO:113)
Antisense B3 R1	5'- ACGAAGCTTAAAAAACGGAGGCGGAGGGCGAGGG-3' (SEQ ID NO:114)
Antisense-C1	
Antisense C1 F1	5'- TAGCTGCAGTGAGCTCATGCTATTCCTATGA-3' (SEQ ID NO:115)
Antisense C1 R1	5'- ACGAAGCTTAAAAAAGGTCCCCAAAGCTGGCTCCT-3' (SEQ ID NO:116)
Antisense-C2	
Antisense C2 F1	5'- TAGCTGCAGTGAGCTCATGCTATTCCTATGA -3' (SEQ ID NO:117)
Antisense C2 R1	5'- ACGAAGCTTAAAAAACTGGGAGCACCGGTGCGC-3' (SEQ ID NO:118)
Antisense-C3	
Antisense C3 F1	5'- TAGCTGCAGGTCCTGGTAACACTATTCATGCACTAACAAGTT-3' (SEQ ID NO:119)
Antisense C3 R1	5'- ACGAAGCTTAAAAAAGGTCCCCAAAGCTGGCTCCT-3' (SEQ ID NO:120)
Antisense-D1	
Antisense D1 F1	5'-TAGCTGCAGGAGAGACAGAGAGAGAGAGG-3' (SEQ ID NO:121)
Antisense D1 R1	5'- ACGAAGCTTAAAAAACGGGTGTGAGGAGCGCGGCG-3' (SEQ ID NO:122)
Antisense-D2	

Antisense D2 F1	5'- TAGCTGCAGGAGAGACAGAGAGAGAGAGG-3' (SEQ ID NO:123)
Antisense D2 R1	5'- ACGAAGCTTAAAAAACGCTGGGGCCAGCCGGGC-3' (SEQ ID NO:124)
Antisense-D3	
Antisense D3 F1	5'- TAGCTGCAGTGAAGGGATAAGTGTCCAGAATCCCTGGATCTG- 3' (SEQ ID NO:125)
Antisense D3 R1	5'- ACGAAGCTTAAAAAACGGGTGTGAGGAGCGCGGCG-3' (SEQ ID NO:126)
Antisense-E1	
Antisense E1 F1	5'- TAGCTGCAGAGGGTATTCAGTATTACTATTT-3' (SEQ ID NO:127)
Antisense E1 R1	5'- ACGAAGCTTAAAAAATGATCAAGCGAAGAATCTTTGTCCATG- 3' (SEQ ID NO:128)
Antisense-F1	
Antisense F1 F1	5'- TAGCTGCAGAATGGATGAATACATAAAATAA-3' (SEQ ID NO:129)
Antisense F1 R1	5'- ACGAAGCTTAAAAAAGCACCTGGGCGCTTCGGTCCACCTCAT- 3' (SEQ ID NO:130)
Antisense-G1	
Antisense G1 F1	5'- TAGCTGCAGTTTATTCATGAGAAAATTCTGG-3' (SEQ ID NO:131)
Antisense G1 R1	5'- ACGAAGCTTAAAAAAGGGGTCATGGGAGACCATGCTTTCTAG- 3' (SEQ ID NO:132)
Sense-3	
Sense 3 F1	5'- ACGCTGCAGGAGTAGGCGCGAGCTTCCAT-3' (SEQ ID NO:133)

Sense 3 R1	5'- TAGAAGCTTAAAAAACTTGAGCCATTCACCTGGCT-3' (SEQ ID NO:134)
Sense-4	
Sense 4 F1	5'- ACGCTGCAGGAGTAGGCGCGAGCTTCCAT-3' (SEQ ID NO:135)
Sense 4 R1	5'- TAGAAGCTTAAAAAACTTGAGCCATTCACCTGGCT-3' (SEQ ID NO:136)
Sense-5	
Sense 5 F1	5'- ACGCTGCAGGAGTAGGCGCGAGCTTCCAT-3' (SEQ ID NO:137)
Sense 5 R1	5'- TAGAAGCTTAAAAAACTTGAGCCATTCACCTGGCT-3' (SEQ ID NO:138)
Sense-6	
Sense 6 F1	5'- ACGCTGCAGGAGTAGGCGCGAGCTTCCAT-3' (SEQ ID NO:139)
Sense 6 R1	5'- TAGAAGCTTAAAAAACTTGAGCCATTCACCTGGCT-3' (SEQ ID NO:140)
Sense-B1	
Sense B1 F1	5'- TAGCTGCAGCGGAGGCGGAGGGCGAGGGG-3' (SEQ ID NO:141)
Sense B1 R1	5'- ACGAAGCTTAAAAAATGGCCTGAGCCTTGAAGAAT-3' (SEQ ID NO:142)
Sense-C1	
Sense C1 F1	5'- TAGCTGCAGGGTCCCCAAAGCTGGCTCCT-3' (SEQ ID NO:143)
Sense C1 R1	5'- ACGAAGCTTAAAAAATGAGCTCATGCTCTTCCTAT-3' (SEQ ID NO:144)
Sense-D1	
Sense D1 F1	5'- TAGCTGCAGCGGGTGTGAGGAGCGCGGCG-3' (SEQ ID NO:145)
Sense D1 R1	5'- ACGAAGCTTAAAAAAGAGAGACAGAGAGAGAGAGG-3' (SEQ ID NO:146)

Antisense <i>TMPRSS2-ETVI-A1</i>	
Antisense <i>TMPRSS2-ETVI-A1</i> F1	5'- TAGCTGCAGTGCCGCATTATGTAAATCGTTCCAAGTTAAAGTC- 3' (SEQ ID NO:147)
Antisense <i>TMPRSS2-ETVI-A1</i> R1	5'- ACGAAGCTTAAAAAACCAGTTGGTGGGAAGAGAGTCCCCCGGG- 3' (SEQ ID NO:148)
Antisense <i>TMPRSS2-ETVI-B1</i>	
Antisense <i>TMPRSS2-ETVI-B1</i> F1	5'- TAGCTGCAGGATTCAGTAATTTCAAAGTTTATTATATTTAAGA- 3' (SEQ ID NO:149)
Antisense <i>TMPRSS2-ETVI-B1</i> R1	5'- ACGAAGCTTAAAAAACAGCCCAGTTACAAGGCATTTGGAATA- 3' (SEQ ID NO:150)
Antisense <i>TMPRSS2-ETVI-C1</i>	
Antisense <i>TMPRSS2-ETVI-C1</i> F1	5'- TAGCTGCAGCTCTGAATAGAAAAATAGAAGTCCATAGTATCAA- 3' (SEQ ID NO:151)
Antisense <i>TMPRSS2-ETVI-C1</i> R1	5'- ACGAAGCTTAAAAAACCTGTCCCAGGAGATCTGGGAAGGTAC- 3' (SEQ ID NO:152)
Antisense <i>TMPRSS2-ETVI-D1</i>	
Antisense <i>TMPRSS2-ETVI-D1</i> F1	5'- TAGCTGCAGATATAAACAAAAAGTGTCAGCATTTGTCTCAACT- 3' (SEQ ID NO:153)
Antisense <i>TMPRSS2-ETVI-D1</i> R1	5'- ACGAAGCTTAAAAAAGATGGGGGTTGGGAGTGTGGCAGGGAG- 3' (SEQ ID NO:154)
Antisense <i>TMPRSS2-ETVI-E1</i>	
Antisense <i>TMPRSS2-ETVI-E1</i> F1	5'- TAGCTGCAGCTCTGAATAGAAAAATAGAAGTCCATAGTATCAA- 3' (SEQ ID NO:155)

Antisense <i>TMPRSS2-ETVI-E1</i> R1	5'- ACGAAGCTTAAAAAAAGGAAAGGTGAATGGTCAATGGAGTGC- 3' (SEQ ID NO:156)
Antisense <i>TMPRSS2-ETVI-F1</i>	
Antisense <i>TMPRSS2-ETVI-F1</i> F1	5'- TAGCTGCAGAGGAGGCCTTAAGTATGATTCAGTGAACACATTA- 3' (SEQ ID NO:157)
Antisense <i>TMPRSS2-ETVI-F1</i> R1	5'- ACGAAGCTTAAAAAAATCATTGCCTTGGTCCCAATTATTACA-3' (SEQ ID NO:158)
Antisense <i>TMPRSS2-ETVI-G1</i>	
Antisense <i>TMPRSS2-ETVI-G1</i> F1	5'- TAGCTGCAGTGTAATTGACTTAGATCTTGAAAGAGTTCTAAAA- 3' (SEQ ID NO:159)
Antisense <i>TMPRSS2-ETVI-G1</i> R1	5'- ACGAAGCTTAAAAAAGGCAACACTGGTCTTTGTTGTAGGTAA- 3' (SEQ ID NO:160)
Antisense <i>TMPRSS2-ETVI-H1</i>	
Antisense <i>TMPRSS2-ETVI-H1</i> F1	5'- TAGCTGCAGTTAACAAACAGCTTAATAAATAAGCTCAGGGATA- 3' (SEQ ID NO:161)
Antisense <i>TMPRSS2-ETVI-H1</i> R1	5'- ACGAAGCTTAAAAAACAATTGCACCCCATCCTCCTACCCACT-3' (SEQ ID NO:162)
Sense <i>TMPRSS2-ETVI-A1</i>	
Sense <i>TMPRSS2-ETVI-A1</i> F1	5'- TAGCTGCAGCCAGTTGGTGGAAGAGAGT-3' (SEQ ID NO:163)

Sense <i>TMPRSS2</i> - <i>ETV1</i> -A1 R1	5'- ACGAAGCTTAAAAAATGCCGCATTATGTAAATC-3' (SEQ ID NO:164)
---	--

[0334] PCR primers used for amplifying the identified *TMPRSS2*-*ERG* genomic DNA breakpoint:

<i>TMPRSS2</i> intron 1	
<i>TMPRSS2</i> genomic bk F1	5'- ATGTGATATTAGTGCGGTTA -3' (SEQ ID NO:165)
<i>TMPRSS2</i> genomic bk F2	5'- GGCTGGGATGTGTCCGTGGA -3' (SEQ ID NO:166)
<i>ERG</i> intron 3	
<i>ERG</i> genomic bk R1	5'- CCATAAGTTTTACTGCGTCT-3' (SEQ ID NO:167)
<i>ERG</i> genomic bk R2	5'- GATACTGAGTGGTAAATTCT-3' (SEQ ID NO:168)

[0335] The rest of PCR primers used for genomic breakpoint analyses are not listed, but their locations are shown in FIG. 18A.

[0336] PCR primers used for amplifying the identified *TMPRSS2*-*ETV1* genomic DNA breakpoint:

<i>TMPRSS2</i>	
<i>TMPRSS2</i> genomic bk pt A1	5'- TCTGCTCGAGCACGGGTCCA-3' (SEQ ID NO:169)
<i>TMPRSS2</i> genomic bk pt A2	5'- AAAACTGCCCCATGTCCAG-3' (SEQ ID NO:170)
<i>TMPRSS2</i> genomic bk pt B1	5'- CAACCTGGGAGGCCCTGCCT-3' (SEQ ID NO:171)
<i>TMPRSS2</i> genomic bk pt B2	5'- CAGCAACAGCACAAGCTTGT-3' (SEQ ID NO:172)

TMPRSS2 genomic bk pt C1	5'- GGCTGGGATGTGTCCGTGGA-3' (SEQ ID NO:173)
TMPRSS2 genomic bk pt C2	5'- TGGTGGTGGTGCTGTCTGGA-3' (SEQ ID NO:174)
TMPRSS2 genomic bk pt D1	5'- CAGGAGAATCACTTGAACCT-3' (SEQ ID NO:175)
TMPRSS2 genomic bk pt D2	5'- AGGCCACTGCACTCCAGCCT-3' (SEQ ID NO:176)
ERG intron 3	
ETV1 genomic bk pt M1	5'-GCAAGTCTCGTTGATCGCCA-3' (SEQ ID NO:177)
ETV1 genomic bk pt M2	5'-TTGCACACGTTTGCGAATCA-3' (SEQ ID NO:178)
ETV1 genomic bk pt N1	5'-AGGGAGAGTTGCTTCCCAGT-3' (SEQ ID NO:179)
ETV1 genomic bk pt N2	5'-GCCGATCTTAGCACATTACT-3' (SEQ ID NO:180)

EXAMPLE 4

CHARACTERIZATION OF CHIMERIC RNAS

[0337] Certain parameters for design of chimeric RNAs are characterized. FIG. 23 shows examples of engineered chimeric RNAs for TMPRSS2-ERG. Each illustration in FIG. 23 demonstrates 3-way junction stem formation between the chimeric RNA and two separate genomic regions of TMPRSS2 and ERG concomitantly. In each example, the 3-way junction produces regions of single-stranded genomic DNA loop or spacer that in at least some cases may be described as a bulge at or near the base of the junction. To examine the parameters, FIG. 24 illustrates different regions of the 3-way junction stem formation that may be variable, including length of the bulge and lengths of the respective regions of the chimeric RNA.

[0338] The optimum size of the bulge for fusion induction was examined in FIGS. 25, 26, and 27 for the corresponding Stem B, C, and D of FIG. 23. As is shown, for fusion gene

induction the optimum size of the bulge is around 35 nucleotides on each side, although other lengths also work.

[0339] It was next examined if there was an optimum chimeric RNA size for fusion induction when the bulge is fixed at 35 nucleotides. FIG. 28 illustrates a schematic for testing of chimeric RNAs at varying sizes when the bulge is 35 nucleotides. FIGS. 29-32 provide results of testing the various sizes of the chimeric RNA when using Stem A, Stem B, Stem C, or Stem D, respectively. For fusion gene induction, an optimum size for the chimeric RNAs is around 50-75 nucleotides on each side.

EXAMPLE 5

CHIMERIC RNA-DRIVEN GENOMIC REARRANGEMENT IN MAMMALIAN CELLS

[0340] The present example and others demonstrate the chimeric RNA technology with respect to five experimentally induced gene fusions, merely as examples. These induced gene fusions involve different gene partners and different cell types, further supporting the embodiment of the disclosure that RNA-induced genomic rearrangement is permissible in mammalian cells using the chimeric RNA technology of the present disclosure.

[0341] Examples in the present disclosure include the following:

[0342] TMPRSS2-ERG, a prostate cancer fusion gene, in LNCaP cells and PNT1A cells

[0343] TMPRSS2-ETV1, a prostate cancer fusion gene, in LNCaP cells

[0344] JAZF1-SUZ12, an endometrial cancer fusion gene, in 293T cells

[0345] BCR-RANGAP1, not known to be disease-relevant but an experimentally induced fusion gene, in 293T cells

[0346] BCR-KCNQ2, not known to be disease-relevant but an experimentally induced fusion gene, in 293T and HL-60 cells

[0347] In the data provided in this example, it is shown that gene fusion can be induced by chimeric RNAs that are designed to target genomic DNA in a sequence-dependent manner. As indicated elsewhere herein, formation of a three-way junction between genomic DNA and chimeric RNA facilitates fusion induction. The three-way junction comprises an RNA/DNA

hybrid and a genomic DNA stem between the two parental genes. The genomic DNA sequences that form the stem can be sense pairing to a sense strand of the gene partners, or sense pairing to an antisense strand of the gene partners.

[0348] This example demonstrates that genomic DNA stems could be formed by DNA repeats, such as the SINE and the LINE repeats, that may constitute stems over 300 nts in length. Genomic DNA stems could also be formed without using the SINE and the LINE repeats that may constitute stems ranging from 10 nts to <100 nts in length.

[0349] The present example indicates that RNA-driven gene fusion can be induced without hormone or other stimuli or can be facilitated by hormone stimulation in at least some cases. RNA-driven gene fusion can occur in adherent and suspension mammalian cell lines such as, but not limited to, 293T and HL-60 cells as examples. Furthermore, the designed chimeric RNAs could have a wide range of GC content, such as, but not limited to from about 27% to 65%.

[0350] FIGS. 39A-39D concern RNA-mediated inter-chromosomal JAZF1-SUZ12 fusion in 293T cells, as one example of RNA-driven gene fusion. FIG. 39A illustrates an interchromosomal fusion between JAZF1 and SUZ12 in which the design of the produced chimeric RNA forms part of a three-way junction between genomic DNA and the chimeric RNA. That is, the three-way junction comprises an RNA/DNA hybrid and a genomic DNA stem between the parental genes. In this specific case, the sense genomic strand of JAZF1 is on the negative strand (of chromosome 7) but SUZ12 is on the positive strand of chromosome 17. For a particular fusion between JAZF1 and SUZ12, FIG. 39B provides an example of the chimeric RNA sequences that induced the JAZF1-SUZ12 fusion. FIG. 39C demonstrates successful generation of produced JAZF1-SUZ12 fusion transcripts; these were confirmed by Sanger sequencing to be a fusion of JAZF1 exon 1 spliced to a cryptic splice site of SUZ12 exon 1. FIG. 39D demonstrates production of a different chimeric RNA of JAZF1-SUZ12 fusion transcript than what is shown in FIG. 39C. In this case, there was production of a different JAZF1-SUZ12 fusion transcript, as confirmed by Sanger sequencing; the fusion transcript was from splicing JAZF1 exon 2 to SUZ12 exon 2.

[0351] FIGS. 40A-40C concern RNA-mediated intra-chromosomal BCR-RANGAP1 fusion by a designed chimeric RNA. FIG. 40A provides an illustration of design using a three-way junction that could be formed between the parental genes BCR and RANGAP1. In this

case, the sense genomic strand of BCR is on the positive strand of chromosome 22 and RANGAP1 is on the negative strand of chromosome 22. FIG. 40B provides one example of the chimeric RNA sequence used to induce fusion of BCR and RANGAP1, and FIG. 40C shows production of the fusion transcript induced by sequence-specific chimeric RNA. The BCR-RANGAP1 fusion transcript sequence was confirmed by Sanger sequencing to comprise BCR exon14 spliced to RANGAP1 exon3 (RANGAP1 isoform 1, NM_001278651.1). The RNA fusion junction is an annotated splice junction for both genes.

[0352] FIGS. 41A-41D regard RNA-mediated inter-chromosomal BCR-KCNQ2 fusions by designed chimeric RNA. FIG. 41A shows a design of chimeric RNA utilizing a three-way junction between the genomic DNA and the parental genes of BCR and KCNQ2. In this case, the sense genomic strand of BCR is on the positive strand of chromosome 22 while KCNQ2 is on the negative strand of chromosome 20. FIG. 41B shows one example of a particular chimeric RNA sequence used to induce BCR-KCNQ2 fusion. Two specific BCR-KCNQ2 fusion transcripts were produced (FIG. 41C), and a schematic of their induced transcripts with particular exon fusion configurations is depicted (FIG. 41D). Sanger sequencing confirmed that BCR-KCNQ2 fusion transcript contains BCR exon16 spliced to KCNQ2 exon8 (KCNQ2 isoform 5, NM_172109.2). RNA fusion junction is an annotated splice junction for both genes. Sanger sequencing confirmed the presence of another BCR-KCNQ2 fusion transcript contains BCR exon18 spliced to KCNQ2 exon8 (KCNQ2 isoform 5, NM_172109.2). RNA fusion junction is an annotated splice junction for both genes.

[0353] FIGS. 42A-42E concern BCR-KCNQ2 fusions induced by chimeric RNAs utilizing or excluding the short interspersed nuclear element (SINE) for three-way junction formation. FIG. 42A shows an illustration of chimeric RNA design utilizing a three-way junction between two SINEs in genomic DNA and the chimeric RNA. In this case, the stem of the junction is formed from DNA repeats, such as SINE/SINE repeats. In FIG. 42B, there is provided one example of a designed chimeric RNA sequence that induced BCR-KCNQ2 fusion RNA. FIG. 42C shows an example of chimeric RNA design targeting sequences of the respective genes in which the genomic stems are formed from regions outside of known DNA repeats, such as SINEs. FIG. 42D provides a specific example of chimeric RNA sequence that was able to induce BCR-KCNQ2 fusion RNA. FIG. 42E shows that fusion of two parental genes can be induced by chimeric RNAs having genomic stems comprising SINE repeats or outside of SINE repeats. Sanger sequencing confirmed that BCR-KCNQ2 fusion transcript contains BCR

exon16 spliced to KCNQ2 exon8 (KCNQ2 isoform 5, NM_172109.2). RNA fusion junction is an annotated splice junction for both genes. Sanger sequencing also confirmed the presence of another BCR-KCNQ2 fusion transcript contains BCR exon18 spliced to KCNQ2 exon8 (KCNQ2 isoform 5, NM_172109.2). RNA fusion junction is an annotated splice junction for both genes.

[0354] A summary of properties of examples of chimeric RNAs is below:

[0355]

Chimeric RNA ID	Fusion gene partners	Sizes of targeting region on each gene (nt)	G-C content (%)	Cell line(s) identified
JS-7	JAZF1-SUZ12	50-50	46%	293T
JS-8	JAZF1-SUZ12	50-50	27%	293T
JS-11	JAZF1-SUZ12	50-50	47%	293T
S1-7	BCR-RANGAP1	38-26	48.4%	293T
S4 30-2	BCR-KCNQ2	46-45	55.4%	HL-60
S4 200-9	BCR-KCNQ2	46-22	50%	HL-60
S1A 200-1	BCR-KCNQ2	50-50	52%	293T
S1A 200-2	BCR-KCNQ2	50-50	52%	293T
S1A 10-9	BCR-KCNQ2	50-50	58%	293T
S1B 10-1	BCR-KCNQ2	48-50	59.2%	293T
S1B 10-2	BCR-KCNQ2	48-50	63.3%	293T
S1B 10-3	BCR-KCNQ2	48-50	55.1%	293T
S1B 10-8	BCR-KCNQ2	51-50	60.4%	293T

[0356] An additional summary of properties of examples of chimeric RNAs is below:

[0357]

Chimeric RNA ID	Fusion gene partners	Sizes of targeting region on each gene (nt)	G-C content (%)	Cell line(s) identified
Antisense-1	TMPRSS2-ERG	78-218	61%	LNCaP
Antisense-2	TMPRSS2-ERG	78-218	61%	LNCaP
Antisense-5	TMPRSS2-ERG	52-75	65%	LNCaP, PNT1A
Antisense-B1	TMPRSS2-ERG	52-75	62%	LNCaP
Antisense-C1	TMPRSS2-ERG	52-75	57%	LNCaP
Antisense-D1	TMPRSS2-ERG	52-75	60%	LNCaP
Sense AZ1 ex16-17	TMPRSS2-ERG	117-103	54%	LNCaP
Antisense-TETV-1	TMPRSS2-ETV1	50-50	49%	LNCaP

REFERENCES

[0358] All patents and publications mentioned in this specification are indicative of the level of those skilled in the art to which the invention pertains. All patents and publications herein are incorporated by reference to the same extent as if each individual publication was specifically and individually indicated to be incorporated by reference in their entirety.

[0359] Aoto, H., Miyake, Y., Nakamura, M., and Tajima, S. (1997). Genomic organization of the mouse AZ1 gene that encodes the protein localized to preacrosomes of spermatids. *Genomics* 40, 138-141.

[0360] Aoto, H., Tsuchida, J., Nishina, Y., Nishimune, Y., Asano, A., and Tajima, S. (1995). Isolation of a novel cDNA that encodes a protein localized to the pre-acrosome region of spermatids. *Eur J Biochem* 234, 8-15.

[0361] Bastus, N.C., Boyd, L.K., Mao, X., Stankiewicz, E., Kudahetti, S.C., Oliver, R.T., Berney, D.M., and Lu, Y.J. (2010). Androgen-induced TMPRSS2:ERG fusion in nonmalignant prostate epithelial cells. *Cancer Res* 70, 9544-9548.

[0362] Boyce, M.J., Baisley, K.J., Clark, E.V., and Warrington, S.J. (2004). Are published normal ranges of serum testosterone too high? Results of a cross-sectional survey of serum testosterone and luteinizing hormone in healthy men. in *BJU Int*, pp. 881-885.

[0363] Coll-Bastus, N., Mao, X., Young, B.D., Sheer, D., and Lu, Y.J. (2015). DNA replication-dependent induction of gene proximity by androgen. *Hum Mol Genet* 24, 963-971.

- [0364] Crick, F. (1970). Central dogma of molecular biology. *Nature* 227, 561-563.
- [0365] Fang, W., and Landweber, L.F. (2012). RNA-mediated genome rearrangement: hypotheses and evidence. *Bioessays* 35, 84-87.
- [0366] Horoszewicz, J.S., Leong, S.S., Chu, T.M., Wajsman, Z.L., Friedman, M., Papsidero, L., Kim, U., Chai, L.S., Kakati, S., Arya, S.K., *et al.* (1980). The LNCaP cell line--a new model for studies on human prostatic carcinoma. *Prog Clin Biol Res* 37, 115-132.
- [0367] Janz, S., Potter, M., and Rabkin, C.S. (2003). Lymphoma- and leukemia-associated chromosomal translocations in healthy individuals. *Genes Chromosomes Cancer* 36, 211-223.
- [0368] Kannan, K., Wang, L., Wang, J., Ittmann, M.M., Li, W., and Yen, L. (2011). Recurrent chimeric RNAs enriched in human prostate cancer identified by deep sequencing. *Proc Natl Acad Sci U S A* 108, 9172-9177.
- [0369] Kimsey, I., and Al-Hashimi, H.M. (2014). Increasing occurrences and functional roles for high energy purine-pyrimidine base-pairs in nucleic acids. *Curr Opin Struct Biol* 24, 72-80.
- [0370] Langabeer, S.E., Walker, H., Rogers, J.R., Burnett, A.K., Wheatley, K., Swirsky, D., Goldstone, A.H., and Linch, D.C. (1997). Incidence of AML1/ETO fusion transcripts in patients entered into the MRC AML trials. MRC Adult Leukaemia Working Party. *Br J Haematol* 99, 925-928.
- [0371] Li, H., Wang, J., Mor, G., and Sklar, J. (2008). A neoplastic gene fusion mimics trans-splicing of RNAs in normal human cells. *Science* 321, 1357-1361.
- [0372] Lieber, M.R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem* 79, 181-211.
- [0373] Lin, C., Yang, L., Tanasa, B., Hutt, K., Ju, B.G., Ohgi, K., Zhang, J., Rose, D.W., Fu, X.D., Glass, C.K., *et al.* (2009). Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. *Cell* 139, 1069-1083.

[0374] Mani, R.S., Tomlins, S.A., Callahan, K., Ghosh, A., Nyati, M.K., Varambally, S., Palanisamy, N., and Chinnaiyan, A.M. (2009). Induced chromosomal proximity and gene fusions in prostate cancer. *Science* 326, 1230.

[0375] Mertz, K.D., Setlur, S.R., Dhanasekaran, S.M., Demichelis, F., Perner, S., Tomlins, S., Tchinda, J., Laxman, B., Vessella, R.L., Beroukhi, R., *et al.* (2007). Molecular characterization of TMPRSS2-ERG gene fusion in the NCI-H660 prostate cancer cell line: a new perspective for an old model. *Neoplasia* 9, 200-206.

[0376] Mitelman, F., Johansson, B., and Mertens, F. (2007). The impact of translocations and gene fusions on cancer causation. *Nat Rev Cancer* 7, 233-245.

[0377] Nowacki, M., Vijayan, V., Zhou, Y., Schotanus, K., Doak, T.G., and Landweber, L.F. (2008). RNA-mediated epigenetic programming of a genome-rearrangement pathway. *Nature* 451, 153-158.

[0378] Perner, S., Demichelis, F., Beroukhi, R., Schmidt, F.H., Mosquera, J.M., Setlur, S., Tchinda, J., Tomlins, S.A., Hofer, M.D., Pienta, K.G., *et al.* (2006). TMPRSS2:ERG fusion-associated deletions provide insight into the heterogeneity of prostate cancer. *Cancer Res* 66, 8337-8341.

[0379] Rowley, J.D., and Blumenthal, T. (2008). Medicine. The cart before the horse. *Science* 321, 1302-1304.

[0380] Rubin, M.A., Maher, C.A., and Chinnaiyan, A.M. (2011). Common gene rearrangements in prostate cancer. *J Clin Oncol* 29, 3659-3668.

[0381] Rudiger, N.S., Gregersen, N., and Kielland-Brandt, M.C. (1995). One short well conserved region of Alu-sequences is involved in human gene rearrangements and has homology with prokaryotic chi. *Nucleic Acids Res* 23, 256-260.

[0382] Rychlik, W., and Rhoads, R.E. (1989). A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in vitro amplification of DNA. *Nucleic Acids Res* 17, 8543-8551.

[0383] Teles Alves, I., Hiltemann, S., Hartjes, T., van der Spek, P., Stubbs, A., Trapman, J., and Jenster, G. (2013). Gene fusions by chromothripsis of chromosome 5q in the VCaP prostate cancer cell line. *Hum Genet* 132, 709-713.

[0384] Tomlins, S.A., Rhodes, D.R., Perner, S., Dhanasekaran, S.M., Mehra, R., Sun, X.W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R., *et al.* (2005). Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* 310, 644-648.

[0385] Watson, J.D., and Crick, F.H. (1953). Genetical implications of the structure of deoxyribonucleic acid. *Nature* 171, 964-967.

[0386] Weier, C., Haffner, M.C., Mosbrugger, T., Esopi, D.M., Hicks, J., Zheng, Q., Fedor, H., Isaacs, W.B., De Marzo, A.M., Nelson, W.G., *et al.* (2013). Nucleotide resolution analysis of TMPRSS2 and ERG rearrangements in prostate cancer. *J Pathol* 230, 174-183.

[0387] Zaphiropoulos, P.G. (2011). Trans-splicing in Higher Eukaryotes: Implications for Cancer Development? *Front Genet* 2, 92.

[0388] Zhang, F., Carvalho, C.M., and Lupski, J.R. (2009). Complex human chromosomal and genomic rearrangements. *Trends Genet* 25, 298-307.

[0389] Although the present disclosure and its advantages have been described in detail, it should be understood that various changes, substitutions and alterations can be made herein without departing from the spirit and scope of the design as defined by the appended claims. Moreover, the scope of the present application is not intended to be limited to the particular embodiments of the process, machine, manufacture, composition of matter, means, methods and steps described in the specification. As one of ordinary skill in the art will readily appreciate from the present disclosure, processes, machines, manufacture, compositions of matter, means, methods, or steps, presently existing or later to be developed that perform substantially the same function or achieve substantially the same result as the corresponding embodiments described herein may be utilized according to the present disclosure. Accordingly, the appended claims are intended to include within their scope such processes, machines, manufacture, compositions of matter, means, methods, or steps.

CLAIMS

What is claimed is:

1. A method of producing genomic DNA rearrangements between two different regions of genomic DNA, comprising the step of:

exposing to the genomic DNA an effective amount of an exogenously provided chimeric RNA, said chimeric RNA comprising fusion of (1) a first part that is reverse complement with respect to a first genomic region and (2) a second part that is reverse complement with respect to a second genomic region, wherein upon exposing of the chimeric RNA to the genomic DNA, the first and second genomic regions become fused through DNA recombination.

2. The method of claim 1, wherein upon exposing the chimeric RNA to the genomic DNA, the chimeric RNA hybridizes with the respective strands of the first and second genomic regions of the genomic DNA through the first and second parts of the chimeric RNA, thereby forming a DNA/RNA hybrid in a sequence-specific manner.

3. The method of claim 2, wherein upon hybridization of the chimeric RNA with the respective strands of the first and second genomic regions of the genomic DNA, a double stranded DNA/DNA stem structure is produced between the first and second genomic regions adjacent to the site of hybridization with the chimeric RNA.

4. The method of claim 3, wherein upon hybridizing of the chimeric RNA with the genomic DNA at the first and second genomic regions, a DNA spacer is produced in each of the respective strands of the first and second genomic regions between the DNA/DNA stem structure and the DNA/RNA hybrid structure.

