



(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2016/015714

(22) International Filing Date:
29 January 2016 (29.01.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/110,034 30 January 2015 (30.01.2015) US

(71) Applicant: **THE REGENTS OF THE UNIVERSITY OF MICHIGAN** [US/US]; 1600 Huron Parkway, 2nd Floor, Ann Arbor, Michigan 48109-2590 (US).

(72) Inventors: **ELENITOBA-JOHNSON, Kojo**; 1600 Huron Parkway, 2nd Floor, Ann Arbor, Michigan 48109-2590 (US). **KIEL, Mark J.**; 1600 Huron Parkway, 2nd Floor, Ann Arbor, Michigan 48109-2590 (US). **ROLLAND, Delphine**; 1600 Huron Parkway, 2nd Floor, Ann Arbor, Michigan 48109-2590 (US). **BETZ, Bryan L.**; 1600 Huron Parkway, 2nd Floor, Ann Arbor, Michigan 48109-2590 (US). **BAILEY, Nathanael G.**; 1600 Huron Parkway, 2nd Floor, Ann Arbor, Michigan 48109-2590 (US). **VELUSAMY, Thirunavukkarasu**; 1600 Huron Parkway, 2nd Floor, Ann Arbor, Michigan 48109-2590 (US). **LIM, Megan**; 1600 Huron Parkway, 2nd Floor, Ann Arbor, Michigan 48109-2590 (US).

(74) Agent: **GOETZ, Robert A.**; CASIMIR JONES S.C., 2275 Deming Way, Ste 310, Middleton, Wisconsin 53562 (US).

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

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

Published:

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

(54) Title: RECURRENT GENE FUSIONS IN CUTANEOUS CD30-POSITIVE LYMPHOPROLIFERATIVE DISORDERS

(57) Abstract: Provided herein are kits, compositions and methods for cancer diagnosis, research and therapy, including but not limited to, cancer markers. In particular, the present invention relates to recurrent gene fusions (e.g., recurrent translocations involving TYK2) as diagnostic markers and clinical targets for cutaneous CD30-positive lymphoproliferative disorders (e.g., lymphomatoid papulosis; primary cutaneous anaplastic large cell lymphoma).



RECURRENT GENE FUSIONS IN CUTANEOUS CD30-POSITIVE LYMPHOPROLIFERATIVE DISORDERS

CROSS-REFERENCE TO RELATED APPLICATIONS

5 The present application claims the priority benefit of U.S. Provisional Patent Application 62/110,034, filed January 30, 2015, which is incorporated by reference in its entirety.

FIELD OF THE INVENTION

10 Provided herein are kits, compositions and methods for cancer diagnosis, research and therapy, including but not limited to, cancer markers. In particular, the present invention relates to recurrent gene fusions (e.g., recurrent translocations involving *TYK2*) as diagnostic markers and clinical targets for cutaneous CD30-positive lymphoproliferative disorders (e.g., lymphomatoid papulosis; primary cutaneous anaplastic large cell lymphoma).

15

BACKGROUND OF THE INVENTION

 Lymphoproliferative disorders (LPDs) refer to several conditions in which lymphocytes are produced in excessive quantities. They typically occur in patients who have compromised immune systems. Examples of LPDs include, but are not limited to, follicular
20 lymphoma, chronic lymphocytic leukemia, acute lymphoblastic leukemia, hairy cell leukemia, lymphomas, multiple myeloma, Waldenstrom's macroglobulinemia, Wiskott-Aldrich syndrome, post-transplant lymphoproliferative disorder, autoimmune lymphoproliferative syndrome (ALPS), lymphoid interstitial pneumonia, and CD30-positive lymphoproliferative disorders. Examples of CD30-positive lymphoproliferative disorders
25 include lymphomatoid papulosis and primary cutaneous anaplastic large cell lymphoma.

