

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
10 September 2010 (10.09.2010)

(10) International Publication Number
WO 2010/102277 A2

(51) International Patent Classification:

G01N 33/574 (2006.01) A61K 38/17 (2006.01)
C07K 19/00 (2006.01) A61P 35/00 (2006.01)
C12N 15/62 (2006.01)

York 14850 (US). **RICKMAN, David, S.** [US/US]; c/o Cornell University, Cornell Center for Technology, Enterprise & Commercialization, 395 Pine Tree Road, Suite 310, Ithaca, New York 14850 (US).

(21) International Application Number:

PCT/US2010/026495

(74) Agent: **GROLZ, Edward, W.**; Scully, Scott, Murphy & Presser, 400 Garden City Plaza, Suite 300, Garden City, NY 11530 (US).

(22) International Filing Date:

8 March 2010 (08.03.2010)

(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, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, 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.

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/158,276 6 March 2009 (06.03.2009) US

(71) Applicant (for all designated States except US): **CORNELL UNIVERSITY** [US/US]; Cornell Center for Technology, Enterprise & Commercialization, 395 Pine Tree Road, Suite 310, Ithaca, New York 14850 (US).

(72) Inventors; and

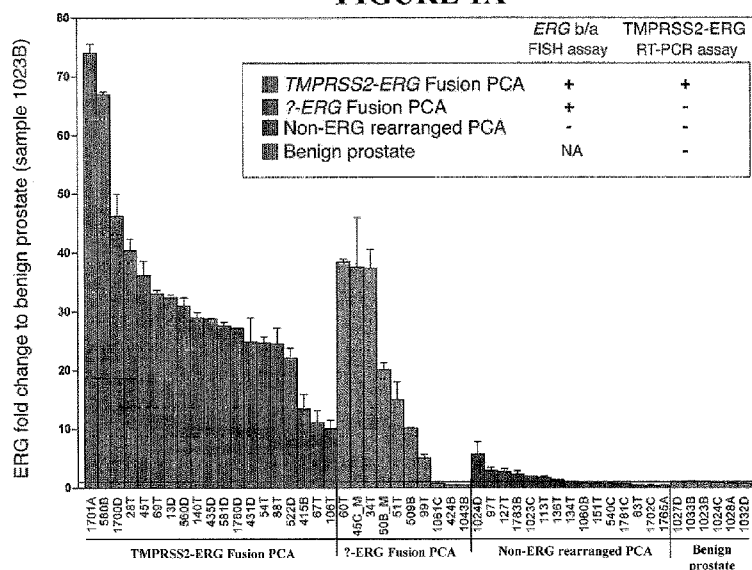
(75) Inventors/Applicants (for US only): **RUBIN, Mark, A.** [US/US]; c/o Cornell University, Cornell Center for Technology, Enterprise & Commercialization, 395 Pine Tree Road, Suite 310, Ithaca, New York 14850 (US). **PFLUEGER, Dorothee** [US/US]; c/o Cornell University, Cornell Center for Technology, Enterprise & Commercialization, 395 Pine Tree Road, Suite 310, Ithaca, New York 14850 (US).

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: METHODS OF DIAGNOSING AND TREATING PROSTATE CANCER CHARACTERIZED BY NDRG1-ERG FUSION

FIGURE 1A



(57) Abstract: An in depth analysis of prostate cancer prostatectomy samples which over-express the ERG oncogene led to the discovery of a novel gene translocation in prostate cancer, between the NDRG1 gene (N-myc downstream regulated gene 1) on chromosome 8 and the ERG oncogene on chromosome 21, leading to the expression of a chimeric NDRG1-ERG protein. Methods and compositions useful for diagnosing and treating prostate cancer characterized by NDRG1-ERG fusion are described.

WO 2010/102277 A2

Published:

- without international search report and to be republished upon receipt of that report (Rule 48.2(g))
- with sequence listing part of description (Rule 5.2(a))

**Methods of Diagnosing and Treating Prostate Cancer
Characterized by *NDRG1-ERG* Fusion**

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This application claims the benefit of priority of U.S. Provisional Application No. 61/158,276, filed on March 6, 2009.

GOVERNMENT FUNDING

[0002] This invention was made with Government support under Grant Number P50CA090381 and R01 CA125612-01 awarded by NIH's National Cancer Institute. The United States Government has certain rights in the invention.

FIELD OF THE INVENTION

[0003] This invention relates to cancer diagnosis and treatment. More specifically, the invention relates to compositions and methods for diagnosing and treating prostate cancer characterized by *NDRG1-ERG* fusion.

BACKGROUND OF THE INVENTION

[0004] The majority of prostate cancers detected through PSA screening harbor an acquired recurrent chromosomal rearrangement (Tomlins et al., *Science*, 310, 644-8, 2005). The promoter region of the androgen-regulated transmembrane protease, serine 2 (*TMPRSS2*) gene is most often fused to the coding region of members of the erythroblast transformation specific (ETS) family of transcription factors, most commonly v-ets erythroblastosis virus E26 oncogene homolog (avian) (*ERG*). The *TMPRSS2-ERG* fusion is observed in around 90% of tumors that over-express the oncogene *ERG*. Other, less common, fusion events occur involving ETS family members (*ETV1*, *ETV4* and *ETV5*) fused to *TMPRSS2* or other 5' partners that differ in their prostate specificity and response to androgen (*SLC45A3*, *HERV-K*, *C15orf21*, *HNRPA2B1*, *FLJ35294*, *DDX5*, *CANT1* and *KLK2*, reviewed by Kumar-Sinha et al., *Nat Rev Cancer* 8(7):497-511, 2008; and more recently, *ACSL3* (Attard et al., *Br J*

Cancer, 99, 314-20, 2008)). Moreover, variations in the structure of the gene fusions in prostate cancer yielding different fusion transcript isoforms have been reported (Wang et al., *Cancer Res*, 66, 8347-51, 2006). ETS rearranged prostate cancer, similar to other translocation tumors, may represent a distinct molecular subclass of prostate cancer based on studies demonstrating characteristic morphologic features (Mosquera et al., *J Pathol* 212: 91-1012007), natural history (Attard et al., *Oncogene* 27: 253-63, 2008; Demichelis et al., *Oncogene* 26: 4596-9, 2007) and specific genomic (Demichelis et al., *Genes Chromosomes Cancer* 48: 366-380, 2009) and expression profiles (Setlur et al., *J Natl Cancer Inst* 100: 815-25, 2008).

SUMMARY OF THE INVENTION

[0005] A novel gene fusion has been identified between NDRG1 (N-myc downstream regulated gene 1) and ERG (v-ets erythroblastosis virus E26 oncogene homolog) in prostate cancer over-expressing ERG. The NDRG1-ERG gene fusion is inducible by androgen and by estrogen, and encodes a fusion-specific protein. Compositions and methods useful for diagnosing and treating cancer including prostate cancer are provided herein.

[0006] In one aspect, the invention provides a method for diagnosing cancer such as prostate cancer based on detecting in a biological sample, the presence of an NDRG1-ERG fusion molecule. The biological sample can be any suitable sample obtained or derived from the patient, including for example, tissue, cells, blood, urine, semen, and prostatic secretions.

[0007] The NDRG1-ERG gene fusion can be detected at the genomic or chromosomal DNA, mRNA or protein level. Fusion nucleic acid molecules can be detected by using various nucleic acid-based techniques, including hybridization, amplification, and sequencing. Fusion proteins can be detected using a variety of assays known for detection of proteins, including, for example, SDS-gel analysis and immunoassays.

[0008] In some embodiments, the NDRG1-ERG fusion is detected at the chromosomal level using a fluorescent in situ hybridization assay (FISH). Either or both of a break apart FISH assay that detects translocation of the NDRG1 gene, and a fusion FISH assay that detects a genomic fusion between NDRG1 and ERG can be used.

[0009] In other embodiments, the NDRG1-ERG fusion is detected at the mRNA level by using a nucleic acid amplification method (e.g., RT-PCR), a nucleic acid hybridization method (Northern blot analysis), or a method that combines nucleic acid amplification and nucleic acid hybridization.

[0010] For detection of an NDRG1-ERG fusion mRNA in an amplification method, one can utilize a first primer specific for a 5' region of an NDRG1 mRNA, and a second primer specific for a 3' region of an ERG mRNA.

[0011] Detection of an NDRG1-ERG fusion mRNA can also be achieved in an amplification or hybridization method by using an oligonucleotide primer or probe specific for the junction of the fusion mRNA. Junctions of two fusion transcript variants are shown in Figure 4A and Figure 5A, and more locally in Figure 2B.

[0012] In still other embodiments, the NDRG1-ERG fusion is detected at the protein level. For example, detection can be directed to an NDRG1-ERG fusion protein containing the amino acid sequence as set forth in SEQ ID NO: 7 or 9. Such fusion protein can be detected in an immunoassay using an antibody, e.g., an antibody which binds specifically to the fusion junction.

[0013] Detection of the NDRG1-ERG fusion can be combined with detection of one or more other fusions associated with cancer such as prostate cancer, including, e.g., fusions between TMPRSS2 and ERG, and between SCL45A3 and ERG.

[0014] Compositions and kits containing one or more nucleic acid primers, probes, and antibodies, suitable for use in the detection of NDRG1-ERG fusion molecules are also provided.

[0015] In another aspect, the present invention provides isolated nucleic acids encoding an NDRG1-ERG fusion protein, and isolated NDRG1-ERG fusion polypeptides, as well as related expression vectors and host cells.

[00016] In a further aspect, the invention provides a method for identifying an agent useful for treating prostate cancer characterized by the presence of the NDRG1-ERG fusion. Such agent can be identified by screening for agents based on the ability to inhibit a biological function or reduce the level of an NDRG1-ERG fusion molecule in a cell which expresses the NDRG1-ERG fusion molecule. An example of a biological function of a NDRG1-ERG fusion protein is to enhance the invasion ability of the cell which expresses the NDRG1-ERG fusion protein.

[00017] In a further aspect, the invention provides a method for treating a patient having prostate cancer characterized by NDRG1-ERG fusion. Such method involves administration of an agent that inhibits a biological function or reduces the level of a NDRG1-ERG fusion molecule.

BRIEF DESCRIPTION OF THE DRAWINGS

[00018] **Figure 1.** ERG mRNA expression in prostate cancer and benign tissue. (A) Quantitative RT-PCR of ERG expression in 29 *ERG* rearranged (including 19 TMRSS2-*ERG* mRNA positive (orange) and unknown mechanism-*ERG* (?-*ERG*, green)), 15 *ERG* non-rearranged (blue) and 6 benign prostate tissue samples (gray). (B) Exon composition and sequence (SEQ ID NO: 3) covering the fusion junction of *SLC45A3-ERG* transcript. (C) FISH images of nuclei with *SLC45A3* rearrangement (upper) and *SLC45A3-ERG* fusion (lower) nucleus with yellow fusion signal.

[00019] **Figure 2.** Identification of *NDRG1-ERG* fusion by RNA sequencing. (A) The schematic shows the linear structure of *NDRG1* and *ERG*. The gene representation shows the "union" transcripts, i.e. the exons of all isoforms are reported and, in the case of overlapping exons, the longest one is shown. Each arc represents one instance of paired reads where one read is mapped to *NDRG1* and the other to *ERG*. The regions of the genes involved in the fusion transcript are highlighted and numbered. (B) RT-PCR products obtained using a forward primer targeting exon 1 of *NDRG1* and a reverse primer targeting exon 6 of *ERG* (positive control:beta actin). Arrows indicate the DNA fragments that were isolated and sequenced. The lower portion provides sequence data from this analysis showing the

NDRG1-ERG transcript exon composition and the sequence covering the fusion junction for the 2 variant mRNAs identified in samples 99_T (top, SEQ ID NO: 4) and 509_B (bottom, SEQ ID NO: 5). (C) Schematic of the FISH *NDRG1* b/a and *NDRG1-ERG* fusion assays. (D) *NDRG1* rearrangement (upper) indicated by separated red and green signals and *NDRG1-ERG* fusion (lower) indicated by an overlap of the red *ERG* and the green *NDRG1* signal in a representative nucleus from case 99_T.

[00020] **Figure 3.** Representative image of a metaphase spread from normal human male lymphocytes displaying the correct chromosome 8q24.22 position of FISH BAC probes targeting the *NDRG1* locus used in the b/a assay.

[00021] **Figure 4A.** Nucleotide sequence (SEQ ID NO: 6) from *NDRG1-ERG* fusion cDNA, variant 1. The *ERG* portion is underlined.

[00022] **Figure 4B.** Protein Sequence of *NDRG1-ERG* chimeric protein (SEQ ID NO: 7) encoded by *NDRG1-ERG* cDNA variant 1, with the *ERG* portion underlined.

[00023] **Figure 5A.** Nucleotide sequence (SEQ ID NO: 8) from *NDRG1-ERG* fusion cDNA, variant 2. The *ERG* portion is underlined.

[00024] **Figure 5B.** Protein Sequence of *NDRG1-ERG* chimeric protein (SEQ ID NO: 9) encoded by *NDRG1-ERG* cDNA variant 2, with the *ERG* portion underlined.

[00025] **Figure 6.** Hormone treatment of LNCaP cells induced *SLC45A3* and *NDRG1* mRNA expression. *SLC45A3* (A and C) or *NDRG1* (B and D) mRNA expression was induced upon stimulation with synthetic androgen (R1881) and 17 β -estradiol (E2). Serum-starved LNCaP cells were stimulated with 1nM R1881, 1nM R1881 in combination with 10 μ M Flutamide (A and B), 10 nM E2 or 10 nM diarylpropionitrile (DPN) (C and D) for 3, 12 and 24h. Samples were run in triplicate and normalized against *TCFL1*. Columns indicate the mean fold change of induction of three biological replicates against vehicle (Ethanol) only treated cells for the respective time points \pm SEM.

[00026] **Figure 7.** Androgen and 17 β -estradiol (E2) signaling of known target genes in LNCaP cells. IGF2R (A) and PSA (B) mRNA expression was induced upon stimulation with synthetic androgen (R1881) and E2. Serum-starved LNCaP cells were stimulated with 1nM R1881, 1nM R1881 in combination with 10 μ M Flutamide, 10 nM E2 or 10 nM diarylpropionitrile (DPN) for 3, 12 and 24h. Total RNA was extracted and used for quantitative RT-PCR using TAQMAN assay. Samples were run in triplicate and normalized against TCFL1. Columns indicate the mean fold change of induction of three biological replicates against vehicle (Ethanol) only treated cells for the respective time points \pm SEM.

[00027] **Figure 8.** NDRG1-ERG protein expression in HEK-293 (embryonic kidney) cells.

[00028] **Figure 9.** NDRG1-ERG protein expression in BPH1 (prostate epithelial) cells.

[00029] **Figure 10.** NDRG1-ERG mRNA over-expression in HEK-293 and BPH1 cell lines. "RQ" stands for relative quantity.

[00030] **Figure 11.** Expression of NDRG1-ERG enhanced cell invasion. mRNA (top) and protein (middle) expression in the indicated cell lines following transient transfection of either NDRG1-ERGflag or NDRG1-ERG retroviral expression systems. Invasion assay (bottom) of HEK293 cells expressing LacZ control (left) or NDRG-1-ERG fusion (right) proteins.

DETAILED DESCRIPTION OF THE INVENTION

[00031] The present inventors have identified a novel gene fusion associated with cancers including prostate cancer. More specifically, a novel gene fusion has been identified in prostate cancer over-expressing ERG, which fusion involves NDRG1 (N-myc downstream regulated gene 1) and ERG (v-ets erythroblastosis virus E26 oncogene homolog). The NDRG1-ERG gene fusion is inducible by androgen and by estrogen, and encodes a fusion-specific protein. Accordingly, the present invention provides compositions and methods useful for diagnosing and treating cancer including prostate cancer characterized by the NDRG1-ERG fusion.

[00032] *NDRG1-ERG Fusion Molecules*

[00033] A "NDRG1-ERG fusion molecule", as referred to herein, can be a chimeric nucleic acid molecule (genomic DNA, cDNA, and RNA) or a chimeric protein molecule.

[00034] Without being bound to any particular theory, it is believed that the fusion between NDRG1 and ERG results from chromosomal rearrangement or translocation which brings together a 5' portion of the NDRG1 gene and a 3' portion of the ERG gene, normally located on separate chromosomes, to create a chimeric gene at one chromosomal location (i.e., a genomic fusion molecule).