5. The method of any one of claims 1-4, wherein the two different regions of genomic DNA are on the same chromosome.

6. The method of any one of claims 1-4, wherein the two different regions of genomic DNA are on different chromosomes.

7. The method of any one of claims 1-6, wherein at least one of the two different regions has active transcription.

8. The method of claim 7, wherein the chimeric RNA is antisense with respect to the at least one region having active transcription.
9. The method of any one of claims 1-8, wherein at least one of the two different regions lacks transcriptional activity.
10. The method of claim 9, wherein one or both of the first part and second part of the chimeric RNA is either antisense or sense with respect to the at least one of the two different regions lacking active transcription.
11. The method of any one of claims 1-10, wherein the length of the first part of the chimeric RNA is in a range of 30-150, 30-125, 30-100, 30-75, 30-50, 50-150, 50-125, 50-100, 50-75, 75-150, 75-125, 75-100, 100-150, 100-125, 125-175, or 125-150 nucleotides in length.
12. The method of claim 11, wherein the length of the first part of the chimeric RNA is in the range of 50-75 nucleotides.
13. The method of any one of claims 1-12, wherein the length of the second part of the chimeric RNA is in a range of 30-150, 30-125, 30-100, 30-75, 30-50, 50-150, 50-125, 50-100, 50-75, 75-150, 75-125, 75-100, 100-150, 100-125, 125-175, or 125-150 nucleotides in length.
14. The method of claim 13, wherein the length of the second part of the chimeric RNA is in the range of 50-75 nucleotides.
15. The method of any one of claims 1-14, wherein the length of the first part of the chimeric RNA and/or second part of the chimeric RNA is 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, or 75 nucleotides.
16. The method of any one of claims 1-15, wherein the length of the chimeric RNA is in the range of 60-300, 60-250, 60-200, 60-150, 60-100, 75-300, 75-250, 75-200, 75-150, 75-100, 100-300, 100-250, 100-200, 100-150, 150-300, 150-250, 150-200, 150-175, 175-300, 175-250, 175-200, 200-300, 200-275, or 200-250 nucleotides.
17. The method of any one of claims 4-16, wherein the DNA spacer is single stranded.
18. The method of any one of claims 4-17, wherein the DNA spacer is double stranded.

19. The method of any one of claims 4-18, wherein the DNA spacer of the first genomic region is a length in a range of 15-50, 15-45, 15-30, 15-25, 15-20, 20-50, 20-45, 20-40, 20-35, 20-30, 20-25, 25-50, 25-45, 25-40, 25-35, 25-30, 30-50, 30-45, 30-40, 30-35, 40-50, 40-45, or 45-50 nucleotides.

20. The method of claim 19, wherein the DNA spacer of the first genomic region is a length in a range of 25-45 nucleotides.

21. The method of any one of claims 4-20, wherein the DNA spacer of the first genomic region is 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 nucleotides.

22. The method of claim 21, wherein the DNA spacer of the first genomic region is a length of 35 nucleotides.

23. The method of any one of claims 4-22, wherein the DNA spacer of the second genomic region is a length in a range of 15-50, 15-45, 15-30, 15-25, 15-20, 20-50, 20-45, 20-40, 20-35, 20-30, 20-25, 25-50, 25-45, 25-40, 25-35, 25-30, 30-50, 30-45, 30-40, 30-35, 40-50, 40-45, or 45-50 nucleotides.

24. The method of claim 23, wherein the DNA spacer of the second genomic region is a length in a range of 25-45 nucleotides.

25. The method of any one of claims 4-17, wherein the single stranded DNA region of the second genomic region is 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 nucleotides.

26. The method of claim 25, wherein the DNA spacer of the second genomic region is a length of 35 nucleotides.

27. The method of any one of claims 1-26, wherein the chimeric RNA has or is designed to have between 27% and 65% GC content.

28. The method of any one of claims 1-27, wherein the genomic DNA is in a cell.

29. The method of claim 28, wherein the cell is an eukaryotic cell.

30. The method of claim 28 or 29, wherein the cell is a mammalian cell.

31. The method of any one of claims 28-30, wherein the cell is a human cell.
32. The method of any one of claims 28-31, wherein the cell is in a tissue or organ or body fluid.
33. The method of any one of claims 28-32, wherein the cell is *in vivo*.
34. The method of any one of claims 1-33, wherein the method occurs *in vivo*.
35. The method of any one of claims 1-33, wherein the method occurs *in vitro*.
36. The method of any one of claims 1-35, wherein the genomic DNA is from a diseased cell or diseased cells.
37. The method of any one of claims 1-36, wherein the genomic DNA is from an individual in need of treatment.
38. The method of any one of claims 1-37, wherein the method occurs at physiological hormone levels.
39. The method of any one of claims 1-38, wherein the chimeric RNA is produced in a cell.
40. The method of claim 39, wherein the cell comprises the genomic DNA.
41. The method of claim 39 or 40, wherein the chimeric RNA is expressed from a vector.
42. The method of claim 41, wherein the vector is a viral vector or a non-viral vector.
43. The method of claim 42, wherein the viral vector is a lentiviral vector, a retroviral vector, an adenoviral vector, or an adeno-associated viral vector.
44. The method of claim 43, wherein the non-viral vector is a plasmid.
45. A method of inducing gene fusion between first and second target genes, comprising the steps of:
 - designing or obtaining a chimeric RNA comprising (1) a first part that is antisense or sense with respect to the first target gene and (2) a second part that is antisense or sense with respect to a second target gene; and

exposing an effective amount of the chimeric RNAs to nucleic acid comprising the first and second target genes, wherein upon exposing of the chimeric RNA to the nucleic acid, the first and second target genes become fused through DNA recombination.

46. The method of claim 45, wherein when the first target gene is actively transcribed, the first part of the chimeric RNA is antisense with respect to the first target gene.

47. The method of claim 45, wherein when the first target gene is not actively transcribed, the first part of the chimeric RNA is sense or antisense with respect to the first target gene.

48. The method of claim 45, wherein when the second target gene is actively transcribed, the first part of the chimeric RNA is antisense with respect to the second target gene.

49. The method of claim 45, wherein when the second target gene is not actively transcribed, the first part of the chimeric RNA is sense or antisense with respect to the second target gene.

50. A method of identifying a cell, group of cells, or tissue at risk for becoming cancerous for an individual, comprising the step of identifying the presence of an endogenous chimeric RNA in the cell or one or more cells of the group of cells or tissue in a sample from the individual.

51. The method of claim 50, wherein the chimeric RNA comprises sequence derived from the parental genes of an oncogenic fusion gene or comprises sequence that is partially complimentary to that of the parental genes.

52. The method of claim 50 or 51, wherein the chimeric RNA is identified with reverse transcriptase-polymerase chain reaction, Northern gel analysis, and/or microarray chips designed for detecting endogenous chimeric RNAs.

53. The method of any one of claims 50-52, wherein the individual is provided an effective amount of a preventative or therapeutic for the cancer.

54. The method of claim 53, wherein the preventative is a vaccine.

55. The method of claim 53, wherein the therapeutic is an agent that inhibits the chimeric RNA directly by binding, or inhibits the chimeric RNA indirectly by reducing its expression in cells.

56. The method of claim 55, wherein the agent that directly inhibits the chimeric RNA is siRNA or antisense oligo DNA.

57. The method of any one of claims 55 or 56, wherein the agent that indirectly inhibits the chimeric RNA is a compound that methylates the promotor that drives the endogenous chimeric RNA expression.

58. A method of treating an individual for cancer, comprising the step of delivering an effective amount of an agent that inhibits a chimeric RNA in cancer cells of the individual.

59. The method of claim 58, wherein the agent is a siRNA.

60. The method of claim 58 or 59 wherein the siRNA is delivered to the individual in a vector.

61. The method of claim 60, wherein the vector is a viral vector or anti-viral vector.

62. The method of claim 61, wherein the viral vector is a lentiviral vector, a retroviral vector, an adenoviral vector, or an adeno-associated viral vector.

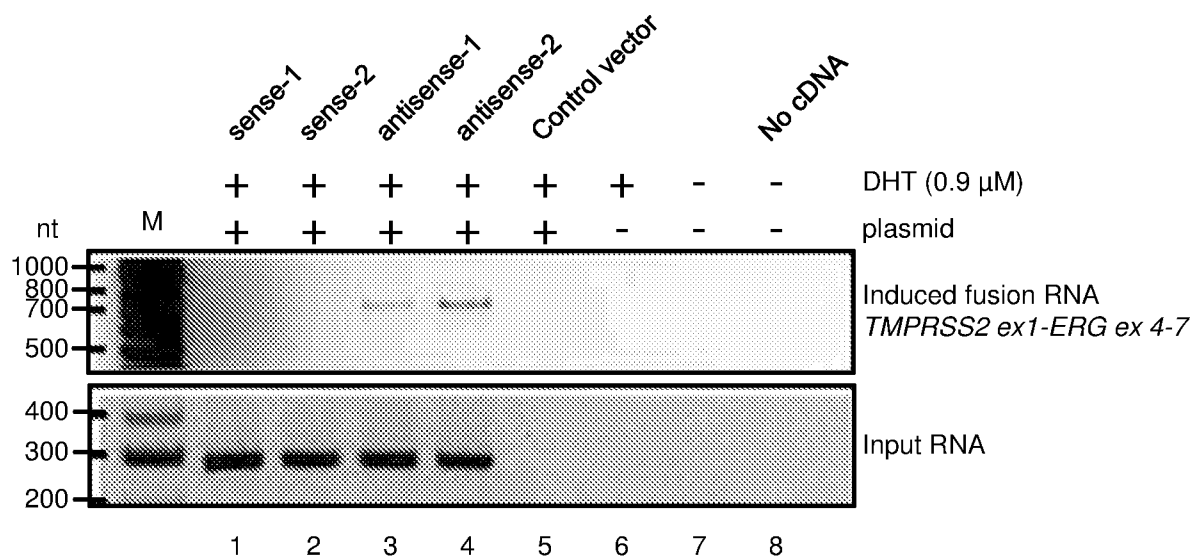
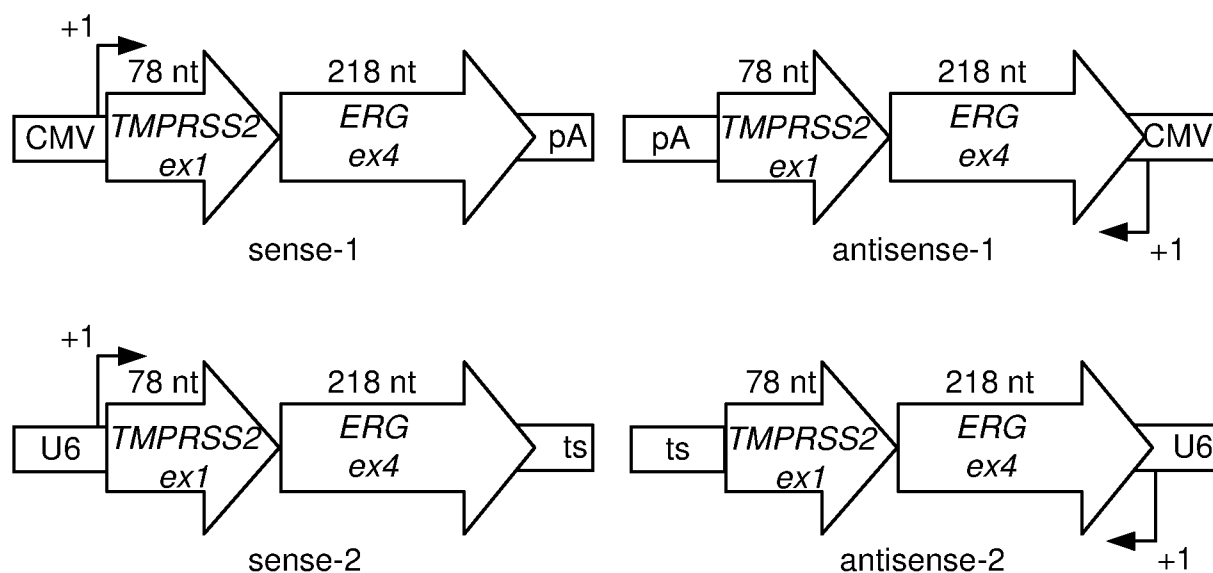


FIG. 1A

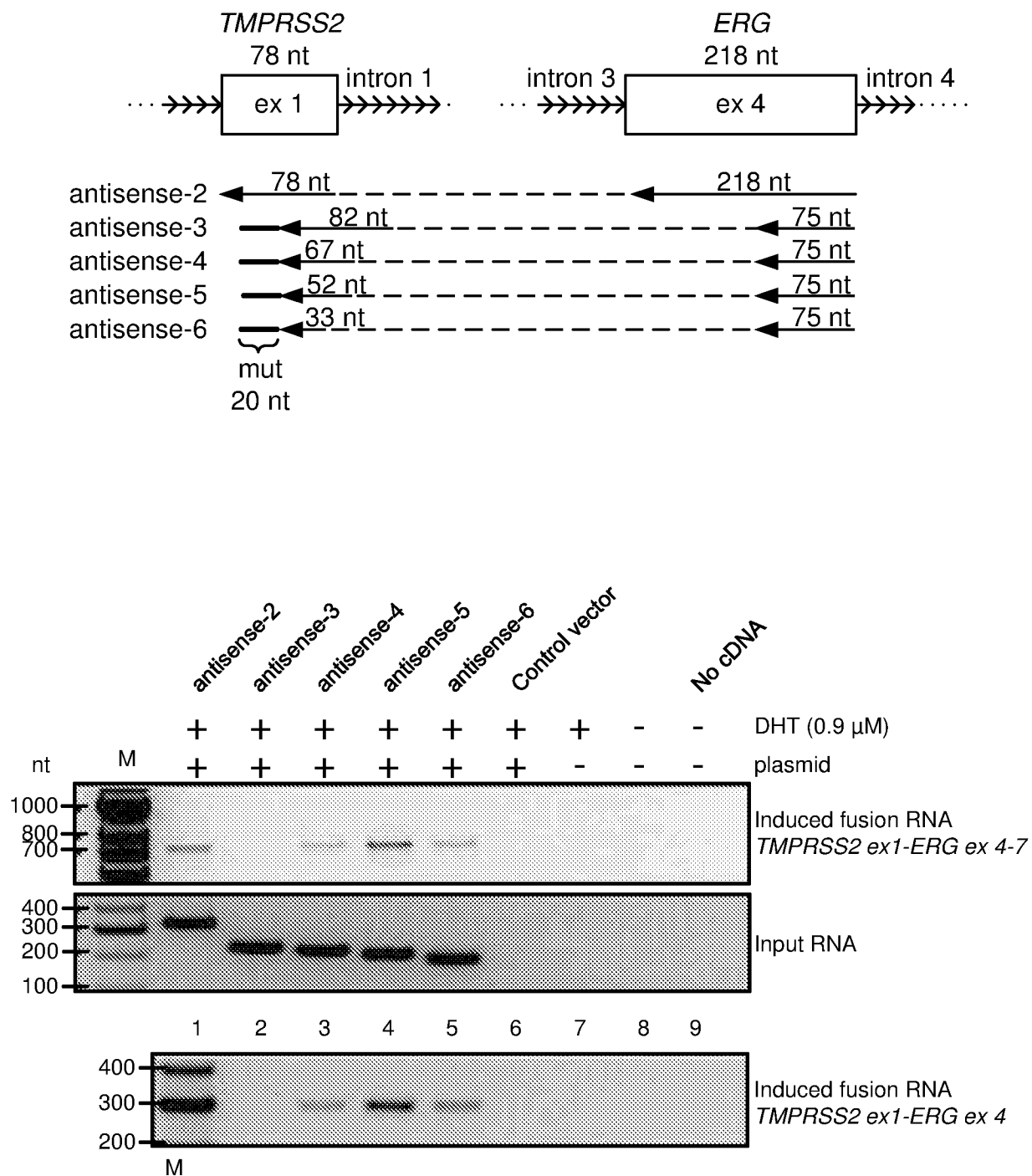


FIG. 1B

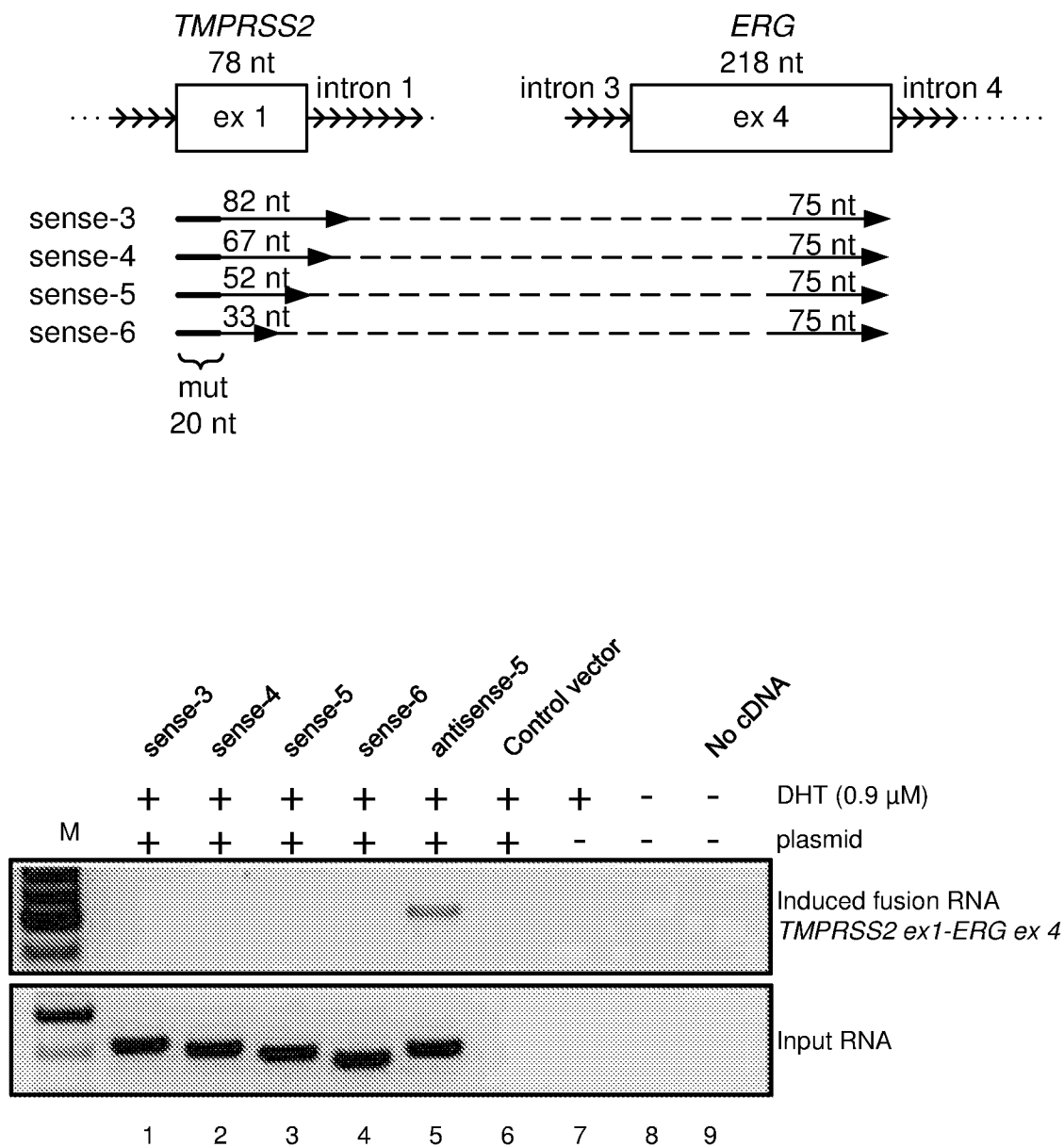


FIG. 1C

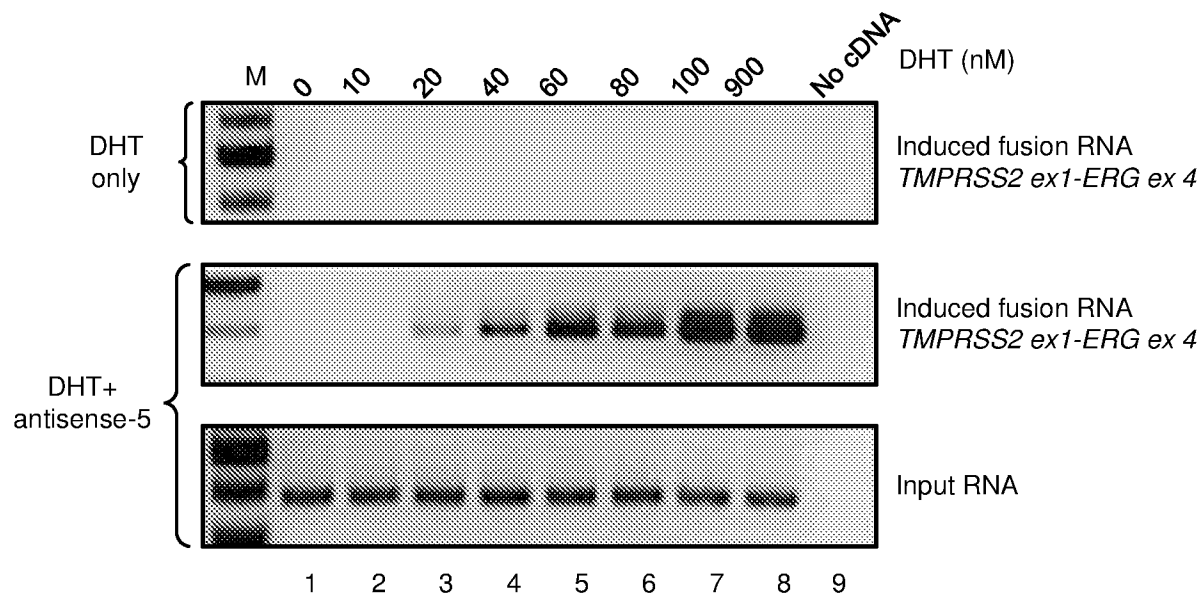


FIG. 1D

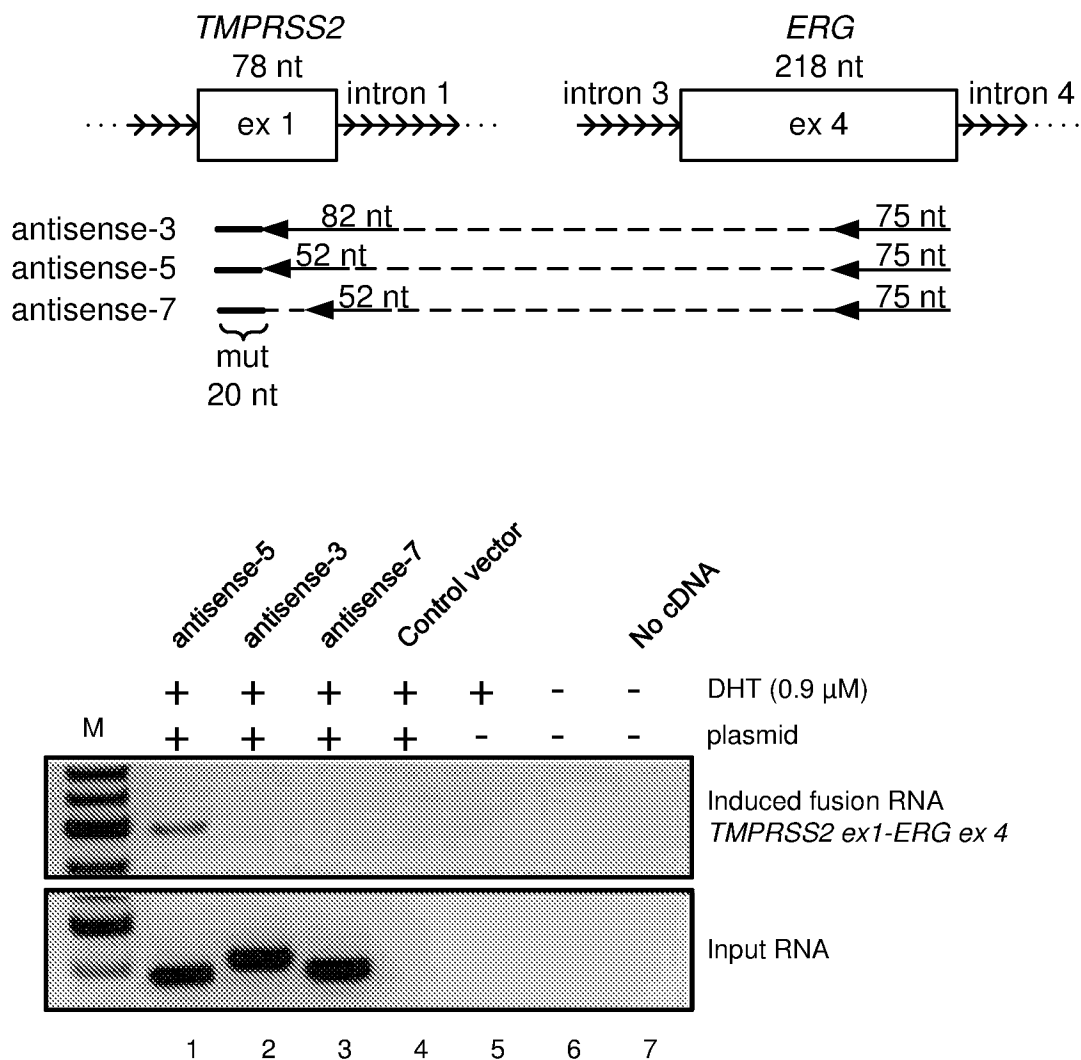


FIG. 1E

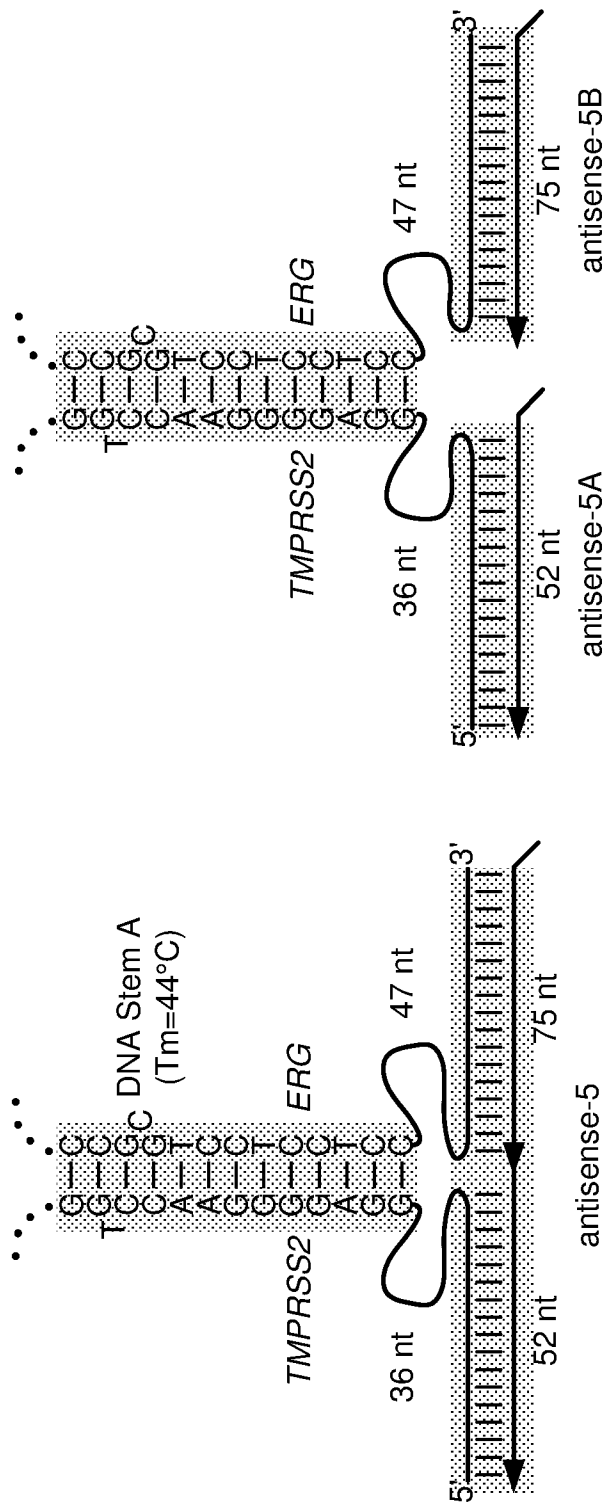


FIG. 2A

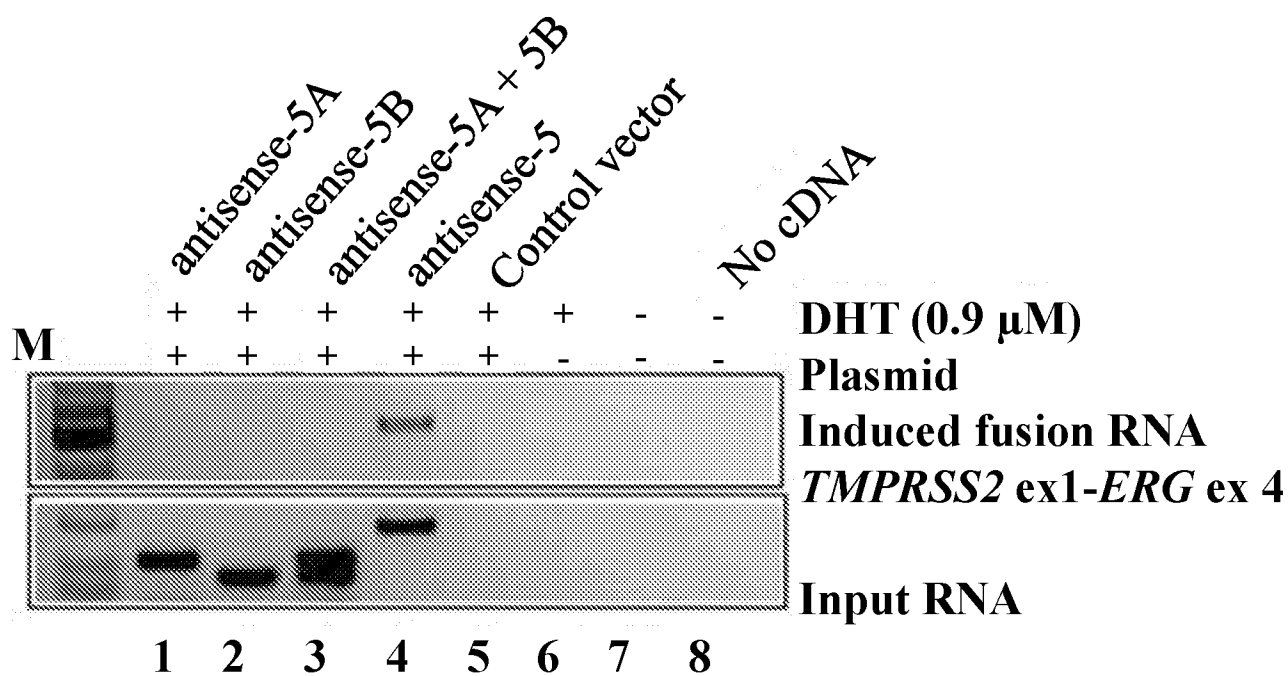


FIG. 2B

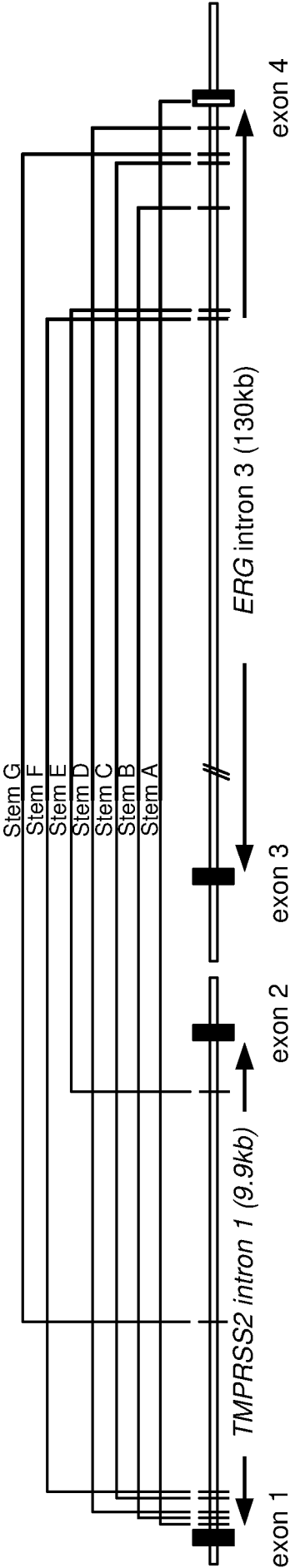


FIG. 2C

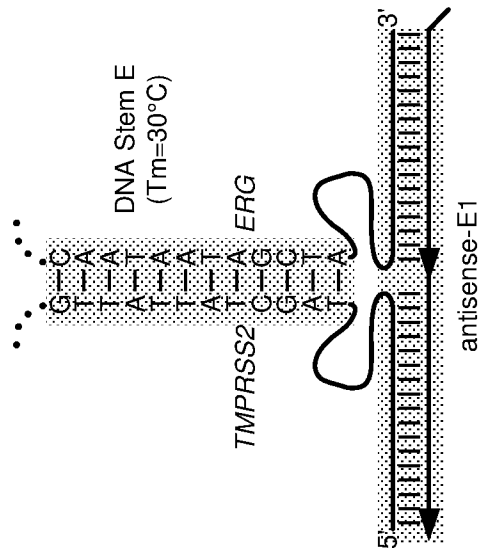
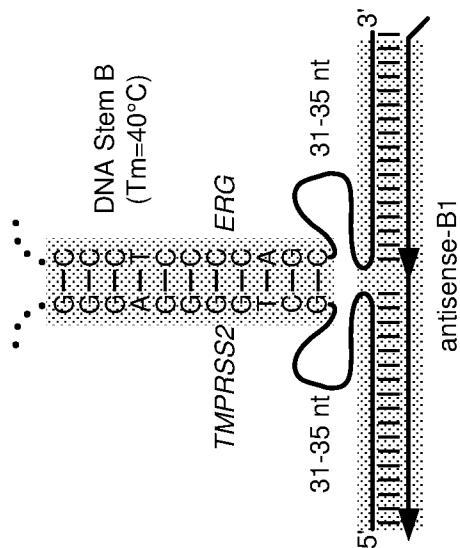
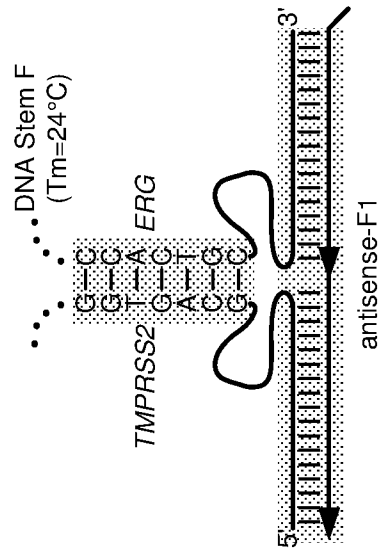
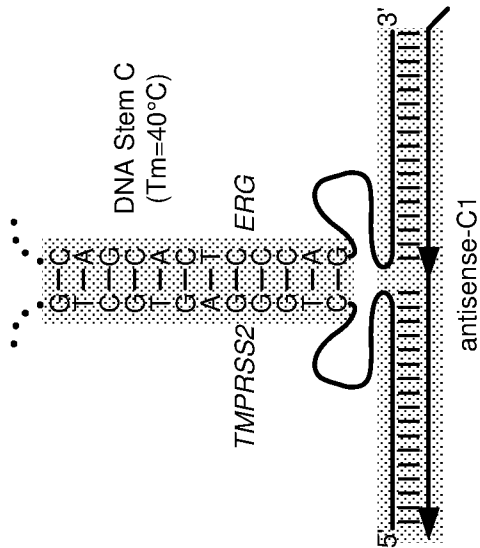
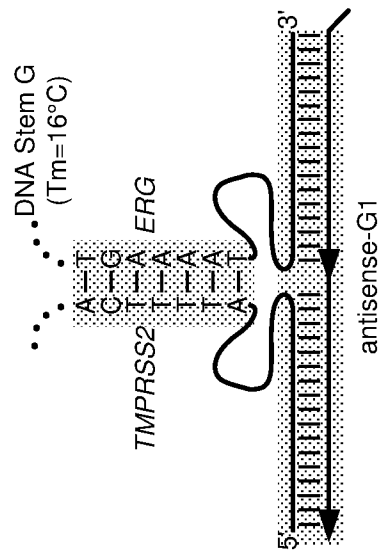
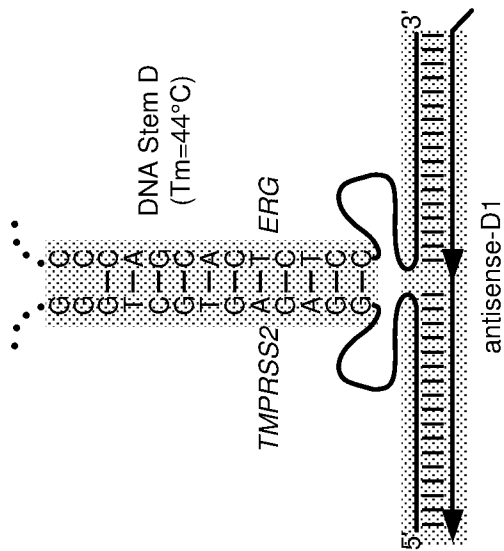


