FIGURES

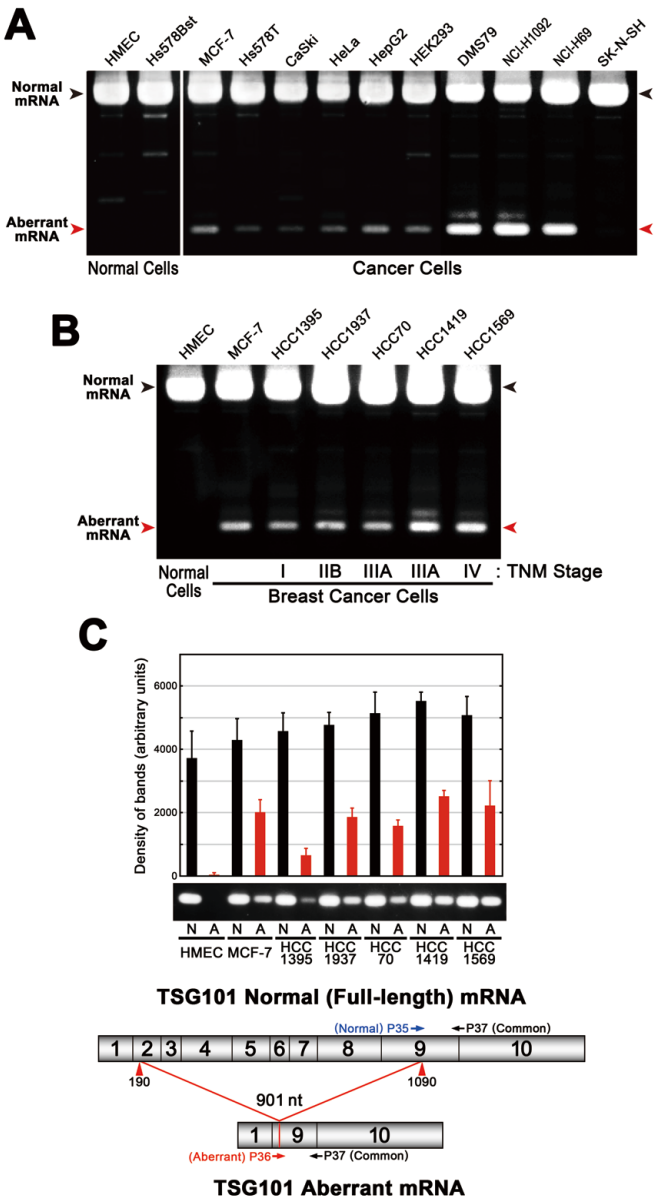

Figure S1. TSG101 pre-mRNA is aberrantly spliced in various cancer cells with different efficiencies. (**A, B**) RT–PCR analysis of TSG101 transcripts in breast cancer cell lines (MCF-7, Hs578T), cervical carcinoma cell lines (CaSki, HeLa), a hepatoma cell line (HepG2), a transformed embryonic kidney cell line (HEK293), small-cell lung cancer cell lines (DMS79, NCI-H1092, NCI-H69), a neuroblastoma cell line (SK-N-SH), a control normal myoepithelial cell line (Hs578Bst), and mammary epithelial cells (HMEC). Black and red arrowheads indicate the normal full-length mRNA and the major aberrant mRNA with a 901-nt deletion, respectively (see Figure 1). Breast cancer cell lines (HCC series) were established by the same contributors from breast cancers with different TNM stages. (**C**) Semi-quantitative RT–PCR analysis of normal TSG101 mRNA (N, black bar) and aberrant TSG101 mRNA (A, red bar). One-round PCR (without nested PCR) was performed with normal/aberrant-specific primer sets (P35–P37, P36–P37, respectively; see Table S1). The PCR bands on the scanned gel were quantitated with ImageJ freeware. Histograms represent the means ± standard deviations (shown above the bars only) of four replicates.