 Lymphomatoid papulosis (LYP) is characterized by multiple papules and nodules, which regress spontaneously. Three histologic subtypes have been described, which represent a spectrum with overlapping features and do not carry prognostic significance. Type A lesions have a few tumor cells in a background of inflammatory cells including neutrophils,
30 eosinophils, and histiocytes. Type B lesions are characterized by epidermotropic lymphocytes with cerebriform nuclei mimicking mycosis fungoides. Type C lesions have sheets of large atypical lymphoid cells with only a few admixed inflammatory cells. The large atypical lymphoid cells are thought to be of T cell origin. Various histologic types may be present in individual patients at the same time. In LYP types A and C the large atypical cells express
35 CD30, CD3, and CD4. CD2 and CD5 are usually expressed. These cells also express the

cytotoxic markers TIA-1 and granzyme. The large atypical cells do not express CD8, CD7, or CD56. The cells may lose expression of CD3. In LYP type B the atypical cells are usually CD30 negative. Five year survival rates for LYP are 100%; however up to 20% of patients develop LYP-associated malignant lymphomas (e.g., mycosis fungoides, Hodgkin lymphoma, systemic or cutaneous CD30+ large T-cell lymphoma), which result in a fatal outcome of 2% of patients.

Primary cutaneous anaplastic large cell lymphoma (ALCL) is composed of large atypical to anaplastic appearing lymphoid cells. This disease mainly affects adults with a peak in the sixth decade, but cases have been reported in children. It usually presents as a solitary rapidly growing nodule. The skin overlying the lesion may ulcerate. Histologically the cells grow in sheets. Mitotic figures are frequent. Clusters of small reactive lymphocytes are found within and around the tumor cells. The malignant cells express CD2, CD3, CD4, CD30, and cytotoxic markers including TIA-1, granzyme, and perforin. Loss of T-cell antigen expression is not infrequent. The malignant cells do not express EMA or ALK. Five year survival rates approach 90%. Interestingly, up to 40% of C-ALCL show spontaneous regression similarly to LYP.

Improved methods for detecting, investigating, and treating CD30-positive lymphoproliferative disorders (e.g., lymphomatoid papulosis; primary cutaneous anaplastic large cell lymphoma) are needed.

SUMMARY OF THE INVENTION

The cutaneous CD30-positive lymphoproliferative spectrum of disorders (LPD) includes lymphomatoid papulosis (LYP) and primary cutaneous anaplastic large cell lymphoma (ALCL). Chromosomal translocations targeting tyrosine kinases in CD30-positive LPD have not been described. Using whole transcriptome sequencing, experiments conducted during the course of developing embodiments for the present invention identified a chimeric fusion involving NPM1 (5q35) and TYK2 (19p13) which encodes an NPM1-TYK2 protein containing the oligomerization domain of NPM1 and an intact catalytic domain in TYK2. Fluorescence in situ hybridization revealed NPM1-TYK2 fusions in 2 of 47 (4%) primary cases of CD30-positive LPDs and was absent in other mature T-cell neoplasms (n = 151). Functionally, NPM1-TYK2 induced constitutive TYK2, STAT1, STAT3 and STAT5 activation. Conversely, a kinase-defective NPM1-TYK2 mutant abrogated STAT1/3/5 signaling. In addition, shRNA-mediated silencing of TYK2 abrogated lymphoma cell growth. Moreover, it was demonstrated that TYK2 is targeted by multiple different translocation

partners in CD30-positive LPD. Such experiments indicate recurrent translocations involving TYK2 and highlights therapeutic opportunities in the treatment of CD30-positive LPDs with TYK2 translocations.

Accordingly, provided herein are kits, compositions and methods for cancer
5 diagnosis, research and therapy, including but not limited to, cancer markers. In particular, the present invention relates to recurrent gene fusions (e.g., recurrent translocations involving TYK2) as diagnostic markers and clinical targets for cutaneous CD30-positive lymphoproliferative disorders (e.g., lymphomatoid papulosis; primary cutaneous anaplastic large cell lymphoma).

10 In certain embodiments, the present invention provides, but is not limited to, methods, compositions, and kits involving gene fusions having a 3' portion from a TYK2 gene and a 5' portion from a TYK2-gene-fusion-partner.

Such embodiments (e.g., composition, methods, kits) are not limited to a particular portion of a TYK2 gene.