FIG. 2D

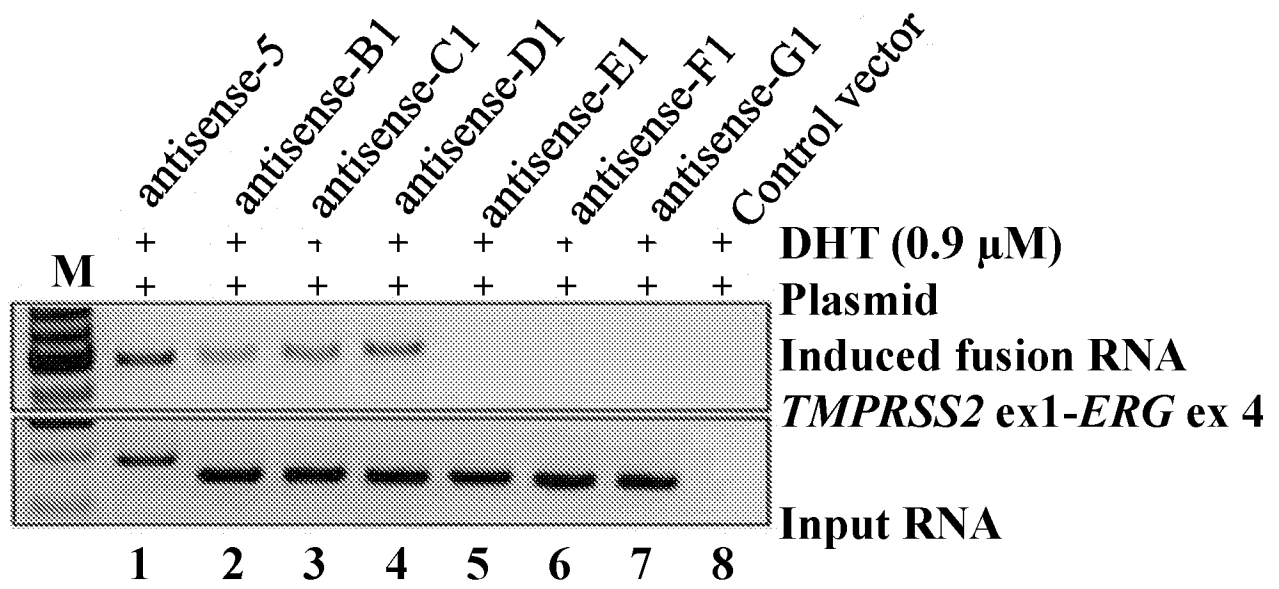


FIG. 2E

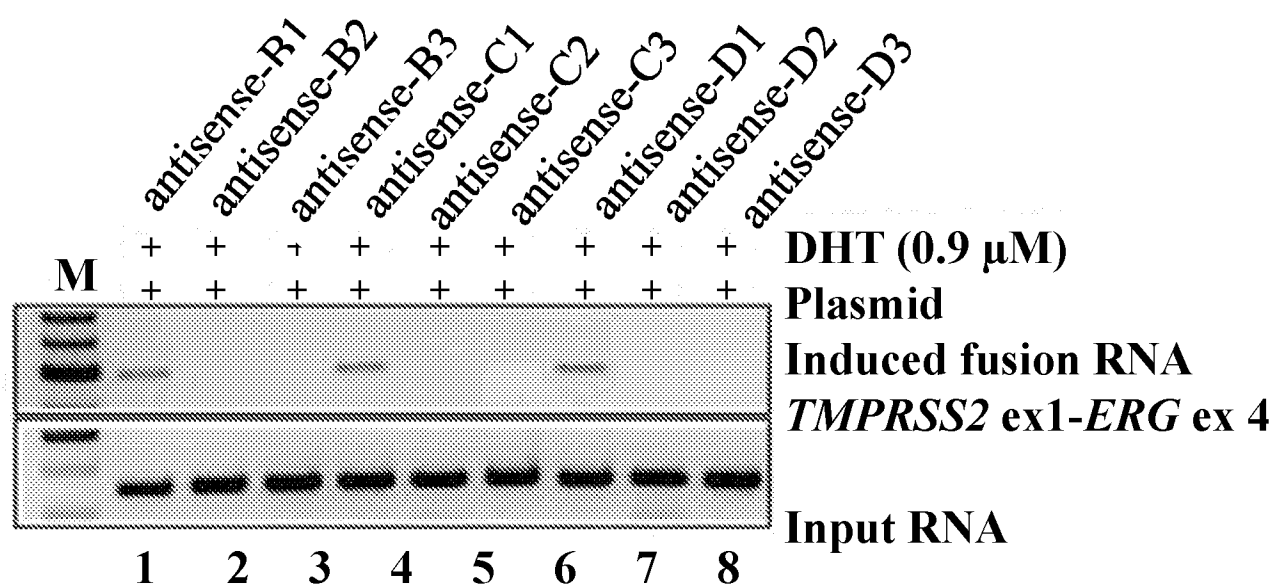


FIG. 2F

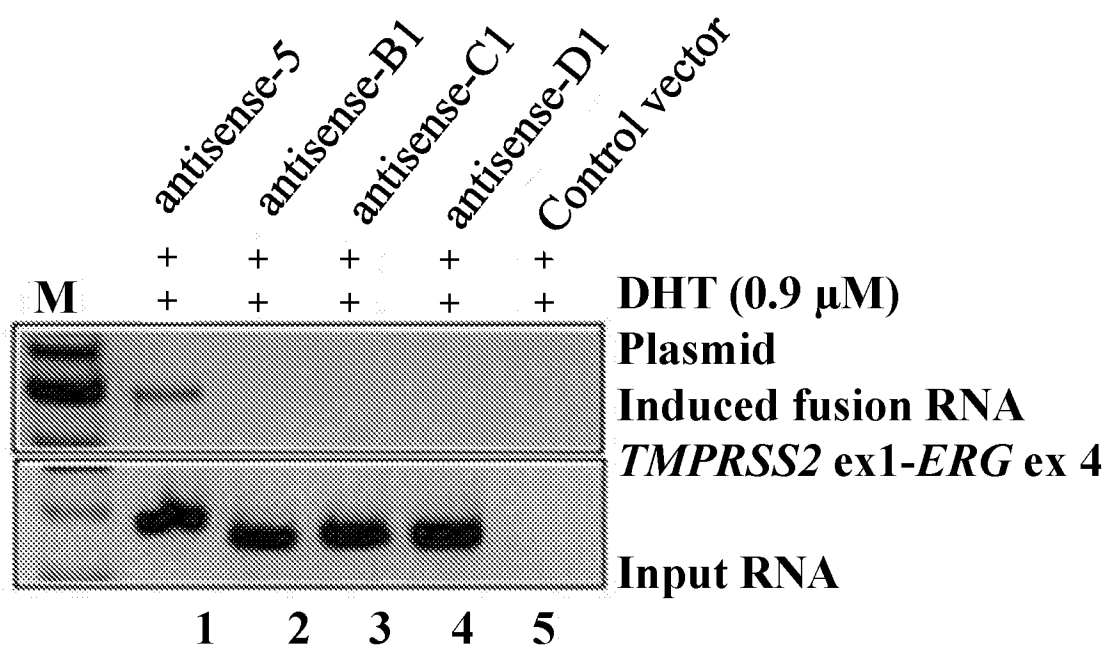


FIG. 2G

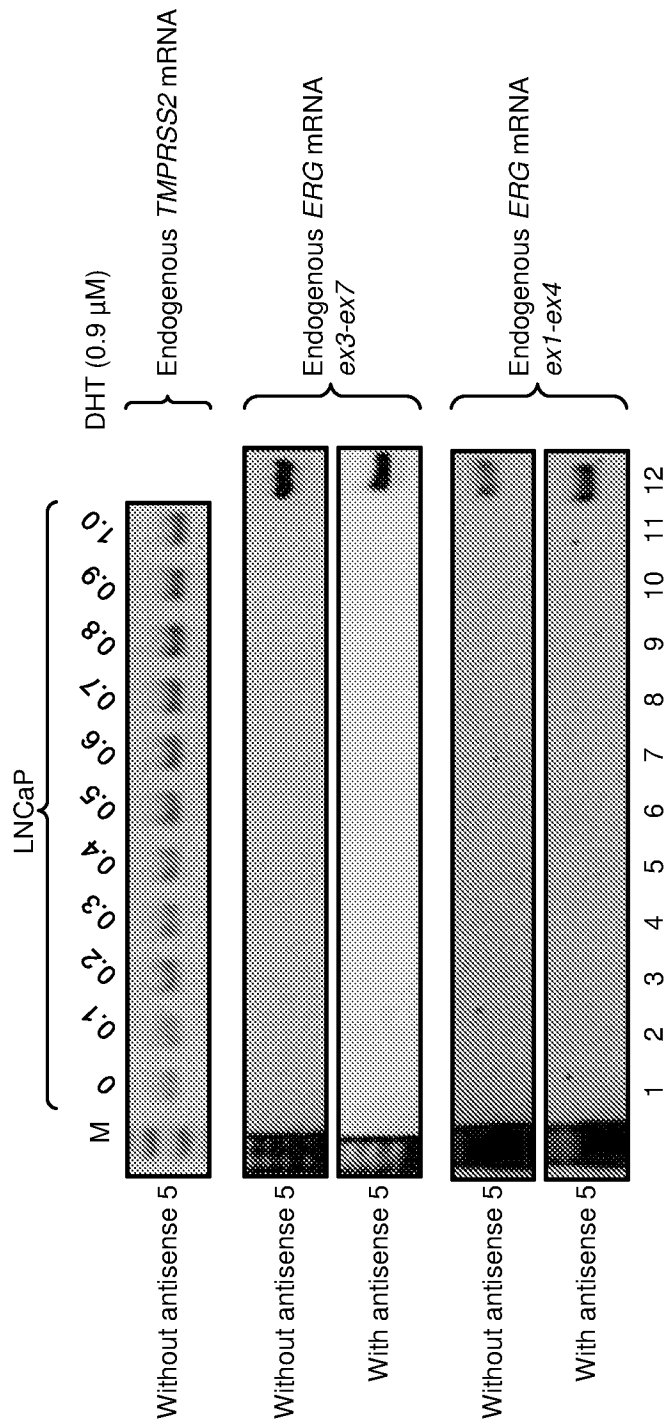


FIG. 3A

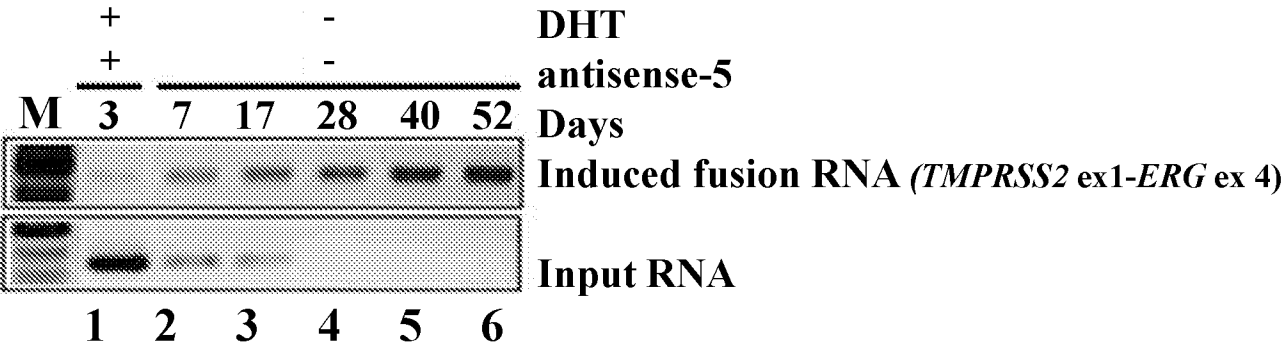
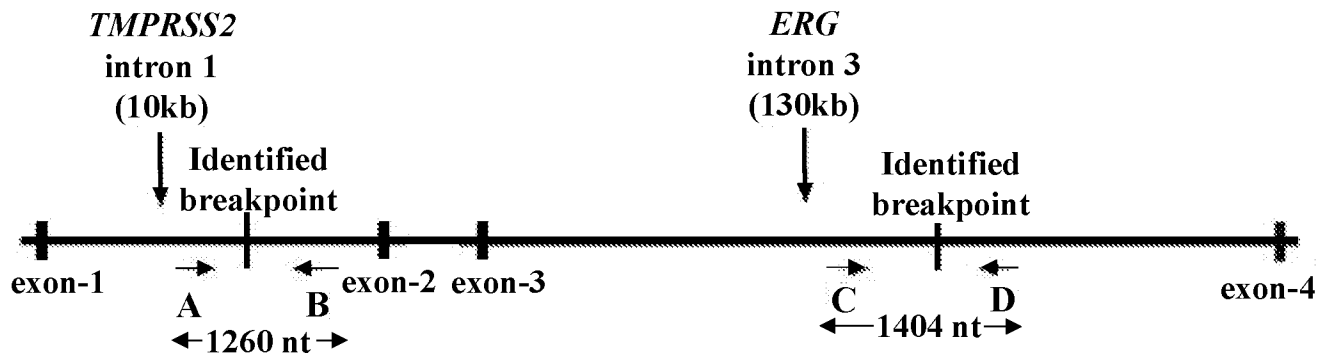