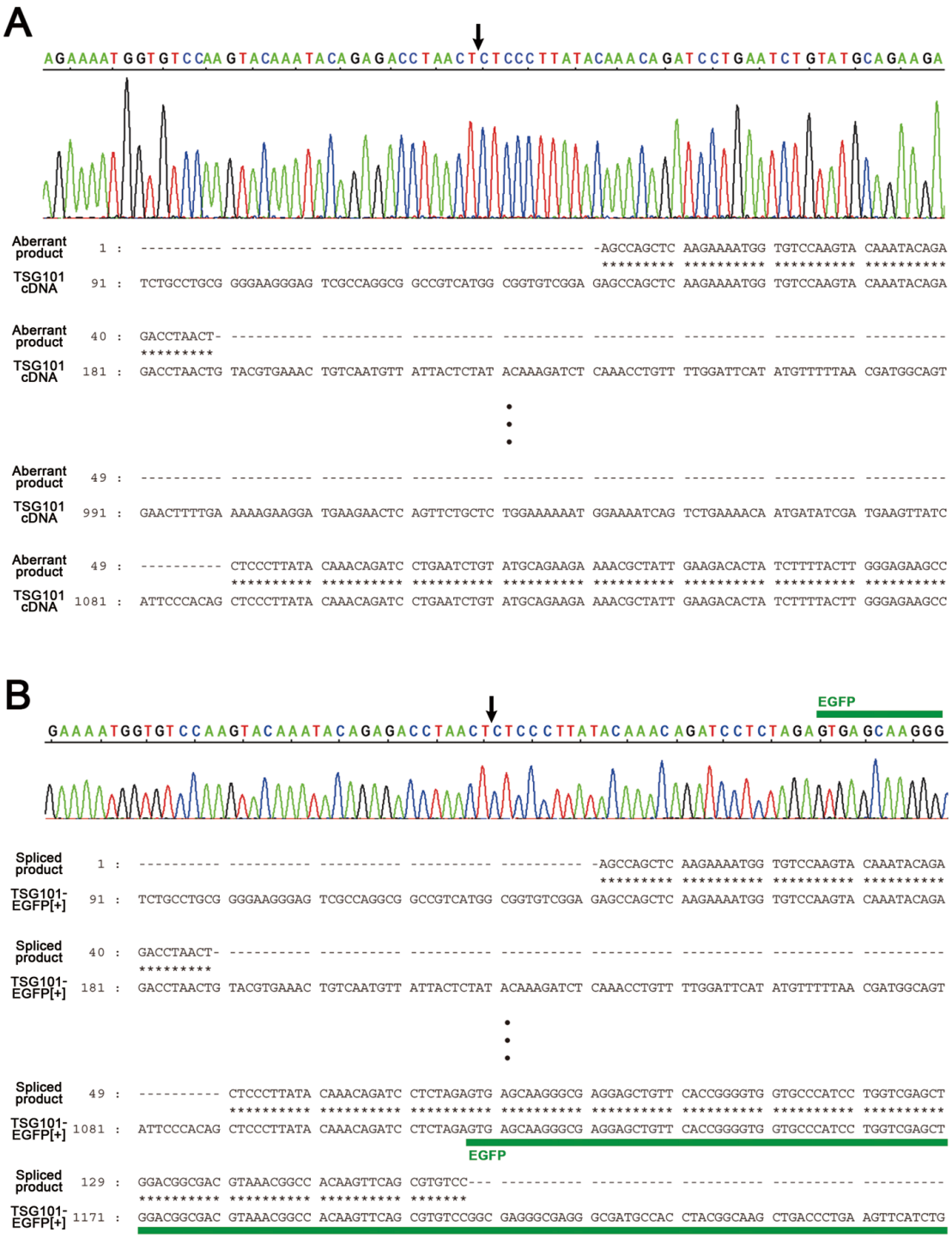

Figure S2. Ectopically expressed TSG101-EGFP[+] was spliced at the same exonic 5' and 3' splice sites used in the aberrant splicing of the endogenous TSG101 pre-mRNA. (**A**) Electropherogram of the aberrantly spliced product that was generated from the endogenous TSG101 pre-mRNA. The alignment of this sequence with that of the TSG101 cDNA reveals the activated alternative 5' and 3' splice sites in exons 2 and 9, respectively. (**B**) Electropherogram of the spliced product that was generated from ectopically expressed TSG101-EGFP[+] in breast cancer cells (MCF-7). The alignment of this sequence with that of TSG101-EGFP[+] reveals the activation of the same exonic 5' and 3' splice sites as those in the endogenous TSG101 mRNA (see panel **A**). The EGFP-coding sequence is indicated with the green bar.