[00035] While the junction of the genomic fusion molecule may vary, the 5' portion of the NDRG1 gene that constitutes the genomic fusion molecule typically includes a portion from the 5' transcription regulatory region of the NDRG1 gene. By "5' transcription regulatory region", it is meant the region upstream of the transcription start site of a gene that controls transcription of the gene, which includes a promoter, a TATA box in many cases, and possibly one or more of other regulatory elements (e.g., an enhancer). In addition to a portion from the 5' transcription regulatory region of the NDRG1 gene, the genomic fusion molecule can also include one or more exons and introns from the 5' region of the NDRG1 gene or portions thereof.

[00036] The 3' portion of the ERG gene that constitutes the genomic fusion molecule typically includes a portion from the 3' region of the ERG gene, for example, the 3' region of the ERG gene coding for the 3' untranslated sequence of an ERG mRNA or a portion thereof, one or more exons or introns from the 3' region of the ERG gene or portions thereof.

[00037] Transcription of a genomic NDRG1-ERG fusion molecule produces a NDRG1-ERG fusion transcript (i.e., chimeric mRNA). A NDRG1-ERG fusion transcript is composed of a 5' portion of an NDRG1 mRNA, joined 5' to a 3' portion of an ERG mRNA.

[00038] The 5' portion of an NDRG1 mRNA that constitutes a fusion transcript typically includes the 5' un-translated region of an NDRG1 mRNA. By "5' un-translated region" it is meant the region of an mRNA that starts at the +1 position (i.e., where transcription begins)

and ends just before the start codon of the coding region. The fusion transcript can also include full length or portions of one or more exons from 5' of an NDRG1 mRNA.

[00039] By a "portion" of an exon, it is meant a contiguous sequence of an exon that is shorter than the entire length of the exon. Generally speaking, a portion of an exon can be at least 5, 10, 15, 20, 25, 30, 35, 40 nucleotides or more in length.

[00040] There are several NDRG1 transcription splice variants and ERG transcription splice variants in human. The cDNA sequence of human NDRG1 transcription variant 2 and the locations of its exons are illustrated in SEQ ID NO: 1.

[00041] In one embodiment, the fusion transcript includes at least exon 1 of an NDRG1 mRNA, or a portion of exon 1. In another embodiment, the fusion transcript includes at least exon 1 and exon 2 or a portion of exon 2 of an NDRG1 mRNA. In still another embodiment, the fusion transcript includes at least exon 1, exon 2, and exon 3 or a portion of exon 3 of an NDRG1 mRNA.

[00042] The 3' portion of an ERG mRNA that constitutes a fusion transcript may include the 3' un-translated region transcribed from the ERG gene. The 3' un-translated region is the section of an mRNA that follows the coding region and is not translated. The 3' un-translated region is typically followed by a poly-A tail. The fusion transcript can also include full length or portions of one or more exons from 3' of an ERG mRNA. Several transcription splice variants have been reported for human ERG. The cDNA sequence of human ERG transcription variant 3 and the locations of its exons are illustrated in SEQ ID NO: 2. The exons in this variant are numbered in SEQ ID NO: 2 consecutively, consistent with the report by Wang et al. (*Cancer Res.* 2006 Sep 1;66(17):8347-51).

[00043] The junction of a NDRG1-ERG fusion transcript may vary, which may result from variations in the junction of NDRG1-ERG fusion at the genomic level, or alternatively from variations in transcription splicing from a genomic fusion molecule. Two NDRG1-ERG fusion transcript variants have been identified in accordance with the invention, as described in details in the following examples. cDNA sequences derived from these mRNA variants

which include the exons and the junction of the fusion are depicted in Figures 4A and 5A, and set forth in SEQ ID NOS: 6 and 8, respectively.

[00044] Upon translation, the fusion transcripts produce fusion proteins. The two NDRG1-ERG fusion transcript variants identified herein are found to encode and produce chimeric NDRG1-ERG fusion proteins, the sequences of which are depicted in Figures 4B and 5B, and set forth in SEQ ID NOS: 7 and 9, respectively.

[00045] Cancer Diagnosis

[00046] According to the present invention, diagnosis of cancer in a subject is based on detection of the NDRG1-ERG fusion. The methods provided by the present invention are applicable to diagnosing cancer, including but not limited to prostate, breast, colon, pancreas, and lung cancers. In one specific embodiment, the methods are directed to diagnosis of prostate cancer.

[00047] The term "subject" being tested includes all mammalian subjects, particularly human subjects.

[00048] The term "diagnosis" or "diagnosing" is meant a determination that the subject has cancer or likely has cancer. The diagnostic method based on detection of NDRG1-ERG fusion molecules can be combined with other diagnostic tests to reduce false positive or false negative results.

[00049] Diagnosis of cancer can be based on detection of the presence of a fusion molecule, either a genomic fusion molecule, a fusion transcript, a fusion protein, or a combination thereof. In some embodiments, detection of the presence of a fusion molecule in a sample, e.g., observation of expected fluorescent signals in a break apart or fusion FISH assay, or observation of a signal in a nucleic acid hybridization or amplification-based assay or an immunoassay, is indicative of the presence of cancer. In other embodiments, the amount of a fusion molecule detected in a sample is quantified and compared to a control, and diagnosis is made based on an elevated level of the fusion molecule in the sample relative to the control. In still some other embodiments, the detection involves the use of reagents (e.g., primers,

probes or antibodies specific for the junction of a fusion molecule) that permits a determination of the composition or identity of the fusion molecule.

[00050] By "control", it refers to the amount of fusion observed in a normal sample, such as sample from benign prostate tissue or normal non-prostate tissue, or urine or blood sample from a normal individual who does not have cancer.

[00051] By "elevated level" it is meant that the level is significantly increased as compared to control. A significant increase is meant an increase by at least 50%, 75%, 100% (twice the normal level), 2 fold, 3 fold, 4 fold, 5 fold, 6 fold, 7 fold, 8 fold, 9 fold, 10 fold, 11 fold, 12 fold, 13 fold, 14 fold, 15 fold, or greater.

[00052] Detection of fusion molecules can employ any suitable sample sources which include any biological specimen that contains fusion molecules for detection as described herein. Examples include tissue (such as prostate tissue), urine, blood, semen, prostatic secretions or prostate cells. In a specific embodiment, a urine sample is collected immediately following a digital rectal examination (DRE), which often causes prostate cells from the prostate gland to shed into the urinary tract. Samples obtained from the above-identified sources can be further processed in order to enrich for the fusion molecules or cells containing the fusion molecules. The processing may include obtaining the serum or plasma portion of blood, obtaining the supernatant or cell pellet portion of urine, homogenization of tissue, lysis of cells, among others, in order to provide materials suitable for assaying the fusion molecules.

[00053] Detection of fusion molecules in a sample can be achieved by using a variety of techniques documented in the art. Fusion nucleic acid molecules can be detected by using various nucleic acid-based techniques, including hybridization (such as solution-phase hybridization, in situ hybridization (ISH), e.g., fluorescent ISH (FISH); microarray, Northern blot and Southern blot), amplification (such as polymerase chain reaction (PCR), reverse transcription polymerase chain reaction (RT-PCR), transcription-mediated amplification (TMA), ligase chain reaction (LCR), strand displacement amplification (SDA), and nucleic acid sequence based amplification (NASBA)), and sequencing. Fusion proteins can be

detected based on a variety of assays known for detection of proteins, including, for example, SDS-gel analysis, immunoassays (such as immunoprecipitation, Western blot, ELISA, immunohistochemistry, immunocytochemistry, and flow cytometry).

[00054] In addition to detection based on samples obtained from a subject, fusion molecules in a subject can also be detected by employing *in vivo* imaging techniques including, e.g., radionuclide imaging; positron emission tomography (PET); computerized axial tomography, X-ray or magnetic resonance imaging method, fluorescence detection, and chemiluminescent detection.

[00055] In one embodiment, detection of NDRG1-ERG fusion is achieved by performing an *in situ* hybridization (ISH) assay. Generally speaking, an ISH assay uses a labeled DNA or RNA strand as a probe that binds to a specific DNA or RNA sequence in a portion or section of tissue (*in situ*), or the entire tissue (whole mount FISH). The probe can be labeled with an isotope, a fluorescent compound, an antigen or any other appropriate label. Sample cells and tissues are usually treated to fix the target nucleic acids in place and to allow for access of the probe. After exposing the sample cells or tissues to the probe under appropriate hybridization conditions, the excess probe is washed away, and the probe bound to the target molecule is located using autoradiography, fluorescence microscopy or immunohistochemistry, depending upon the nature of the label.

[00056] In a specific embodiment, detection of NDRG1-ERG fusion is achieved by performing a fluorescent *in situ* hybridization (FISH) assay using a fluorescent-labeled probe.

[00057] In specific embodiments, a break-apart FISH assay is performed for detection of translocation of a gene of interest. Such break-apart assay uses a pair of probes, as illustrated in **Figures 1C** and **2C**. One of the probes specifically binds to a chromosomal region on the centromeric side of the gene of interest and is labeled to generate a first fluorescent color, and the other probe binds to a chromosomal region on the telomeric side of the gene of interest and is labeled to generate a second fluorescent color different from the first color. In preferred embodiments, the probes do not overlap with sequences of the gene of interest. With normal chromosomes without rearrangement of the gene of interest, juxtaposition or superimposition

of the two colors is observed. On the other hand, the two colors will split and appear on separate derivative chromosomes in cases of a reciprocal translocation involving the gene of interest; or alternatively, a single color generated by the centromeric probe will appear in cases of a translocation with a deletion of the telomeric region.

[00058] In a specific embodiment, a break-apart FISH assay is performed for detection of translocation of the NDRG1 gene using a centromeric probe and a telomeric probe flanking the NDRG1 gene. Observation of a split of the fluorescent colors generated from the two probes is indicative of translocation of the NDRG1 gene, and hence the presence of cancer.

[00059] In other specific embodiments, a fusion FISH assay is performed for detection of a gene fusion. Such fusion FISH assay also uses a pair of probes, as illustrated in **Figure 1C**. One of the probes specifically binds to a chromosomal region upstream of the 5' partner of the gene fusion and is labeled to generate a first fluorescent color, and the other probe binds to a chromosomal region downstream of the 3' partner of the gene fusion and is labeled to generate a second, different fluorescent color. With normal chromosomes without the gene fusion, the two colors will appear on separate derivative chromosomes; whereas juxtaposition or superimposition of the two colors will be observed if the gene fusion has occurred.

[00060] In a specific embodiment, a fusion FISH assay is performed for detection of the NDRG1-ERG fusion using a pair of probes. One of the probes specifically binds to a chromosomal region upstream of the NDRG1 gene and is labeled to generate a first fluorescent color, and the other probe binds to a chromosomal region downstream of the ERG gene and is labeled to generate a second, different fluorescent color. Observation of juxtaposition or superimposition of the two colors is indicative of the fusion and hence the presence of cancer.

[00061] In certain embodiments, FISH assays are performed using fluorescence-labeled bacterial artificial chromosomes (BACs) as probes. BAC clones containing specific BACs are available from distributors that can be located through many sources, e.g., National Center for Biotechnology Information (NCBI). Each BAC clone from the human genome has been given a reference name that unambiguously identifies such clone. These names can be used to find a corresponding GenBank sequence and to order copies of the clone from a distributor.

Non-limiting examples of BAC clones suitable for use in the diagnostic methods of the invention are listed in **Table 3**.

[00062] In another embodiment, detection of NDRG1-ERG fusion is achieved by using a nucleic acid-amplification based technique. Both genomic DNA and mRNA can be obtained from a suitable sample and used as template in an amplification reaction, permitting detection of genomic fusion molecules and fusion transcripts.

[00063] For example, NDRG1-ERG genomic fusion molecules can be detected by PCR using primers including a first primer which is specific for a 5' region of the NDRG1 gene (for example, the 5' regulatory region, the genomic region encoding the 5' untranslated region of NDRG1 mRNA, exons in the 5' region such as exons 1, 2 or 3), and a second primer specific for a 3' region of the ERG gene (for example, exons in the 3' region such as exon 4 or any other downstream exon, the genomic region encoding the 3' untranslated region of ERG mRNA).

[00064] NDRG1-ERG fusion transcripts can be detected by RT-PCR using primers including a first primer which is specific for a 5' region of a NDRG1 mRNA (such as the 5' untranslated region, exons 1, 2 or 3), and a second primer specific for a 3' region of an ERG mRNA (such as exon 4 or any other downstream exon, and the 3' untranslated region).

[00065] When referring to an oligonucleotide primer or probe as "specific for" a region, it is meant that such primer or probe has sufficient identity with a sequence within the region or its complementary strand, such that the primer or probe specifically hybridizes to the sequence or its complementary strand under stringent conditions. Stringency is dictated by temperature, ionic strength, and the presence of other compounds such as organic solvents. For example, "high stringency conditions" can encompass hybridization at 42°C in a solution consisting of 5x SSPE (43.8 g/l NaCl, 6.9 g/l NaH₂PO₄H₂O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.5% SDS, 5x Denhardt's reagent and 100µg/ml denatured salmon sperm DNA, followed by washing in a solution comprising 0.1 x SSPE, 1.0% SDS at 42°C or higher (e.g., 55°C, 60°C or 65°C). "Medium stringency conditions" can encompass hybridization at 42°C in a solution consisting of 5x SSPE with NaOH), 0.5% SDS, 5x Denhardt's reagent and

100µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 1.0 x SSPE, 1.0% SDS at 42°C.

[00066] In some embodiments, NDRG1-ERG fusion transcripts are detected based on hybridization or amplification types of assays (such as RT-PCR, FISH, among others) that utilize a primer or probe specific for the junction of an identified fusion transcript variant, alone or in combination with primer or probe not specific to the junction. Junction-specific oligonucleotides are specific for the junction of a fusion nucleic acid, and permits differentiation of a fusion nucleic acid versus native nucleic acids (e.g., native NDRG1 or ERG gene or mRNA).

[00067] A junction-specific primer or probe can be designed based on the sequence surrounding the point of fusion between the NDRG1 portion and the ERG portion in a fusion variant. Generally speaking, a junction-specific oligonucleotide primer or probe should be at least about 14 or 15 nucleotides in length, or 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or more nucleotides in length. A junction-specific oligonucleotide primer or probe is designed to have sufficient identity to a junction such that it hybridizes specifically to the junction under stringent conditions, but not to native nucleic acids without fusion. In specific embodiments, a junction specific primer or probe includes at least 3, 5, 6, 7, 8, 9, 10, 11, or 12 nucleotides from either side of the point of junction. If a fusion junction contains one or more nucleotides that are common to the two joining nucleic acids, a junction-specific primer should include the shared or common nucleotide or nucleotides, and additionally, at least 3, 5, 6, 7, 8, 9, 10, 11, or 12 nucleotides from either side of the shared nucleotide(s). In other embodiments, especially for amplification-based detection, a junction-specific primer is designed to target more of the 5' partner of the fusion than the 3' partner to minimize hybridization of the primer to native, non-fusion NDRG1 or ERG mRNA. In other words, the primer has a bigger 5' portion that hybridizes to one side of the junction sequence than the 3' portion of the primer which hybridizes to the other side of the junction. For example, a junction specific primer of 18 nucleotides in length can include a 5' portion of 12-14 nucleotides that corresponds to one side of the junction sequence, and a 3' portion of 4-6 nucleotides that corresponds to the other side of the junction sequence.

[00068] In a further embodiment, NDRG1-ERG fusion is detected based on analysis of chimeric NDRG1-ERG proteins. For such detection, peptides can be synthesized based on the junction amino acid sequences of identified chimeric variants and used to generate antibodies which specifically recognize the chimeric fusion protein, and not the native protein without fusion. Generally, a junction-specific peptide is at least 6 or 7 amino acids, preferably 8 or 9 amino acids, in length to be immunogenic. In some embodiments, a junction-specific peptide contains 7, 8, 9, 10, 12, 13, 14, 15, 16, or more amino acids of the junction of a fusion variant, and depending on the length of the peptide, may include at least 1, 2, 3, 4, 5, 6, 7, 8 or more amino acids from each side of the junction. In a specific embodiment, a junction-specific peptide includes at least 2 or 3 amino acids from each fusion partner. In other embodiments, full length chimeric NDRG1-ERG proteins or fragments thereof can be used as immunogens to generate antibodies, which are screened to identify those antibodies that only bind chimeric fusion proteins but not native NDRG1 or ERG protein.