FIG. 3B

**FIG. 3C**

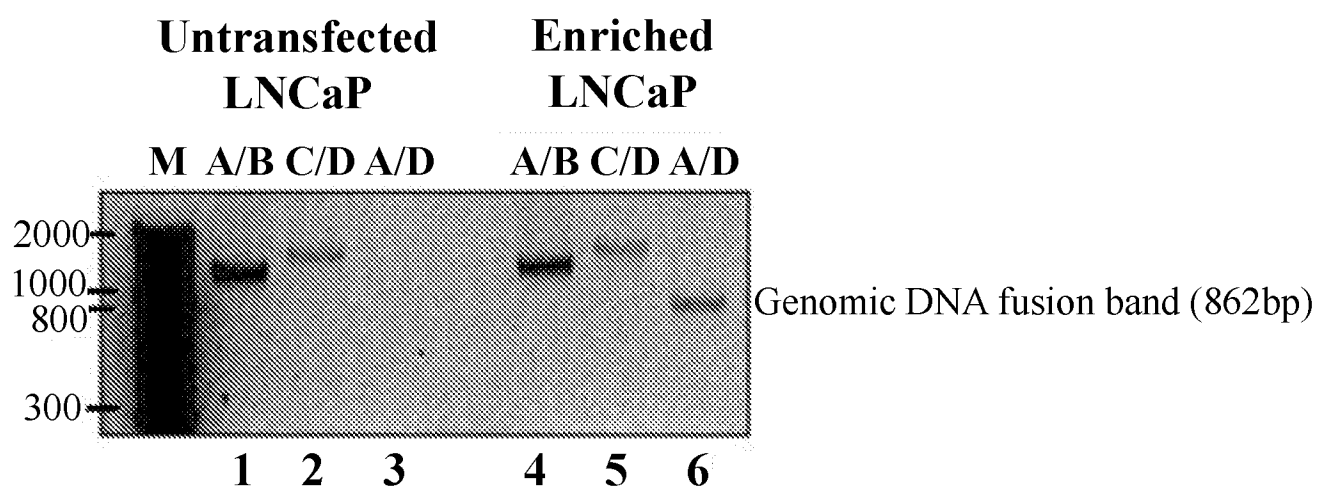


FIG. 3D

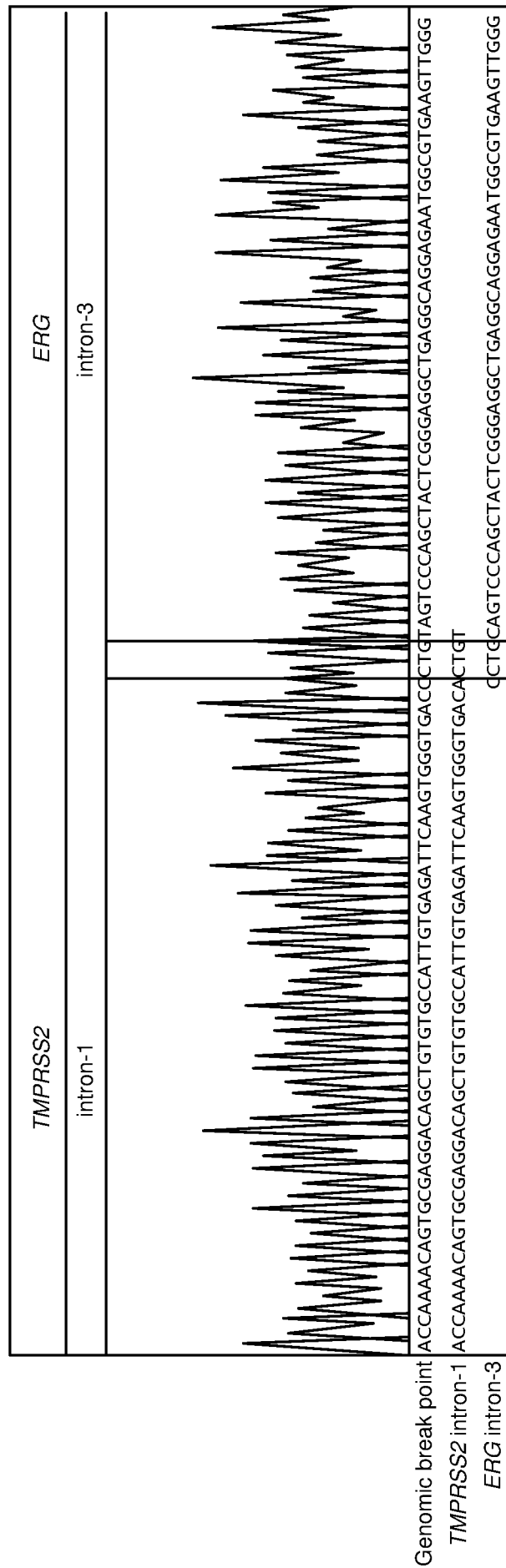


FIG. 3E

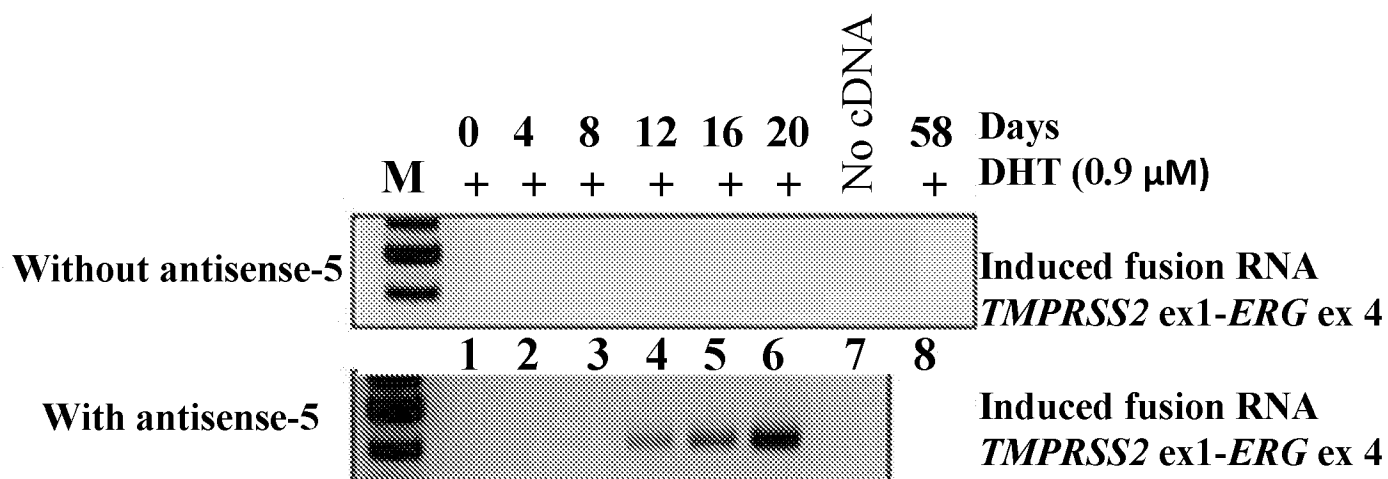


FIG. 3F

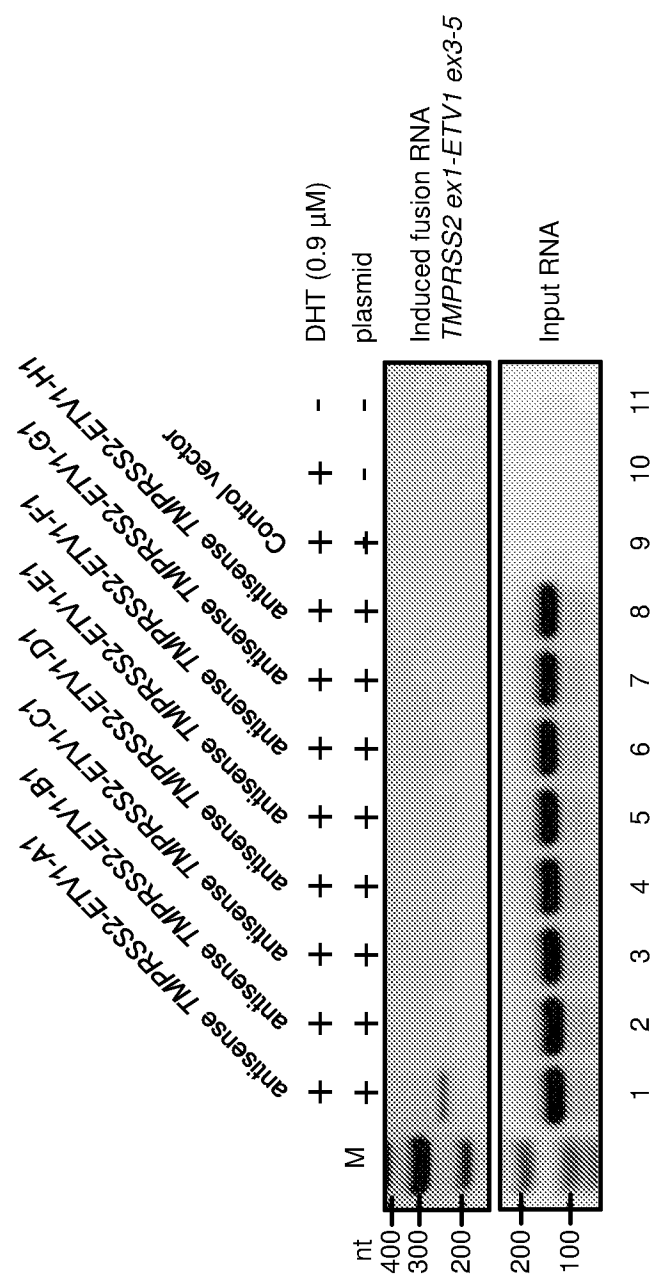


FIG. 4A

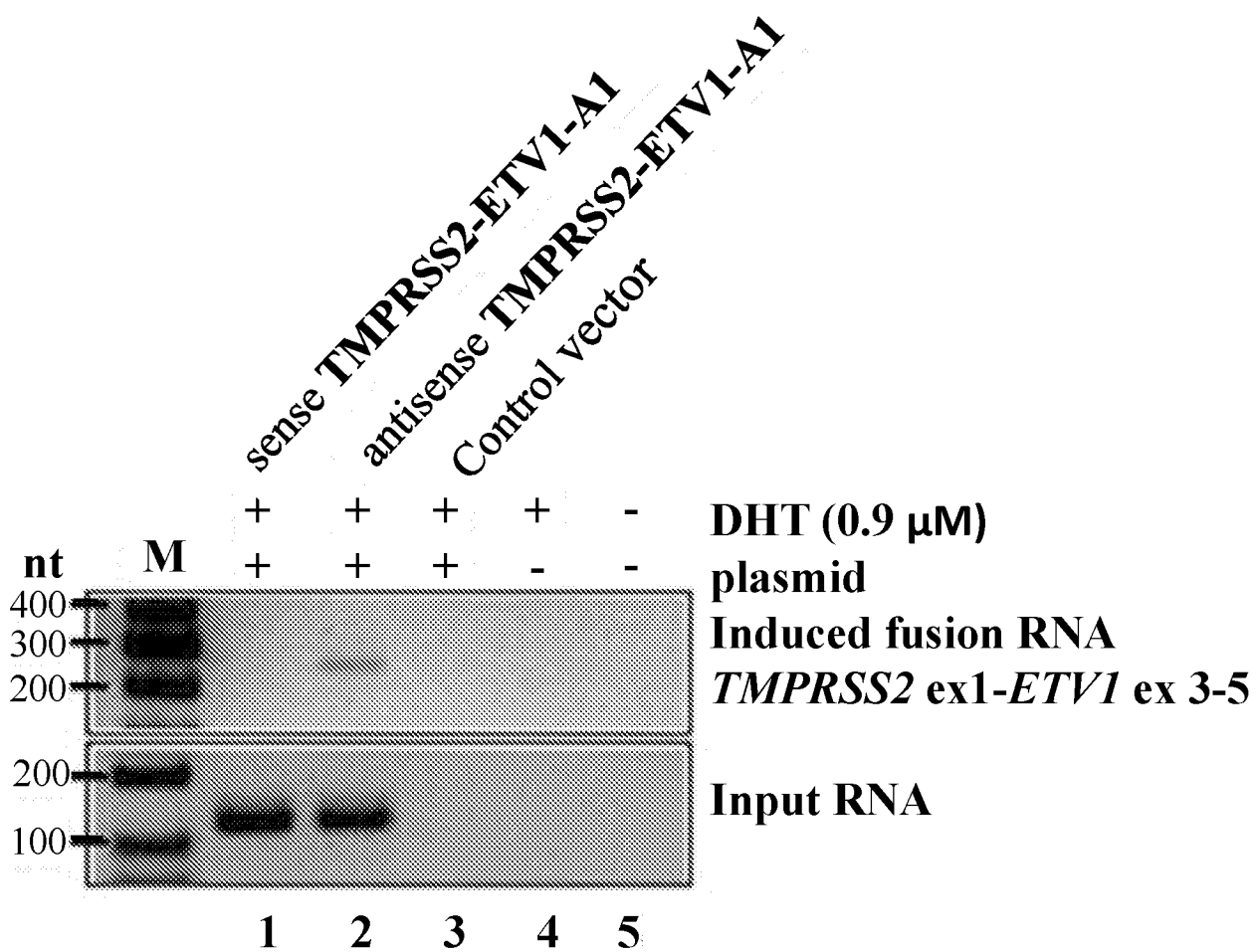


FIG. 4B

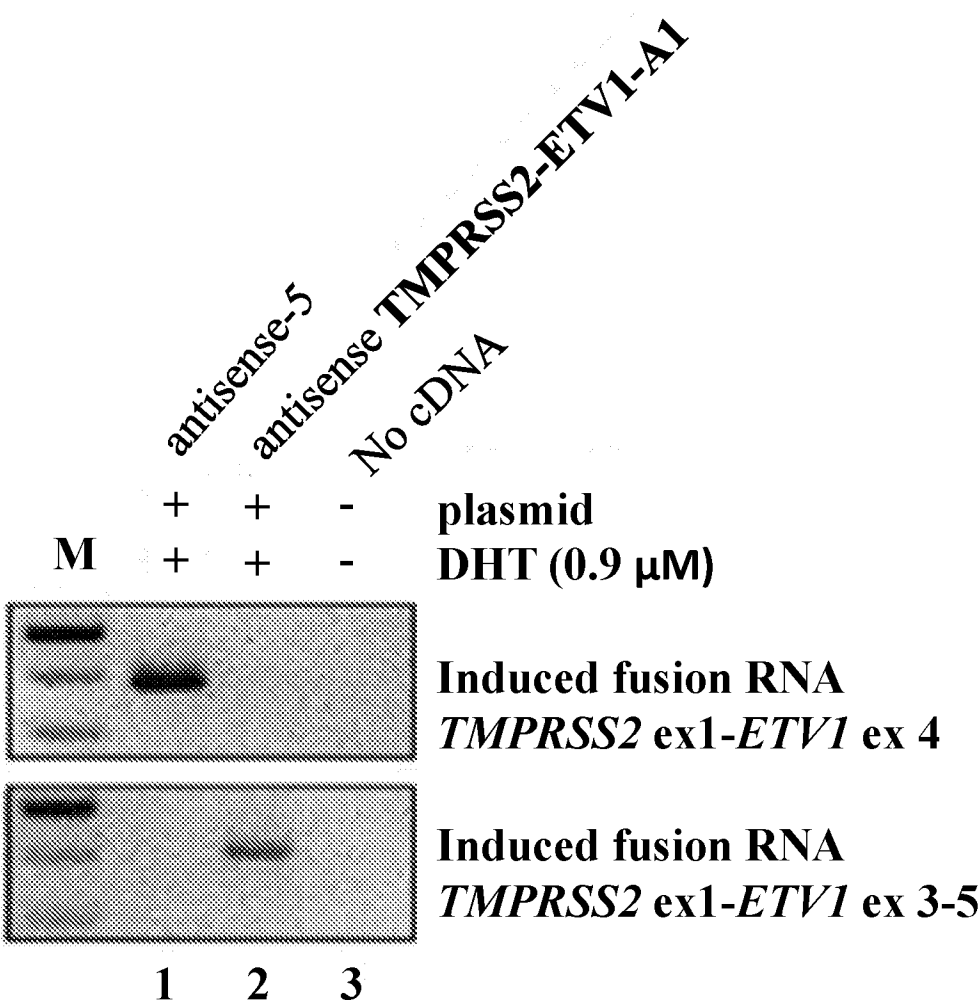


FIG. 4C

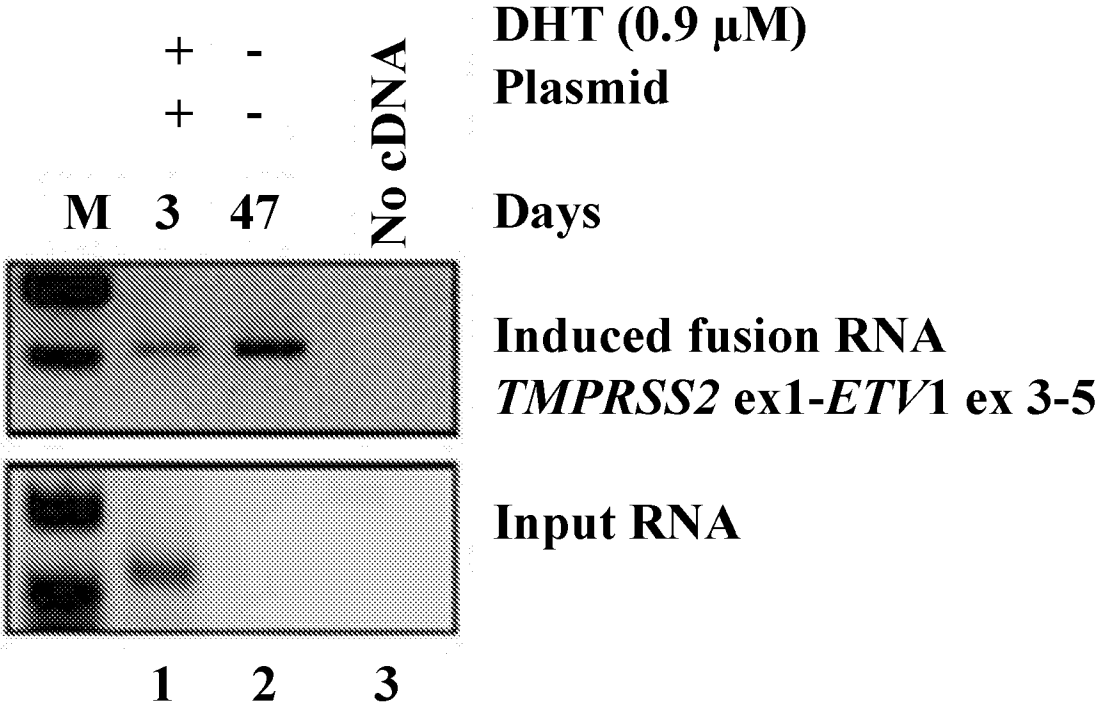


FIG. 4D

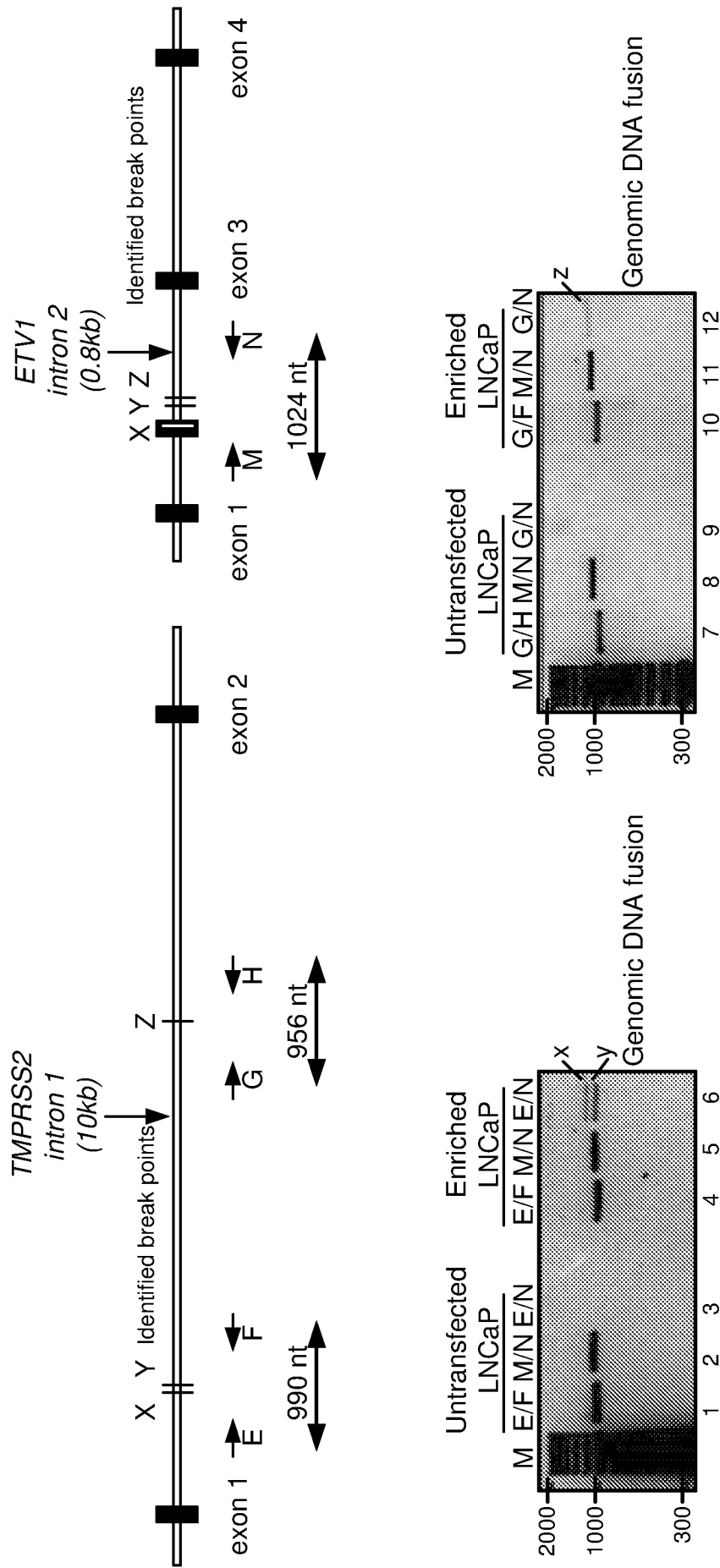
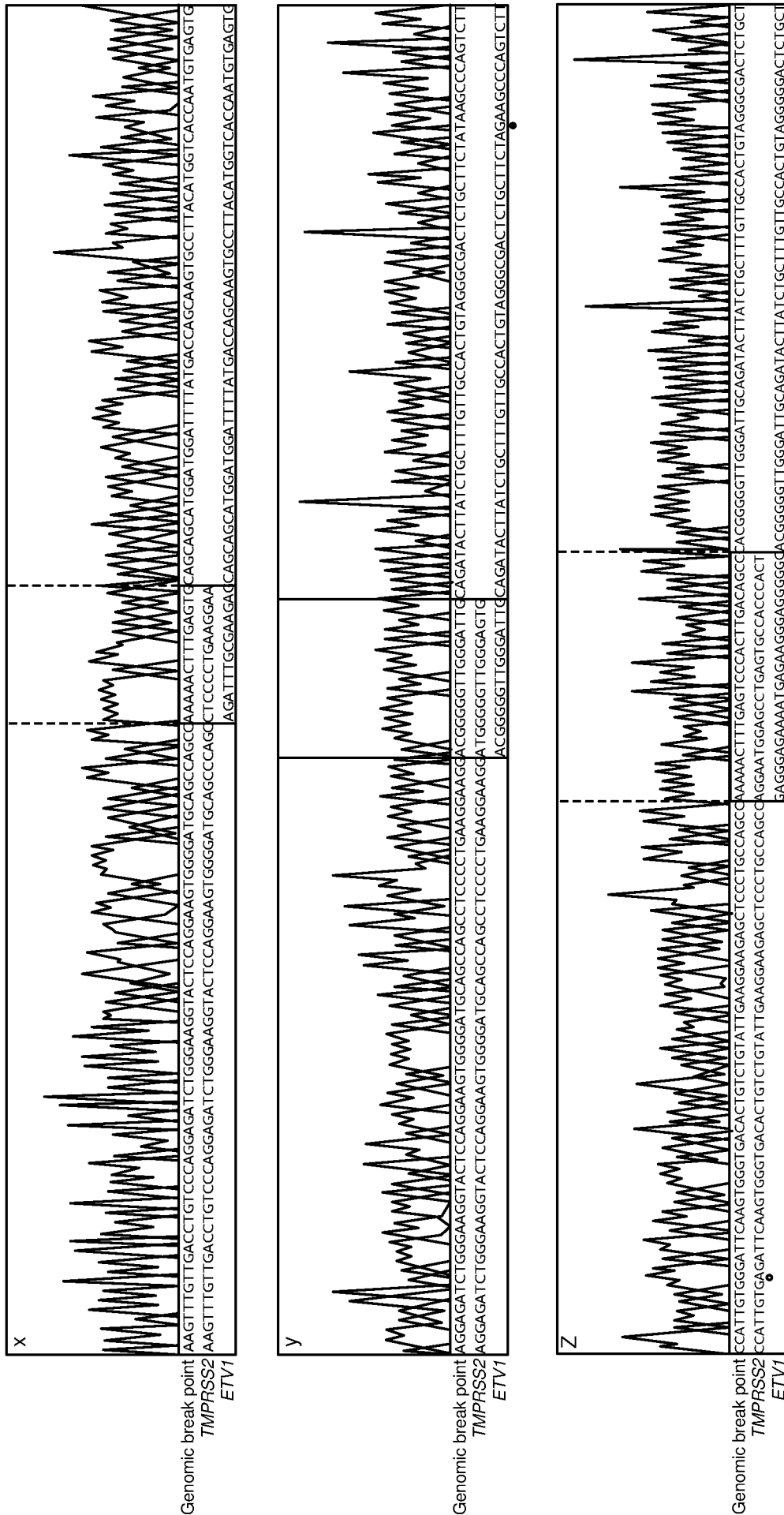
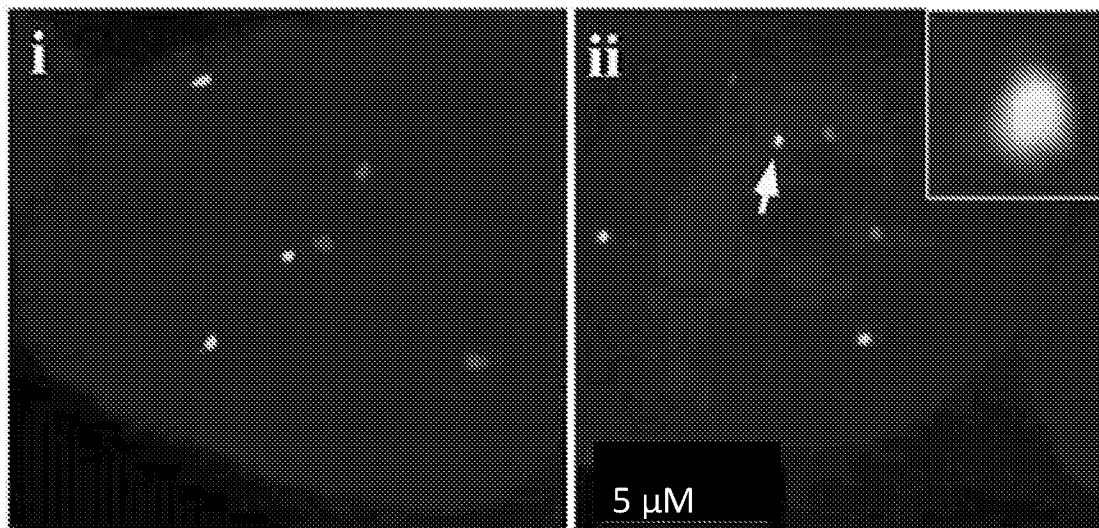


FIG. 4E



	Untransfected LNCaP cells	Enriched LNCaP population
Total No. of imaged cells	620	3301
No. of cells positive for fusion genes	0	30
%	0	0.91

**FIG. 4G**

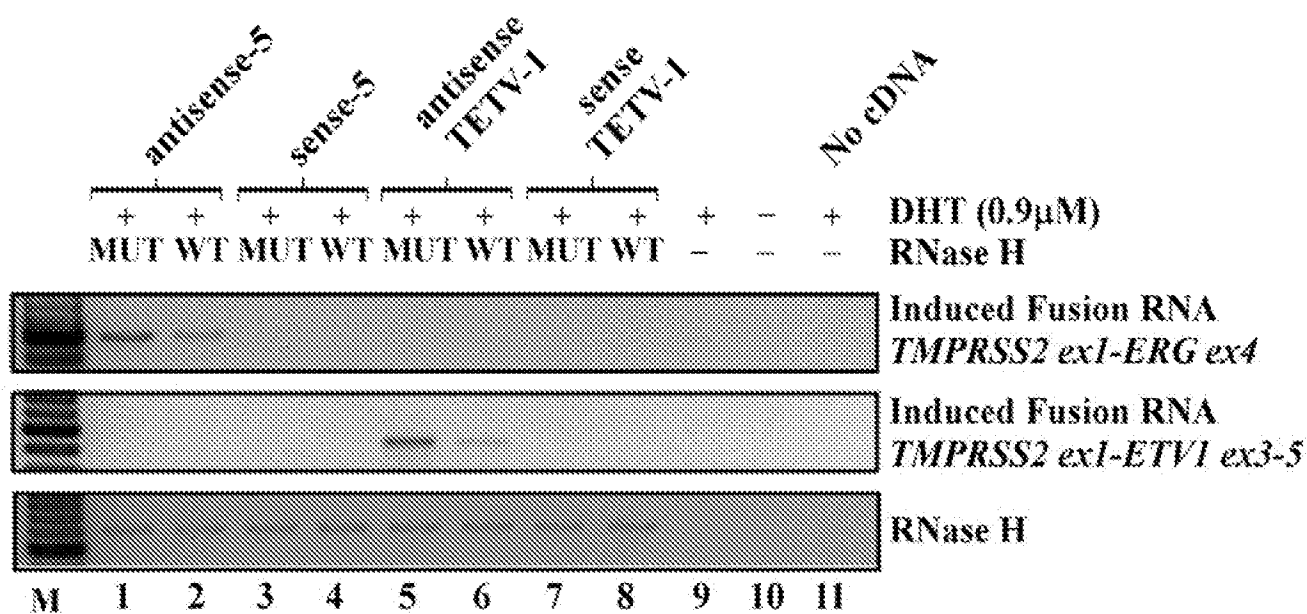
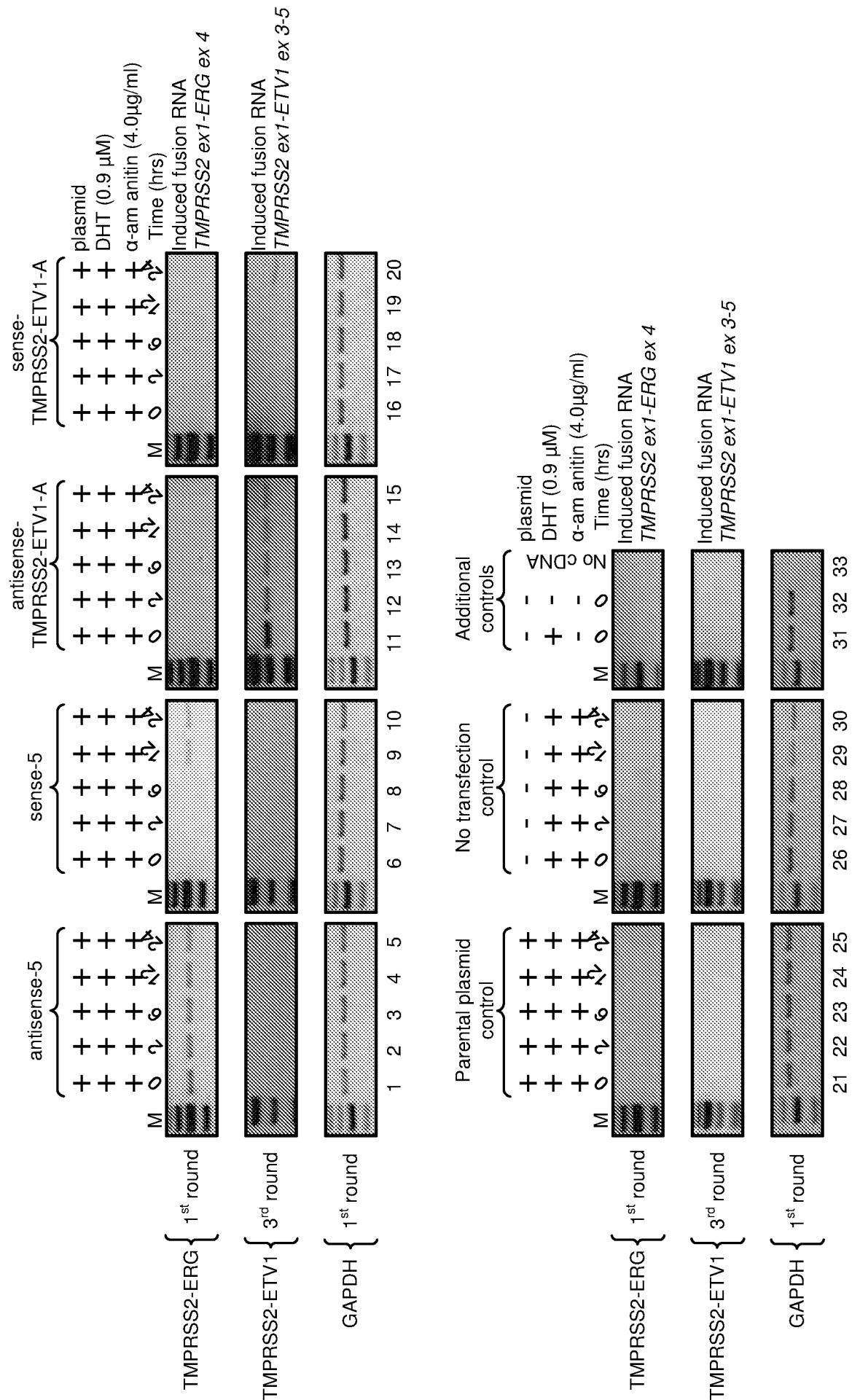


FIG. 5A



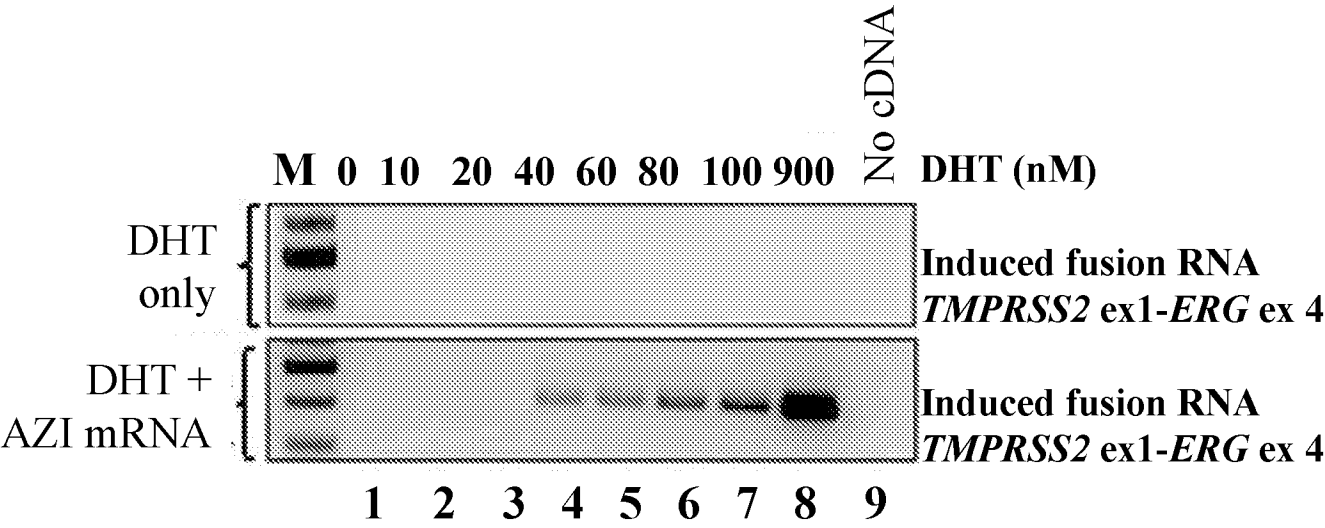


FIG. 6A

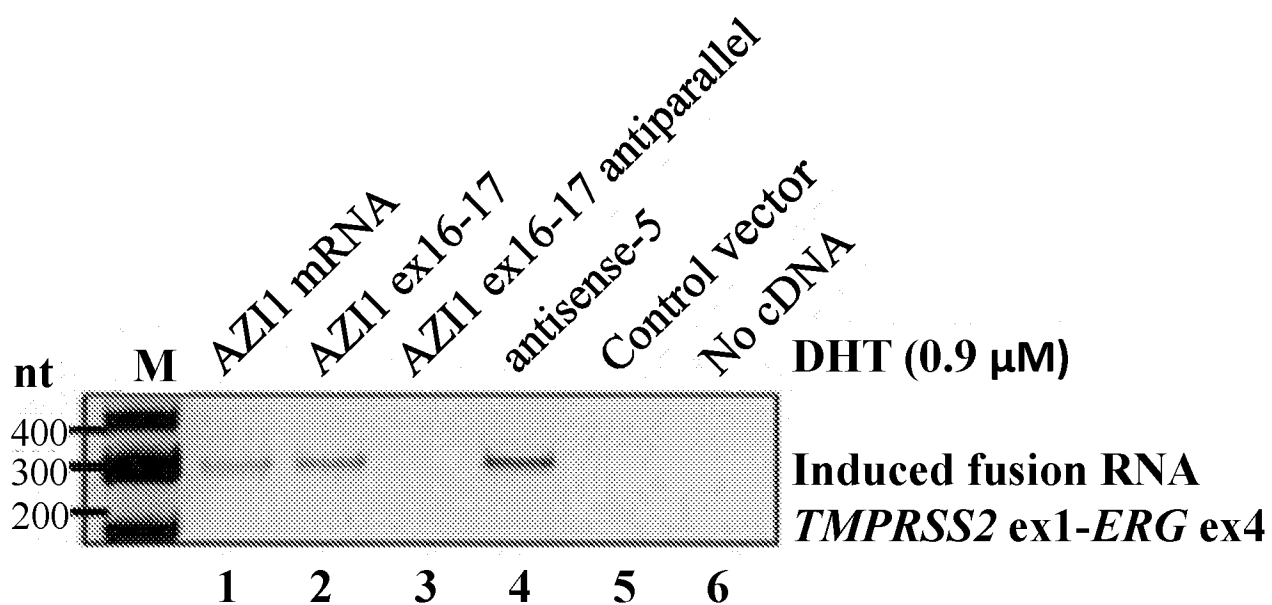


FIG. 6B

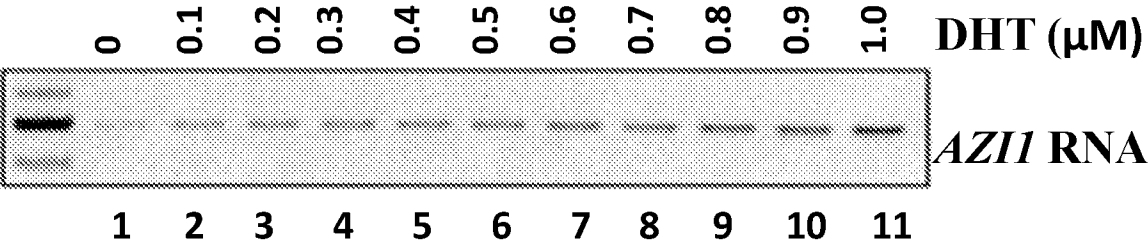
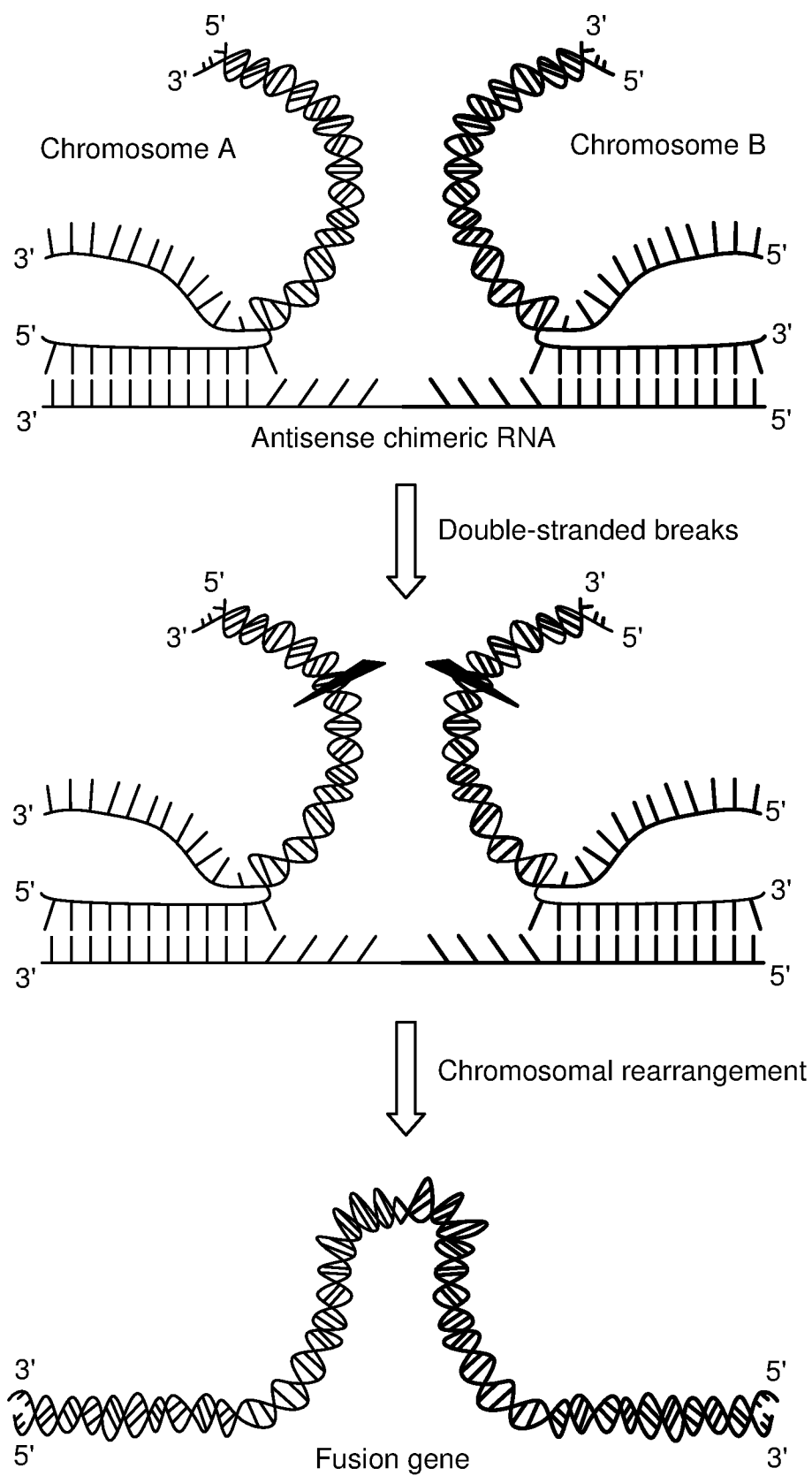
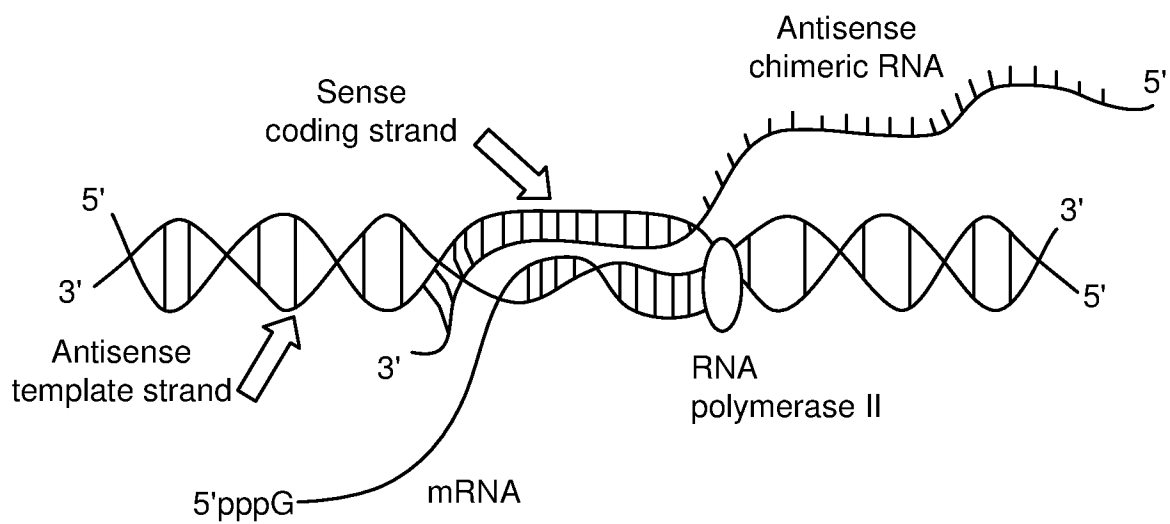
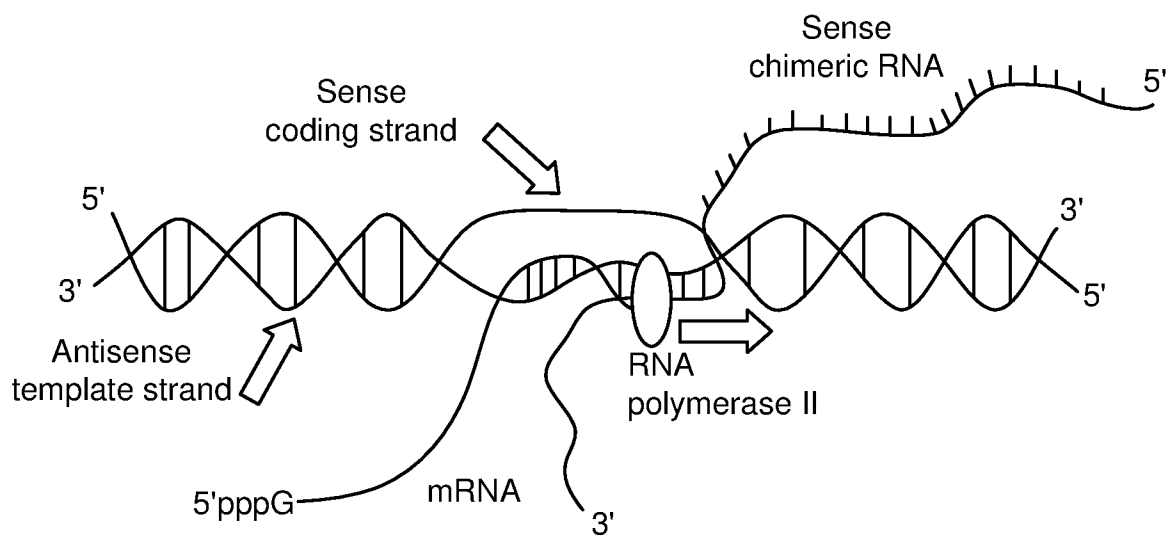
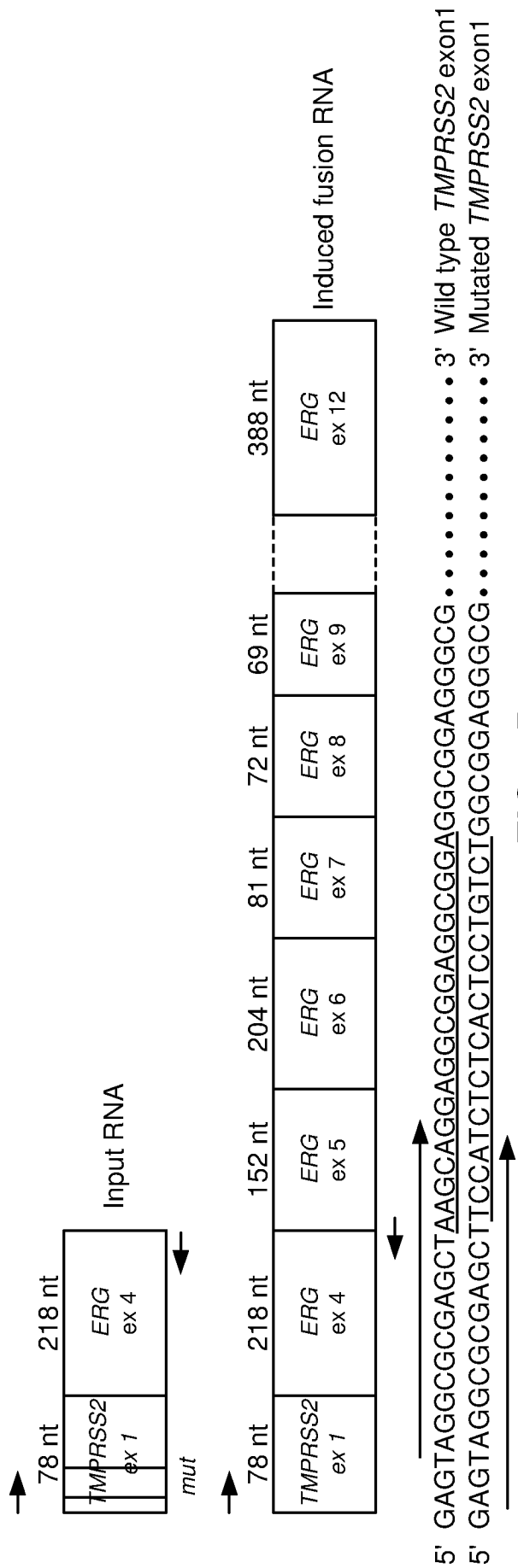
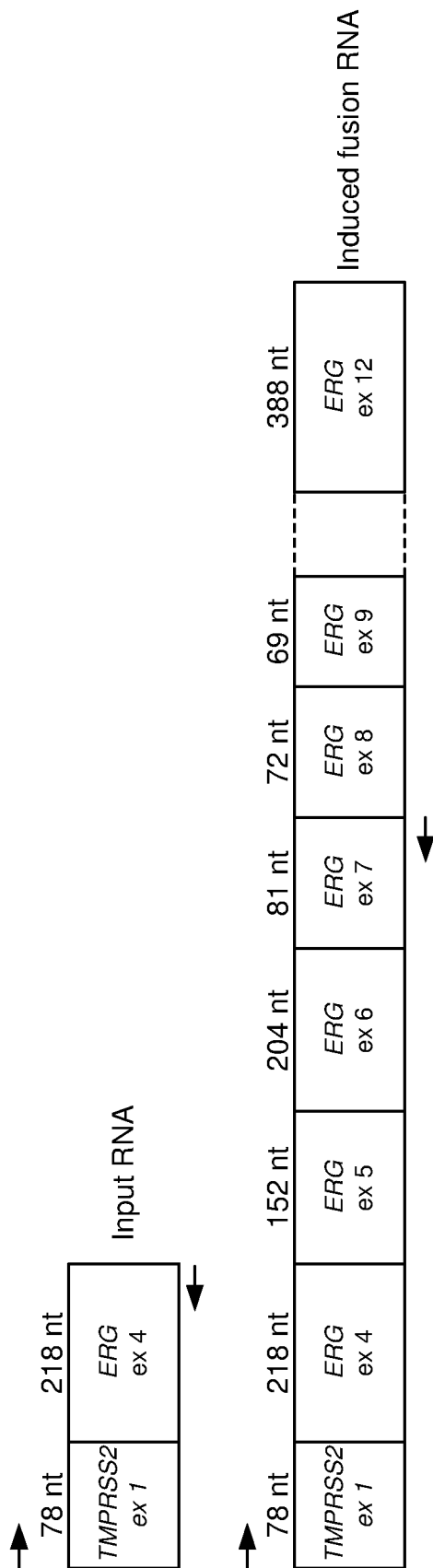
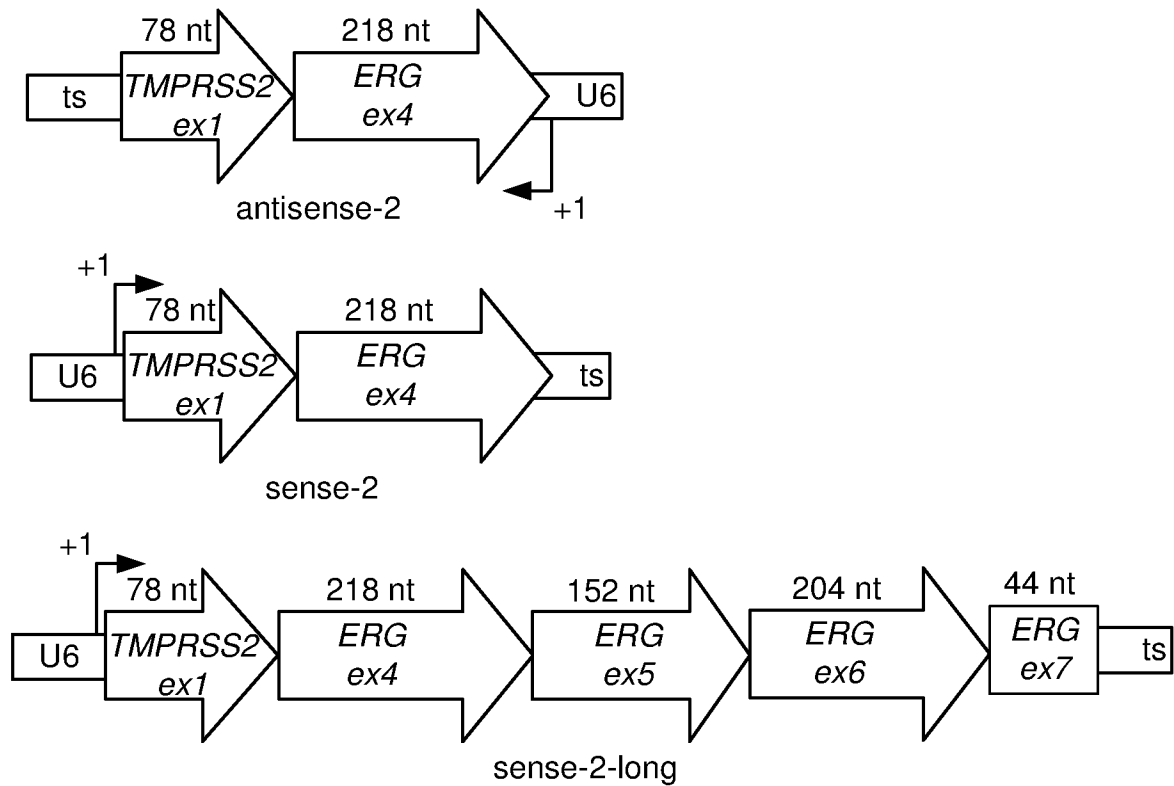
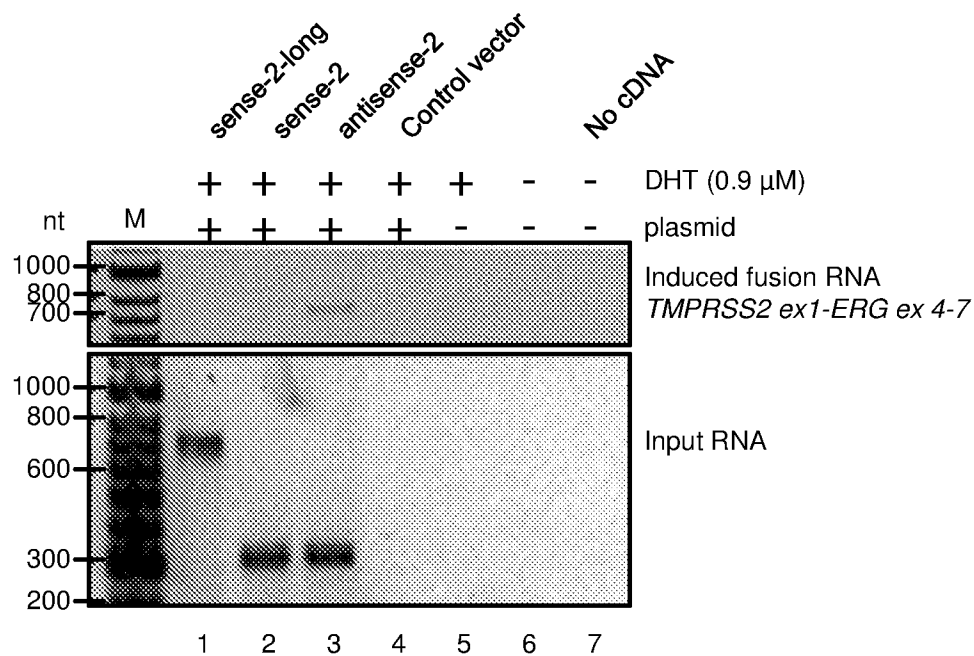