TABLE

Table S1. DNA sequences of the primers used in the RT–PCR amplification of the TSG101 and FHIT transcripts

RT primers to detect TSG101/FHIT exonic lariats across branch point	
R1 (for TSG101)	5'-GGATAGGATGCCGAAATAGGACGAGAGAAG-3'
R2 (for FHIT)	5'-TTCGGAGTCCTCAGTGGC-3'

RT primers to detect FHIT lariat RNA	
R3 (for exons)	5'-TGTTTTTCCACCACTGTCCCG-3'
R4 (for intron 5)	5'-CCACATCACAAAGGCACCACTC-3'

Detection of TSG101 aberrant splicing	
P1	5'-CGGTGTCGGAGAGCCAGCTCAAGAAA-3'
P2	5'-CCTCCAGCTGGTATCAGAGAAGTCAGT-3'
P3 (inner)	5'-AGCCAGCTCAAGAAAATGGTGTCCAAG-3'
P4 (inner)	5'-TCACTGAGACCGGCAGTCTTTCTTGCTT-3'
P5	5'-TCTTTGCTCAGGGCGGACTGGGTG-3'
P6 (inner)	5'-GGACACGCTGAACTTGTGGCCG-3'

Detection of TSG101 Exon 1	
P7	5'-CCGACTTCCTGTTGTTTGAGGC-3'
P8	5'-CTTGGACACCATTTTCTTGAGCTG-3'

Detection of TSG101 Exon 9–10	
P9	5'-TTGAGAAGGGGCGTGATAGACC-3'
P10	5'-TGGGCACCTACTGATAAAAGGAAG-3'

Detection of TSG101 mature mRNA	
P11	5'-TAGGGATGGCACAATCAGCGAG-3'
P11' (inner)	5'-CTGGAAGAGATGGTTACCCGTTTAGA-3'
P12	5'-TTAGCAGTCCCAACATTCAGCACA-3'
P12' (inner)	5'-CGGCAGTCTTTCTTGCTTTTTGC-3'

Detection of TSG101 lariat Intron 7	
P13	5'-GGCTTTGGTGATTCCTTACTCCAG-3'
P13' (inner)	5'-GGTGATTCCTTACTCCAGTCTCAGCA-3'
P14	5'-GCAGCATTTTCAGGGTTTTCCAC-3'
P14' (inner)	5'-CAGCATTTTCAGGGTTTTCCACTCC-3'

Detection of TSG101 lariat Intron 8	
P15	5'-GGTTCAAGCGATTCTCCTGACTC-3'
P15' (inner)	5'-TGGTGGCAAGAGGAAATGAGG-3'
P16	5'-GGGGATGAAAATGGAGCGTG-3'
P16' (inner)	5'-CATAGGGCTGACTTACACACGGAC-3'

Detection of TSG101 exonic lariat	
P17	5'-ACTTGTTGGGGCTTATTCAGGTC-3'
P17' (inner)	5'-TGAACCTCCAGTCTTCTCTCGTCC-3'
P18	5'-TCATCCGCCATCTCAGTTTGTC-3'
P18' (inner)	5'-ATGGTGTCCTCGCTGATTGTGC-3'

Detection of TSG101 pre-mRNA / large lariat (targeting Exon 7)	
P19	5'-TGTTGCTTCTGATGCTGTTGGATTG-3'
P19' (inner)	5'-TTCCGCTGCTATGAGGGTGACTC-3'
P20	5'-AAGGAACTGCTGAAGTGATGCCC-3'
P20' (inner)	5'-AGATAGGCATAGGTGAAGGCTGGAG-3'