15 In some embodiments, the portion of the TYK2 gene comprises any portion of TYK2 enabling tyrosine kinase activity upon expression as a polypeptide. In some embodiments, the portion of the TYK2 gene comprises at least a portion (e.g., 1%, 10%, 25%, 50%, 75%, 80%, 95%, 99%, 100%) of the C-terminal kinase domain of TYK2. In some embodiments, the portion of the TYK2 gene comprises the entire C-terminal kinase domain of TYK2. In some
20 embodiments, the portion of the TYK2 gene comprises at least a portion of the TYK2 pseudokinase domain. In some embodiments, the portion of the TYK2 gene comprises at least a portion (e.g., 1%, 10%, 25%, 50%, 75%, 80%, 95%, 99%, 100%) of the C-terminal kinase domain of TYK2 and/or at least a portion (e.g., 1%, 10%, 25%, 50%, 75%, 80%, 95%, 99%, 100%) of the TYK2 pseudokinase domain.

25 Such embodiments (e.g., composition, methods, kits) are not limited to a particular TYK2-gene-fusion-partner.

In some embodiments, the TYK2-gene-fusion-partner is any gene such that upon fusion with TYK2 the resulting gene fusion encodes a polypeptide capable of having constitutively activated tyrosine kinase activity when expressed in a cell.

30 In some embodiments, the TYK2-gene-fusion-partner is any gene such that upon fusion with TYK2 the resulting gene fusion encodes a polypeptide that when expressed in a cell is capable of constitutively activating downstream targets of TYK2 (e.g., STAT1, STAT3, STAT5) within the cell.

In some embodiments, the TYK2-gene-fusion-partner is any gene such that upon fusion with TYK2 the resulting gene fusion encodes a polypeptide that when expressed in a cell results in increased cellular proliferation of that cell.

5 In some embodiments, the TYK2-gene-fusion-partner is any gene such that upon fusion with TYK2 the resulting gene fusion encodes a polypeptide, wherein 1) the polypeptide is capable of having constitutively activated tyrosine kinase activity when expressed in a cell, 2) expression of the polypeptide in a cell results in constitutive activation of downstream targets of TYK2 (e.g., STAT1, STAT3, STAT5), and 3) expression of the polypeptide in a cell results in increased cellular proliferation of the cell.

10 In some embodiments, the TYK2-gene-fusion-partner is any portion of NPM1.

In some embodiments, the TYK2-gene-fusion-partner comprises at least a portion (e.g., 1%, 10%, 25%, 50%, 75%, 80%, 95%, 99%, 100%) of the NPM1 oligomerization domain. In some embodiments, the TYK2-gene-fusion-partner comprises the entire NPM1 oligomerization domain.

15 In some embodiments, the TYK2-gene-fusion-partner comprises at least a portion (e.g., 1%, 10%, 25%, 50%, 75%, 80%, 95%, 99%, 100%) of the NPM1 histone binding domains. In some embodiments, the TYK2-gene-fusion-partner comprises the entire NPM1 histone binding domains.

In some embodiments, the TYK2-gene-fusion-partner comprises at least a portion (e.g., 1%, 10%, 25%, 50%, 75%, 80%, 95%, 99%, 100%) of the distal portion of the NPM1 DNA/RNA binding domains. In some embodiments, the TYK2-gene-fusion-partner comprises the entire distal portion of the NPM1 DNA/RNA binding domains.

20 In some embodiments, the TYK2-gene-fusion-partner comprises the entire NPM1 oligomerization domain, the entire NPM1 histone binding domain, and the entire distal portion of the NPM1 DNA/RNA binding domain.

In certain embodiments, the present invention provides methods for detecting cutaneous CD30-positive lymphoproliferative disorders (e.g., lymphomatoid papulosis; primary cutaneous anaplastic large cell lymphoma) in a subject (e.g., a human patient) comprising, consisting essentially of, or consisting of: providing a sample from the patient; and, detecting the presence or absence in the sample of a gene fusion having a 3' portion from a TYK2 gene and a 5' portion from a TYK2-gene-fusion-partner, wherein the presence in the sample of the gene fusion is indicative of a cutaneous CD30-positive lymphoproliferative disorder (e.g., lymphomatoid papulosis; primary cutaneous anaplastic large cell lymphoma) in the subject.

In some embodiments, the sample includes biopsy tissue, tissue, blood, plasma, serum, urine, stool, and/or cells.

Detecting the presence or absence in the sample of such a gene fusion may comprise detecting chromosomal rearrangements of genomic DNA having a 5' portion from a TYK2-gene-fusion-partner (e.g., NPM1) and a 3' portion from a TYK2 gene. A variety of techniques may be used for detecting the chromosomal rearrangements of genomic DNA, including nucleic acid sequencing, nucleic acid hybridization, and nucleic acid amplification. Nucleic acid hybridization techniques include in situ hybridization (ISH), microarray, and Southern blot. Nucleic acid amplification techniques include 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).