FIG. 6C

**FIG. 7A**

**FIG. 7B****FIG. 7C**



**FIG. 9A****FIG. 9B**

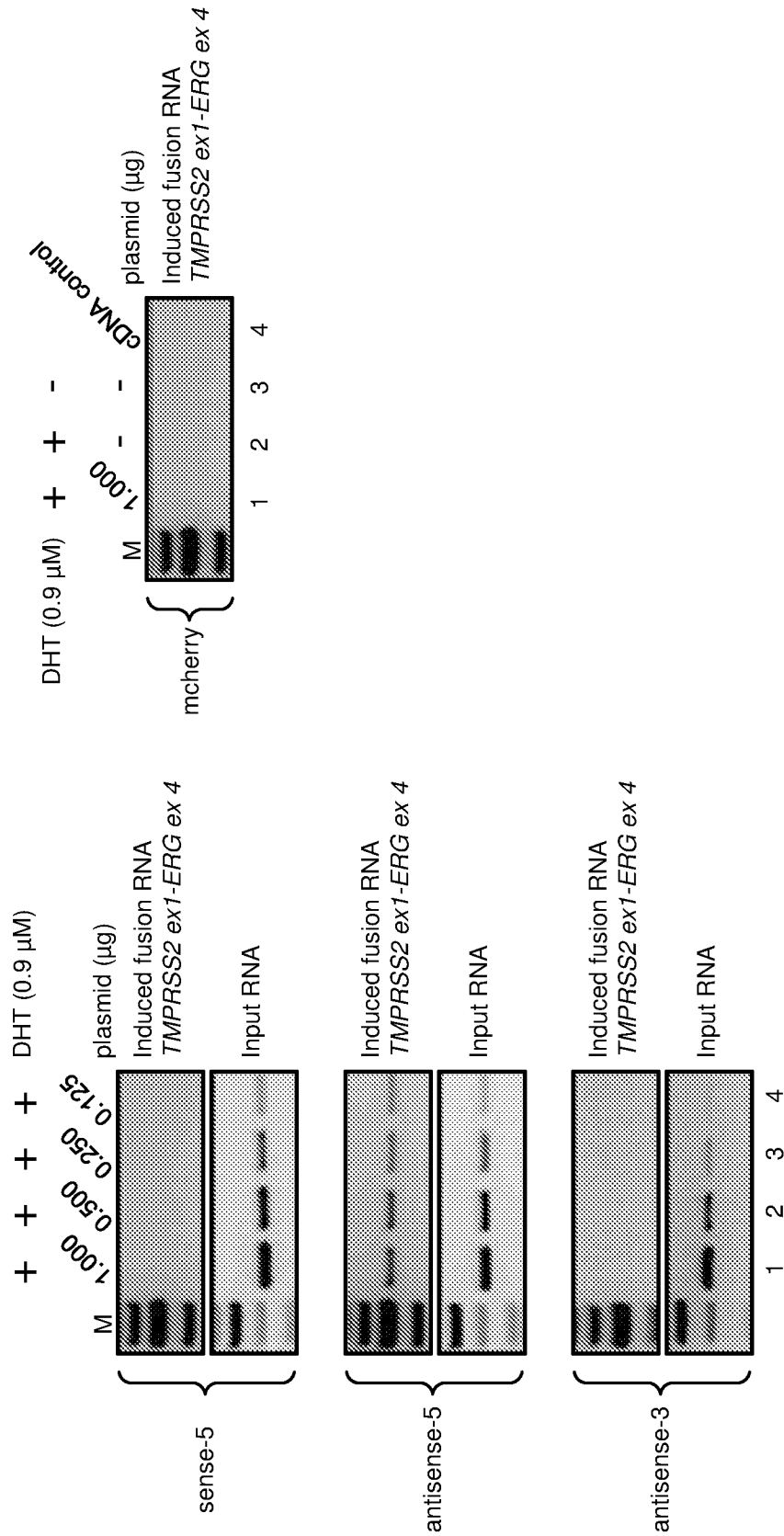
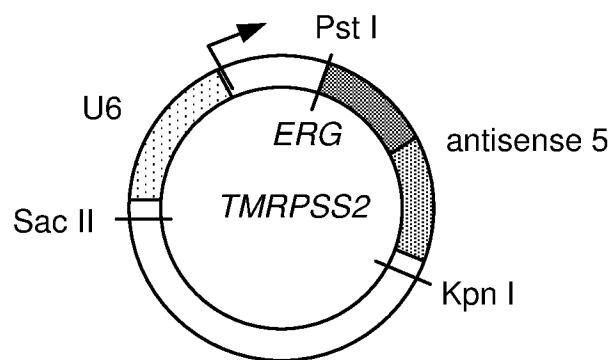
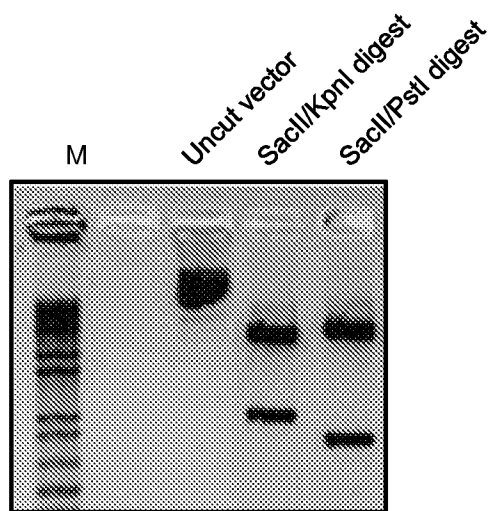
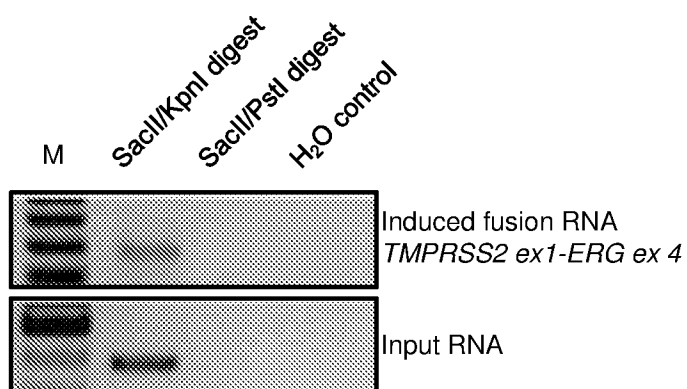


FIG. 10

**FIG. 11A**

Digested DNA fragments

FIG. 11B**FIG. 11C**

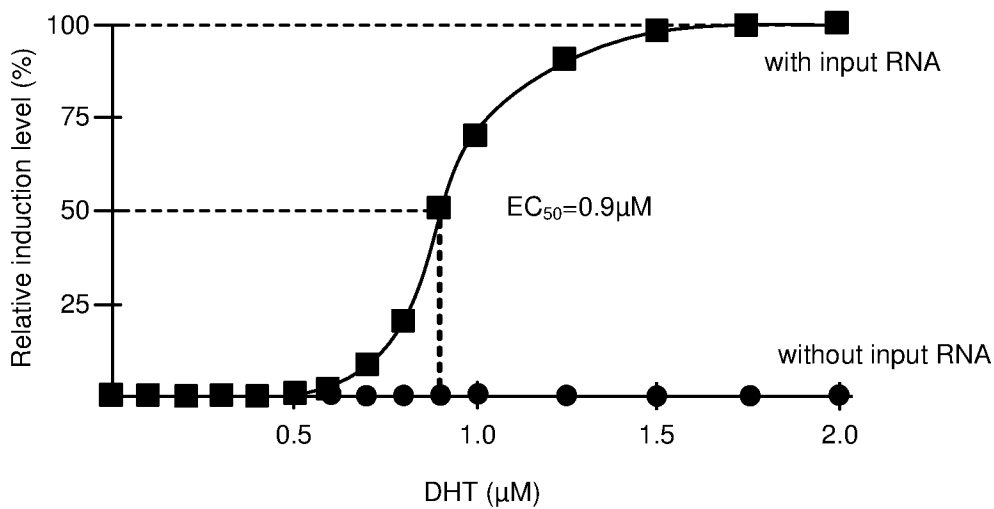
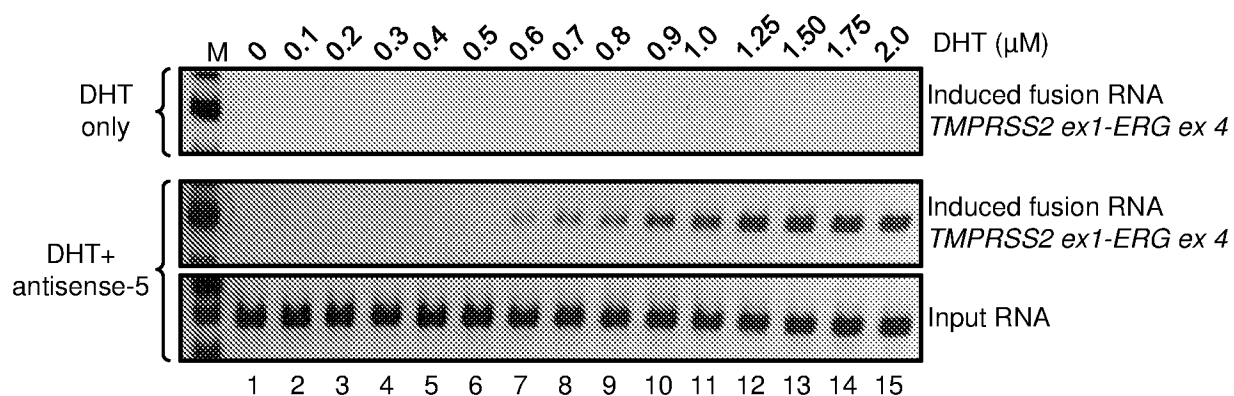
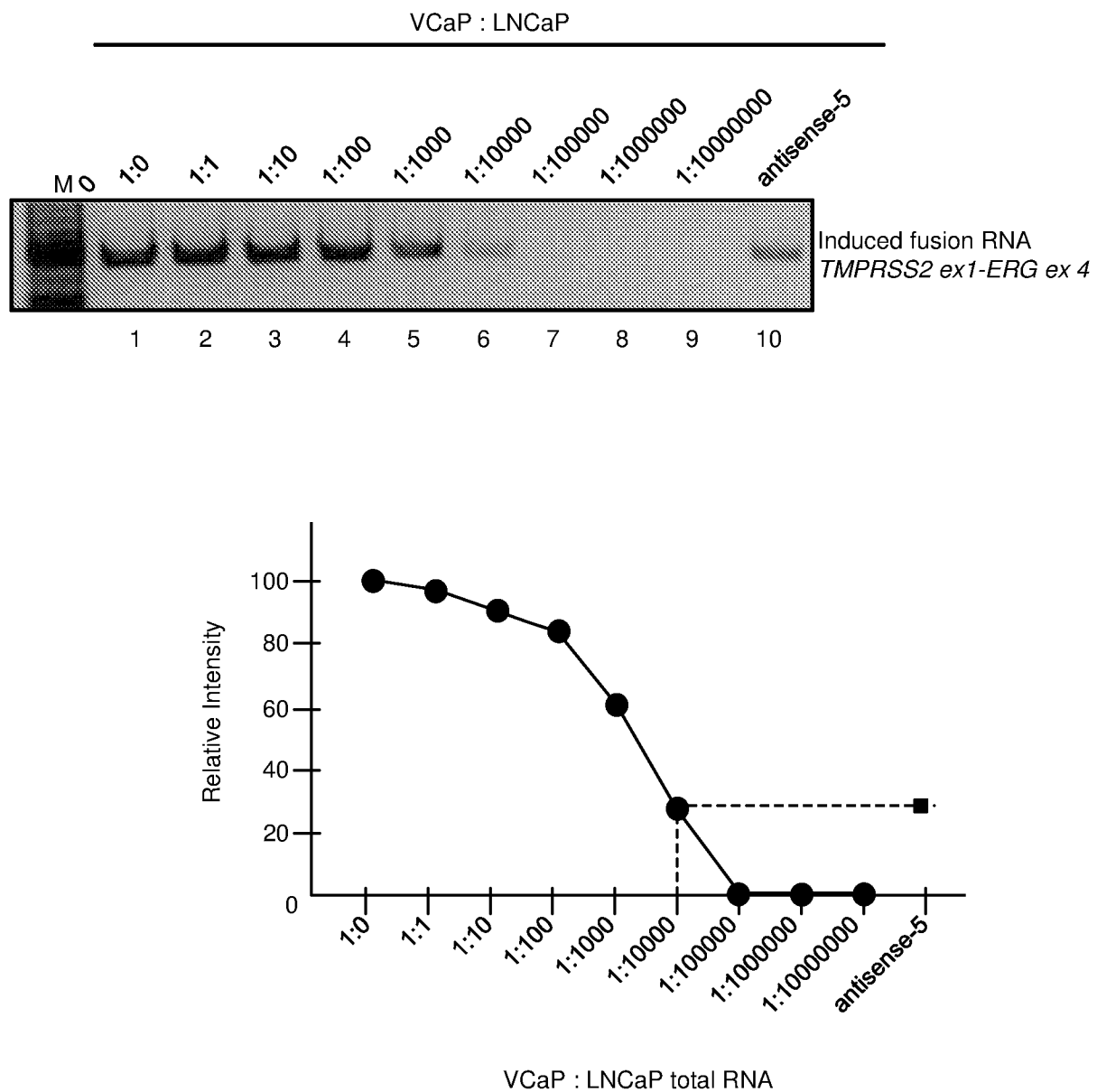