Detection of TSG101 pre-mRNA / large lariat (targeting Exon 8)	
P21	5'-CTGTCCTGTTGGTAAAGGGTGAAG-3'
P21' (inner)	5'-TTAGCAGCGAAGCCTGATGTTG-3'
P22	5'-TGATTCTGAGGTCTCCCCAGC-3'
P22' (inner)	5'-CTCCTACTATGCCCACAACATCCAG-3'

Detection of TSG101 exonic lariat across branch point	
P23	5'-CCAGTAGGGATGGCACAATCAGCG-3'
P23' (inner)	5'-GGATGAAGAACTCAGTTCTGCTCTGG-3'
P24	5'-ACAAGTCTGACTGTGGGTGTTTCCATTCA-3'
P24' (inner)	5'-CAGGTTTGAGATCTTTGTATAGAGTAATAACATTG-3'

Detection of FHIT aberrant splicing	
P25	5'-GCTCCCTCCCTCTGCCTTTCATTC-3'
P26	5'-GGGAAACCTCAAATCTGCCTGTCTGA-3'
P25' (inner)	5'-TCCCTCTGCCTTTCATTCCCAGC-3'
P26' (inner)	5'-CAAATCTGCCTGTCTGAGCCGTTTAG-3'

Detection of FHIT lariat Intron 5	
P27	5'-AAAGTCTTCTGAGCCTAACCAGCC-3'
P27' (inner)	5'-GCCTCCAGTCTGTCCAAAAAGG-3'
P28	5'-ATTCACCTCCAAAGCCACGG-3'
P28' (inner)	5'-TCCAAAGCCACGGTTGCTAAG-3'

Detection of FHIT pre-mRNA / large lariat (targeting Exon 5)	
P29	5'-TCCCTCAAAAGGAAGACGCAG-3'
P29' (inner)	5'-CTCCGTTTTGCTCCTTCCTTCT-3'
P30	5'-TTACCTTTTTGGACAGACTGGAGG-3'
P30' (inner)	5'-TCATTTGGCTGGTTAGGCTCAG-3'

Detection of FHIT exonic lariat	
P31	5'-GAAGGGAGAGAAAGAGAAAGAAGGTATC-3'
P31' (inner)	5'-GGGAGAGAAAGAGAAAGAAGGTATCC-3'
P32	5'-CTGGGTCGTCTGAAACAAATCG-3'
P32' (inner)	5'-CCACTTCATCAGGACGCAGG-3'

Detection of FHIT of exonic lariat across branch point	
P33	5'-GTGGCCGATTTGTTTCAGACGACC-3'
P33' (inner)	5'-GAGAGTCGGGACAGTGGTGGAAAAAC-3'
P34	5'-GCTTTTATAGAGGGTCTAAGCAGGCAGG-3'
P34' (inner)	5'-TAGAGGGTCTAAGCAGGCAGGTATTCCTAG-3'

Specific detection of normal / aberrant TSG101 mRNA (for semi-quantitation)	
P35 (Normal)	5'-GAAGTTATCATTCCCACAGCTC-3'
P36 (Aberrant)	5'-TACAAATACAGAGACCTAACTCTCCC-3'
P37 (Common)	5'-GACGTACATGCTTCAGGAAGAC-3'

MATERALS AND METHODS (Full Descriptions)

Splice sites scoring methods

The URLs of the nine computer programs listed in Table 1 are as follows: http://ibis.tau.ac.il/ssat/SpliceSiteFrame.htm for S&S (Shapiro and Senapathy Score) and ΔG (Free energy); http://www.uni-duesseldorf.de/rna/html/hbond_score.php for H-Bond (Hydrogen bonding); http://www.fruitfly.org/seq_tools/splice.html for NN (Neural Network); http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html for MAXTENT (Maximum Entropy Model), MM (Markov Model), MDD (Maximum Dependence Deposition), and WMM (Weight Matrix Model); http://www.med.nagoya-u.ac.jp/neurogenetics/SD_Score/sd_score.html for SD (SD score).