Detecting the presence or absence in the sample of such a gene fusion may alternatively comprise detecting chimeric mRNA transcripts having a 5' portion from a TYK2-gene-fusion-partner (e.g., NPM1) and a 3' portion from a TYK2 gene. A variety of techniques may be used for detecting the chimeric mRNA, including nucleic acid sequencing, nucleic acid hybridization, and, nucleic acid amplification. Nucleic acid hybridization techniques include in situ hybridization (ISH) (e.g., Fluorescence in situ hybridization (FISH)), microarray, and Northern blot). Nucleic acid amplification techniques include, 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).

Detecting the presence or absence in the sample of a gene fusion may also alternatively comprise detecting an amino-terminally truncated TYK2 protein resulting from a fusion of a TYK2-gene-fusion-partner (e.g., NPM1) to a TYK2 gene, or detecting a chimeric protein having an amino-terminal portion from a a TYK2-gene-fusion-partner (e.g., NPM1) and a carboxy-terminal portion from an TYK2 gene. A variety of techniques may be used for detecting the truncated TYK2 protein or chimeric protein: protein sequencing; and, immunoassay. Immunoassay techniques include immunoprecipitation, Western blot, ELISA, immunohistochemistry, immunocytochemistry, flow cytometry, and immuno-PCR.

The present invention further provides, but is not limited to, compositions and kits for detecting / diagnosing cutaneous CD30-positive lymphoproliferative disorders (e.g., lymphomatoid papulosis; primary cutaneous anaplastic large cell lymphoma) in a subject (e.g., human patient). The compositions and kits may comprise:

a single labeled probe comprising a sequence that hybridizes to the junction at which a 5' portion from a TYK2-gene-fusion-partner (e.g., NPM1) fuses to a 3' portion from a TYK2 gene;

a pair of labeled probes wherein the first labeled probe comprises a sequence that
5 hybridizes to a TYK2-gene-fusion-partner (e.g., NPM1) and the second labeled probe comprises a sequence that hybridizes to a TYK2 gene;

a pair of amplification oligonucleotides wherein the first amplification
oligonucleotide comprises a sequence that hybridizes to a TYK2-gene-fusion-partner (e.g.,
NPM1) and the second amplification oligonucleotide comprises a sequence that hybridizes to
10 a TYK2 gene;

an antibody to an amino-terminally truncated TYK2 protein resulting from a fusion of
a TYK2-gene-fusion-partner (e.g., NPM1) to a TYK2 gene; and/or,

an antibody to a chimeric protein having an amino-terminal portion from a TYK2-
gene-fusion-partner (e.g., NPM1) and a carboxy-terminal portion from a TYK2 gene.

15 In certain embodiments, the present invention provides methods for treating
cutaneous CD30-positive lymphoproliferative disorders (e.g., lymphomatoid papulosis;
primary cutaneous anaplastic large cell lymphoma) in a subject (e.g., a human patient)
comprising: administering to the patient an agent that inhibits at least one biological activity
of such a gene fusion having a 5' portion from a TYK2-gene-fusion-partner (e.g., NPM1) and
20 a 3' portion from a TYK2 gene. In some embodiments, the agent inhibits tyrosine kinase
activity resulting from the gene fusion. In some embodiments, the agent inhibits STAT1,
STAT3, and/or STAT5 activity resulting from the gene fusion. In some embodiments, the
agent inhibits increased cellular proliferation resulting from the gene fusion. In some
embodiments, the agent inhibits one or more of tyrosine kinase activity resulting from the
25 gene fusion, STAT1, STAT3, and/or STAT5 activity resulting from the gene fusion, and
increased cellular proliferation resulting from the gene fusion. In some embodiments, the
agent inhibits expression of the gene fusion. In some embodiments, the agent inhibits any
expression and/or activity resulting from the gene fusion. The agent may be a small molecule,
an siRNA, an antisense nucleic acid, or an antibody. In some embodiments, the agent is co-
30 administered with an additional agent known to treat cancer (e.g., an agent for treating
cutaneous CD30-positive lymphoproliferative disorders (e.g., lymphomatoid papulosis;
primary cutaneous anaplastic large cell lymphoma)).