[00069] In accordance with the present invention, detection of NDRG1-ERG fusion molecules can be combined with other tests in order to achieve more accurate diagnostic results. Other diagnostic tests include, for example, detection of other fusions associated with cancer, including gene fusions between TMPRSS2 and ERG, between SLC45A3 and ERG gene, as described in e.g., U.S. Published Application 2007/0212702. In the experiments described in the following examples, NDRG1-ERG fusion has been observed in prostate cancers overexpressing ERG yet negative for TMPRSS2 or SLC45A3 rearrangement. Accordingly, detection of NDRG1-ERG fusion molecules may provide a useful complement to other diagnostic tests based on fusion detection. For example, a multiplex panel can be utilized which detects TMPRSS2-ERG, SCL45A3-ERG and NDRG1-ERG fusions.

[00070] Drug Screening

[00071] In a further embodiment, the invention provides a method of screening for inhibitors of NDRG1-ERG fusion. Specifically, candidate compounds can be screened for their ability to reduce the level of expression or to inhibit a biological function of an NDRG1-ERG fusion molecule. The method can be performed *in vitro* using a cell line having elevated levels of a

NDRG1-ERG fusion molecule, e.g., a cell line transfected to express an NDRG1-ERG fusion molecule. Candidate compounds can include nucleic acid molecules, small organic molecules, and antibodies, for example. The identified compound may reduce either the mRNA or the protein level of an NDRG1-ERG fusion molecule, or inhibiting a biological function of such fusion. Biological functions of NDRG1-ERG fusion proteins include, e.g., enhancing cell migration or cell invasion, which are properties frequently observed with cancerous cells. Cell invasion and cell migration can be assessed by using known assays and techniques, such as the Boyden chamber assay well documented in the art.

[00072] Cancer Treatment

[00073] The present invention also provides methods for treating cancers associated with NDRG1-ERG fusion, including but not limited to prostate, breast, colon, pancreas, and lung cancers. By "treating" it is meant eliminating or at least inhibiting or reducing the growth or metastasis of cancerous cells. Treatment may also reduce or prevent the occurrence of cancer (e.g., in subjects predisposed to developing cancer associated with NDRG1-ERG fusion), or reduce or prevent reoccurrence of cancer associated with NDRG1-ERG fusion.

[00074] The treatment involves administration to a subject an agent that inhibits a biological function of a NDRG1-ERG fusion molecule, or reduces the level of the fusion molecule. The agent can be any one of a small molecule compound, an siRNA, an antisense nucleic acid, or an antibody, or a combination thereof.

[00075] In one embodiment, the treatment employs an inhibitor of an NDRG1-ERG fusion protein, e.g., a compound that inhibits a biological function (e.g., the function of conferring enhanced cell invasion potential) of an NDRG1-ERG fusion protein.

[00076] In another embodiment, the treatment employs an siRNA molecule. The term "siRNAs" refers to small interfering RNAs, which may include a double-stranded region of about 18-30, or 20-25 nucleotides. One strand of the double-stranded region is identical or substantially homologous to a target RNA molecule. The double-stranded region can be formed by two separate RNA strands, or a singled RNA molecule (e.g., a hairpin shape). In

some embodiments, siRNAs are designed to target the junction region of a NDRG1-ERG fusion transcript.

[00077] *Compositions and Kits*

[00078] Isolated or recombinant NDRG1-ERG fusion nucleic acid molecules, including genomic DNA, mRNA and cDNA fusion molecules, are provided by the present invention. In one embodiment, the nucleic acid molecule encodes a chimeric NDRG1-ERG fusion protein. In a specific embodiment, the nucleic acid molecule encodes a chimeric NDRG1-ERG fusion protein having the amino acid sequence set forth in SEQ ID NO: 7 or SEQ ID NO: 9. Examples of such nucleic acid molecules include those having the nucleotide sequence as set forth in SEQ ID NO: 6 or SEQ ID NO: 8.

[00079] Isolated or recombinant NDRG1-ERG fusion proteins are also provided by the present invention. Examples of NDRG1-ERG fusion proteins include those having the amino acid sequence as set forth in SEQ ID NO: 7 or SEQ ID NO: 9.

[00080] Modified fusion nucleic acid or protein molecules, where one or more nucleotides or amino acids have been substituted, added or deleted, are also contemplated by the invention, so long as the modified molecules are substantially identical to the fusion molecules prior to the modification. A substantial identity is measured by a substantial sequence identity (i.e., at least 75%, 80%, 85%, 90%, 95%, 98%, 99% or greater), or by substantial functional identity (i.e., the modified molecule retains at least 70%, 75%, 80%, 85%, 90%, 95%, or greater of a biological function of a fusion molecule prior to modification). Fusion molecules can also be modified to include additional features, such as labels or compounds capable of generating a detectable signal, additional sequences corresponding to an epitope tag or a restriction endonuclease site, among others.

[00081] The invention also provides expression vectors for expressing a chimeric NDRG1-ERG fusion protein in a host cell. Such expression vectors contain a nucleic acid which encodes a chimeric NDRG1-ERG fusion protein, and the coding sequence of the chimeric protein is operably linked to a promoter at 5', and to a termination sequence at 3'. Any promoter which can direct the expression of a chimeric NDRG1-ERG protein in a desirable

host cell can be used, and can be a constitutive or inducible promoter, including e.g., the native human NDRG1 promoter. Numerous promoters suitable for directing expression in bacterial, fungal or mammalian cells have been documented in the art.

[00082] Host cells transformed with any such expression vector are also provided by the invention. Suitable host cells include any bacterial, fungal, and mammalian cells suitable for propagation of the expression vector or recombinant expression of fusion molecules.

[00083] In additional embodiments, the present invention provides oligonucleotide primers and probes, peptides, and antibodies, useful for practicing the diagnostic methods described herein. One or more such components or reagents can be provided in a diagnostic kit.

[00084] Oligonucleotide primers or probes suitable for use in the detection, whether specific for a fusion junction or otherwise, can include additional features in addition to the sequence binding region, such as a sequence that does not bind to the junction sequence (e.g., a tag sequence or a promoter sequence) and does not interfere with binding to the intended target sequence in the junction. The primers or probes can also include non-nucleic acid moieties such as labels that do not interfere with target binding.

[00085] Similarly, peptides or antibodies, whether specific for a fusion junction or otherwise, can include additional features, such as additional amino acids that are not part of a fusion protein, labels, among others.

[00086] In the following examples, reference is made to the accompanying drawings that form a part hereof, and in which is shown by way of illustration specific embodiments which may be practiced. These embodiments are described in detail to enable those skilled in the art to practice the invention, and it is to be understood that other embodiments may be utilized and that changes may be made without departing from the scope of the present invention. The following description of exemplified embodiments is, therefore, not to be taken in a limited sense.

EXAMPLES

[00087] **Example 1.** Analysis of ERG Overexpression in Prostate Cancer Prostatectomy Samples

[00088] Prostate cancer prostatectomy samples from 101 men were screened for *ERG* gene rearrangement using a FISH break-apart (b/a) assay. In total, 44 cases were positive for *ERG* rearrangement. Given the heterogeneity of TMPRSS2-*ERG* mRNA expression level (as reported by Wang et al., *Cancer Res*, 66, 8347-51, 2006) in prostate cancer, TMPRSS2-*ERG* mRNA variant expression was screened using conventional RT-PCR and DNA sequencing. Of the 44, 34 (77%) expressed 7 different variants of TMRPSS2-*ERG* mRNA described by Wang et al. (2006), *supra*. In order to determine the level of ERG mRNA over-expression, quantitative PCR was performed using cDNA from 29 cases (19 that were TMRPSS2-*ERG* mRNA positive and 10 TMPRSS2-*ERG* mRNA negative), 15 cases that did not show *ERG* rearrangement and 6 benign prostate tissue samples (**Figure 1A**). ERG mRNA was over-expressed up to 75 times (median of 27) in *ERG* rearranged cases compared to baseline levels in benign prostate tissue and cases negative for both *ERG* rearrangement and TMPRSS2-*ERG* mRNA. Contrary to findings by Wang et al. (2006), *supra*, TMPRSS2-*ERG* mRNA isoform expression was not associated with ERG over-expression or with prostate cancer progression (Gleason score, pathologic stage, or surgical margin status, as shown in **Table 1**).

[00089] TMPRSS2-*ERG* mRNA was absent in 10 (23% of 44) *ERG* rearranged cases, of which 7 expressed high ERG mRNA levels (5-38 times). To confirm the absence of *TMPRSS2* rearrangement in these cases, a *TMPRSS2* b/a FISH assay was performed. *TMPRSS2* rearrangement was observed in 2/10 cases (60T and 51T), indicating a novel *TMPRSS2-ERG* fusion that was not detected using standard RT-PCR approaches. To screen for other possible fusion events with *ERG*, RT-PCR analysis was performed targeting known ETS family fusion partners (*SLC45A3*, *HERV-K*, *C15ORF21*, *HNRPA2B1*, *DDX5*, *CANT1*, *KLK2* and *ACSL3*). This screening revealed that exon 4 of *ERG* was fused to exon 1 of *SLC45A3* in 3 ERG mRNA over-expressed cases (34T, 150B_M, 145C_M, **Figure 1B**). The predicted open reading frame is identical to what is encoded by the most common TMPRSS2

(exon 1)-ERG (exon 4) mRNA transcript. This was confirmed *in-situ* using *SLC45A3* and *ERG* b/a-assays and an *SLC45A3-ERG* fusion assay (**Figure 1C**).

[00090] Patient information, the materials and methods used in the above experiments are as follows.

[00091] *Patient Population* - The study is composed of 101 men with localized and locally advanced prostate cancer who underwent radical prostatectomy as a monotherapy. All prostate cancer cases were collected as part of institutional review board–approved research protocols.

[00092] *Sample processing for RNA Analyses* – Hematoxylin and eosin slides were prepared from formalin-fixed paraffin-embedded material and evaluated for cancer extent and tumor grade (Gleason score). Hematoxylin and eosin slides were prepared from the corresponding frozen tissue block and evaluated for the extent of cancer involvement. To ensure for a high concentration of cancer cells and minimized benign tissue, tumor isolation was performed by first selecting for high-density cancer foci (<10% stromal and other nontumor tissue contamination) and then taking 1.5-mm tissue cores from the frozen tissue block for RNA extraction. Sections for fluorescence in situ hybridization (FISH) evaluation were taken from the frozen tissue block used for molecular analysis. The cancer foci selected for RNA extraction were well characterized by FISH to evaluate the ERG rearrangement status throughout the entire focus. Special care was taken to extract the RNA from a single cancer focus to exclude the problem of heterogeneity when looking for putative fusion transcripts. RNA was isolated from frozen tissue using TRIzol LS reagent (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. After DNase treatment (Invitrogen), RNA concentration was measured using a NanoDrop 8000 spectrophotometer (Thermo Scientific, Wilmington, DE). Quality was assessed using the Bioanalyzer 2100 (Agilent Technologies, Inc, Santa Clara, CA). The qualitative detection of fusion transcripts in the cases was performed using conventional reverse transcription–polymerase chain reaction (RT-PCR), agarose gel fractionation/purification, and subsequent complementary DNA (cDNA) sequencing. For this amplified DNA, fragments corresponding to the expected sizes of fusion transcripts were gel-extracted using the MinElute Gel Extraction Kit (Qiagen) and sequenced

at the Q4 Life Sciences Core Laboratories Center's DNA sequencing facility of Cornell University (Ithaca, NY). Quantitative ERG and TMPRSS2-ERG RT-PCR was performed using QuantiTect SYBR Green PCR Kit (Qiagen). Each sample was run in duplicate. The amount of each target gene relative to a control gene was determined using the comparative Ct method (ABI Bulletin 2; Applied Biosystems). Ct values Q5 for ERG were first normalized using the average Ct values obtained for SART3 and TCFL1/VPS72 and then calibrated using normalized Ct values obtained from benign prostate. The protocols and primers for all RT-PCR assays used are shown in **Table 2**.

[00093] *Assessment of ERG, TMPRSS2, SLC45A3 and NDRG1 Rearrangements Using Two-Color FISH Assays* – To assess for rearrangement of *ERG*, *TMPRSS2*, *SLC45A3* and *NDRG1*, break-apart (b/a) FISH assays (essentially as described by Perner et al., *Am J Surg Pathol*, 31, 882-8, 2007) were performed on sections from the corresponding frozen tissue blocks. For designing a FISH break-apart assay, the inventors tested 5-10 Bacterial Artificial Chromosomes (BAC) probes flanking a gene of interest (GOI) on the centromeric side and 5-10 BAC probes flanking the telomeric side of the GOI, ideally not overlapping with sequences of the GOI. The BAC probes were hybridized on metaphase spreads of fixed cells to evaluate their target sequence specificity and selectivity (correct chromosomal location and no cross-hybridization to other chromosomes), fluorescence signal intensity and compatibility with the hybridization protocol. The probes which best matched all these requirements were selected for the assay. The centromeric probes for *ERG*, *TMPRSS2*, *SLC45A3* and *NDRG1* were RP11-24A11, RP11-354C5, RP11-249H15 and RP11-185E14, respectively. The telomeric probes for *ERG*, *TMPRSS2*, *SLC45A3* and *NDRG1* were RP11-372O17, RP11-891L10, RP11-131E5 and RP11-1145H17, respectively. Probes RP11-131E5 (*SLC45A3*) and RP11-24A11 (*ERG*) were used for the *SLC45A3-ERG* fusion assay. Additional information regarding these probes is provided in **Table 3**. Correct chromosomal probe localization was confirmed on normal lymphocyte metaphase preparations, as exemplified in **Figure 3** which displays metaphase results for BACs targeting the *NDRG1* locus. For each sample a minimum of 100 nuclei were analyzed. The b/a assays used for *ERG*, *TMPRSS2* and *SLC45A3* are schematically represented in **Figure 1C** and that for *NDRG1* is shown in **Figures 2C-2D**.

Table 1. ERG rearrangements, Over-Expression, TMRPSS2-ERG mRNA Variants and Clinical Information of Prostatectomy Samples Analyzed

sample ID	ERG rearrangement	TMRPSS2-ERG isoform	ERG expression	Gleason score	Stage	Surgical margin	Seminal vesicle invasion
1701_A	+	III	74.04	3+4	3a	-	-
580_B	+	III, VI	66.97	3+4	2c	-	-
1700_D	+	III	46.18	4+4	3a	+	-
28_T	+	III	40.40	3+4	2c	-	-
45_T	+	III, VI	36.20	2+3	2c	+	-
69_T	+	III, VI	32.98	4+3	3a	-	-
13_D	+	III	32.35	3+3	2c	-	-
560_D	+	III, VI	30.96	4+3	3a	-	-
140_T	+	II	28.91	3+4	3a/b	-	+
435_D	+	III	28.75	3+4	3a	-	-
581_D	+	III	27.56	4+5	3b	-	+
1780_D	+	III, VI	27.10	3+4	2c	-	-
431_D	+	III, VI	24.92	3+3	2	-	-
54_T	+	III	24.69	3+4	2c	+	-
88_T	+	III	24.52	3+5	3b	-	+
522_D	+	IV	22.06	3+4	2c	-	-
415_B	+	III	13.46	4+5	3a	-	-
67_T	+	II, III, VI, VIII	11.08	3+4	2c	-	-
106_T	+	I, V	10.06	3+5	3b	-	+
60_T	+	-	38.40	3+4	2c	+	-
145_C_M	+	-	37.54	4+5	3a	-	-
34_T	+	-	37.37	3+4	3b	-	+
150_B_M	+	-	20.06	2+3	2c	-	-
51_T	+	-	15.01	3+4	3a	-	-
509_B	+	-	10.12	3+4	2a	-	-
99_T	+	-	5.07	3+3	3a	+	-
1061_C	+	-	0.75	4+3	2c	-	-
424_B	+	-	0.53	3+4	2c	-	-
1043_B	+	-	0.50	3+3	2c	-	-
1024_D	-	-	5.74	4+5	3b	-	+
97_T	-	-	2.93	2+3	2c	-	-
127_T	-	-	2.72	3+3	2a	-	-
1783_B	-	-	2.26	4+4	2c	-	-
1023_C	-	-	1.97	3+3	2c	-	-
113_T	-	-	1.77	4+5	3b	+	+
136_T	-	-	1.36	3+3	3a	+	-
134_B	-	-	1.04	3+2	2b	+	-
1060_B	-	-	0.79	4+3	2c	-	-
151_T	-	-	0.77	2+3	2c	-	-
540_C	-	-	0.74	3+4	3a	-	-
1781_C	-	-	0.71	3+4	2c	+	-
63_T	-	-	0.45	2+4	2c	-	-
1702_C	-	-	0.43	3+4	3a	-	-
1765_A	-	-	0.41	3+4	2a	+	-
1027_D	na	-	1.07	na	na	na	na
1033_B	na	-	1.04	na	na	na	na
1023_B	na	-	1.04	na	na	na	na
1024_C	na	-	0.90	na	na	na	na
1028_A	na	-	0.76	na	na	na	na
1032_D	na	-	0.73	na	na	na	na

Table 3. DNA probes based on hg19 (unless otherwise indicated)

Gene	centromeric probe (labeled with red fluorescence)		telomeric probe (labeled with green fluorescence)	
ERG	RP11-24A11	Chromosome: chr21 Start: 39546498 End: 39733869 Length: 187372 Strand: + Score: 1000 Bands: 21q22.13-21q22.2	RP11-372O17	Chromosome: chr21 Start: 40367344 End: 40557436 Length: 190093 Strand: + Score: 1000 Band: 21q22.2
TMPRSS2	RP11-354C5	Chromosome: chr21 Start: 42439601 End: 42635437 Length: 195837 Strand: + Score: 1000 Bands: 21q22.2-21q22.3	RP11-891L10	Chromosome: chr21 Start: 43409124 End: 43594929 Length: 185806 Strand: + Score: 1000 Band: 21q22.3
SLC45A3	RP11-249H15 based on hg18	Chromosome: chr1 Start: 203724624 End: 203787895 Length: 63271 Band: 1q32.1	RP11-131E5 based on hg18	Chromosome: chr1 Start: 203910487 End: 204074037 Length: 163550 Band: 1q32.1
NDRG1	RP11-185E14	Chromosome: Chr8 Start: 134024919 End: 134198328 Length: 173410 Strand: - Score: 1000 Band: 8q24.22	RP11-1145H17	Chromosome: Chr8 Start: 134333724 End: 134466739 Length: 133016 Strand: + Score: 1000 Band: 8q24.22

Table 2. Oligonucleotide primers and cycling conditions for RT-PCR assays.