FIG. 12

**FIG. 13**

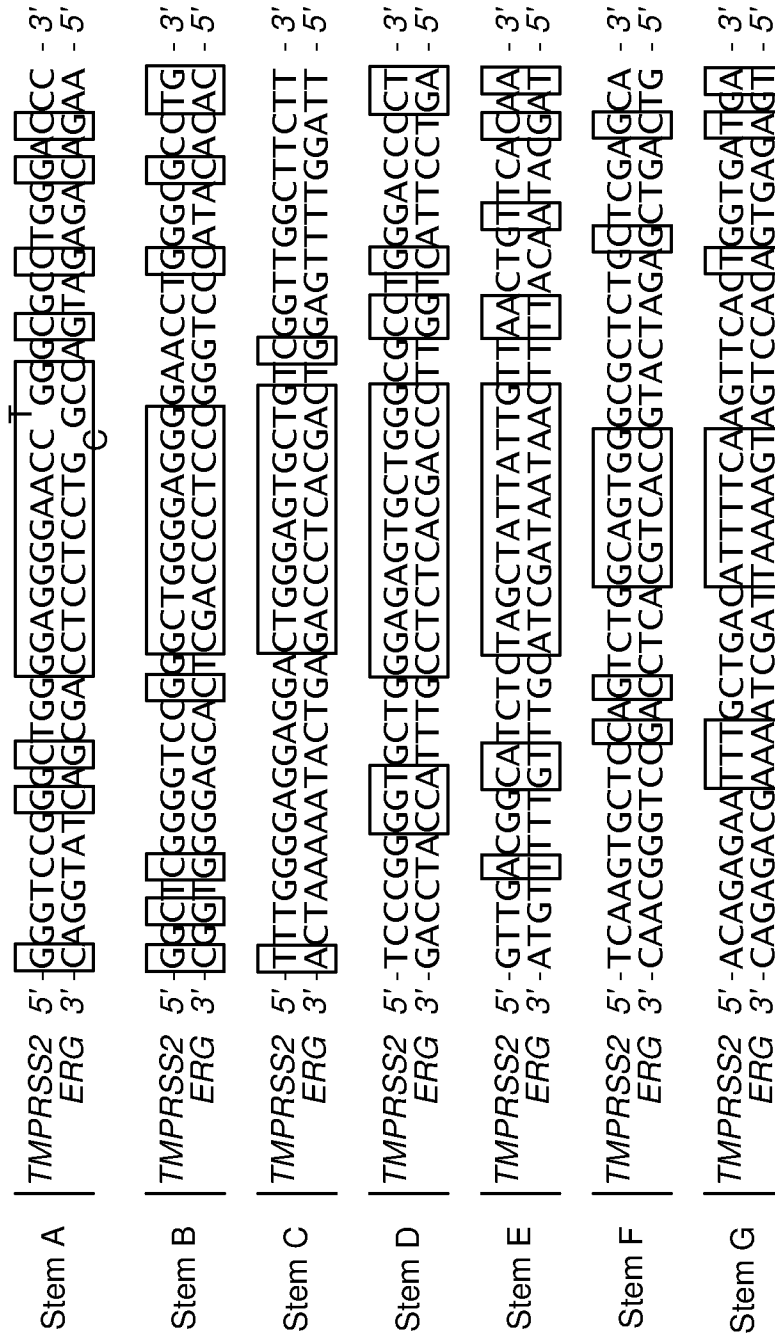


FIG. 14

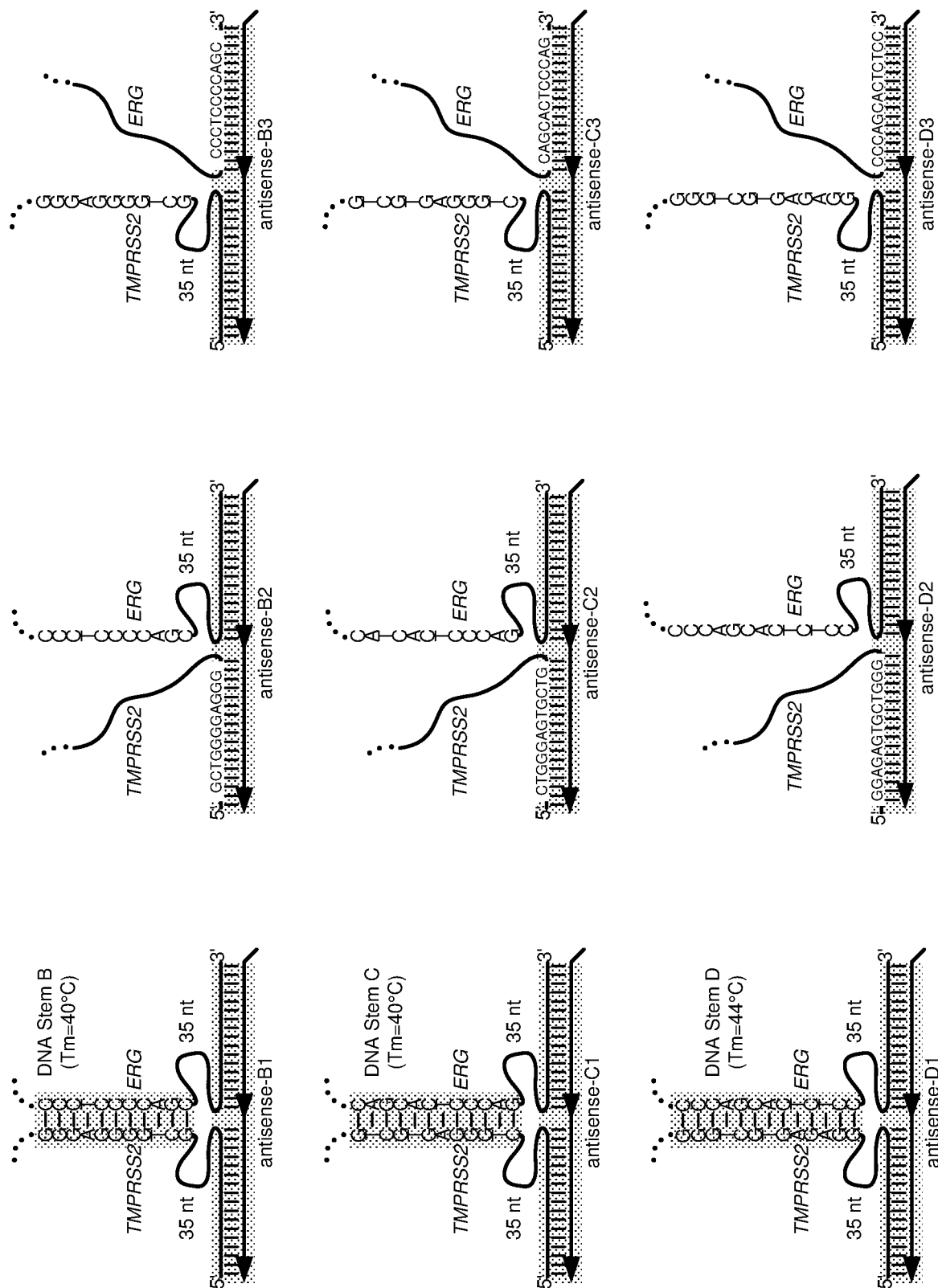


FIG. 15

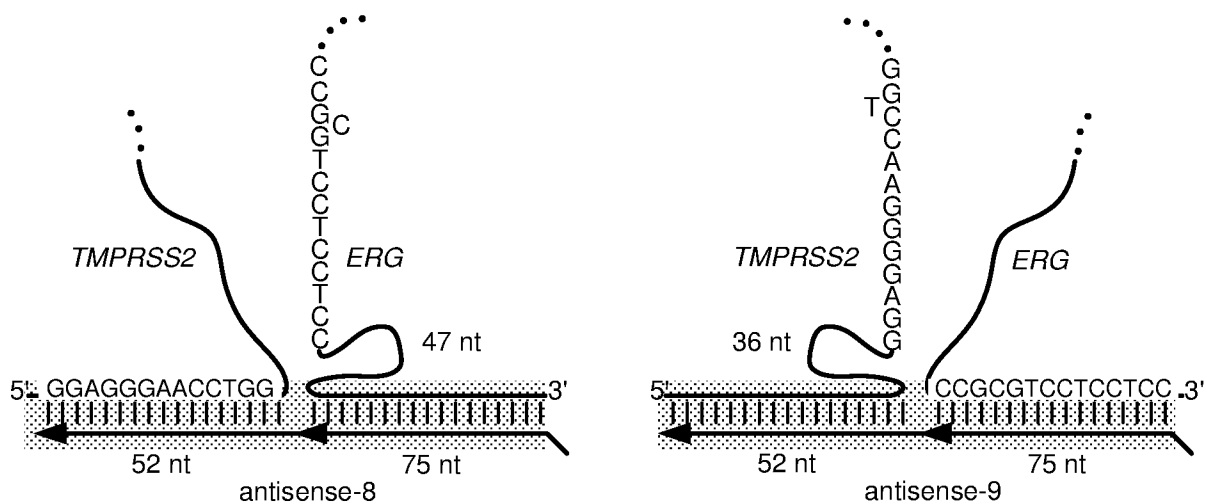


FIG. 16A

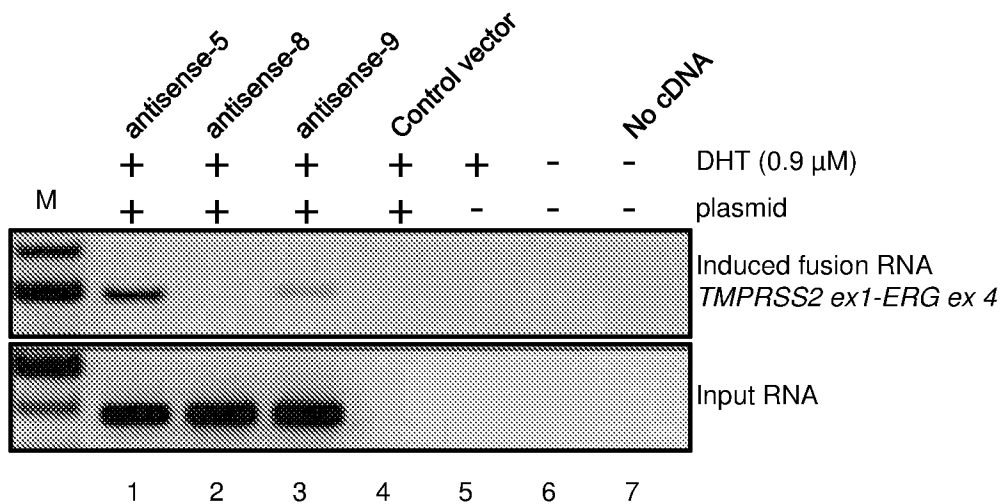


FIG. 16B

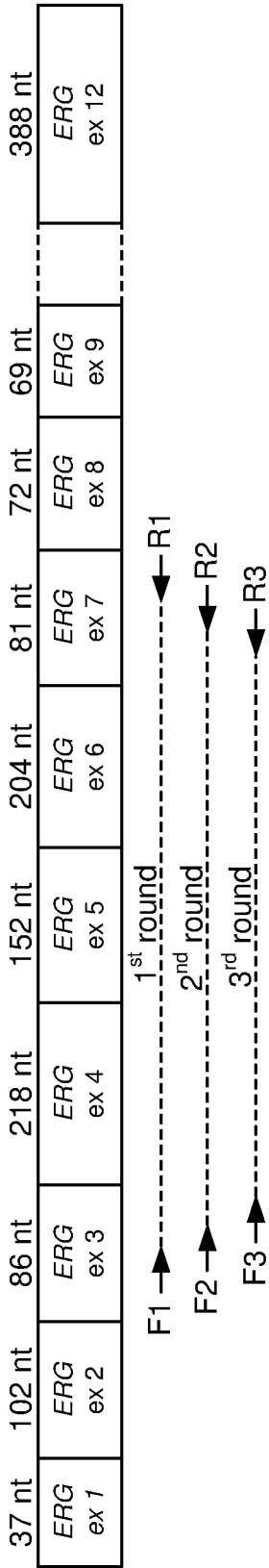


FIG. 17A

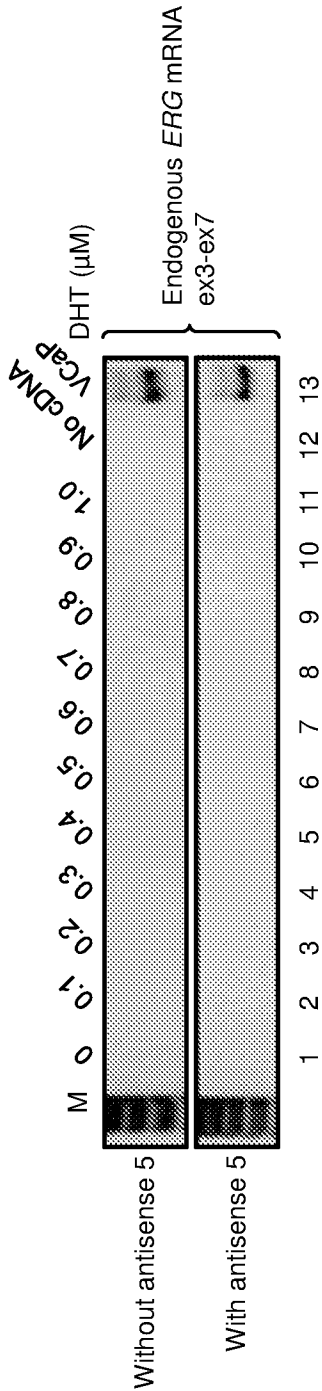


FIG. 17B

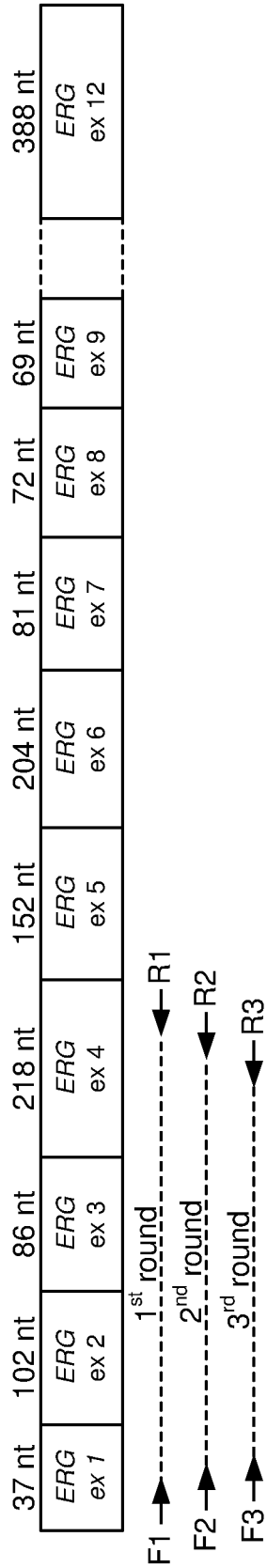


FIG. 17C

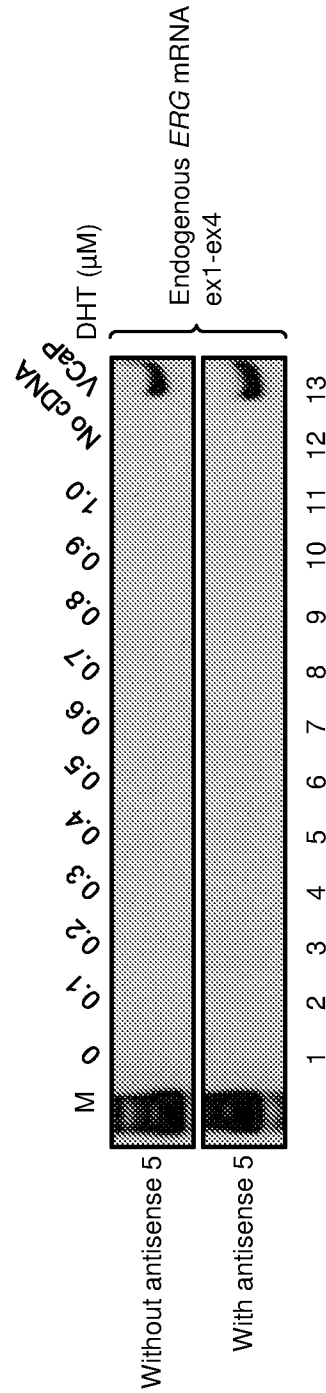


FIG. 17D

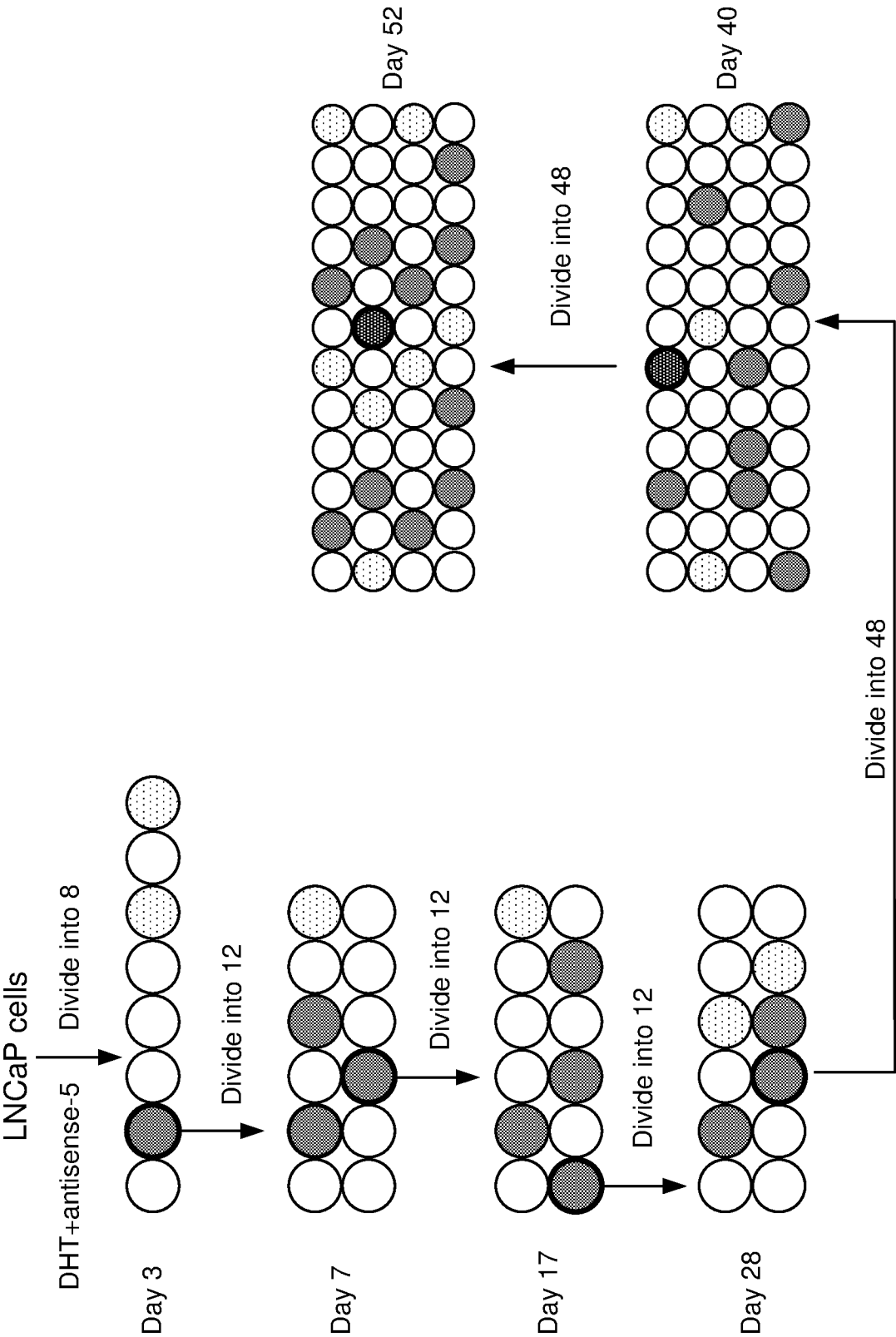
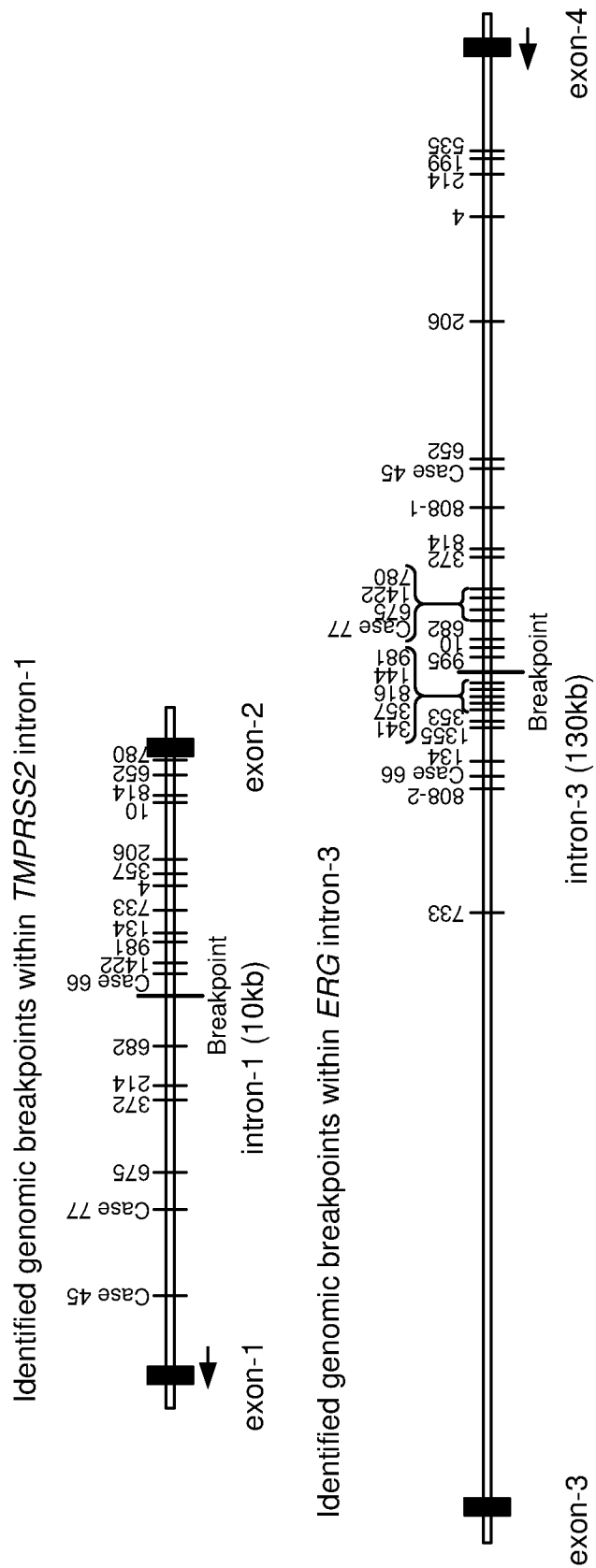
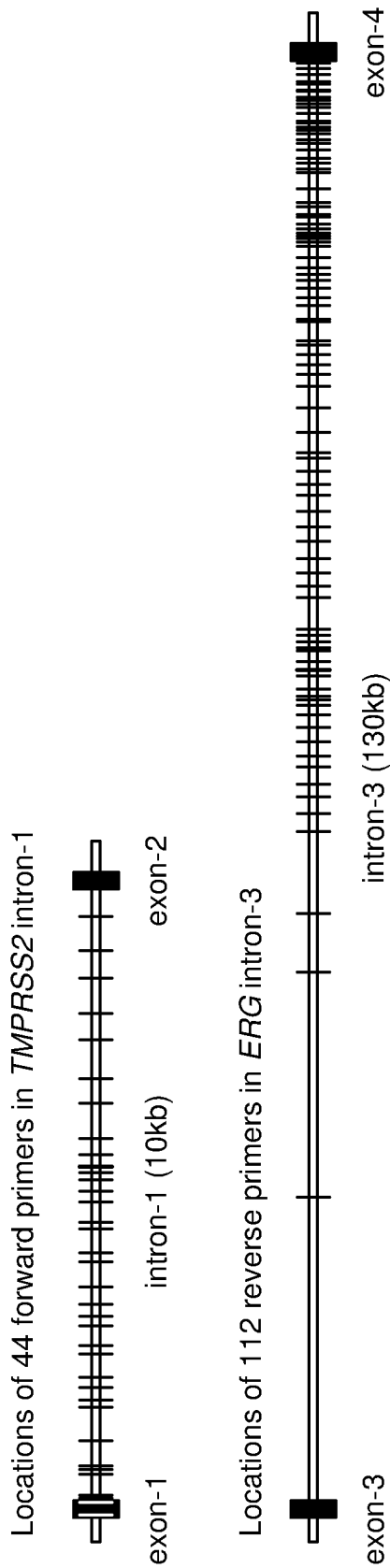


FIG. 18



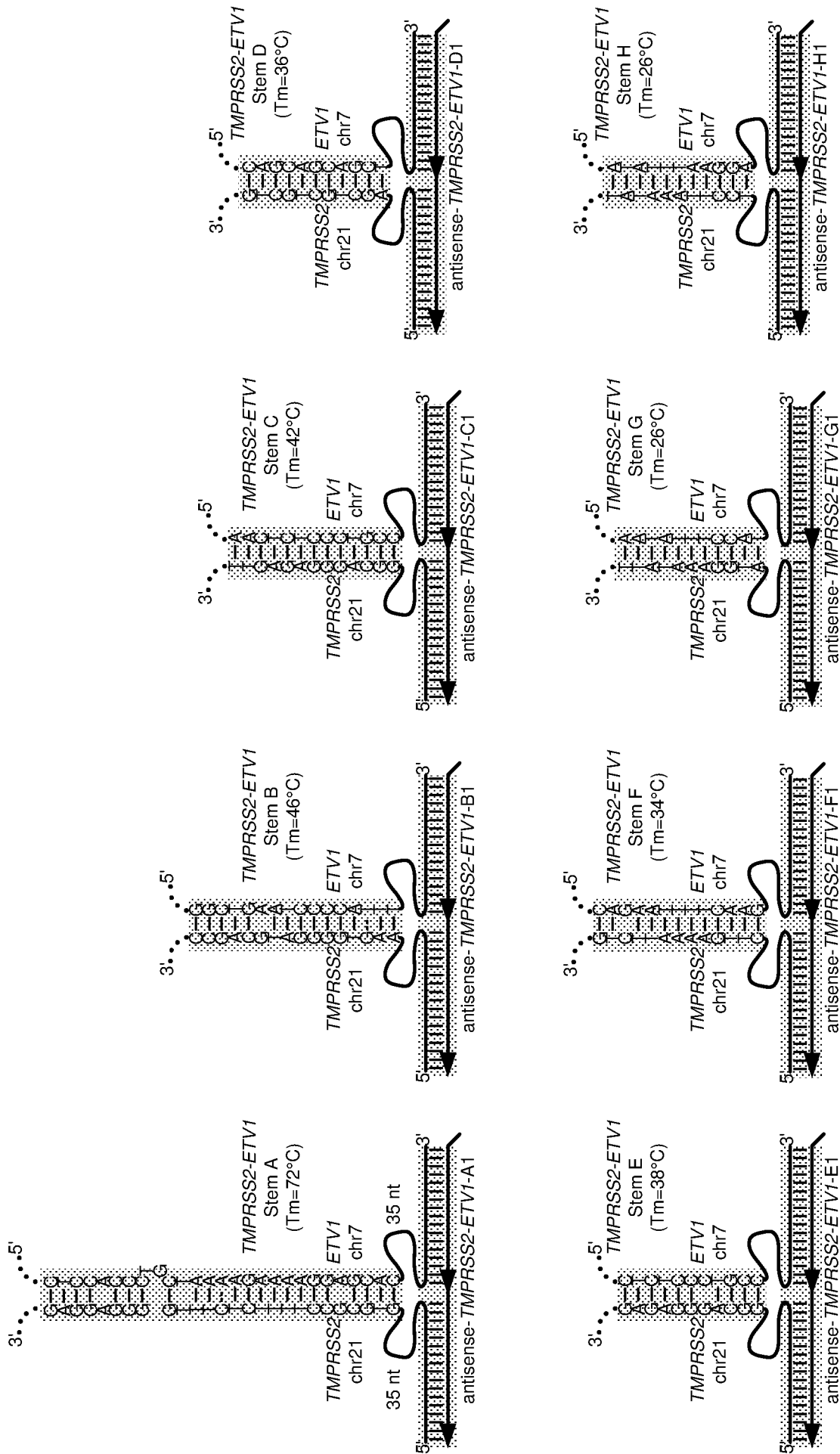
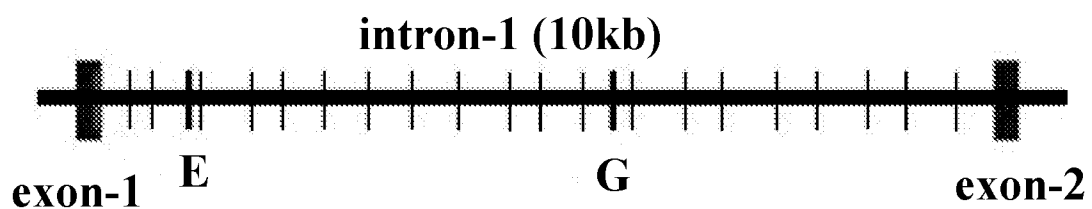


FIG. 20

Locations of 23 forward primers in *TMPRSS2* intron-1



Locations of 7 reverse primers in *ETV1* intron-2

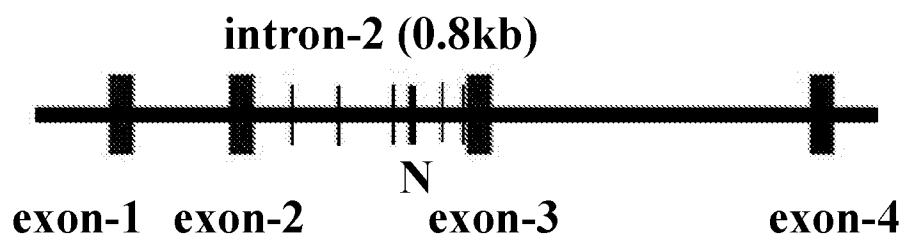


FIG. 22

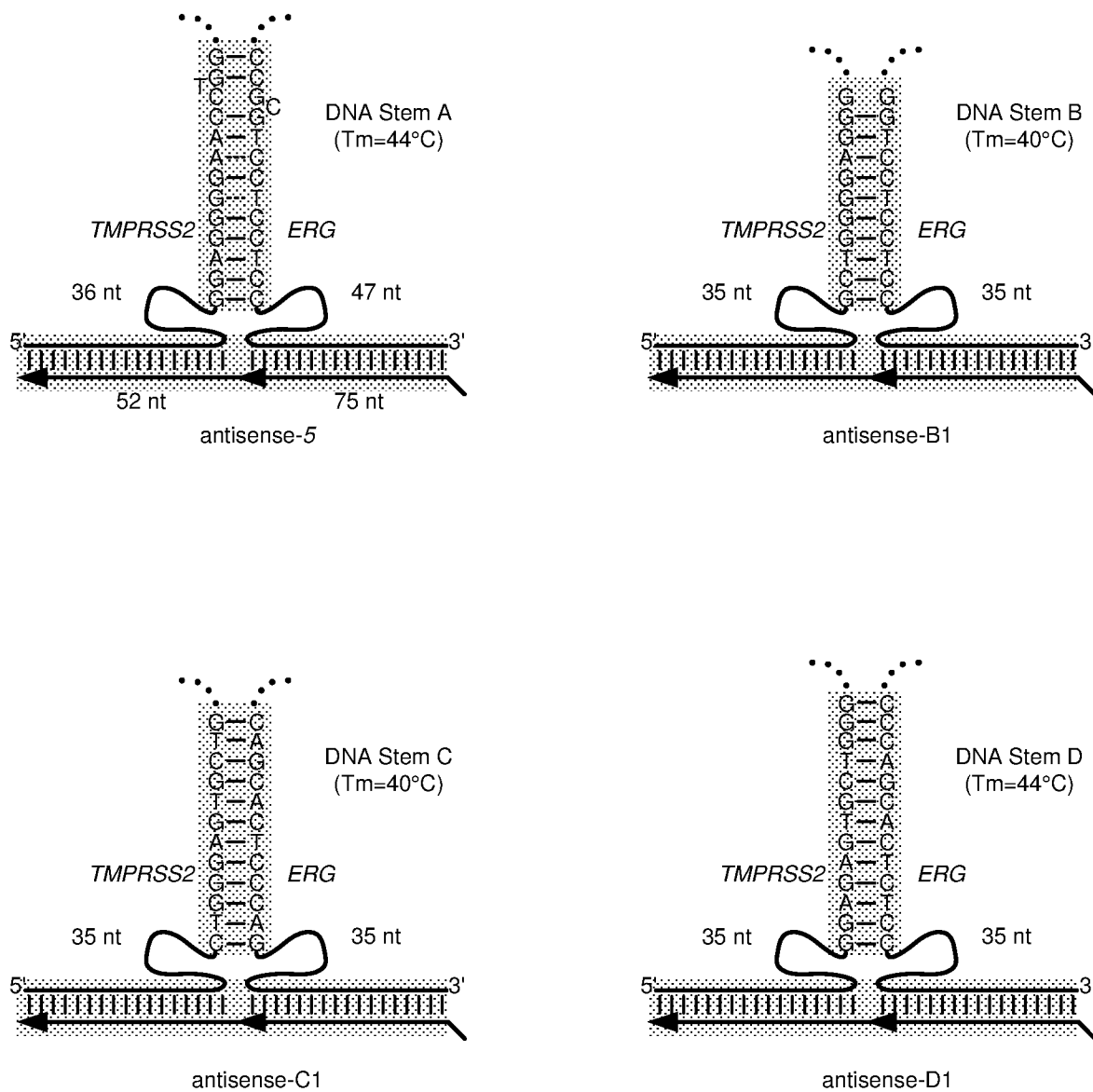
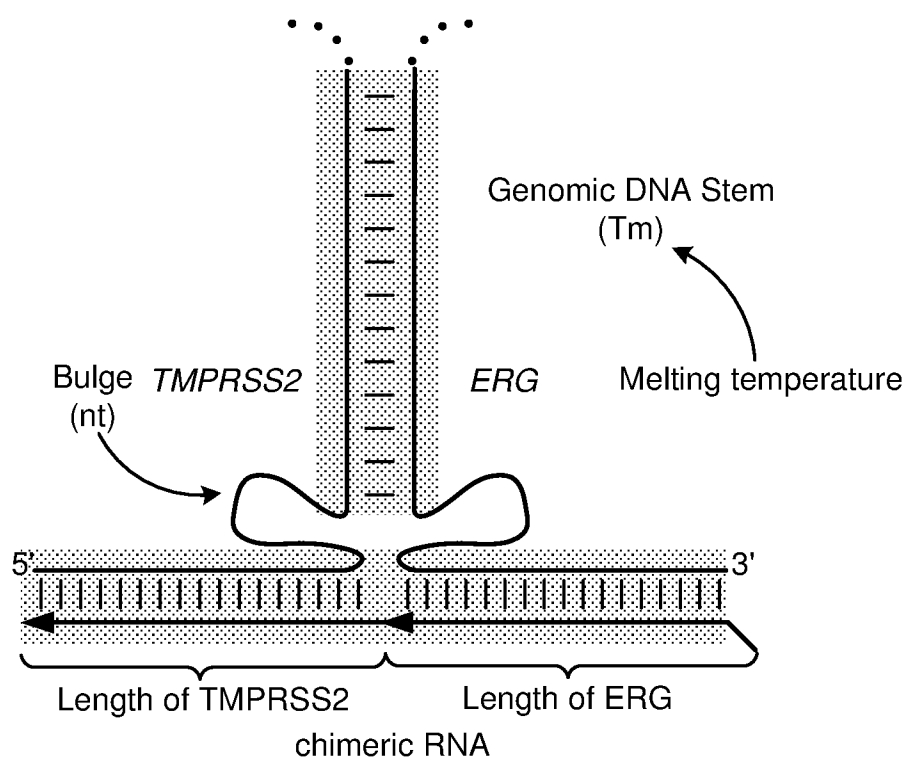
3-way junction stem formation TPRSS2-ERG

FIG. 23

3-way junction stem formation

**FIG. 24**

Stem B- Size of Bulge

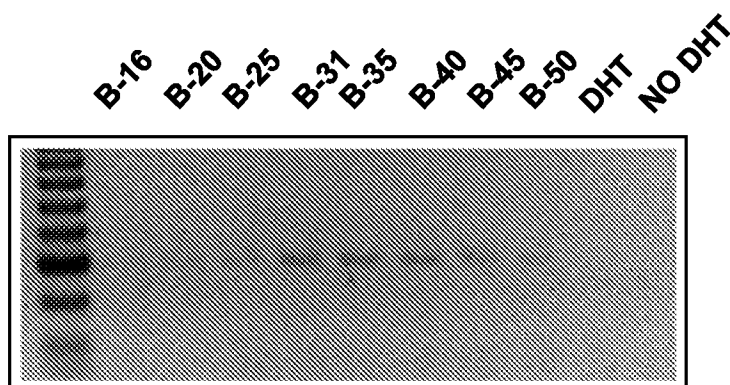
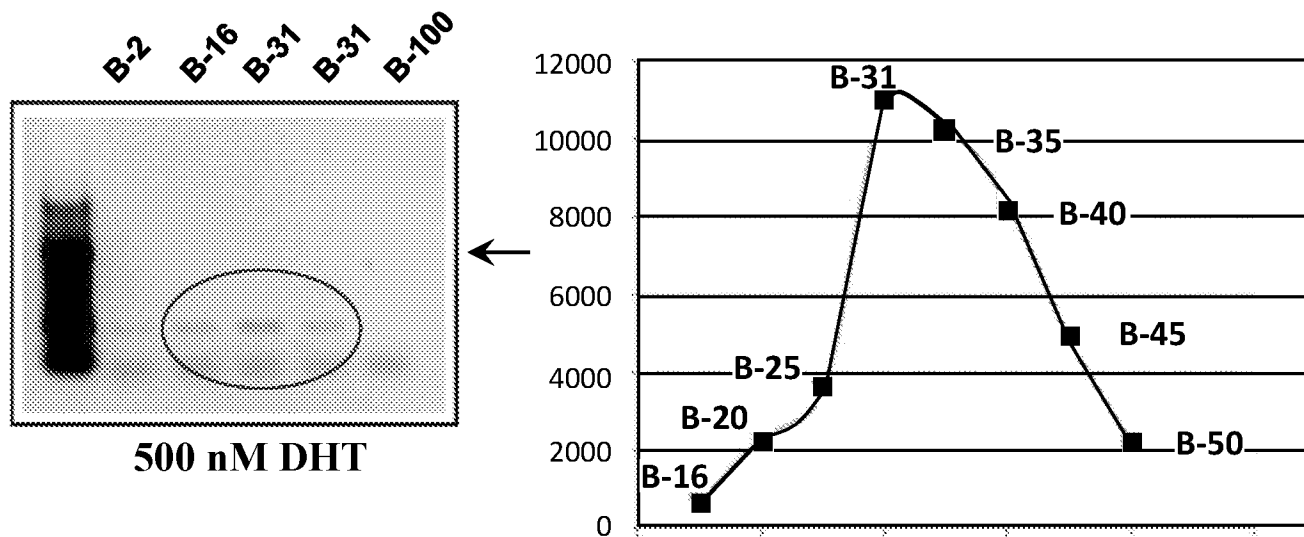