In certain embodiments, the present invention provides kits for detecting gene fusions
associated with CD30-positive lymphoproliferative disorders (e.g., lymphomatoid papulosis;

primary cutaneous anaplastic large cell lymphoma) in a subject, comprising or consisting essentially of or consisting of detecting the presence of one or more gene fusions involving TYK2. In some embodiments, such kits comprise, consist essentially of, or consist of informative reagents for detecting a gene fusion having or resulting from the fusion of, rearrangement of, or genomic deletion between, a 3' portion from a TYK family member gene and a 5' portion from a TYK2-gene-fusion-partner (e.g., NPM1), wherein detecting the presence in a sample (e.g., a sample from a human subject) (e.g., biopsy tissue, tissue, blood, plasma, serum, urine, stool, and/or cells) of the gene fusion between a TYK2 gene and a TYK2-gene-fusion-partner (e.g., NPM1) identifies or characterizes cutaneous CD30-positive lymphoproliferative disorders (e.g., lymphomatoid papulosis; primary cutaneous anaplastic large cell lymphoma).

Such embodiments are not limited to a particular reagent(s). Examples include but are not limited to, a probe that specifically hybridizes to the fusion junction of a TYK2 gene and a TYK2-gene-fusion-partner (e.g., NPM1), a pair of primers that amplify a fusion junction of a gene fusion between a TYK2 gene and a TYK2-gene-fusion-partner (e.g., NPM1) (e.g., a first primer that hybridizes to a TYK2 nucleic acid and second primer that hybridizes to nucleic acid from a TYK2-gene-fusion-partner (e.g., NPM1)), an antibody that binds to the fusion junction of a fusion polypeptide of TYK2 and a TYK2-gene-fusion-partner (e.g., NPM1) (e.g., a NPM1-TYK2 fusion polypeptide), a sequencing primer that binds to a fusion between TYK2 and a TYK2-gene-fusion-partner (e.g., NPM1) (e.g., a NPM1-TYK2 fusion polypeptide) and generates an extension product that spans the fusion junction of said gene fusion between TYK2 and a TYK2-gene-fusion-partner (e.g., NPM1) (e.g., a NPM1-TYK2 fusion polypeptide), or a pair of probes wherein the first probe hybridizes to TYK2 gene and the second probe hybridizes to a TYK2-gene-fusion-partner (e.g., NPM1).

Further embodiments of the present invention provide uses and methods for characterizing / detecting / identifying cutaneous CD30-positive lymphoproliferative disorders (e.g., lymphomatoid papulosis; primary cutaneous anaplastic large cell lymphoma) in a subject (e.g., human subject) using the aforementioned kits or other components.

For example, in some embodiments, the present invention provides a method for characterizing / detecting / identifying cutaneous CD30-positive lymphoproliferative disorders (e.g., lymphomatoid papulosis; primary cutaneous anaplastic large cell lymphoma) in a subject (e.g., human subject), comprising: (a) contacting a biological sample from a subject (e.g., biopsy tissue, tissue, blood, plasma, serum, urine, stool, and/or cells) with a nucleic acid or polypeptide detection assay comprising at least a first gene fusion informative

reagent for identification of a NPM1-TYK2 gene fusion under conditions that the presence of such a gene fusion is detected; and (b) characterizing / detecting / identifying cutaneous CD30-positive lymphoproliferative disorders (e.g., lymphomatoid papulosis; primary cutaneous anaplastic large cell lymphoma) in the subject when the NPM1-TYK2 gene fusion
5 is present in the sample.

In some embodiments, step (a) comprises detecting chromosomal rearrangements of genomic DNA having a 5' portion from NPM1 and a 3' portion from TYK2. In some embodiments, step (a) comprises detecting chromosomal rearrangements of genomic DNA using a nucleic acid sequencing technique. In some embodiments, step (a) comprises
10 detecting chromosomal rearrangements of genomic DNA using a nucleic acid hybridization technique. In some embodiments, step (a) comprises detecting chromosomal rearrangements of genomic DNA using a nucleic acid hybridization technique (e.g., including but not limited to, in situ hybridization (ISH), microarray or Southern blot). In some embodiments, step (a) comprises detecting chromosomal rearrangements of genomic DNA using a nucleic acid
15 amplification method (e.g., including but not limited to, polymerase chain reaction (PCR), reverse transcription polymerase chain reaction (RT-PCR), transcription-mediated amplification (TMA), ligase chain reaction (LCR), strand displacement amplification (SDA), or nucleic acid sequence based amplification (NASBA)). In some embodiments, step (a) comprises detecting chimeric mRNA transcripts having 5' portion from NPM1 and a 3'
20 portion from TYK2.