Assay	Gene	Accession number	Exon(s)	Primer	Sequences 5' → 3'	Cycling
TMPRSS2-ERG exon 4	TMPRSS2	NM_006656.2	4-17	TMPRSS2-ERG_f	TAGGGGCGAGCTTAAGCAGCAG	94°C 2min; 94°C 30s; 55°C 30s; 72°C 1min; 10 cycles; 72°C 10min; 35 cycles
	ERG	NM_004449.3	275-282	ERG_r	GTAGGCGACAGCAAGAAAGAGACTGG	
TMPRSS2-ERG exon 6	TMPRSS2	NM_005656.2	4-17	TMPRSS2-ERG_f	TAGGCGGAGCTTAAGCAGGAG	94°C 2min; 94°C 30s; 63°C 30s; 68°C 2min; 68°C 10min; 40 cycles
	ERG	NM_004449.3	659-636	ERG_Exon 5-6_r	CCATATCTTTCACCCGCCACTCC	
TMPRSS2-ERG exon 9	TMPRSS2	NM_005656.2	4-17	TMPRSS2-ERG_f	TAGGCGGAGCTTAAGCAGGAG	94°C 2min; 94°C 30s; 55°C 30s; 68°C 2min; 68°C 10min; 40 cycles
	ERG	NM_004449.3	945-928	ERG_Exon10_r	CGATTTGCGGCTGAGGAG	
TMPRSS2-ERG exon 12	TMPRSS2	NM_005656.2	4-17	TMPRSS2-ERG_f	TAGGCGGAGCTTAAGCAGGAG	94°C 2min; 94°C 30s; 55°C 30s; 68°C 2min; 68°C 10min; 40 cycles
	ERG	NM_004449.3	1575-1549	ERG_Exon13_r	TTAGTAGTAGTGCCCGAGATGAGAAG	
ERG qPCR	ERG	NM_004449.3	574-537	ERG_Exon 5-6_f	GGCAGAGATTATGGTCCAGGAGAG	50°C 2min; 95°C 15min; 94°C 1min; 50°C 1min; 72°C 1min; 41 cycles
	ERG	NM_004449.3	659-636	ERG_Exon 5-6_r	CCATATCTTTCACCCGCCACTCC	
SART3	SART3	NM_014706.3	535-558	SART3_f	GCCGCGCAGCAGATGATCAAAATC	melting curve; 70-90°C every 0.2°C for 1s
	SART3	NM_014706.3	899-866	SART3_r	ACCAAGACGAGCAGAGCAGCTTC	
TCFL1P572	TCFL1P572	NM_005097.1	778-801	TCFL1P572_f	ATTGACTGCTCATGCTGGGACTGG	
	TCFL1P572	NM_005097.1	948-927	TCFL1P572_r	CGGTATAGGGCTGGAGGAGTGGG	
TMPRSS2-ERG qPCR	TMPRSS2	NM_005656.2	4-17	TMPRSS2-ERG_f	TAGGCGGAGCTTAAGCAGGAG	50°C 2min; 95°C 15min; 94°C 1min; 50°C 1min; 72°C 1min; 41 cycles
	ERG	NM_004449.3	276-252	ERG_r1	GTAGGCACACTAAACAAGCACTGG	melting curve; 70-90°C every 0.2°C for 1s
HMBS	HMBS	NM_000190.3	711-730	HMBS_f	CCATCATCTCTGGCAACAGGT	
	HMBS	NM_000190.3	790-772	HMBS_r	GCATTCCTCAGGGTCCAGG	
SLC45A3-ERG	SLC45A3	NM_033102.2	74-51	SLC45A3_f	GATGGGCTTGGGGTGGAG	94°C 2min; 94°C 30s; 55°C 30s; 68°C 2min; 68°C 10min; 40 cycles
	ERG	NM_004449.3	659-636	ERG_Exon 5-6_f	CCATATCTTTCACCCGCCACTCC	
HERV-K-ERG	HERV-K_22q11.23	BC020811.1	305-327	HERV-K_f	GTCCCAAGTAC3TCCACCGGTCAG	94°C 2min; 94°C 30s; 55°C 30s; 68°C 2min; 68°C 10min; 40 cycles
	ERG	NM_004449.3	659-636	ERG_Exon 5-6_r	CCATATCTTTCACCCGCCACTCC	
HNRPA2B1-ERG	HNRPA2B1	NM_007371.2	136-155	HNRPA2B1_f	TCCGCGAAATGGGCTCAAG	94°C 2min; 94°C 30s; 55°C 30s; 68°C 2min; 68°C 10min; 40 cycles
	ERG	NM_004449.3	659-636	ERG_Exon 5-6_r	CCATATCTTTCACCCGCCACTCC	
C15ORF21-ERG	C15ORF21	NM_001005266	313-336	C15ORF21_f	CAATAACACTGGGGTTCCTGAG	94°C 2min; 94°C 30s; 55°C 30s; 68°C 2min; 68°C 10min; 40 cycles
	ERG	NM_004449.3	659-636	ERG_Exon 5-6_r	CCATATCTTTCACCCGCCACTCC	
CANT1-ERG	CANT1-E1E	NM_004449.3	n/a	CANT1-E1E_f	AGCCAAGCCCGCCGATC	94°C 2min; 94°C 30s; 55°C 30s; 68°C 2min; 68°C 10min; 40 cycles
	ERG	NM_004449.3	945-928	ERG_Exon 5-6_r	CCATATCTTTCACCCGCCACTCC	
CANT1-ERG	CANT1-E1	NM_138793.2	48-65	CANT1-E1_f	AGCCAAGCCCGCCGATC	94°C 2min; 94°C 30s; 55°C 30s; 68°C 2min; 68°C 10min; 40 cycles
	ERG	NM_004449.3	659-636	ERG_Exon 5-6_r	CCATATCTTTCACCCGCCACTCC	
K1K2-ERG	K1K2	NM_005531.3	65-77	K1K2_f	TGTTGCAATGGGCTGCTGTG	94°C 2min; 94°C 30s; 55°C 30s; 65°C 2min; 68°C 10min; 40 cycles
	ERG	NM_004449.3	659-636	ERG_Exon 5-6_r	CCATATCTTTCACCCGCCACTCC	
DDX5-ERG	DDX5	NM_004396.2	423-443	DDX5_f	AGAGTCAACAAGTCCCGCAAG	94°C 2min; 94°C 30s; 55°C 30s; 68°C 2min; 68°C 10min; 40 cycles
	ERG	NM_004449.3	659-636	ERG_Exon 5-6_r	CCATATCTTTCACCCGCCACTCC	
KDRG1-ERG	KDRG1	NM_002098.7	30-57	KDRG1_f	GTCAGCTGCTGAGTTCACATCC	94°C 2min; 94°C 30s; 61°C 30s; 65°C 2min; 68°C 10min; 40 cycles
	ERG	NM_004449.3	945-928	ERG_Exon 10_r	CGACTGGGCGTGGGGTG	
ERG 5'RACE	ERG	miRNA	601-578	ERG_GSP1	ATCCTAGCACCGGTGGCTATTGAG	Reverse Transcription reaction at 55°C
	ERG	NM_004449.3	571-552	ERG_GSP2	CGTTCGTGGTCATGTTGGG	94°C 2min; 94°C 30s; 60°C 1min; 72°C 1min; 72°C 7min; 35 cycles (45 cycles for GSP3-PCR)
ERG 5'RACE	ERG	NM_004449.3	470-448	ERG_GSP3	GCCCACTGCAATTCATCAGGAGA	

[00094] Example 2. Massively Parallel RNA-seq Discovers NDRG1-ERG Fusion

[00095] Having characterized all but two ERG over-expressing/*ERG* rearranged cases (509B, 99T), paired-end RNA-seq was used to identify potential 5' partners. Fusion transcripts were explored by looking for paired reads where each pair mapped to regions that were either greater than 2MB apart and less than 5MB apart, or mapped to different chromosomes (see **Table 4**). The utility of this approach was confirmed by limiting this analysis to matches with high numbers of reads. First, in prostate cancer cases known to harbor the *TMPRSS2-ERG* fusion (e.g., case 1701A), numerous *TMPRSS2-ERG* transcripts were detected. Second, *SLC45A3-ELK4* transcript could also be detected in case 1701A as observed in an independent study. Finally, in 1 case (99T) with ERG over-expression but no *SLC45A3* or *TMPRSS2* rearrangement as determined by RT-PCR and FISH, RNA-seq demonstrated 17 copies of a fusion transcript that mapped paired reads to *ERG* exons and to exons of *NDRG1* (**Figure 2A**). This was confirmed by conventional RT-PCR (**Figure 2B**). Sequence analysis of *NDRG1-ERG* cDNA (**Figure 4A**, SEQ ID NO: 6) indicates that this fusion, as with *BCR-ABL1* fusion gene in patients with chronic myeloid leukemia, encodes a chimeric protein containing 33 amino acids from *NDRG1* as well as the conserved protein domains of wild type *ERG* (Sterile alpha motif/Pointed domain and ETS domain) (**Figure 4B**, SEQ ID NO: 7). Screening other *TMPRSS2-ERG*, *SLC45A3-ERG* mRNA negative cases revealed another, slightly different, *NDRG1-ERG* transcript variant (variant 2) in 509B. *NDRG1-ERG* variant 2 mRNA (**Figure 5A**, SEQ ID NO: 8) is also predicted to encode a chimeric protein including the first 21 amino acids of *NDRG1* and the same conserved domains of *ERG* as in the protein encoded by *NDRG1-ERG* variant 1 (**Figure 5B**, SEQ ID NO: 9). Sequences for *NDRG1-ERG* variant 1 and variant 2 have been submitted to GenBank (acc. # FJ627786 and # FJ627787). The chromosomal translocation which resulted in *NDRG1-ERG* fusion was confirmed at the genome level using an *NDRG1* b/a and *NDRG1-ERG* fusion FISH assays (**Figures 2C-2D**).

[00096] RT-PCR analysis and b/a FISH assays were performed following the protocols described in Example 1. RNA-seq data analysis was performed as follows.

[00097] *RNA Sequencing Data Analysis* – The Illumina Genome Analyzer II was used for paired-end RNA sequencing. This provided a pair of approximately 30-36 base reads, from each end of a transcript fragment of relatively well-defined length (about 330 nucleotides). The paired reads were aligned independently to the human genome (hg18 assembly in the UCSC genome browser using "eland," a short-read alignment tool included in the Genome Analyzer software suite. For each read, eland provides the coordinate(s) of the alignment to the reference genome, allowing for up to two mis-matches in the sequence. Only the reads that are mapped uniquely to the genome were kept, although they might have up to two mis-matches. In order to search for novel translocations involving *ERG*, two strategies were applied. First, mapped paired reads were selected that were more than 2MB and less than 5MB apart. This allowed the identification of translocations similar to *TMPRSS2-ERG* in which the two genes are approximately 3MB apart. Second, paired reads mapping to different chromosomes were also selected as potential candidates. Because the focus was on novel *ERG* partners, paired reads were selected where one of the reads lay within *ERG*. This allowed us to identify several candidate fusion transcripts spanning all chromosomes. Finally the chromosome with the highest number of reads was selected and checked if those reads lay within a gene.

Table 4. RNA-Seq Data

Sample ID	# reads	# mappable reads	% mappable reads	Comments
1701_A	8,542,482	3,108,222	36.39%	(T2-ERG fusion positive)
1783_B	3,080,154	1,330,949	43.21%	(T2-ERG fusion negative)
99_T	2,844,879	1,180,781	41.51%	(NDRG1-ERG fusion positive)

[00098] Example 3. TMPRSS2-, SLC45A3-, and NDRG1-ERG Are Regulated by Androgen and Estrogen

[00099] ERG mRNA expression in cases positive for *SLC45A3-ERG* or *NDRG1-ERG* is similar in magnitude to those measured for *TMPRSS2-ERG* positive cases. *TMPRSS2* (Lin et al., *Cancer Res*, 59, 4180-4, 1999), *SLC45A3* (Xu et al., *Cancer Res*, 61, 1563-8, 2001), and *NDRG1* (Segawa et al., *Oncogene*, 21, 8749-58, 2002; Tu et al., *Mol Cell Proteomics*, 6, 575-88, 2007) are all known androgen induced genes. This was confirmed by treating LNCaP with a synthetic androgen (R1881, 1nM) (**Figures 6A-6B**). Androgen regulation of *NDRG1* is supported by the observation of an AR binding site ~30kb upstream of the start site (chr8:134407748-134408779) in LNCaP cells. The induction of gene expression was abrogated in the presence of Flutamide. If *KLK3 (PSA)* mRNA was considered a surrogate read-out of androgen signaling, it would be expected to find similar profiles between PSA and ERG mRNA levels in *TMPRSS2-ERG*, *SLC45A3-ERG* or *NDRG1-ERG* mRNA positive prostate cancer cases. PSA mRNA levels, however, did not mimic the pattern of ERG mRNA levels in *TMPRSS2-ERG*, *SLC45A3-ERG* or *NDRG1-ERG* mRNA positive cases, indicating an additional mechanism for the regulation of the fusion transcripts.

[000100] *TMPRSS2-ERG* has been shown to be regulated by estrogen (Setlur et al., *J Natl Cancer Inst*, 100, 815-25, 2008). Chromosome-immunoprecipitation data indicates the presence of an estrogen receptor (ER) binding site within the *SLC45A3* gene, an ER binding site in the first intron of *NDRG1* (chr8:134373799-134375086) and at ~60kb upstream of the start site (chr8:134441414-134442401). Similar data show that FoxA1, a known ER cofactor, binding sites overlap with the ER binding sites. To test this, the levels of *SLC45A3* or *NDRG1* mRNA in LNCaP cells were measured at different time points as a function of estrogen treatment. Induction of *SLC45A3* mRNA and *NDRG1* mRNA was observed at 3 hours (**Figure 6C**) and 12 hours (**Figure 6D**), respectively, following 17 β -estradiol treatment, but not with the ER α receptor agonist DPN, similar to *IGF1R* mRNA, a known estrogen-induced gene in LNCaP cells (Pandini et al., *Cancer Res*, 67, 8932-41, 2007) (**Figure 7**). This data indicate that, like *TMPRSS2-ERG*, *SLC45A3-ERG* and *NDRG1-ERG* fusion genes are also estrogen-regulated through ER α . This provides another mechanism for *ERG* over-

expression when fused to *SLC45A3* or *NDRG1*, particularly in the case of castrate-resistant prostate cancer.

[000101] Materials and methods used in the experiments of this Example are as follows.