Construction of plasmids

All the synthetic oligonucleotides used as primers were purchased (Invitrogen). The following PCRs were performed with Ex *Taq* HS DNA polymerase (Takara Bio). An enhanced green fluorescent protein (EGFP) fragment lacking the initiation codon was amplified by PCR from the pBI-EGFP plasmid (Clontech) using the forward primer 5'-GCtctagaGTGAGCAAGGGCGAGGAGCTG-3' (lower-case letters indicate the XbaI site) and the reverse primer 5'-GGgaattcTTACTTGTACAGCTCGTCCATGCCG-3' (lower-case letters indicate the EcoRI site). The amplified products were subcloned into the pGEM-T Easy vector (Promega) to generate the pGEM-T-EGFP plasmid.

The TSG101 fragments (including the full-length sequence, 1–1111 bp and 1–1112 bp of the ORF for EGFP, EGFP[+], and EGFP[–], respectively) were constructed by PCR from the TSG101 cDNA (MGC full-length cDNA collection; Open Biosystems) using the forward primer 5'-GgaattcCACCATGGCGGTGTCGGAGAGCCA-3' (lower-case letters indicate the EcoRI site) and the reverse primers 5'-GCtctagaGTAGAGGTCACTGAGACCGGCAGTC-3', 5'-GCtctagaGGATCTGTTTGTATAAGGGAGCTGTGG-3', and 5'-GCtctagaAGGATCTGTTTGTATAAGGGAGCTGTG-3' (lower-case letters indicate the XbaI site), respectively. The amplified products were subcloned into the pGEM-T Easy vector. The TSG101 fragments were excised with SpeI/XbaI and ligated into the corresponding sites on the pGEM-T-EGFP plasmid. The TSG101-EGFP fusion plasmids generated were cleaved with EcoRI, and subcloned into the same site on the pCXN2 mammalian expression vector (28; see REFERENCES in the text) to produce the pCXN2-TSG101-EGFP, pCXN2-TSG101-EGFP[+], and pCXN2-TSG101-EGFP[–] plasmids.

The FHIT fragment, comprising 116–829 bp from the transcription start site, was constructed by PCR from the first-strand cDNA of HMEC (see the RT–PCR method described below) using the forward primer 5'-GactagtgaattcTCCCTCCCTCTGCCTTTCATTCCC-3' (lower-case letters indicate the SpeI/EcoRI sites) and the reverse primer 5'-GCtctagaGAAAAACATCTGTGTCACTGAAAGTAGACCCG-3' (lower-case letters indicate the XbaI site), respectively. As described above for the TSG101 series plasmids, the amplified products were fused to the EGFP fragment and subcloned into the pCXN2 vector to produce the pCXN2-FHIT-EGFP plasmid.

Cell lines, transient transfection of cells, and preparation of total cellular RNA

Normal human mammary epithelial cells (HMEC; no. CC-2551) were purchased from Lonza. MCF-7 cells were supplied by the Cell Resource Center for Biomedical Research (at the Institute of Development, Aging and Cancer, Tohoku University). CaSki, HepG2, and SK-N-SH cells were purchased from the Cell Bank of the RIKEN BioResource Center. The breast cancer cell lines Hs 578Bst (no. HTB-125), HCC1395 (no. CRL-2324), HCC1937 (no. CRL-2336), HCC70 (no. CRL-2315), HCC1419 (no. CRL-2326), and HCC1569 (no. CRL-2330); and the small-cell lung cancer cell lines DMS 79 (no. CRL-2049), NCI-H1092 (no. CRL-5855), and NCI-H69 (no. HTB-119) were purchased from and the American Type Culture Collection (ATCC). The Hs 578T cell line was kindly provided by Dr R. Reeves. These cells were grown as recommended by the manufacturers.

The MCF-7 cells (70%–80% confluence) were transiently transfected with plasmids using Lipofectamine LTX (Invitrogen), according to the manufacturer's instructions. At 24 h after

transfection, the cells were examined for the expression of GFP fluorescence, and at 48 h after transfection, the total RNA was prepared from the cells.