FIG. 25

Stem C- Size of Bulge

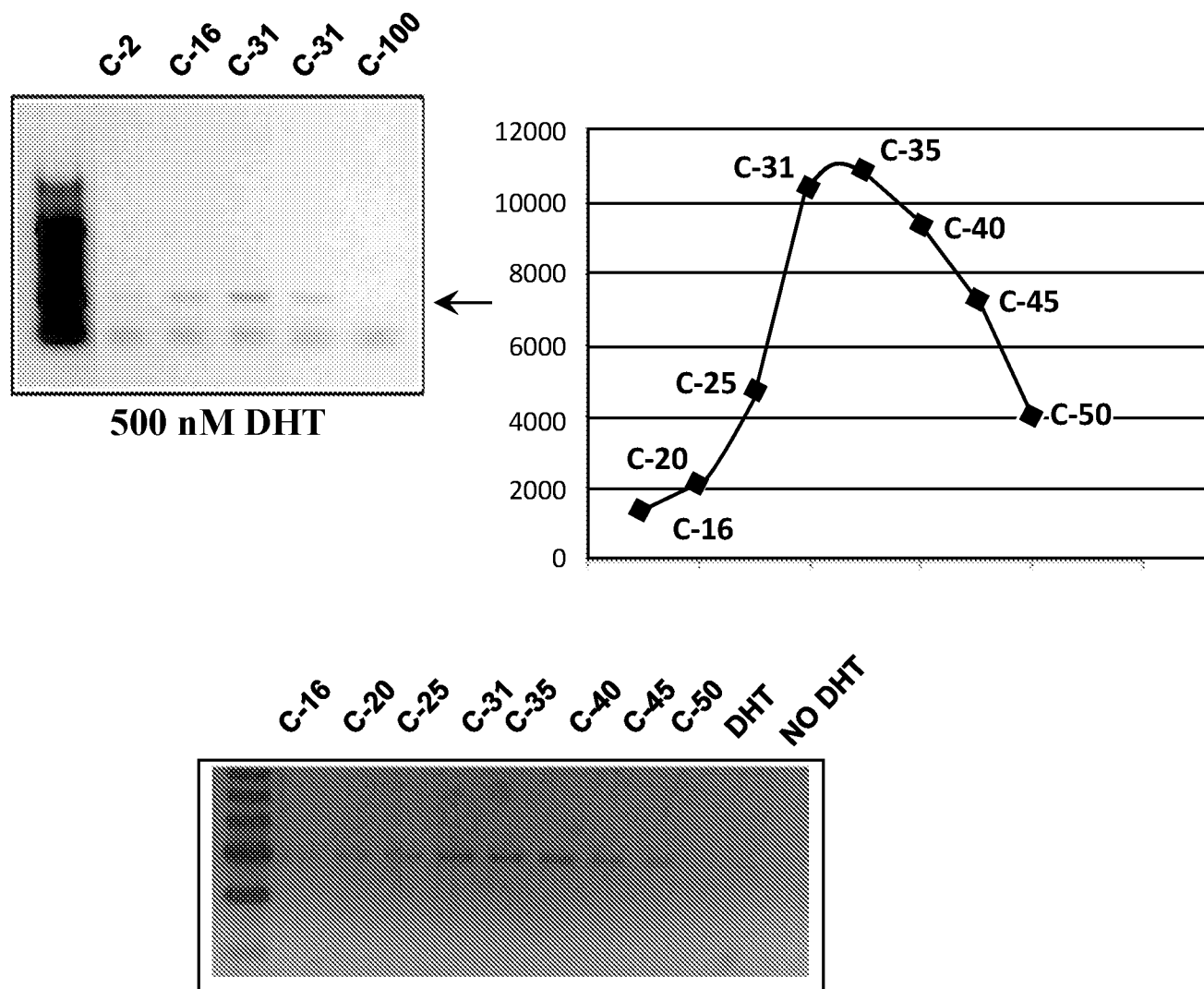


FIG. 26

Stem D- Size of Bulge

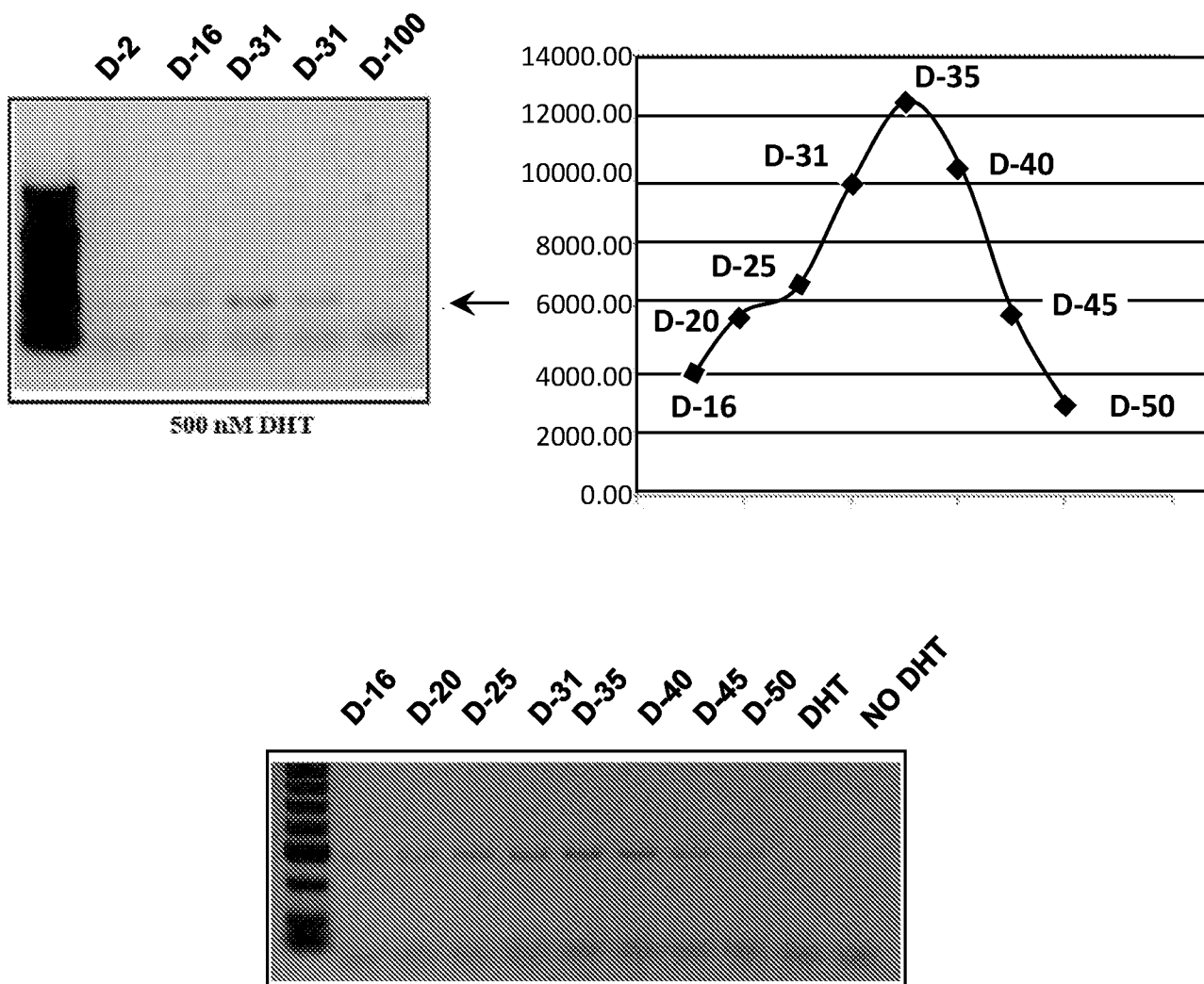
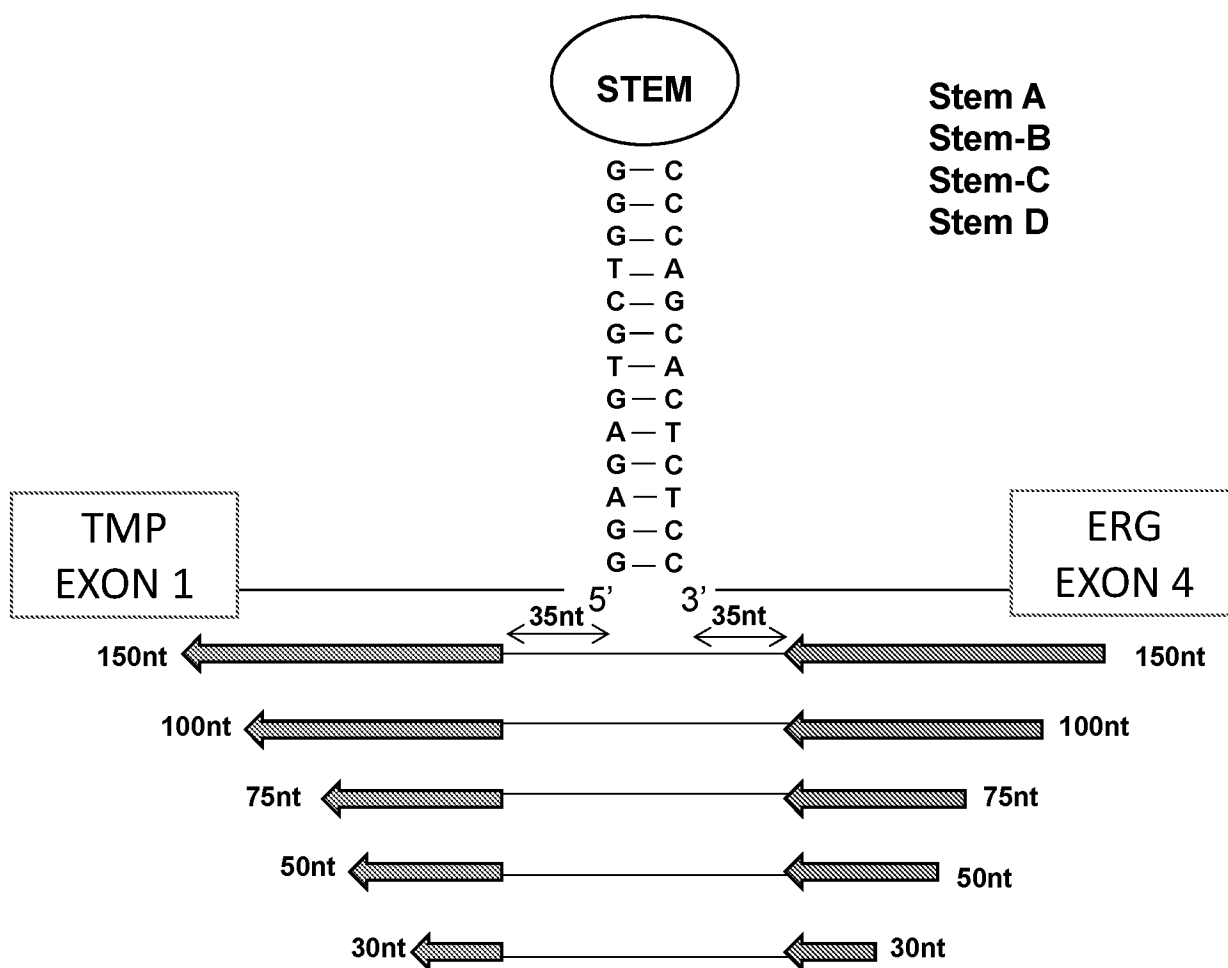


FIG. 27

**FIG. 28**

Stem A

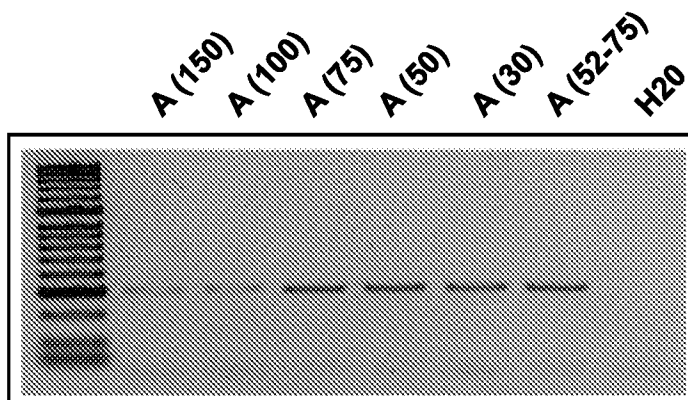
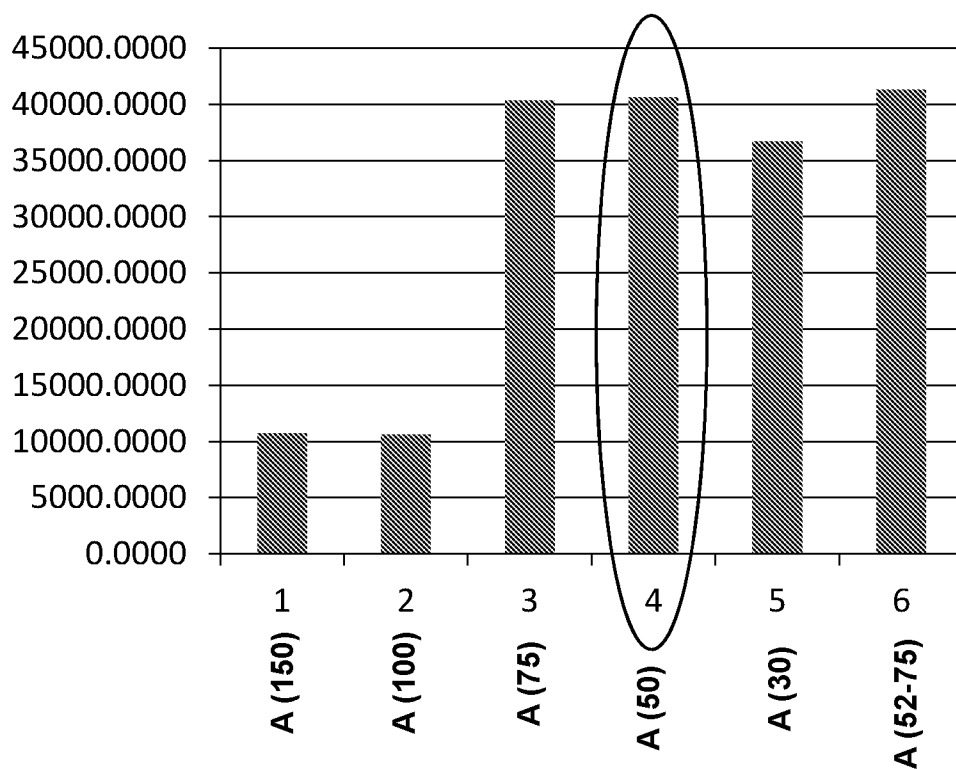


FIG. 29

Stem B

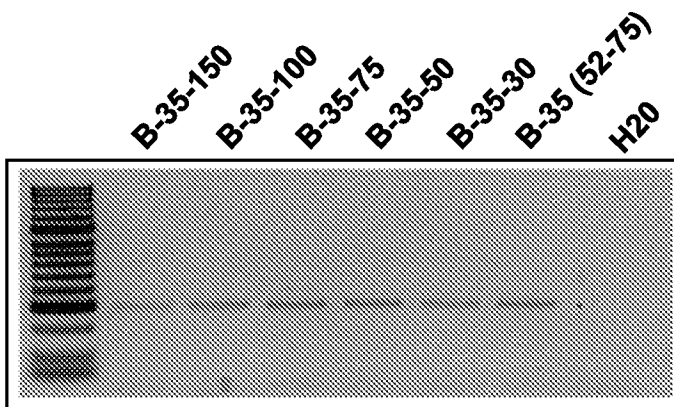
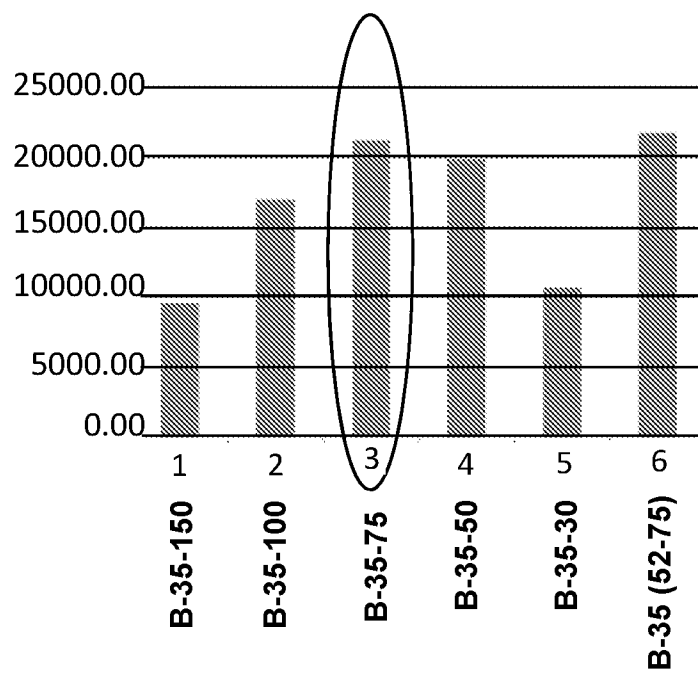


FIG. 30

Stem C

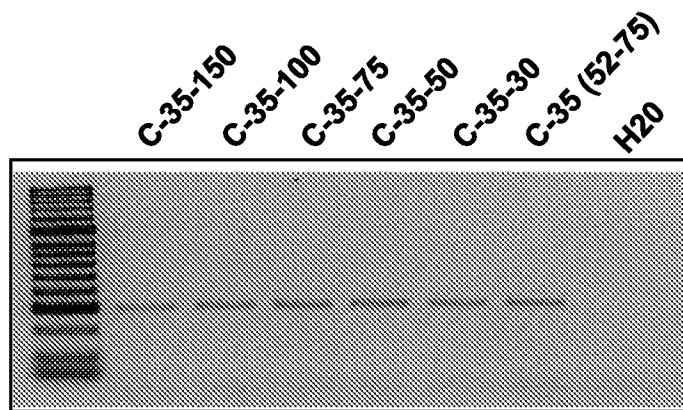
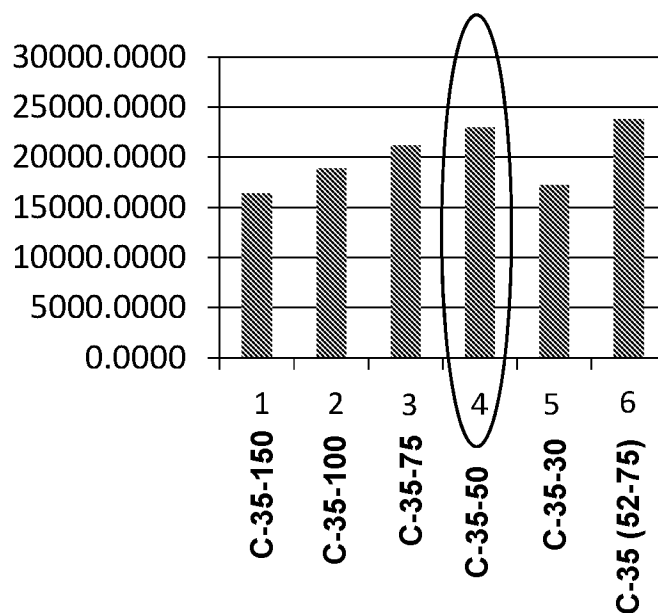


FIG. 31

stem D

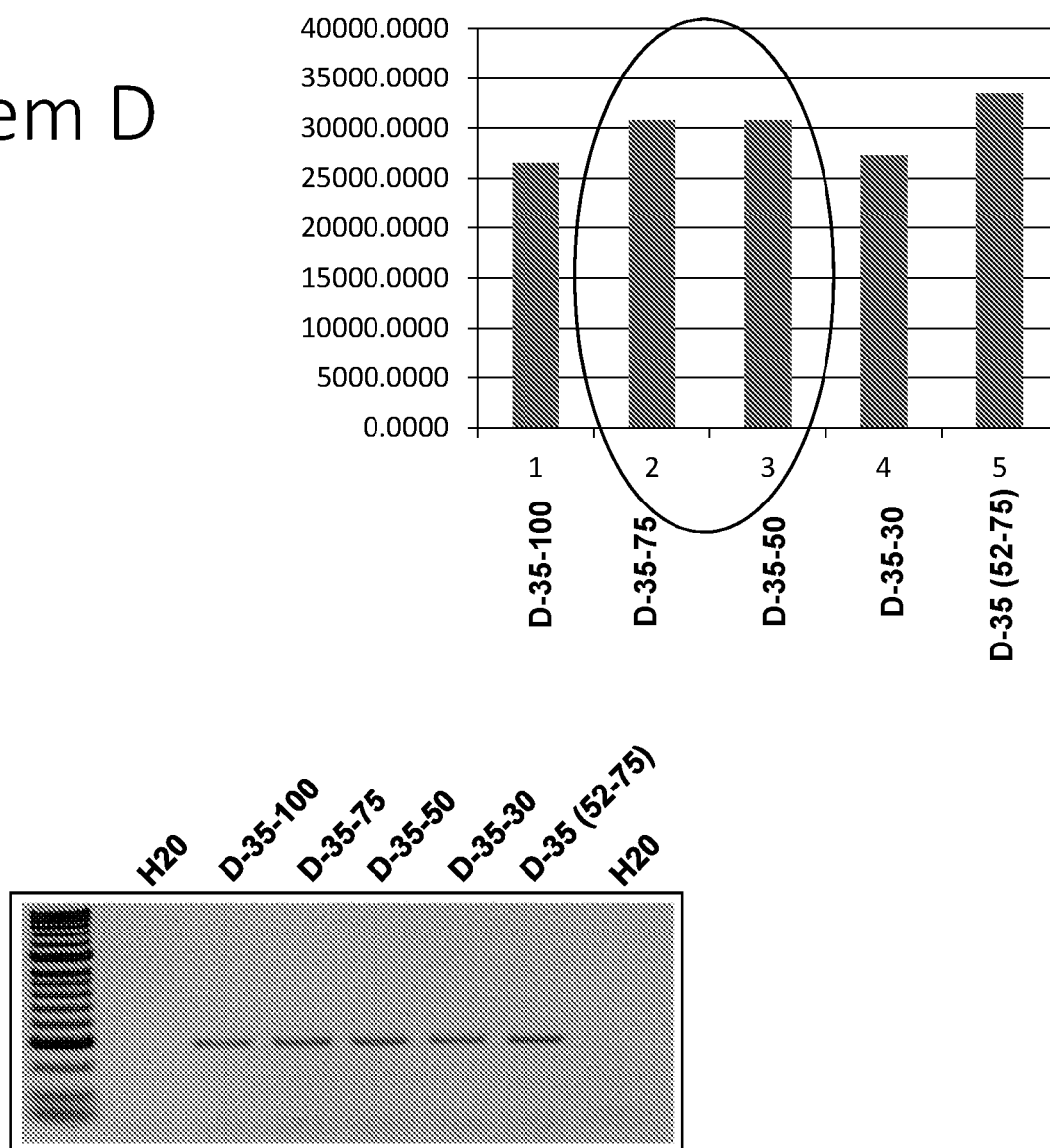
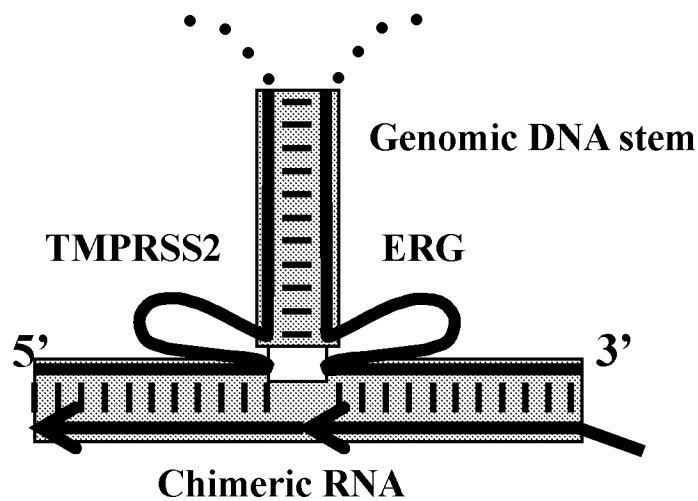
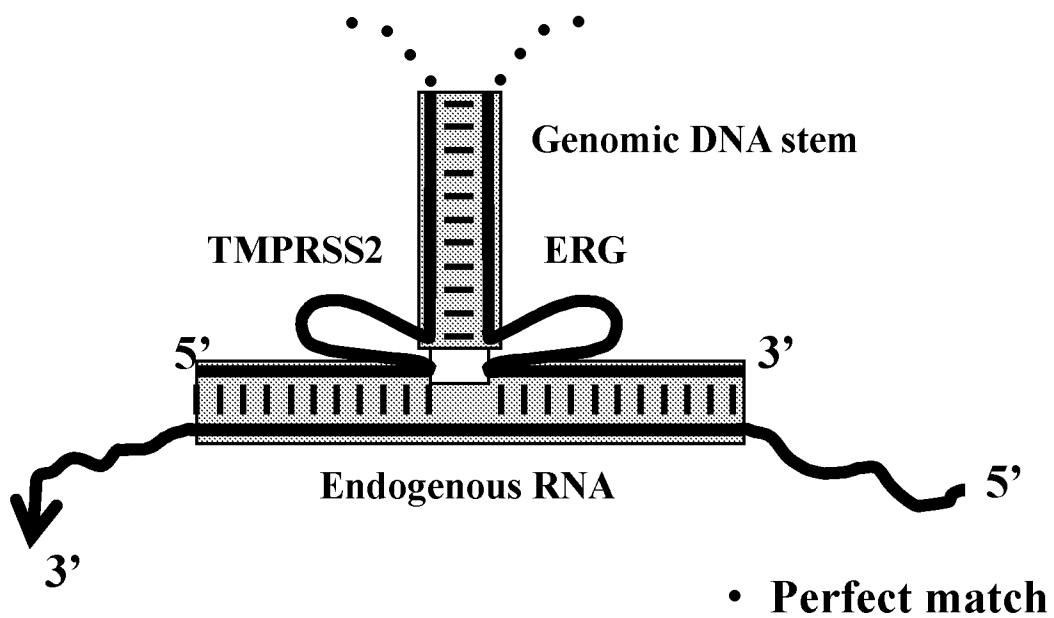


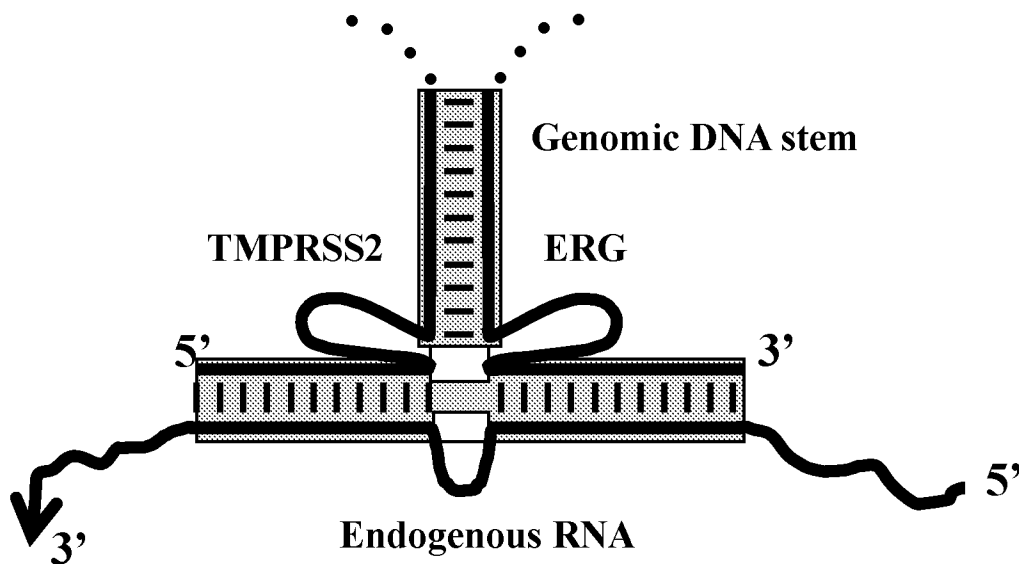
FIG. 32

Man-made designer chimeric RNA

- Perfect match
- Formed of two different sequences

FIG. 33

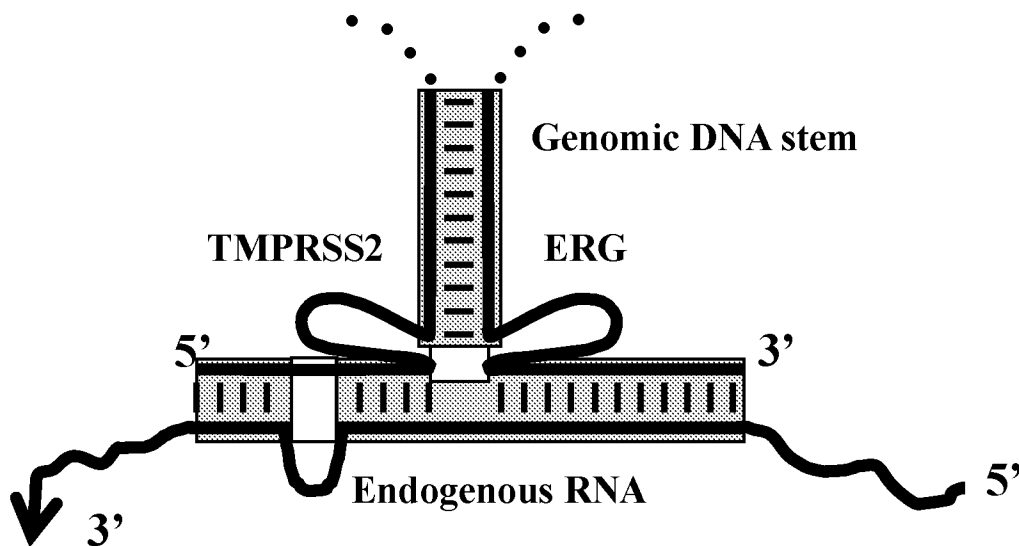
Endogenous initiator chimeric RNA**FIG. 34**

Endogenous initiator chimeric RNA

- Can have loop or secondary structure in middle

FIG. 35

Endogenous initiator chimeric RNA



- Can have loop or secondary structure at one side

FIG. 36

Endogenous initiator chimeric RNA

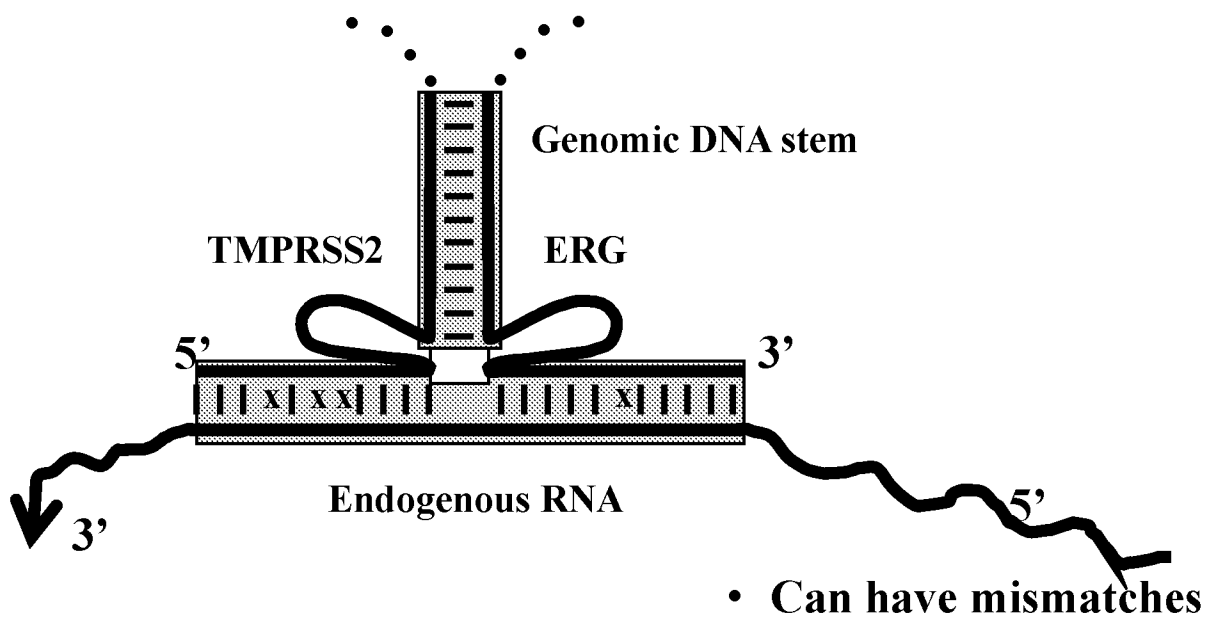
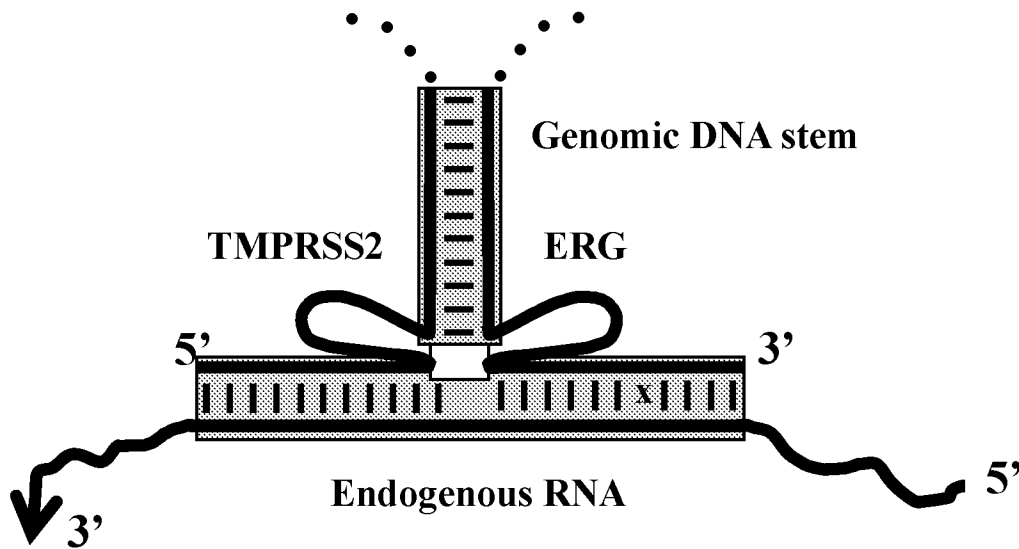


FIG. 37

Endogenous initiator chimeric RNA



- Can be perfect match
- Formed by trans-splicing of two different parental endogenous RNAs