In some embodiments, step (a) comprises detecting chimeric mRNA transcripts using a nucleic acid sequencing technique. In some embodiments, step (a) comprises detecting chimeric mRNA transcripts using a nucleic acid hybridization technique. In some
25 embodiments, step (a) comprises detecting chimeric mRNA transcripts using a nucleic acid hybridization technique (e.g., including but not limited to, in situ hybridization (ISH), microarray or Northern blot). In some embodiments, step (a) comprises detecting chimeric mRNA transcripts using a nucleic acid amplification method (e.g., including but not limited to, polymerase chain reaction (PCR), reverse transcription polymerase chain reaction (RT-PCR), transcription-mediated amplification (TMA), ligase chain reaction (LCR), strand
30 displacement amplification (SDA), or nucleic acid sequence based amplification (NASBA)).

In some embodiments, the method further comprises the step of collecting the sample from the subject. In some embodiments, the sample is, for example, biopsy tissue, tissue, blood, plasma, serum, urine, stool, and/or cells.

In some embodiments, the described methods further comprise the step of detecting the level of expression of the gene fusion in the sample, wherein detecting an enhanced level of expression of the gene fusion in the patient sample relative to the level of expression of the gene fusion in a normal sample (e.g., relative to the level in normal cells, increase or decrease in level relative to a prior time point, increase or decrease relative to a pre-established threshold level, etc.) indicates the presence of neoplastic cells or cells predisposed to the onset of a neoplastic state in the sample.

Further embodiments provide the step of determining a treatment course of action based on the presence or absence of the gene fusion. For example, in some embodiments, the treatment course of action comprises administration of an inhibitor of activity resulting from such a gene fusion (e.g., NPM1-TYK2) (e.g., tyrosine kinase activity resulting from the gene fusion, STAT1, STAT3, and/or STAT5 activity resulting from the gene fusion, and increased cellular proliferation resulting from the gene fusion) when the gene fusion is present in the sample. In some embodiments, the agent inhibits expression of the gene fusion. In some embodiments, the agent inhibits any expression and/or activity resulting from the gene fusion. The agent may be a small molecule, an siRNA, an antisense nucleic acid, or an antibody.

In certain embodiments, the present invention provides a method for diagnosing a neoplastic disorder or susceptibility to a neoplastic disorder comprising detecting the presence of a NPM1-TYK2 gene fusion in a sample from a subject (e.g., a human patient), wherein the NPM1-TYK2 gene fusion is a fusion between chromosome 5, exon 9 of NPM1 and chromosome 19, exon 15 of TYK2 (e.g., N5;T19), and diagnosing the subject as having or being susceptible to a neoplastic disorder when the gene fusion is present. In some embodiments, the neoplastic disorder is a cutaneous CD30-positive lymphoproliferative disorder (e.g., lymphomatoid papulosis; primary cutaneous anaplastic large cell lymphoma).

In certain embodiments, the present invention provides a method of determining the NPM1-TYK2 gene fusion status of a human by: (a) determining the presence or absence of the N5;T19 gene fusion in both alleles of the of the NPM1-TYK2 fusion gene of the human in a nucleic acid sample obtained from the human, and (b) identifying the human (i) as being homozygous for N5;T19 gene fusion in the NPM1-TYK2 fusion gene when one of the gene fusions is present in both alleles of the NPM1-TYK2 fusion gene, or (ii) as being heterozygous for the N5;T19 gene fusion in the NPM1-TYK2 fusion gene when one of the gene fusions is present in one of the alleles of the NPM1-TYK2 fusion gene, or (iii) as having no alteration in the NPM1-TYK2 fusion gene caused by the N5;T19 gene fusion when each of the gene fusions is absent from both alleles of the NPM1-TYK2 fusion gene.