[000102] *Hormone Treatment of LNCaP* – The prostate cancer cell line LNCaP was obtained from ATCC (Manassas, VA; cat.# CRL-1740) and maintained according to the suppliers instructions. For hormonal treatment, cells were plated (500,000 cells/10 cm²) in the presence of complete growth medium supplemented with 1% Penicillin/Streptomycin. Cells were starved for 48h in charcoal-stripped (CS) medium (RPMI-1640 1x, 5% charcoal-stripped FBS, 1% Penicillin/Streptomycin) and then treated with R1881 (1 nM), 17 β -estradiol (10 nM), diarylpropionitrile (DPN, 10 nM) or ethanol vehicle for 3, 12, and 24 hours. RNA was extracted using the TRIzol Reagent (Invitrogen, Carlsbad, CA), subjected to DNase treatment (DNA-freeTM Kit, Applied Biosystems) according to the manufacturer's instructions. To test for the specificity of androgen-stimulation, cells were treated with 10 μ M Flutamide for 2 hours and then treated with R1881 as described above. TAQMAN expression assays (Applied Biosystems) were used to quantify relative levels of SL45A3, NDRG1, PSA (KLK3) and IGF1R. See **Table 5**.

Table 5. TAQMAN expression assays

Gene	assay ID	reporter dye	location
SLC45A3	Hs00263832_m1	FAM	ex 5
TCFL1	Hs00195618_m1	FAM	ex 6
IGF1R	Hs99999020_m1	FAM	ex 2 and 3
KLK3	Hs02576345_m1	FAM	ex 1 and 2, detects all KLK3 transcript variants
NDRG1	Hs00608387_m1	FAM	ex 11 and 12

[000103] *Expression profiling of ERG and 3 androgen-regulated genes* – A subset of 65/101 samples were processed using Illumina HumanWG-6 v2.0 bead-arrays. A heatmap was constructed showing relative expression levels of *ERG*, *TMPRSS2* and *SLC45A3*. The gene expression levels in a given sample have been color coded where orange to blue indicates high to low levels of expression. The samples have been grouped according to *TMPRSS2-ERG (T2-ERG)* fusion status as determined by RT-PCR and then ordered by the level of *ERG* microarray feature level normalized intensity.

[000104] **Example 4.** Expression of NDRG1-ERG Chimeric Protein Enhances Cell Invasion

[000105] HEK-293 (embryonic kidney) and BPH1 (prostate epithelial) cell lines were transiently transfected with a vector carrying the NDRG1-ERG fusion variant 1 cDNA. Expression of the NDRG1-ERG chimeric protein was observed by immunostaining the cells with an anti-ERG antibody, as shown in **Figures 8-9**.

[000106] Overexpression of the NDRG1-ERG mRNA in HEK-293 and BPH1 cell lines is shown in **Figure 10**.

[000107] In a separate experiment, HEK-293 and BPH1 cell lines were transiently transfected with an NDRG1-ERGflag or NDRG1-ERG retroviral expression vector, or a vector expressing LacZ as control. The levels of mRNA in transfected HEK-293 cells were quantified (**Figure 11**, top). Immunostaining of the transfected cells expressing NDRG1-ERGflag demonstrates that the chimeric NDRG1-ERG protein was produced in the cells (**Figure 11**, middle). To assess the function of the chimeric protein, an invasion assay was performed using Boyden chambers coated with matrigel (BD Biosciences), with 10% fetal calf serum as chemoattractant, and HEK293 cells expressing LacZ control or NDRG-1-ERG fusion protein. As shown in **Figure 11**, bottom, expression of the NDRG-1-ERG fusion protein enhanced cell invasion.

WHAT IS CLAIMED IS:

1. A method for diagnosing prostate cancer in a patient, comprising providing a biological sample from said patient, detecting the presence or absence of an NDRG1-ERG fusion molecule, wherein the presence of said fusion molecule is indicative of the presence of prostate cancer cells in the patient.
2. The method of claim 1, wherein said sample is selected from the group consisting of prostate tissue, prostate cells, blood, urine, semen, and prostatic secretions.
3. The method of claim 1, wherein the NDRG1-ERG fusion molecule being detected is a genomic fusion molecule on a chromosome comprising a 5' portion of the NDRG1 gene and a 3' portion of the ERG gene, wherein the 5' portion of the NDRG1 gene includes a portion of the 5' transcription regulatory region of the NDRG1 gene.
4. The method of claim 3, wherein the genomic fusion molecule is detected by using a nucleic acid amplification method, a nucleic acid hybridization method, or a method that combines nucleic acid amplification and nucleic acid hybridization.
5. The method of claim 4, wherein said nucleic acid hybridization is fluorescence in situ hybridization (FISH).
6. The method of claim 5, wherein the FISH assay is performed using a pair of break-apart probes flanking the NDRG1 gene, wherein one probe is specific for a region on the centromeric side of the NDRG1 gene, and the other probe is specific for a region on the telomeric side of the NDRG1 gene.
7. The method of claim 5, wherein the FISH assay is performed using a pair of probes that detect chromosomal rearrangement which creates an NDRG1-ERG fusion, wherein one probe is specific for the upstream chromosomal region of the NDRG1 gene, and the other probe is specific for the downstream chromosomal region of the ERG gene.

8. The method of claim 1, wherein the NDRG1-ERG fusion molecule being detected is a fusion mRNA molecule comprising a 5' portion of an NDRG1 mRNA and a 3' portion of an ERG mRNA.
9. The method of claim 8, wherein said fusion mRNA molecule is encoded by a cDNA comprising the nucleotide sequence of SEQ ID NO: 6 or 8.
10. The method of claim 8, wherein the fusion mRNA molecule is detected by using a nucleic acid amplification method, a nucleic acid hybridization method, or a method that combines nucleic acid amplification and nucleic acid hybridization.
11. The method of claim 10, wherein the fusion mRNA molecule is detected by using a nucleic acid amplification method selected from the group consisting of polymerase chain reaction (PCR), reverse transcription polymerase chain reaction (RT-PCR), transcription mediated amplification (TMA), ligase chain reaction (LCR), strand displacement amplification (SDA) and nucleic acid sequence based amplification (NASBA).
12. The method of claim 10, wherein the nucleic acid amplification method is performed using a first primer specific for a 5' region of an NDRG1 mRNA, and a second primer specific for a 3' region of an ERG mRNA.
13. The method of claim 12, wherein said 5' region of the NDRG1 mRNA comprises the 5' untranslated region, exon 1, exon 2 and exon 3 of the NDRG1 mRNA.
14. The method of claim 12, wherein said 3' region of the ERG mRNA comprises the 3' untranslated region, exon 4 and exons downstream of exon 4.
15. The method of claim 11, wherein the nucleic acid amplification method is performed using primers comprising at least a primer specific for the junction of said fusion mRNA.

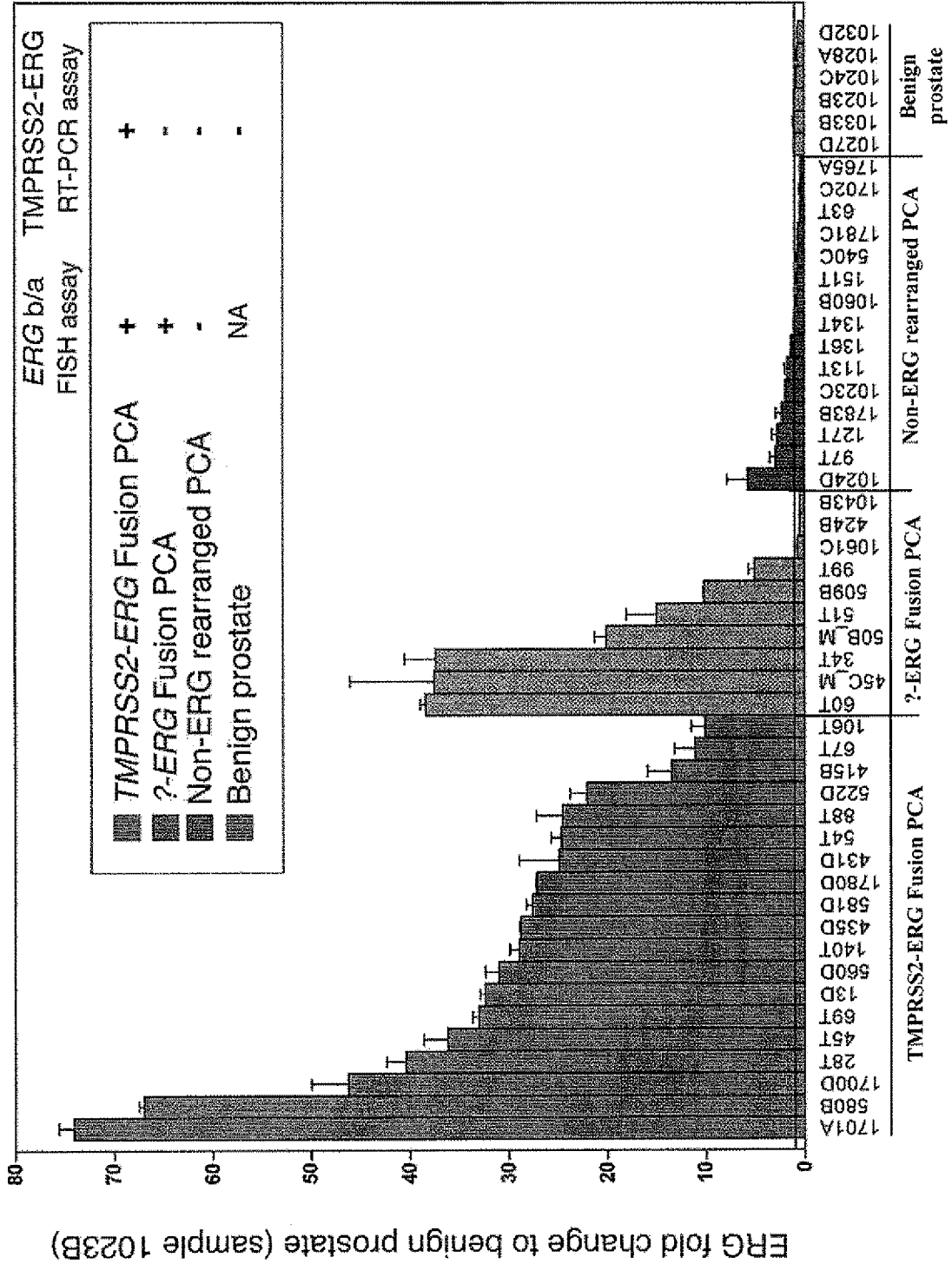
16. The method of claim 15, wherein said junction comprises the sequence as set forth in SEQ ID NO: 4 or 5.
17. The method of claim 10, wherein said hybridization method is selected from the group consisting of in situ hybridization, hybridization to a microarray, solution phase hybridization, and Northern blot hybridization.
18. The method of claim 17, wherein the hybridization method is performed using an oligonucleotide probe specific for the junction of said fusion mRNA.
19. The method of claim 18, wherein said junction comprises the sequence as set forth in SEQ ID NO: 4 or 5.
20. The method of claim 1, wherein the NDRG1-ERG fusion molecule being detected is a NDRG1-ERG fusion protein comprising an N-terminal sequence of an NDRG1 protein and a C-terminal sequence of an ERG protein.
21. The method of claim 20, wherein said NDRG1-ERG fusion protein comprises the amino acid sequence as set forth in SEQ ID NO: 7 or 9.
22. The method of claim 20, wherein the fusion protein is detected in an immunoassay using an antibody that binds to said fusion protein.
23. The method of claim 22, wherein said antibody is specific for the fusion junction of said fusion protein.
24. The method of claim 1, further comprising detecting a TMPRSS2-ERG fusion, a SCL45A3-ERG fusion, or a combination thereof.

25. A composition for detecting a fusion molecule associated with prostate cancer comprising at least one of the following:
- (a) a first nucleic acid probe specific for a region on the centromeric side of the NDRG1 gene; and a second nucleic acid probe specific for a region on the telomeric side of the NDRG1 gene.
 - (b) a first nucleic acid probe specific for the upstream chromosomal region of the NDRG1 gene, and a second nucleic acid probe specific for the downstream chromosomal region of the ERG gene;
 - (c) a first oligonucleotide specific for a 5' region of the NDRG1 genomic sequence and a second oligonucleotide specific for a 3' region of the ERG genomic sequence;
 - (d) a first oligonucleotide specific for a 5' portion of an NDRG1 mRNA and a second oligonucleotide specific for to a 3' portion of an ERG mRNA;
 - (e) an oligonucleotide specific for the junction of an NDRG1-ERG fusion mRNA; and
 - (f) an antibody specific for an amino acid sequence at the junction of an NDRG1-ERG fusion protein.
26. The composition of claim 25, wherein the nucleic acid probes of (a) comprise BAC clones designated as RP11-185E14 and RP11-1145H17.
27. The composition of claim 25, wherein the nucleic acid probes of (b) comprise BAC clones designated as RP11-1145H17 and RP11-24A11.
28. The composition of claim 25, wherein in (d) said 5' portion comprises the 5' untranslated region, exon 1, exon 2 and exon 3 of said NDRG1 mRNA; and said 3' portion comprises the 3' untranslated region, exon 4 and exons downstream of exon 4 of said ERG mRNA.
29. The composition of claim 25, wherein the oligonucleotide of (e) is specific for the junction of the NDRG1-ERG fusion variant 1 comprising SEQ ID NO: 6, or the junction of the NDRG1-ERG fusion variant 2 comprising SEQ ID NO: 8.

30. The composition of claim 25, wherein the antibody of (f) is specific for the junction of the NDRG1-ERG fusion variant 1 protein comprising SEQ ID NO: 7, or the junction of the NDRG1-ERG fusion variant 2 protein comprising SEQ ID NO: 9.
31. An isolated nucleic acid, coding for an NDRG1-ERG fusion protein which comprises the amino acid sequence as set forth in SEQ ID NO: 7 or SEQ ID NO: 9.
32. The isolated nucleic acid of claim 31, comprising the nucleotide sequence as set forth in SEQ ID NO: 6 or SEQ ID NO: 8.
33. An expression vector comprising the isolated nucleic acid of claim 31 or 32, operably linked to a promoter.
34. A host cell transformed with the expression vector of claim 33.
35. An isolated NDRG1-ERG fusion polypeptide, comprising the amino acid sequence as set forth in SEQ ID NO: 7 or SEQ ID NO: 9.
36. A method of identifying an agent useful for treating prostate cancer in a patient, comprising providing a cell which expresses an NDRG1-ERG fusion molecule, exposing said cell to candidate agents, and identifying an agent that inhibits a biological function or reduces the level of said fusion molecule.
37. The method of claim 36, wherein said biological function is enhanced cell invasion.
38. The method of claim 36, wherein said cell is transformed to express an NDRG1-ERG fusion protein which comprises the amino acid sequence as set forth in SEQ ID NO: 7 or SEQ ID NO: 9.

39. A method for treating a patient having prostate cancer characterized by NDRG1-ERG fusion, comprising administering to the patient an agent that inhibits a biological function or reduces the level of a NDRG1-ERG fusion molecule.