FIG. 38

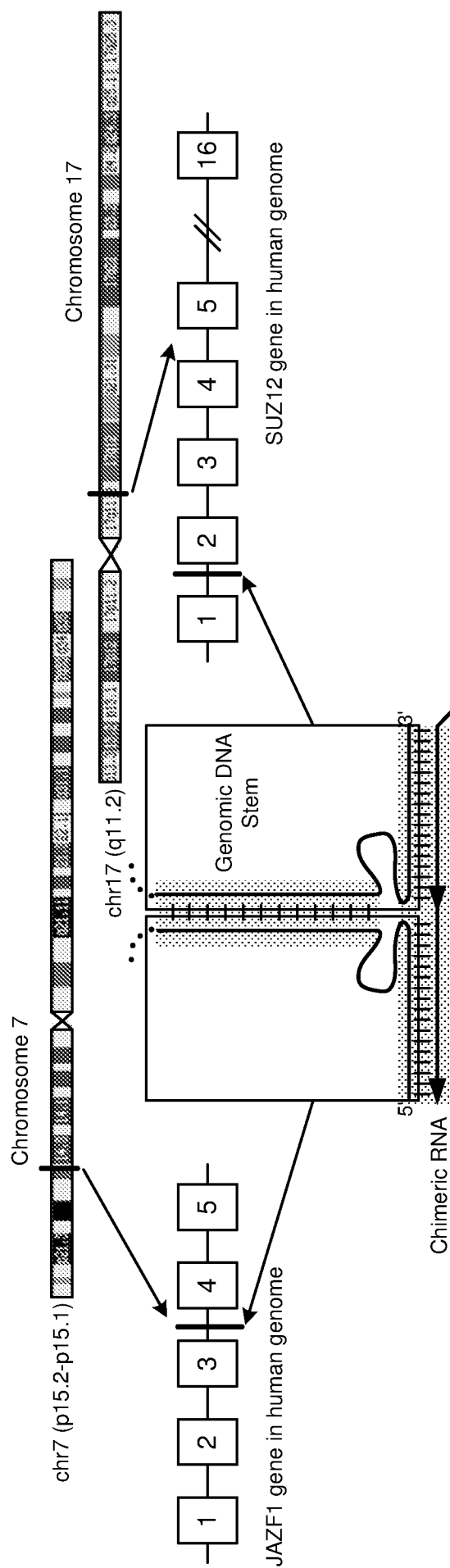


FIG. 39A

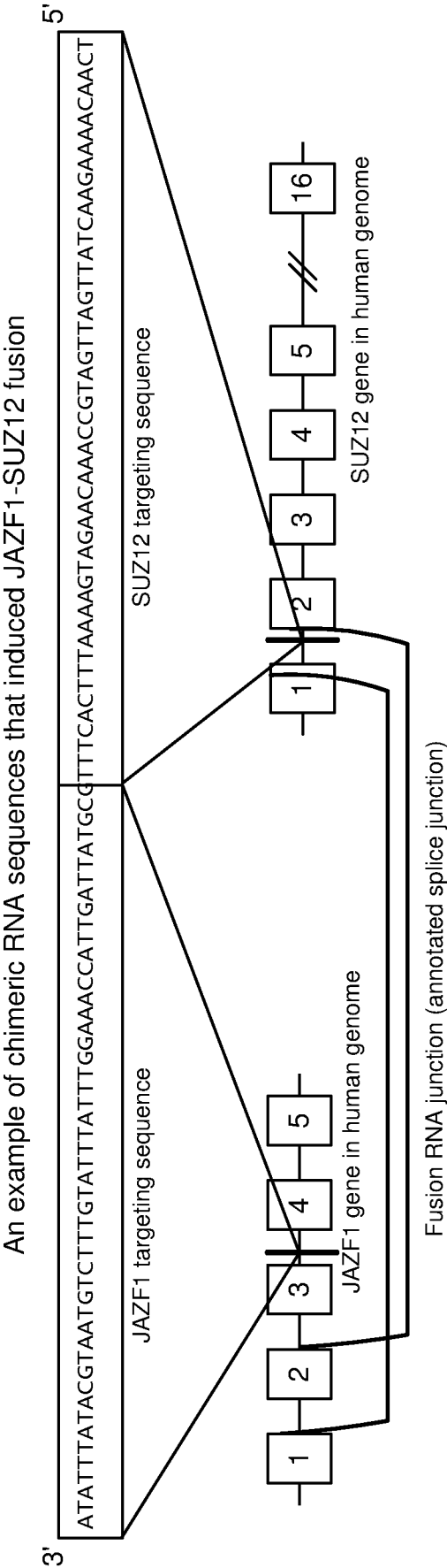
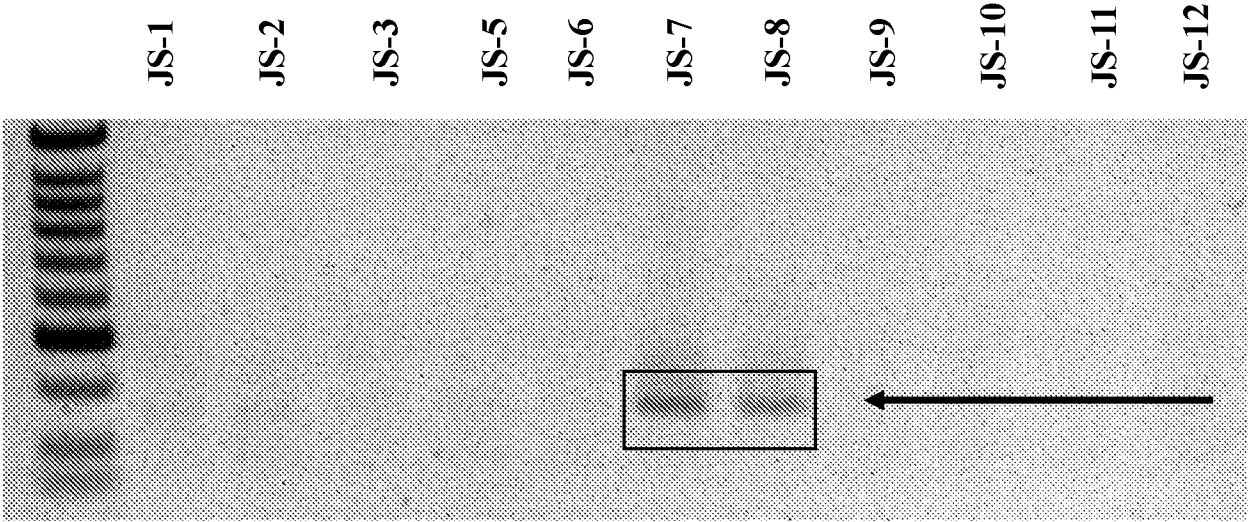


FIG. 39B



JAZF1		SUZ12	
1	1	2	3

FIG. 39C

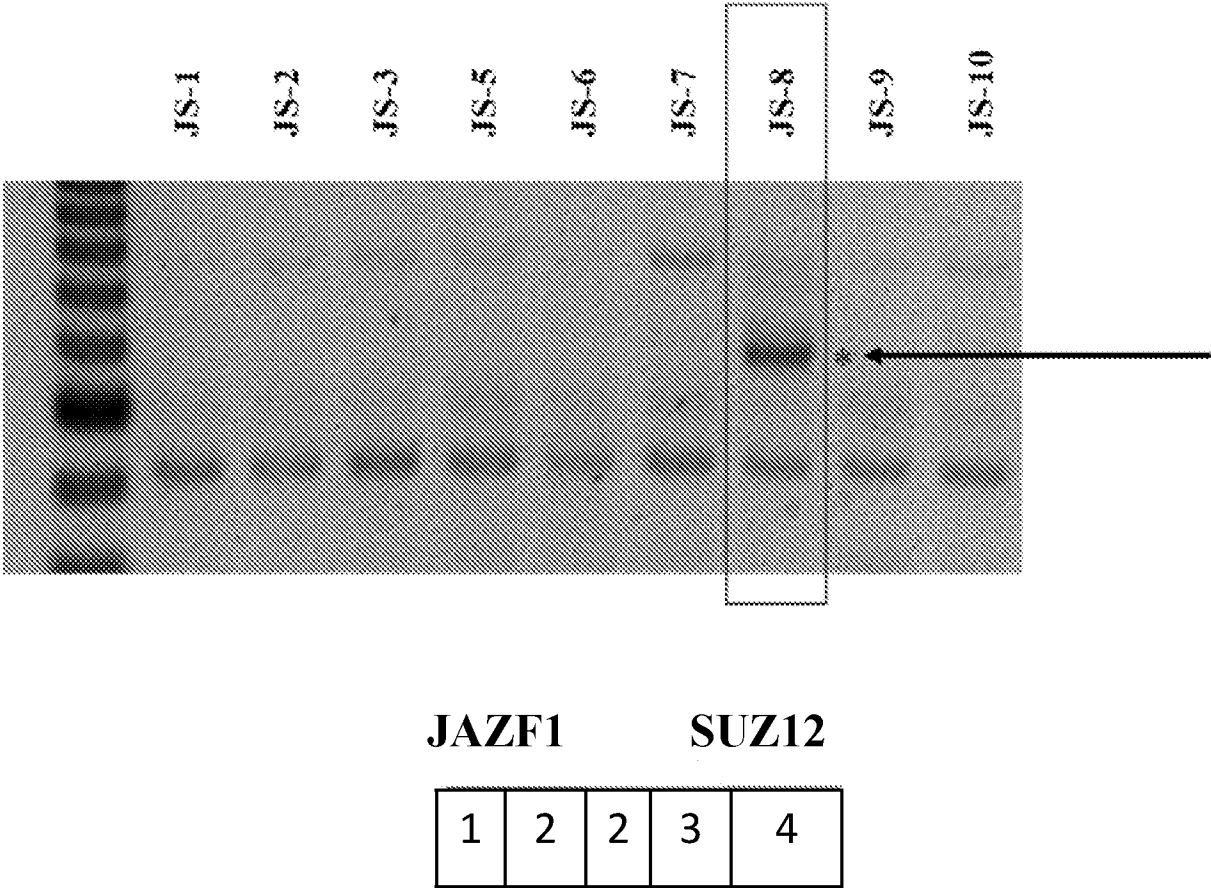


FIG. 39D

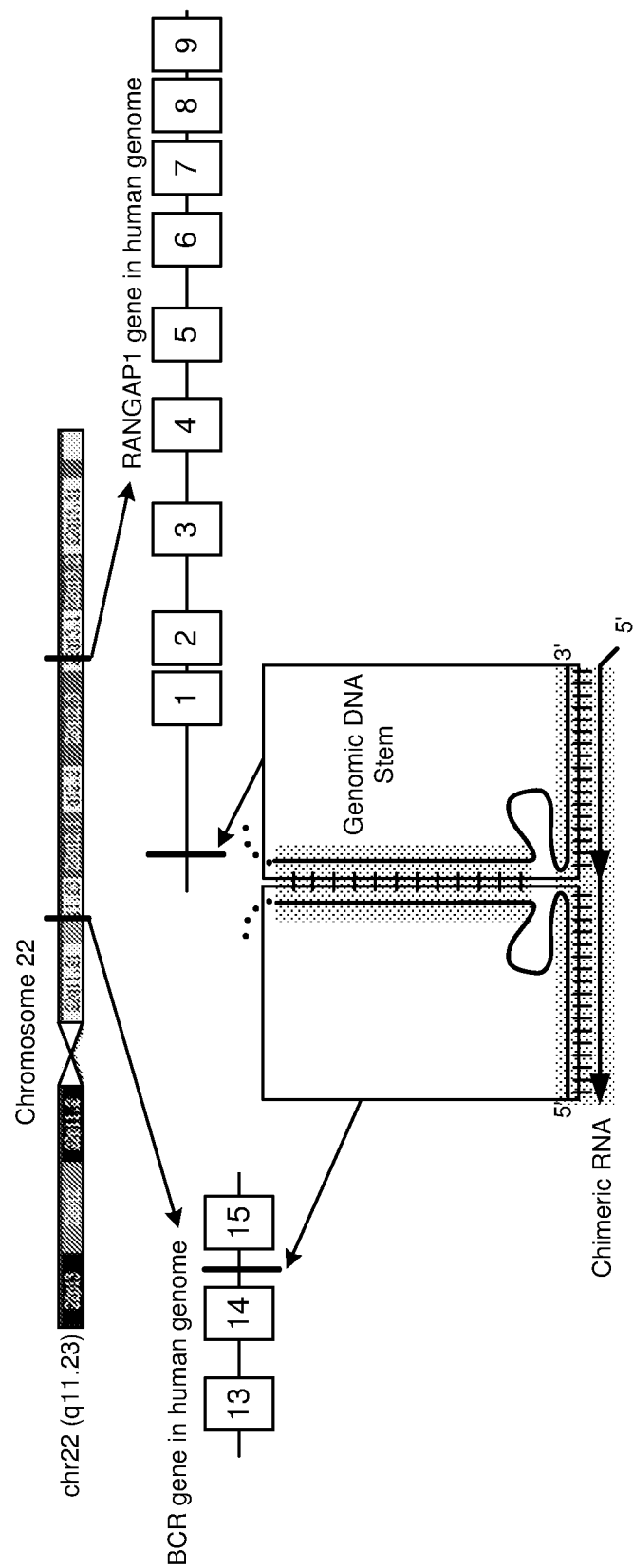


FIG. 40A

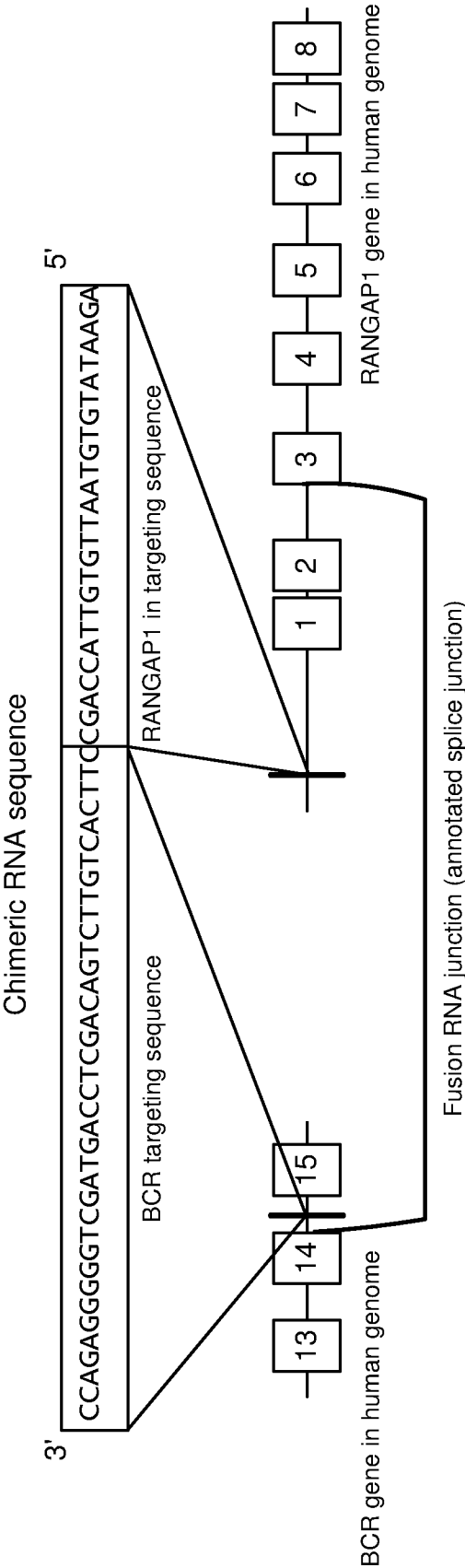


FIG. 40B

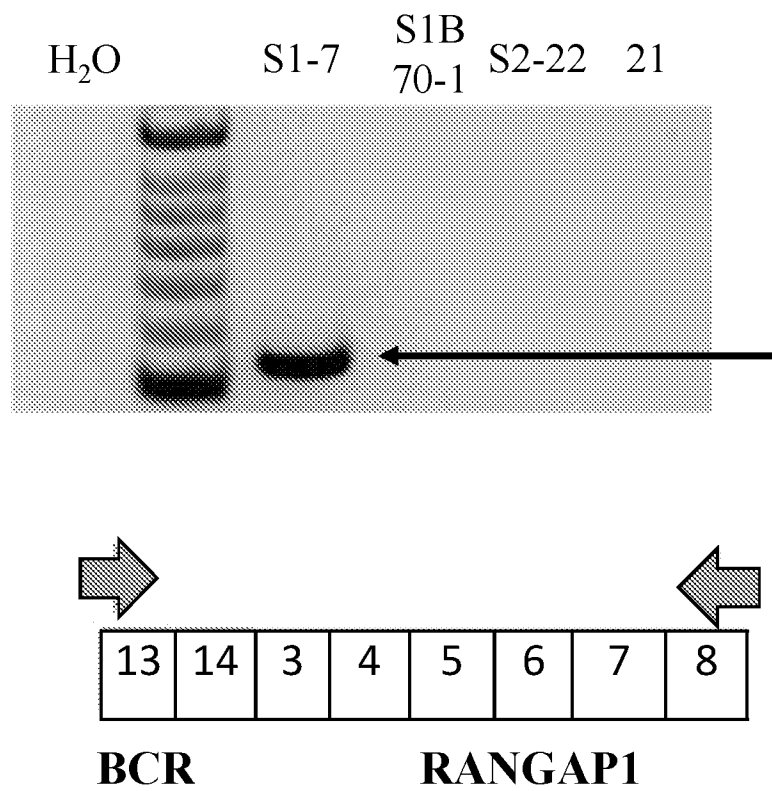


FIG. 40C

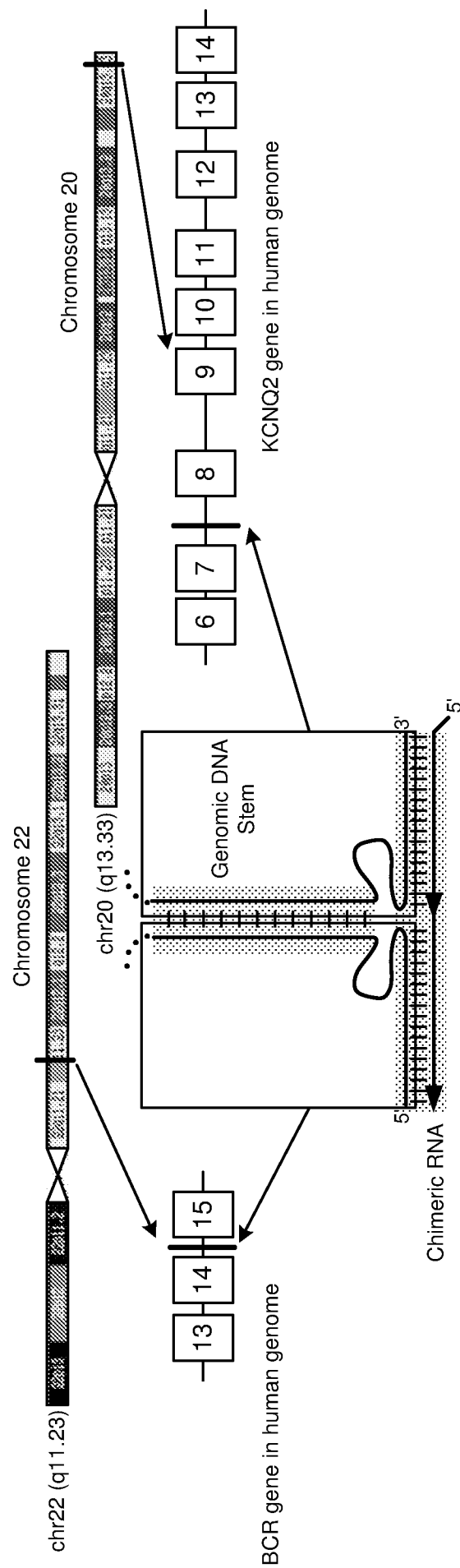


FIG. 41A

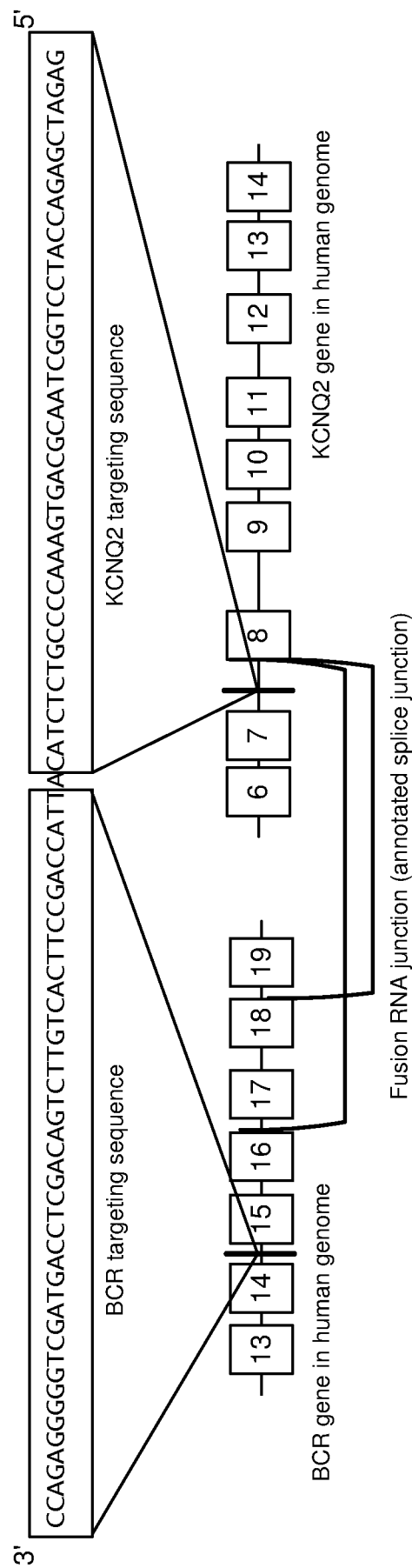


FIG. 41B

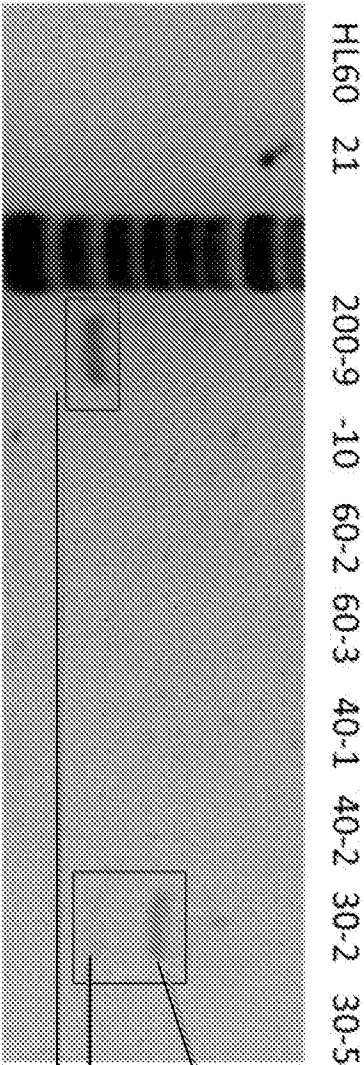


FIG. 41D

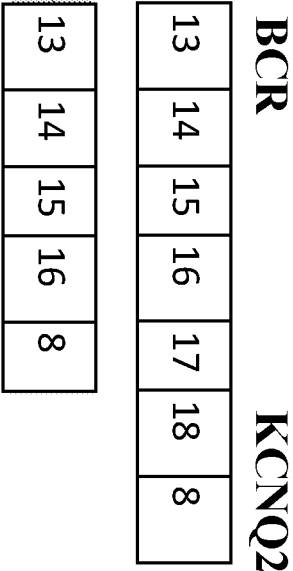


FIG. 41C

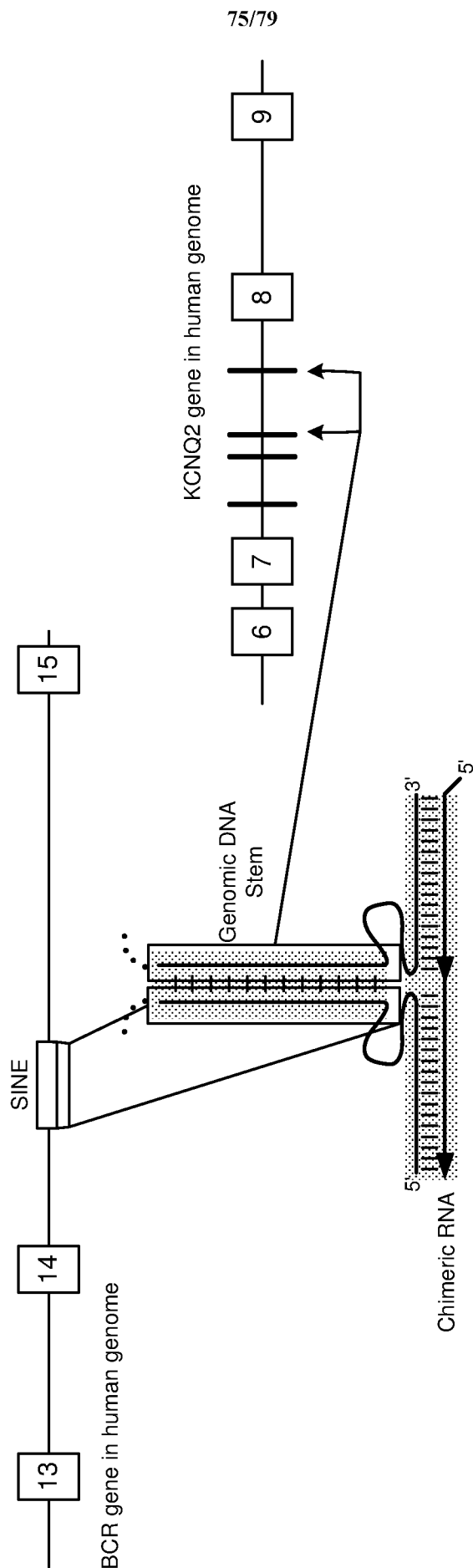


FIG. 42A

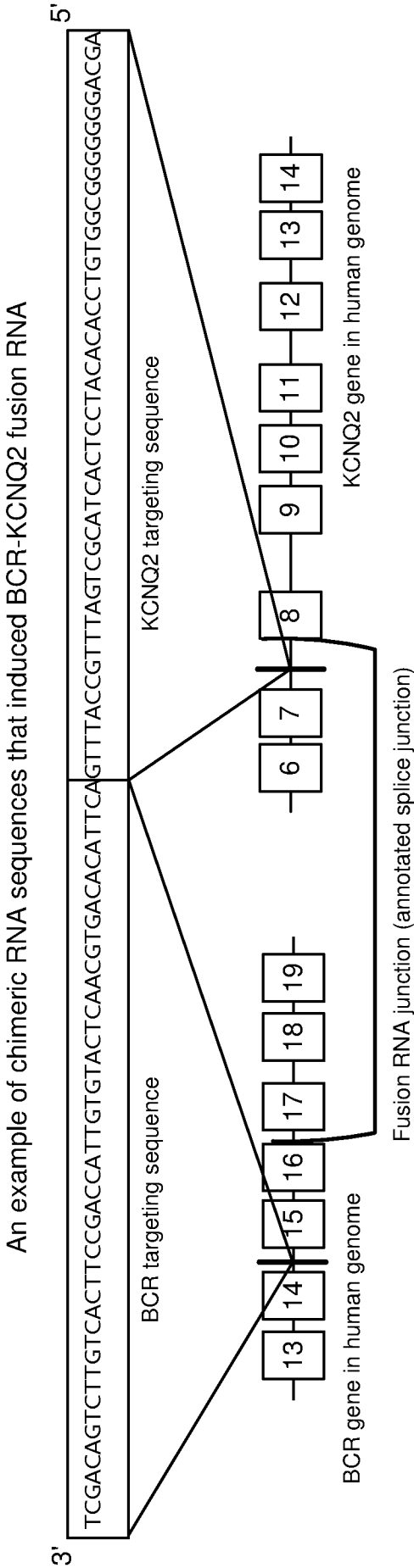


FIG. 42B

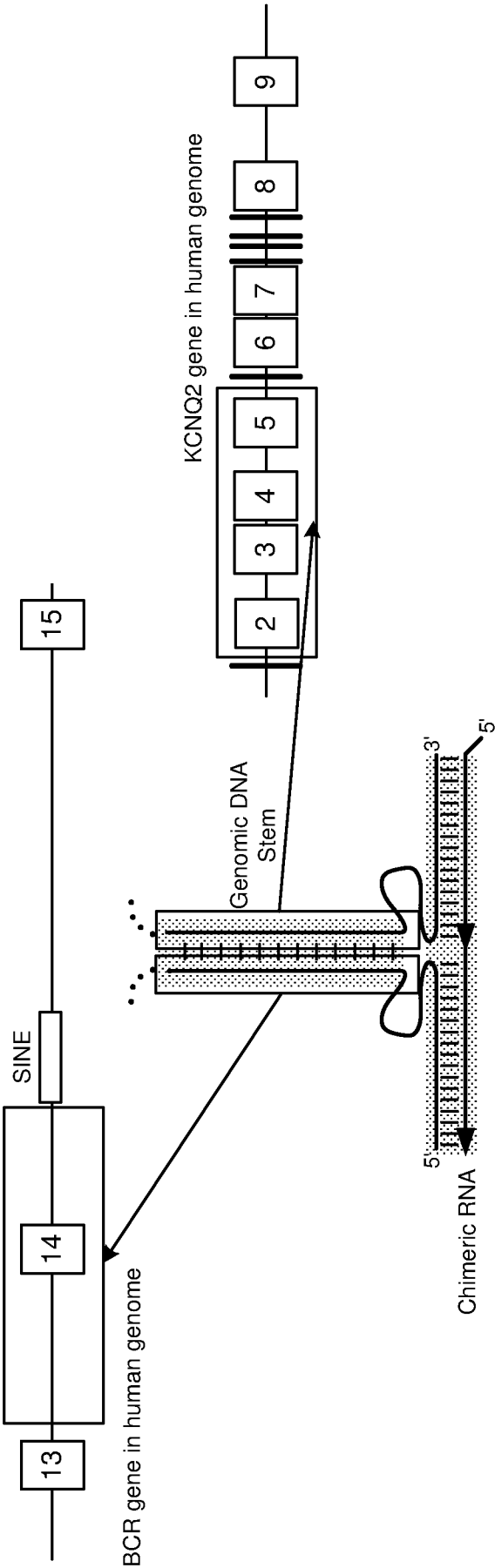


FIG. 42C

An example of chimeric RNA sequences that induced BCR-KCNQ2 fusion RNA

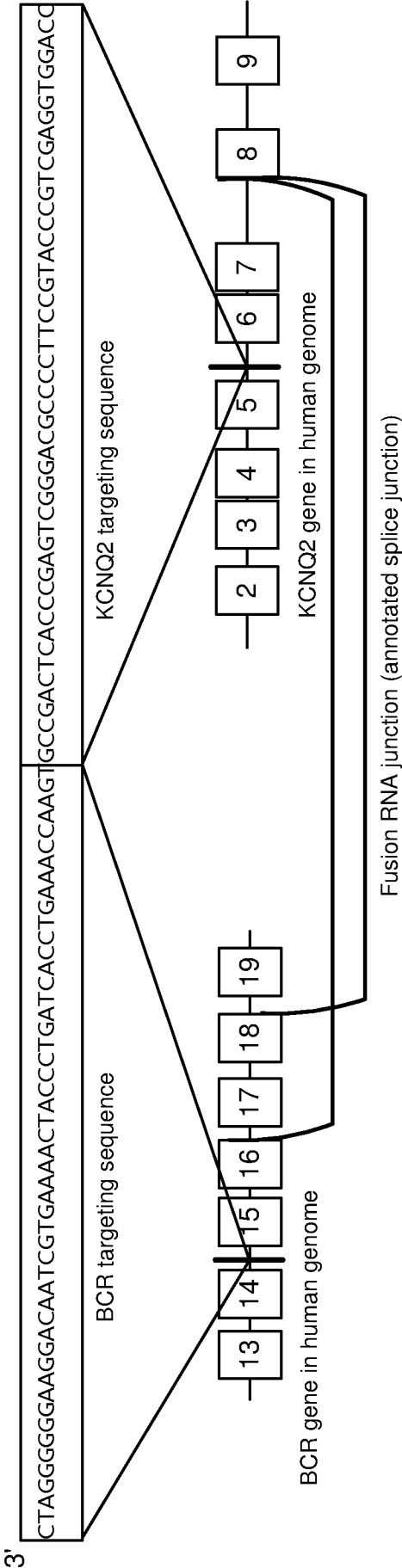


FIG. 42D

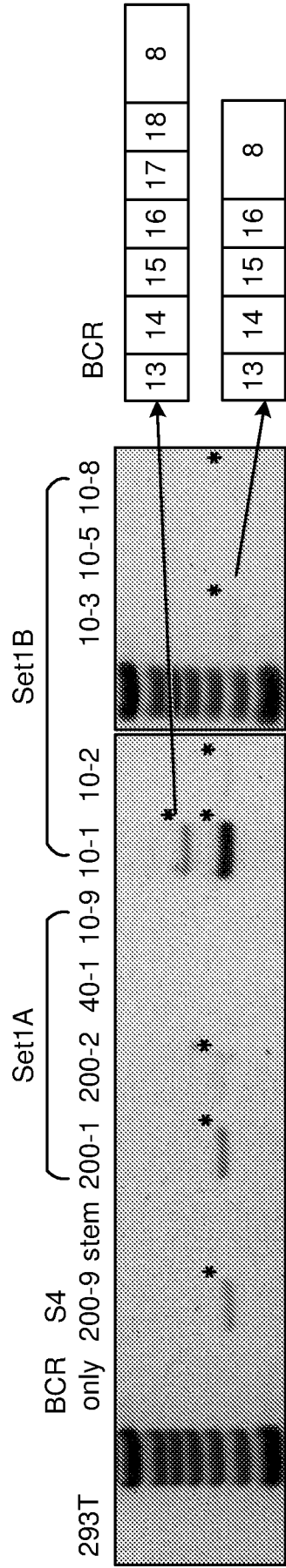


FIG. 42E