In some embodiments, the NPM1-TYK2 gene fusion is N5;T19. An NPM1-TYK2 gene fusion may be detected by amplifying any of SEQ ID NOs: 1, 2, 3, or 4, or diagnostic fragments thereof. In one embodiment, the method includes detecting one or more additional NPM1-TYK2 gene fusions in a sample from the subject.

5 The NPM1-TYK2 fusion may be detected by assessing sample nucleic acid by PCR, RT-PCR, and/or nucleic acid hybridization. In one embodiment, the sample is amplified by reverse transcriptase polymerase chain reaction (RT-PCR). In one embodiment, the amplifying employs a detectably labeled primer. In one embodiment, the detecting is accomplished with electrophoresis. In one embodiment, the detecting is accomplished using a
10 real-time PCR-based detection system, such as TaqMan®.

The present invention also provides a method for diagnosing a neoplastic disorder or susceptibility to a neoplastic disorder by detecting the presence or absence of an NPM1-TYK2 fusion protein in a sample from a subject, wherein the NPM1-TYK2 fusion protein is a
15 fusion between exon 9 of NPM1 and exon 15 of ALK, and diagnosing the subject has having or being susceptible to a neoplastic disorder when the fusion protein is present.

The invention also provides kits for detecting a NPM1-TYK2 fusion mutations which include one or more oligonucleotides (e.g., a primer) for amplifying a fragment of a nucleic acid sample which contains the N5;T19 mutation, if present. Optionally, the kit further contains one or more mutation-specific oligonucleotide probes.

20 Additional embodiments of the present invention are provided in the description and examples below.

DESCRIPTION OF THE FIGURES

25 FIG. 1: FISH Results and Clinical Data for Cutaneous CD30-Positive Lymphoproliferative Disorders.

FIG. 2A-F: Identification of TYK2 translocations in cutaneous T cell lymphoma-derived cell line and primary CD30 positive LPD patient samples. (A) Read support from RNA sequencing illustrating the forward fragment reads spanning the breakpoint between 5' NPM1 (NM_002520 c.1016) component and 3' TYK2 (NM_003331 c.2554) component and
30 confirmation by Sanger sequencing. (B) Protein domain and exon diagram illustrating the preservation of N-terminal oligomerization and histone and DNA/RNA binding domains of NPM1 and the C-terminal kinase domain of TYK2 in the NPM1-TYK2 fusion protein. (C) Cloning of the genomic breakpoint of the t(5;19) fusion event at chr5:170,832,813 and

chr19:10,469,815 and confirmation by Sanger sequencing. (D) The NPM1-TYK2 fusion joins the positive strand of the NPM1 locus up to and including exon 9, with the inverted negative strand of the TYK2 locus between exons 15 and 16. (E) FISH studies. A TYK2 break-apart assay shows a TYK2 rearrangement in the MyLa cell line and in a primary cutaneous CD30-positive LPD (upper panel). NPM1-TYK2 fusion FISH reveals that NPM1 is the partner gene in both. A case that is negative for TYK2 translocation is also illustrated (lower panel). (F) NPM1-TYK2 fusion probe FISH assay shows the occurrence of t(5;19) translocation in a cutaneous ALCL primary sample.

FIG. 3: Summary of RNAseq data for MyLa.

FIG. 4: SYBR Green- based qRT-PCR validation of NPM1-TYK2 fusion transcript expression in MyLa and other hematological cell lines.

FIG. 5: Conventional gel based RT-PCR validation of NPM1-TYK2 fusion transcript expression in MyLa and other hematological cell lines.

FIG. 6: Agarose gel electrophoresis of PCR products for genomic breakpoint mapping from NPM1-TYK2 positive MyLa and the negative control HH. The cartoon depicts the location of genomic breakpoint in NPM1 and TYK2 genes and their juxtaposition. The bottom panel shows the Sanger sequencing of the PCR product which shows the exact genomic fusion point.