FIGURE 1A



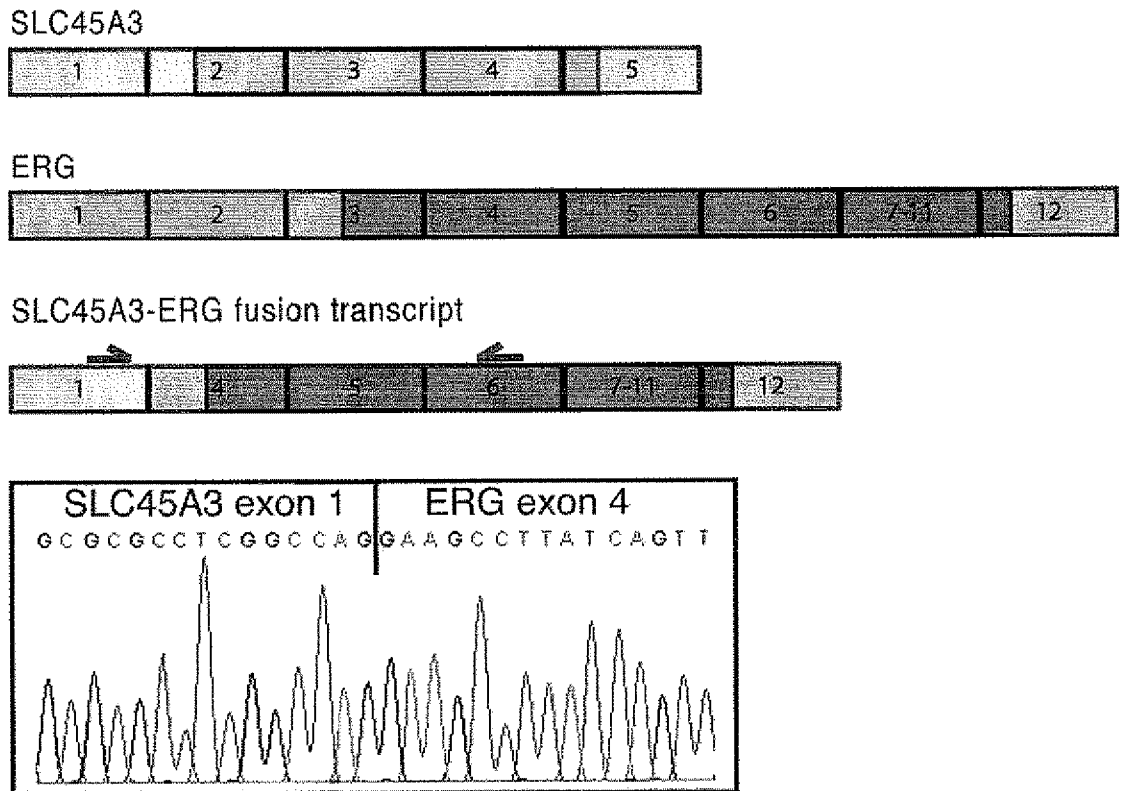


FIGURE 1B

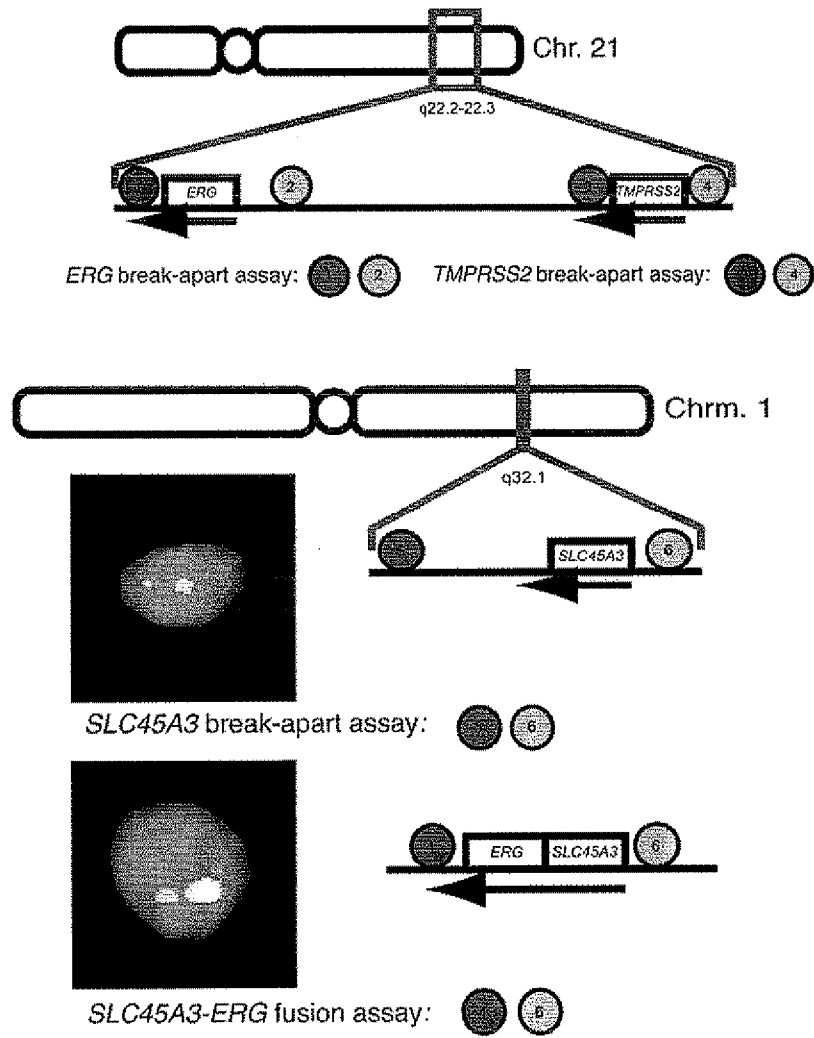


FIGURE 1C

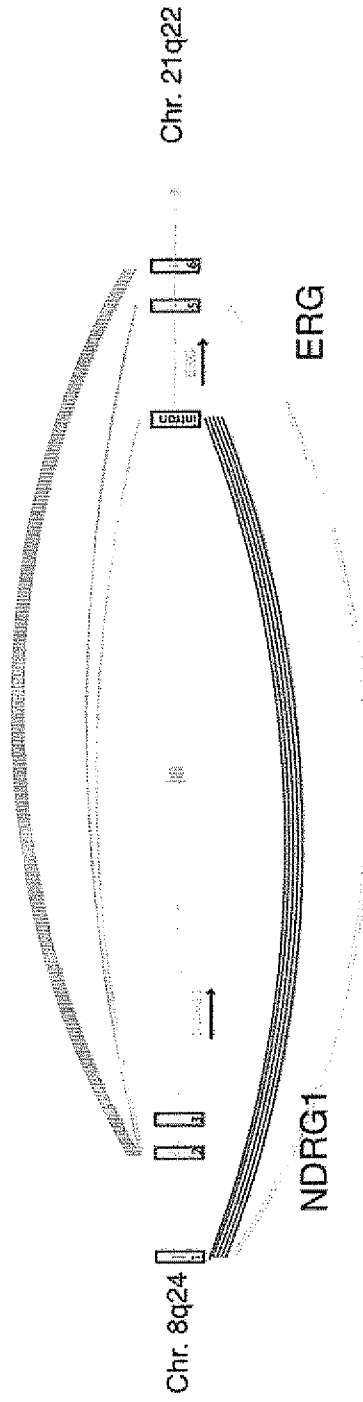
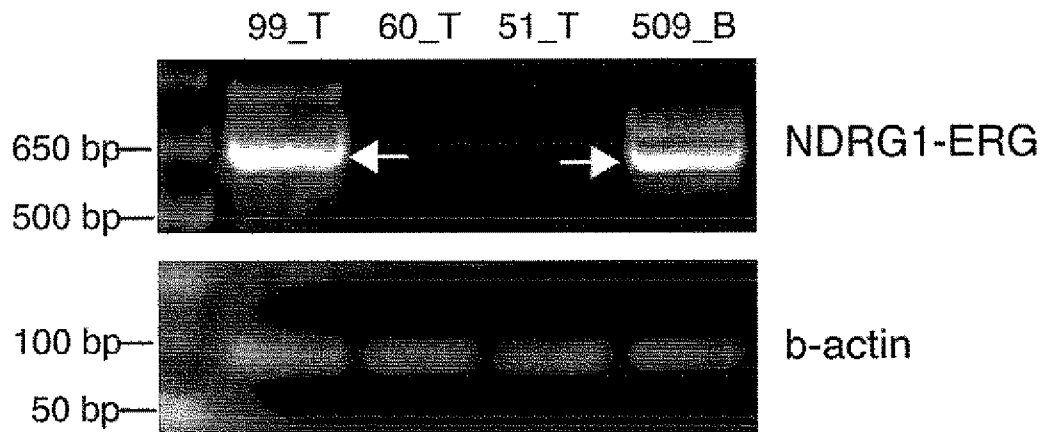
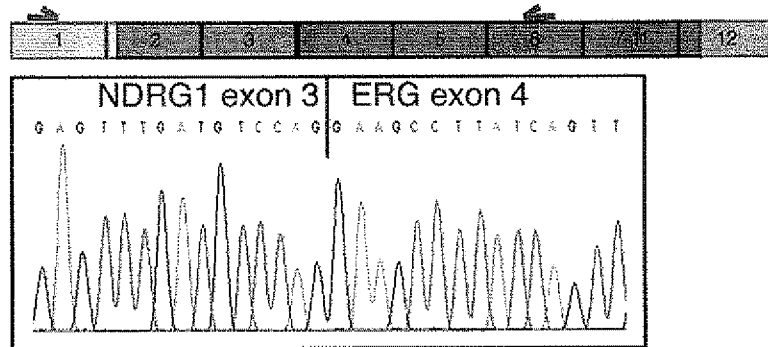


FIGURE 2A



NDRG1-ERG fusion transcript v1



NDRG1-ERG fusion transcript v2

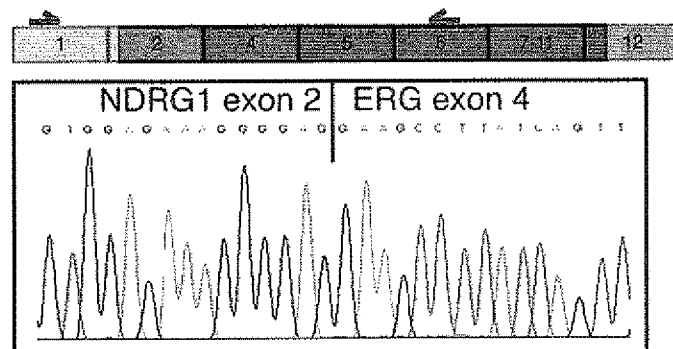


FIGURE 2B

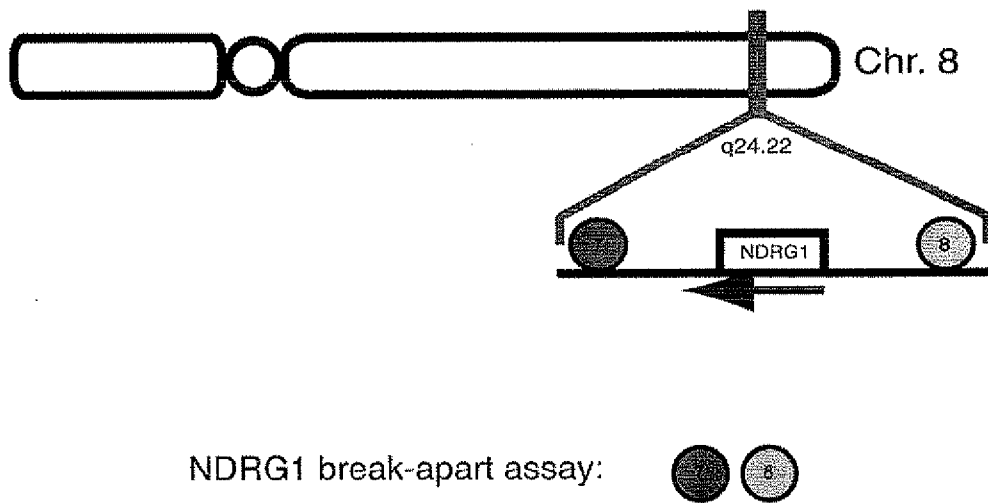


FIGURE 2C

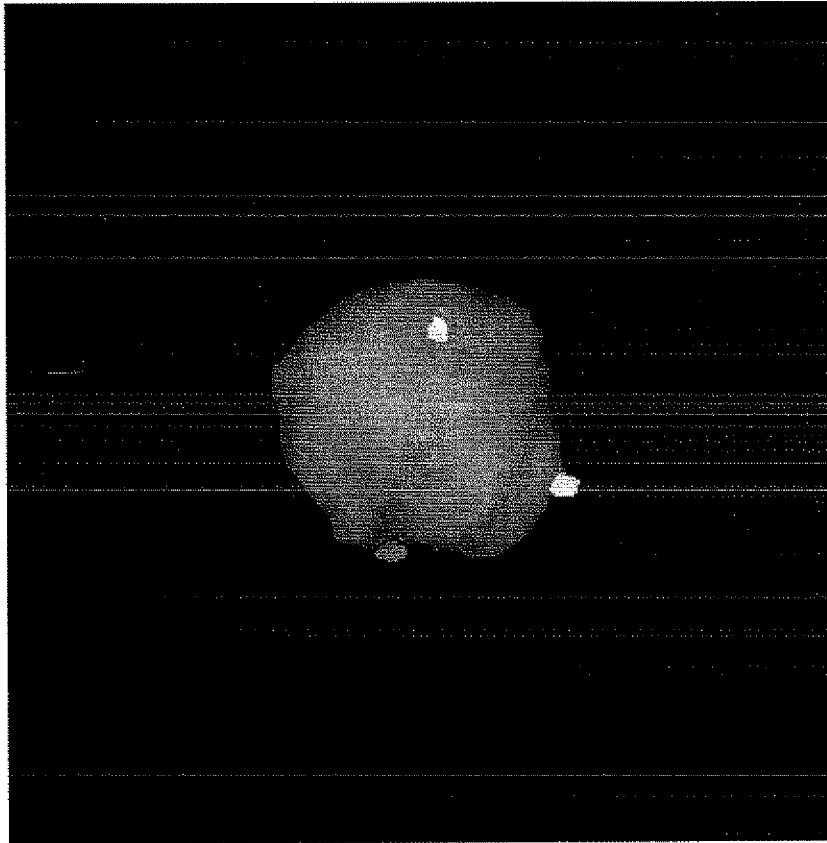


FIGURE 2D

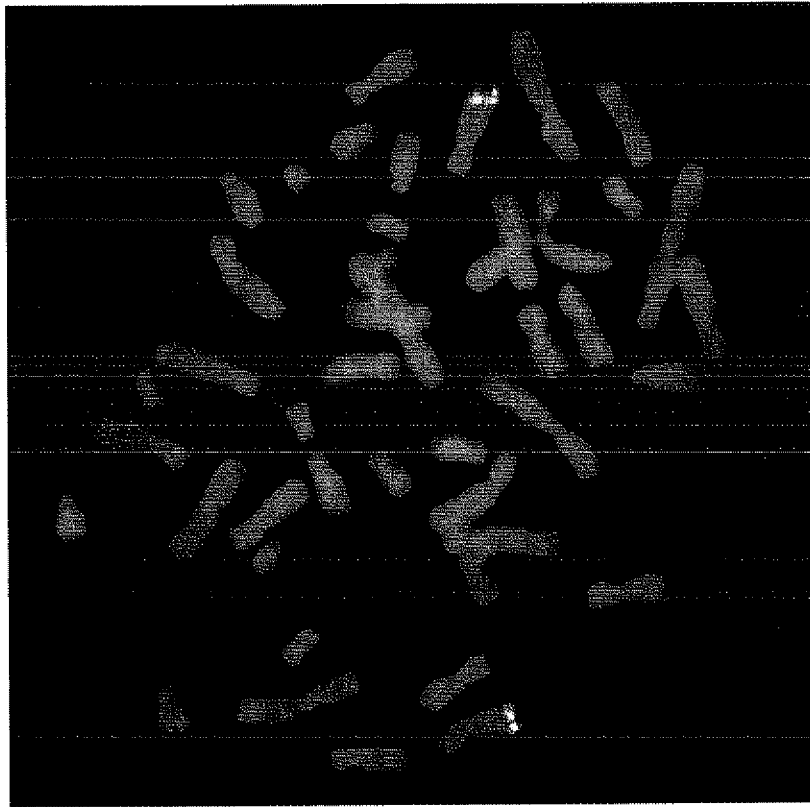


FIGURE 3

Figure 4A

NDRG1 ex1-3 – ERG ex4-12 (SEQ ID No: 6)