The total cellular RNA was prepared from cells on 35-mm dishes (either separate dishes or six-well plates) using 1 mL of TRIzol reagent (Invitrogen) and digested at 37°C for 10 min with 10 units of recombinant DNase I (Takara Bio) in a 50 μL reaction mixture containing 40 mM Tris-HCl (pH 7.5), 8 mM $MgCl_2$, 1 mM $CaCl_2$, 5 mM dithiothreitol (DTT), and 20 units of RNase inhibitor (New England Biolabs).

RT–PCR detection of TSG101 and FHIT mRNAs

The prepared total cellular RNA (5 μg) was reverse transcribed in a 20 μL reaction mixture with oligo(dT) primer (Invitrogen) and PrimeScript reverse transcriptase (Takara Bio), according to the manufacturer's instructions. The sequences of all the RT–PCR primers (R1–R4, P1–P37, P11'–P34'; see below) are listed in Table S1.

cDNA solutions (1 μL) were used for PCR amplification in 20 μL volumes with 1 unit of Ex *Taq* HS DNA polymerase (Takara Bio) and 1 μM primers, according to the manufacturer's instructions. To detect TSG101 mRNA (Figure 2B, Figure S1A, B), the PCR mixtures with the primer set P1–P2 were incubated at 95°C for 5 min, followed by 24 PCR cycles of; 98°C for 10 sec, 65°C for 30 sec, and 72°C for 2 min, with a final incubation at 72°C for 10 min. To detect FHIT mRNA (Figure 5B), the PCR mixtures with the primer set P25–P26 were incubated at 95°C for 5 min followed by 22 PCR cycles of: 95°C for 30 sec, 62°C for 30 sec, and 72°C for 1.5 min, with a final incubation at 72°C for 10 min. The amplified TSG101 and FHIT products were purified with Nucleospin Extract II (Macherey-Nagel), and 1 μL of the 50 μL of eluted DNA solution was used for the second nested PCR (performed under the same conditions as the first PCR) with the adjacent inner sets of primers, P3–P4 and P25'–P26', respectively.

To detect the spliced products from the TSG101-EGFP- and FHIT-EGFP-transfected cells (Figure 2E, 5C), 1 μL of 1:100-diluted cDNA solutions, prepared with SuperScript III RNase H^- reverse transcriptase (Invitrogen), were used instead. The first PCR (30 and 22 cycles for TSG101-EGFP series and FHIT-EGFP, respectively) were performed with the primer sets P1–P5 and P25–P5, respectively. The amplified products were purified with Nucleospin Extract II, and 1 μL aliquots of the 1:100-diluted solutions were used for the second nested PCR (performed under the same conditions as the first PCR) with the two inner primer sets, P3–P6 and P25'–P6, respectively.

All the PCR products were electrophoresed on 2% agarose gels and visualized them with GelRed (Nacalai Tesque).

Preparation of genomic DNA and PCR of the *TSG101* gene

Cells on 35-mm dishes were lysed in 0.5 mL of lysis buffer containing 10 mM Tris-HCl (pH 8.0), 150 mM NaCl, 10 mM EDTA (pH 8.0), 0.1% SDS, and 144 μg/mL proteinase K (Roche) at 50°C for 90 min. The lysate was extracted with phenol/chloroform and precipitated with ethanol, and the pellet was dissolved in buffer containing 10 mM Tris-HCl (pH 8.0) and 1mM EDTA to a concentration of 1 μg/μL.

The genomic DNA solution (1 μL) was used for a two-round PCR under exactly the same RT–PCR conditions as described above, with the primer sets P1–P2 and P3–P4 (Figure 2B). For the positive control, a one-round PCR (30 cycles) was performed with the primer sets P7–P8 and P9–P10 (Figure 2B) and the cycling program described above, except for an annealing temperature of 57.5°C for 30 sec and 72°C for 1 min (instead of 65°C for 30 sec and 72°C for 2 min, respectively). The PCR products were electrophoresed on 2% agarose gels and visualized them with GelRed

Detection/identification of TSG101 splicing products

To prepare an RNA sample lacking linear RNA species, RNase R digestion was performed, essentially as described previously (29; see REFERENCES in the text). In brief, 5 μg of total cellular RNA was digested in a 50 μL reaction mixture containing 20 mM Tris-HCl (pH 8.0), 100 mM KCl, 0.25 mM $MgCl_2$, and 1 mM DTT, with (+) or without (–) 1 μg of purified recombinant RNase R (a generous gift from Dr A. Malhotra; commercially available from Epicentre Biotechnologies). The reaction mixtures

were incubated at 37°C for 1 h, extracted with phenol/chloroform, and precipitated with ethanol. This RNase R digestion was repeated once more under the same conditions.