aacaaacctcgccctggctccagctggctgclgaagctcgtcagttcaccatecgcctcggcttcogcggggcgtggccgccagcctcggcaccgtccttcccttccctcgcgttaggcagg
tgacagcaggacatgctcgggagatgcaggatgtagacctcgtgagggtgaagcctttggggagaaaggggagaccatcaccggcctcctgcaagatllgatgtccaggaagccfta
fcagttgtgagtgaggaccagtcttcttggagtgctcctacgggaacgccacacctggctaagacagagatgaccgcgtcctct
ccagcgaactatggacagacttccaagatgagccacgcgtcctcagcaggattggctgtctcaacccccagccagggtcacc
atcaaaatggaatgtaaccctagccaggtgaatggcicaaggaactctcctgatgaatgcagtggtgccaagggcgggaagat
ggtggcagcccagacaccgttgggatgaactacggcagctacatggaggagaagacatgccacccccaaacatgaccac
gaacgagcgcagagttatcgtgccagcagatcctacgctatggagtacagaccatgtcggcagtgctggatggcgggtg
aaagaatagccttccagacgtcaacatcttattccagaacatcgtatgggaaggaaactgtcaagatgaccaaggacgact
tccagaggcacccccagctacaacgccgacatcctctctcaatctccactacctcagagagactcctctccacattgactt
cagatgatgttgataaagccttacaacacctccacggtaatgcatgctagaaacacaggggtgcagctttatttccaaata
cttcagttatctcgaagctacgcaagaattacaactaggccagattaccatagagccccagagatcagcctggaccgg
tcacggccaccccacgcccagtcgaaagctgtcaaacatctcctccacagtgcccaaaactgaagaccagcgtcctcagtt
agatcctatcagatcttggaccaacaagtagcccttgc aaatccaggcagtgccagatccagcttggcagtlcctcctgg
agctcctgtcggacagctccaactccagctgcatcaccctgggaaggcaccacggggagtcaagatgacggatcccagca
ggtggcccggcgtggggagagcgggaagagcaaacccaacatgaactacgataagctcagccggcctccgttactactat
gacaagaacatcatgaccaaggtccatgggaagcgtacgcctacaagttcgaactccacgggatcggccagccctccagcc
ccacccccggagtcctctctgtacaagttaccctcagacctcccgtacatgggctctctacacgccacccacagaagatgaa
ctttgtggcgcccaccctccagcccctcccgtgacatctccagtttttctgcccacaaccatactggaattaccaactgg
gggtatatacccaaacactaggtcccaccagccatagccttctcactctggcacttactactaaagacctggcggagccttt
cccatcagcgtgcaaccagcccacgccacaacactctatcggagaacatgaatcaaaagtgcctcaagaggatgaaaaa
gctttactgggctggggaagggaagccggggaagagatccaagactctgggagggagttactgaagtcttactacagaat
gagggagatgctaaaaatgtcacgaatgacatcatctgtgactgacctgtaaaagacagtgatgtagaagcatgaagt
cttaaggacaaagtccaaaagaaagtgtcttaagaaatgtataaactttagatgaggttggaatcccactaatgcaactggg
atgaaactaaagcaatagaacaacacagttttgacctaacataccgtttataatgccatttaaggaaaactacctgtattaaaaat
agaaacatatacaaaaacaagagaaaagacagagagagactgtggccatcaacagacgttgatatgcaactgcatggcatgt
gctgtttgttgaatcaaatcattccgtttgatggacagctgtcagctttctcaactgtgaagatgacccaaagttccaactcc
ttacagttaccgggactatgaactaaaaggtgggactgaggatgtgtatagagtgcgctgtgattgtagacagagggtga
agaaggaggaggaagagccagagaaggagggagaccaggctgggaaagaaacttcaagcaatgaagactggactcagg
acatttgggactgtgtacaatgagttatggagactcggggttcatgcagtcagttataccaaacccagctttaggagaagg
acacagcgtaatggagaaggggaagtagtagaatcagaacaaaaatcgcatctcttcttctgtcaaatgaaaatttaact
ggaattgtctgatattaaagagaacatcaggacctcattatgtggggccttctcaccagggcaggtlaagagatggcc
ttcttggctgccacaatcagaatcacgcaggcatttggtagcggcctccagtttcttcttggagtcgcaacctgtgcgtttg
cagaatgaagtatacaagtcatttttccccctttatataataattatataacttatgattatatacactacaggttgatctcggcca
gcaaaagacacacgacaaaagagacaatcgaataatgtggccttgaatttaactctgtatcctaatgtttacaatatgaagtatt
agttctagaatgcagaatgtatgtafaaaaaaagccttggcctagcatggcaaatcagattatatacaggagctgcatttgcactttt
ttagtgactaaaagttcctaatgaaaacatgtgctgaatgttggatttgggtataatftacttctcagggaactgtgcaaggga
gagccaaggaaataggatgttggcacc

Figure 4B

Variant 1: NDRG1 ex1-3 – ERG ex4-12 (SEQ ID No: 7)

MSREMQDVDLAEVKPLVEKGETITGLLQEFDVQEALSVVSEDOSLFECAYGTPH
LAKTEMTASSSSDYGOTSKMSPRVPOODWLSOPPARVTIKMECNPSOVNGSRNS
PDECSVAKGGKMVGSPDTVGMNYGSYMEEKHMPPPNMTTNERRVIVPADPTL
WSTDHVRQWLEWAVKEYGLPDVNILLFONIDGKELCKMTKDDFORLTPSYNAD
ILLSHLHYLRETPLPHLTSDDVDKALONSPRLMHARNTGGAAFIFPNTSVYPEAT
QRITTRPDLPEPPRRSAWTGHGHPTQSKAAQSPSTVPKTEDORPOLDPYQILG
PTSSRLANPGSGOIQLWQFLELLSDSSNSSCITWEGTNGEFKMTDPDEVARRWG
ERKSKPNMNYDKLSRALRYYYDKNIMTKVHGKRYAYKFDFFHGIAOALOPHPPE
SSLYKYPSDLPYMGSYHAHPQKMNFVAPHPPALPVTSSSFFAAPNPYWNSPTGGI
YPNTRLPTSHMPSHLGTY

Figure 5A

NDRG1 ex 1-2 -ERG ex 4-12 *Variant 2* : (SEQ ID No: 8)

aacaaacctgcctggctcccagctggctgctgaagctcgtcagttcaccatccgccctcggcttccggggcctggccgccagcctcggcaaccgcttcttctccctcgcgttaggcagg
tgacagcagggacatgctcgggagatgcaggatgtagacctcgtgaggtgaaaccttgggggagaaggggaggaagccttatcagttgtgagtgaggaccagtc
gttctttgagtgcttacggaacgccacacctggctaagacagagatgaccgctcctcctccagcactatggacagactcc
aagatgagcccacgctcctcagcaggtggctgtctcaacccccagccagggtcaccatcaaaatggaatgtaacctag
ccaggatgaaatggctcaaggaactctctgatgaatgcagtggtgccaaaggcgggaagatggtggcagcccagacaccgt
gggatgaaclacggcagctacatggaggagaagcacatgccacccccaaacatgaccacgaacgagcgcaggtatctgtg
ccagcagatctacgctatggagfacagaccatgtgcggcagtgctggagtgccgggtgaaagaatatggcctccagacgt
caacatctgttatccagaacatcgtgggaaggaactgtgcaagatgaccaaggacgactccagaggctcacccccagcta
caacgccgacatccttctcacaatctccactacctcagagagactcctctccacattgacttcagatgatgtgataaagccttac
aaaactctccacggtaatgcatgctagaacacagggggtgcagcttttatttccaaatacttcagtatacctgaagctacgca
aagaatfacaactaggccagattaccatagagccccagagatcagcctggaccggtcacggccaccccacgcccagt
cgaagctgtcfaaccatctctccacagtgcccaaaactgaagaccagcgtcctcagttagatccttatcagattcttgacca
acaagtagccgcttgcfaatccagcagtgccagatccagcttggcagttcctcctggagctcctgtcggacagctccaact
ccagctgcatcaccgtgggaagcaccacggggagttcaagatgacggatcccgacagagtgcccggcgtggggagag
cggaaagacaaacccaacatgaactacgataagctcagccggccctccgtfactactatgacaagaacatcatgaccaaggf
ccatgggaagcgtctacgctacaagttcagctccacgggatcggccaggccctccagccccaccccggatcctctgt
acaagtaccctcagacctcccgtacatggctcctatcacgccaccacagaagatgaactttgtggcggcccacctccag
ccctccccgtgacatctccagtttttctgctccccaaaccatactggaattcaccactgggggtatatacccacaactaggc
tcccaccagccatgctctcctctctggcacttactactaaagacctggcggaggctttccatcagcgtgcaatccaccg
ccatcgccacaacctctatcggagaacatgaatcaaaagtcctcaagaggaatgaaaaagctttactgggctggggaag
aagccggggaagagatccaaagactctgggagggaactgactgaagcttactacagaatgaggagatgctaaaaatgtca
cgaatgagacatctctgtgactgacctgtaaaagacagtgatgtagaagcatgaagctttaaaggacaaagtccaaag
aaagtgtcttaagaaatgataaactttagatgagtttggaaatcccactaatgcaaacctggatgaaactaaagcaatgaaa
caacacagttttgacctaacataaccgttataatgccatttaaggaaaactacctgtatttaaaaatgaaaacatatacaaaacaag
agaaaagacacgagagagactgtggccatcaacagacgttgatgcaactgcatggcatgtgctgtttgtgaaatcaaat
acattccgttgatggacagctgtcagcttctcaactgtgaagatgacccaagttccaactcctttacagtattaccgggactat
gaaactaaaagtgaggactgaggatgtgtatagagtgagcgtgtgattgtagacagaggggtgaaagaggaggaggaagg
cagagaaaggaggagaccagcgtggaaagaaacttctcaagcaatgaagactggactcaggacatttgggactgtgtcaat
gagttatggagactcgagggttcatgcaagtgatgataccaaaccaggttaggagaaaaggacacagcgtaatggagaaag
ggaagtagtagaafccagaacaaaaatgcgcatcttcttctgttctcaaatgaaaatttaactggaattgtctgatattagag
aaacattcaggacctcatcattatgtggggcttgtctccacagggtcaggtaagagatggccttctggctgccacaatcaga
aatcacgcaggcatttgggtagcggcctccagtttctttagctgcggaacgctgtgctgttctcagaatgaagtatacagtc
aatgttttcccccttttataataattataaacttatgcaattatacactacaggtgatctcggccagccaaagacacacgacaaa
agagacaatcgaataatgtgacctgaatttaactctgtatgcttaatgtttacaatatgaagttattagttcagaatgcagaatg
atgtaataaaaataagcctggcctagcatgcaaaatcagattatfacaggagctcgtcattgcacttttttagtactaaaagtgtttaa
tgaaaacatgtcgtgaatgttggattttgtgtataattactttgcccaggaactgtgcaaggagagccaaggaaataggatg
ttggcacc

Figure 5B

Variant 2: NDRG1 ex 1-2 - ERG ex 4-12 (SEQ ID No: 9)

MSREMQDVDLAEVKPLVEKGEEALSVVSEDOSLFECAYGTPHLAKTEMTASSSS
DYGOTSKMSPRVPOQDWLSOPPARVTIKMECNPSOVNGSRNSPDECSVAKGGK
MVGSPDTVGMNYGSYMEEKHMPPPNMTTNERRVIVPADPTLWSTDHVROWLE
WAVKEYGLPDVNILLFONIDGKELCKMTKDDFORLTPSYNADILLSHLHYLRETP
LPHLTSDDDVDKALONSPRLMHARNTGGAAFIFPNTSVYPEATORITTRPDLPEP
PRRSAWTGHGHPTPOSKAAOPSPSTVPKTEDORPOLDPYQILGPTSSRLANPGSG
QIOLWQFLLELLSDSSNSSCITWEGTNGEFKMTDPDEVARRWGERKSKPNMNYD
KLSRALRYYYDKNIMTKVHGKRYAYKDFDHGIAQALOPHPPESSLYKYPSDLPY
MGSYHAHPQKMNFVAPHPALPVTSSSFFAAPNPYWNSPTGGIYPNTRLPTSHM
PSHLGTY

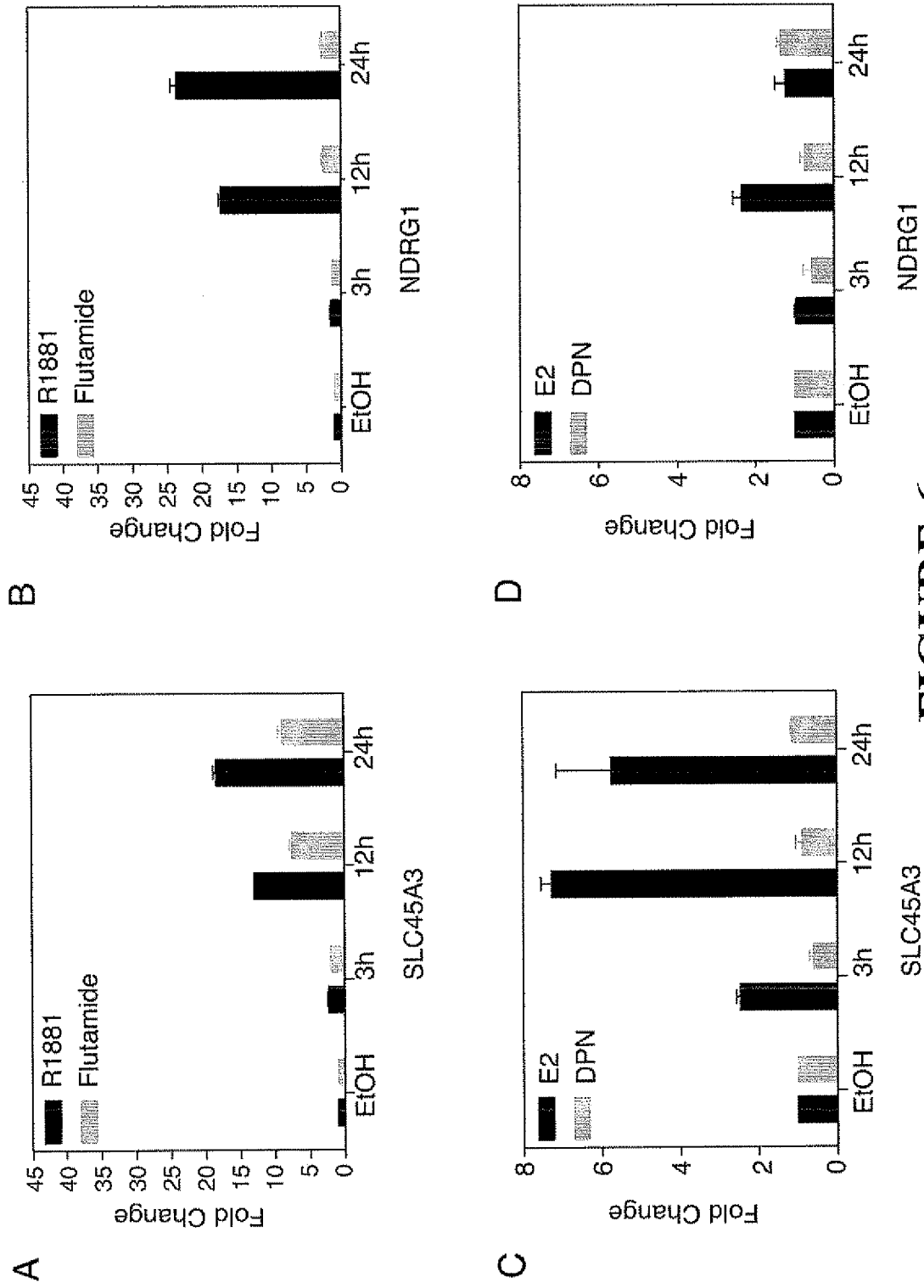


FIGURE 6

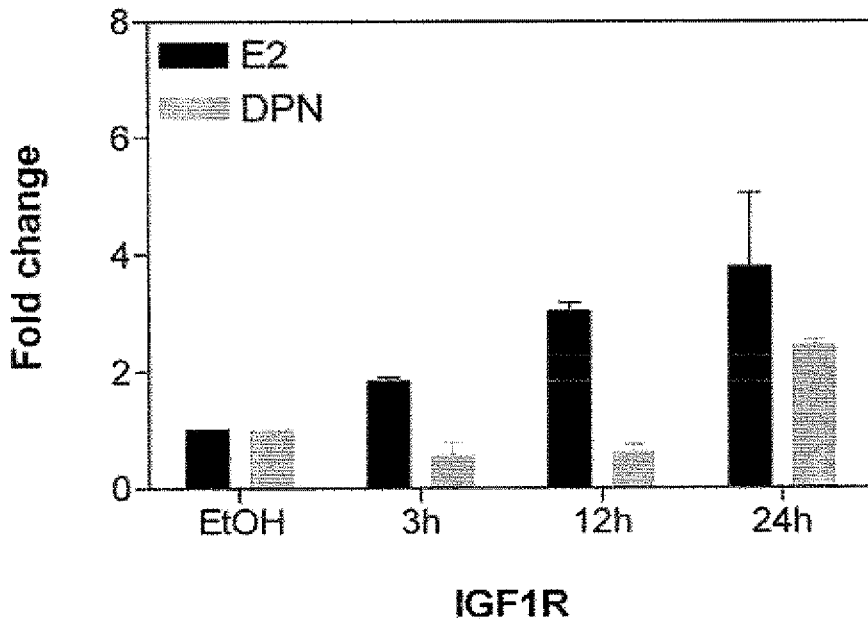
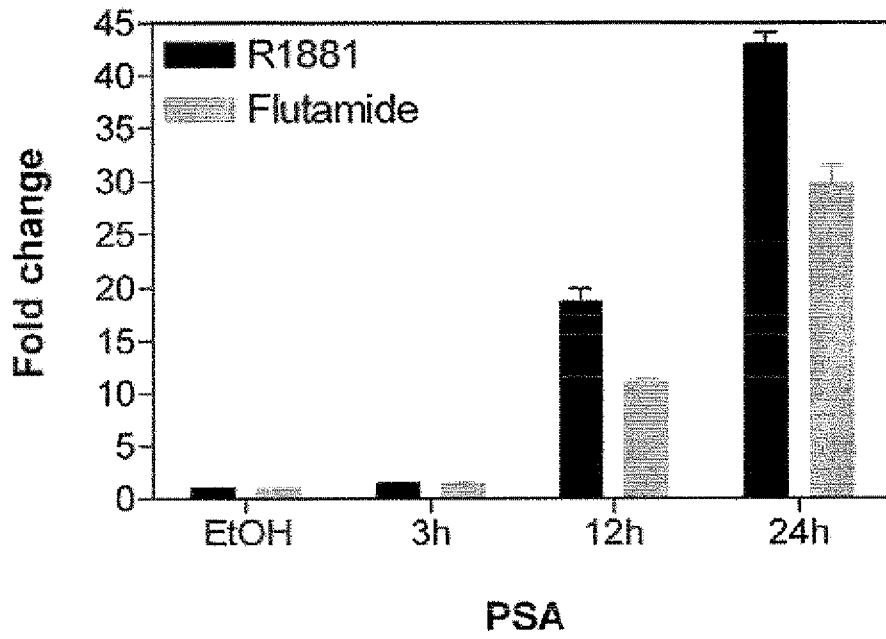