Each RNA sample was reverse transcribed in 20 μL using the PrimeScript reverse transcriptase (Takara Bio) or SuperScript III reverse transcriptase (Invitrogen) with random hexamer primers (Takara Bio). To detect the branched region of the TSG101 exonic lariat (Figure 4A), the specific primer R1 was used together with random hexamer primers. The cDNA solution (1 μL) was used as the template for a PCR in 20 μL using PrimeSTAR Max or Ex *Taq* HS DNA polymerase (Takara Bio), according to the manufacturer's instructions. The first PCR mixtures with target-specific primers P11–P24 were incubated at 95°C for 5 min, followed by 35 PCR cycles of: 98°C for 10 sec, 55°C for 15 sec, and 72°C for 1 min, with a final incubation at 72°C for 10 min (Figure 3B, 4A). To detect lariat intron 7 and exon 7 in the large lariat RNA (pre-mRNA), an annealing temperature of 58°C for 30 sec (instead of 55°C for 15 sec) and 36 PCR cycles (instead of 35 cycles) were used (Figure 3B). The amplified products were purified as described above and used for the second nested PCR (35 cycles) with the inner primers P11'–P24'.

All the PCR products were electrophoresed on 2% agarose gels and visualized them with GelRed. The isolated DNA fragments were subcloned into the pGEM-T Easy vector and sequenced with the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems).

Detection/identification of the FHIT splicing products

The RT–PCR procedure was essentially same as that for the TSG101 splicing products (see above), unless otherwise specified below.

Total cellular RNA (5 μg) was digested once with 2 μg of recombinant RNase R at 37°C for 1 h and reverse transcription was performed with random hexamer primers and the specific primers R4, R4, and R3, to detect the lariat intron 5, exon 5 in the large lariat RNA (pre-mRNA), and the exonic lariat, respectively (Figure 6). The PCR mixtures with the primers P27–P32 were incubated at 95°C for 5 min, followed by 40 PCR cycles of: 98°C for 10 sec, 56°C for 30 sec, and 72°C for 1 min, with a final incubation at 72°C for 10 min. This PCR was followed by a second nested PCR (same cycle number) with the inner primers P27'–P32'.

Total cellular RNA (30 μg) was digested twice with 1 μg of RNase R at 37°C for 1 h and reverse transcription was performed with random hexamer primers and the specific primer R2 to detect the branched region of the FHIT exonic lariat (Figure 7A). The two-round PCRs with primers P33–P34 and P33'–P34' were as described above, except for an annealing temperature of 60°C for 30 sec (instead of 56°C for 30 sec).

All the PCR products were electrophoresed on 2% agarose gels and visualized them with GelRed.

Semi-quantitative RT–PCR analysis of the normal and aberrant TSG101 mRNAs

One-round PCR (no nested PCR) was performed with either the normal mRNA-specific primer set P35–P37 or the aberrant mRNA-specific primer set P36–P37 (Figure S1C). The PCR mixtures were incubated at 95°C for 5 min, followed by 35 PCR cycles of: 98°C for 10 sec, 62°C for 30 sec, and 72°C for 30 sec, with a final incubation at 72°C for 5 min. The RT–PCR procedure was essentially the same as that for the TSG101 splicing products (see above).

For the semi-quantitative analysis, the PCR products were electrophoresed on 2% agarose gels, visualized them with GelRed, and scanned with the BioDoc-It Imaging System (Ultra-Violet Products). The intensity of the PCR bands was analyzed with ImageJ freeware. The means ± standard deviation of four replicates were plotted as histograms.