FIGURE 7

FIGURE 8
NDRG1-ERG protein expression in HEK-293 (embryonic kidney) cells

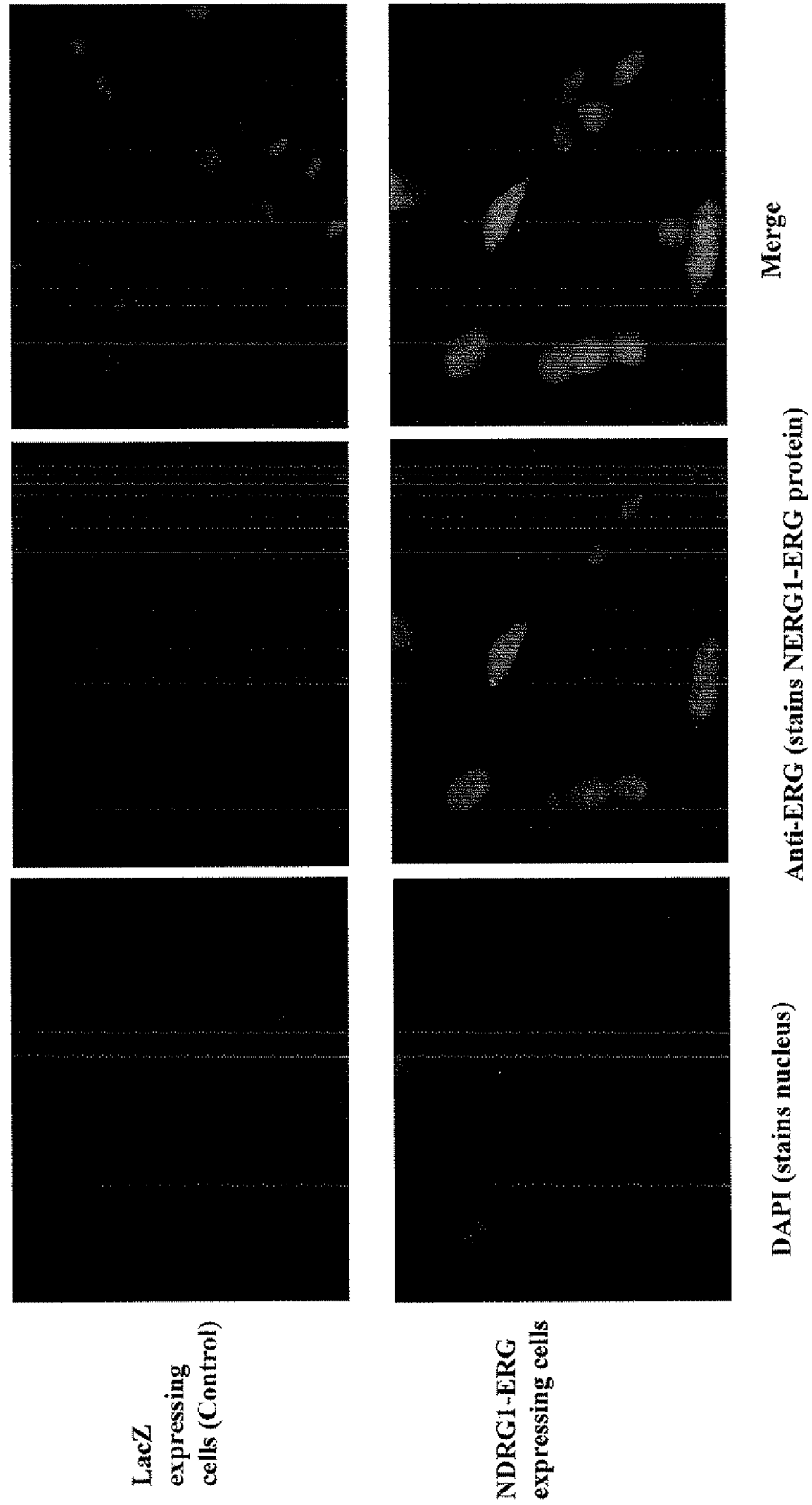
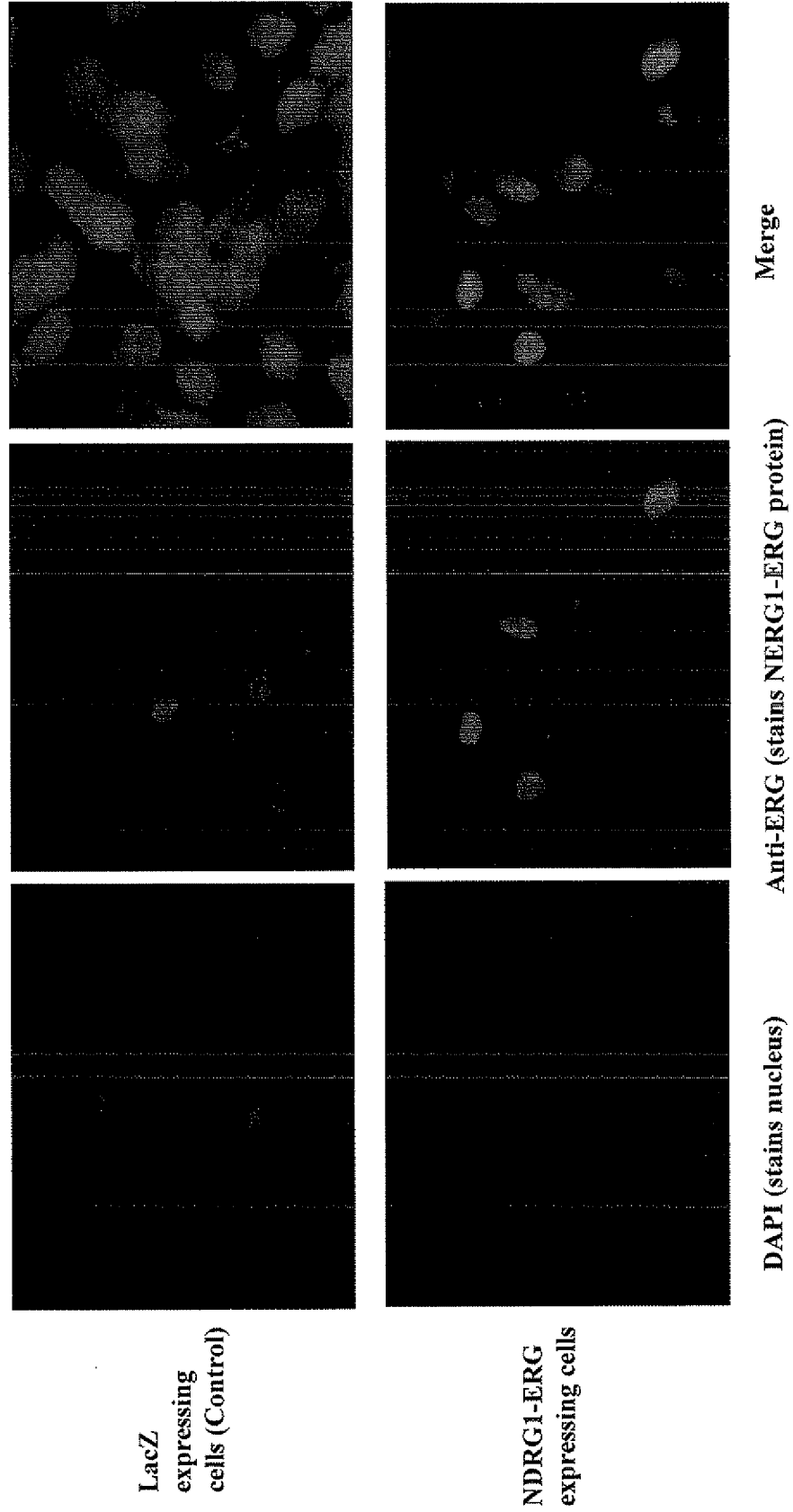
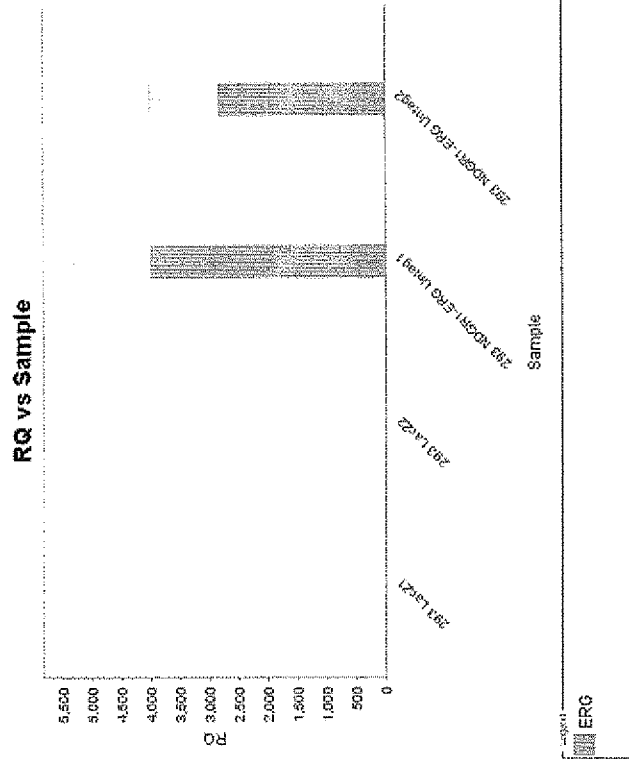


FIGURE 9
NDRG1-ERG protein expression in PHI (prostate epithelial) cells



HEK-293 Cell



BPH1 Cell

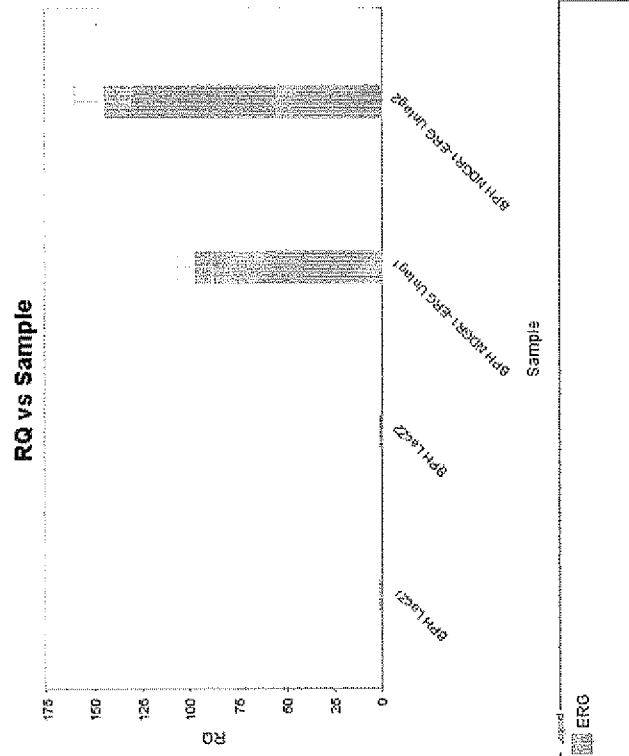
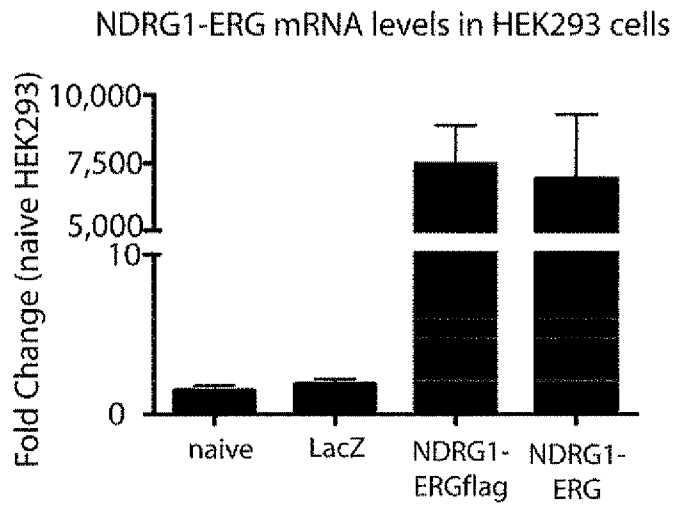
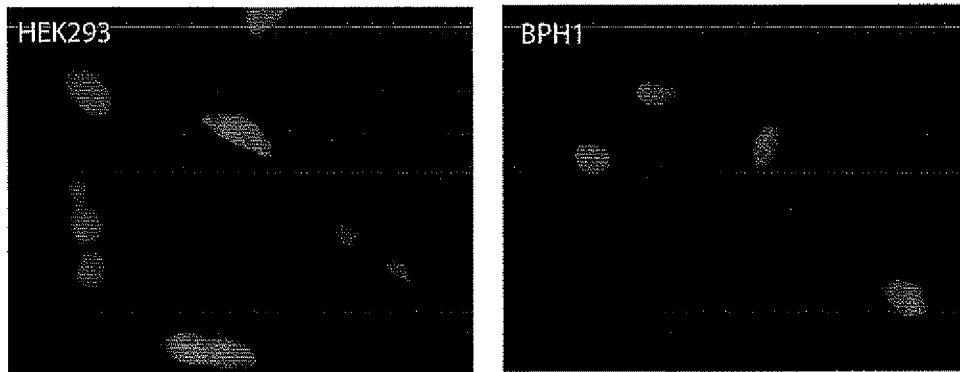


Figure 10. NDRG1-ERG mRNA over-expression in different cell lines



Immunofluorescence staining of NDRG-ERGflag expressing cells



Invasion assay of NDRG-ERGflag HEK293 cells

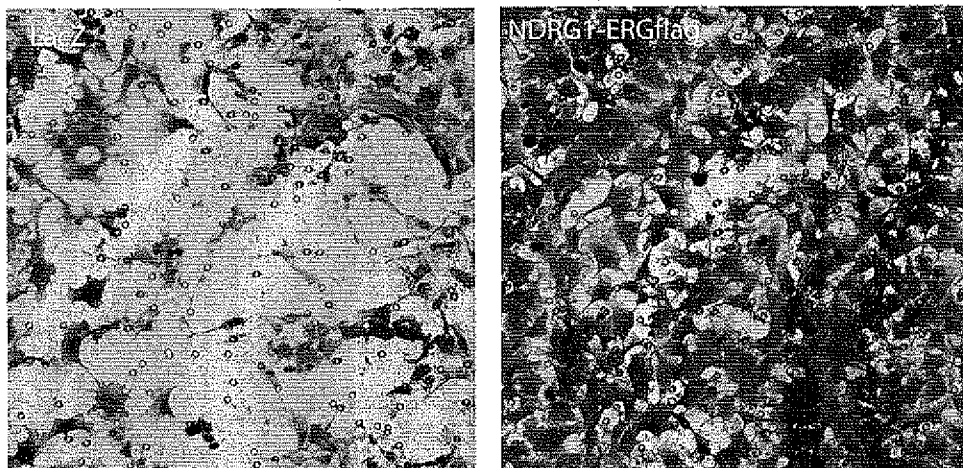


FIGURE 11