FIG. 7A-F: Oncogenic potential of NPM1-TYK2 fusion gene product. (A) Presence of the NPM1-TYK2 fusion protein in MyLa and not in other T-cell lines. Individual NPM1 and TYK2 western blot assays show shift in the size of the protein in MyLa as a result of the fusion. (B) Hyperactivation of TYK2 and elevated STAT pathway activation in MyLa cells endogenously expressing NPM1-TYK2 fusion protein as compared with other T-cell lines. The Mac1 cell line that has a JAK2 rearrangement is used as a positive control. HUT78 cell line exhibits constitutive STAT5 activation. (C) Ectopic expression of NPM1-TYK2 fusion protein in HEK293T cells reveals activation of STAT proteins in Western blot assays. Note the significantly reduced levels of STAT activation in cells expressing NPM1-TYK2 kinase-defective mutant K462R. (D) Exogenously expressed NPM1-TYK2 fusion protein in HEK293FT cells leads to transcriptional activation of STAT1/3/5. Cells expressing kinase-defective mutant K462R NPM1-TYK2 fusion protein show reduced levels of STAT activation indicating a specific effect of TYK2 kinase activity on downstream STAT activation. (E) Diminished STAT pathway activation following knockdown of TYK2 protein in MyLa cell line by shRNA knockdown. (F) ShRNA-mediated silencing of TYK2 reduces proliferation of MyLa cells demonstrating oncogenic potential of NPM1-TYK2.

FIG. 8: Immunohistochemical staining for NPM1 in primary patient samples. In the absence of NPM1-TYK2, NPM1 expression is restricted to the nucleus (left panel). In a patient with NPM1-TYK2 fusion, there is atypical cytoplasmic localization of NPM1 (right panel). (Anti-NPM1, 500x magnification).

5 FIG. 9: Immunofluorescence staining demonstrates a marked expression of cytoplasmic and nuclear phosphorylated STAT 1/3/5 in NPM-TYK2 positive MyLa cell line as compared to fusion negative HH cell line.

FIG. 10: Immunohistochemical staining for pSTAT5 in primary patient samples. In the absence of NPM1-TYK2, there is weak expression of pSTAT5, predominantly limited to the cytoplasm (left panel). In a patient with NPM1-TYK2 fusion, expression of pSTAT5 is stronger and increased nuclear expression is present (right panel). (Anti-pSTAT5, 400x magnification).

FIG. 11A-E: A) wild type cDNA sequence for TYK2; B) wild type ORF nucleic acid sequence for TYK2; C) wild type amino acid sequence for TYK2; D) wild type nucleic acid sequence for TYK2; and E) TYK2 nucleic acid intron involved in TYK2 translocation.

FIG. 12A-E: A) wild type cDNA sequence for NPM1; B) wild type ORF nucleic acid sequence for NPM1; C) wild type amino acid sequence for NPM1; D) wild type nucleic acid sequence for NPM1; and E) NPM1 nucleic acid intron involved in NPM1 translocation.

FIG. 13A-D: A) cDNA sequence for NPM1-TYK2 gene fusion; B) coding region for NPM1-TYK2 gene fusion; C) amino acid sequence for NPM1-TYK2 gene fusion product; and D) NPM1-TYK2 introns involved in translocation.

DEFINITIONS

To facilitate an understanding of the present invention, a number of terms and phrases are defined below:

As used herein, the term "gene fusion" refers to a chimeric genomic DNA, a chimeric messenger RNA, a truncated protein or a chimeric protein resulting from the fusion of at least a portion of a first gene to at least a portion of a second gene. The gene fusion need not include entire genes or exons of genes.

30 As used herein, the term "sensitivity" is defined as a statistical measure of performance of an assay (e.g., method, test), calculated by dividing the number of true positives by the sum of the true positives and the false negatives.

As used herein, the term “specificity” is defined as a statistical measure of performance of an assay (e.g., method, test), calculated by dividing the number of true negatives by the sum of true negatives and false positives.

As used herein, the term “informative” or “informativeness” refers to a quality of a
5 marker or panel of markers, and specifically to the likelihood of finding a marker (or panel of markers) in a positive sample.

As used herein, "TYK2" refers to a member of the Janus Kinase (JAK) family of tyrosine kinases. An example of TYK2 polypeptide is the human polypeptide sequence described in Genbank accession number P29597.3. An exemplary TYK2 coding sequence is
10 the human sequence described in Genbank accession number X54637. TYK2 also includes species homologs, splice variants, polymorphic variants, and conservatively modified variants of TYK2 that retain TYK2 tyrosine kinase activity. Fig. 11 provides: A) human wild type cDNA sequence for TYK2; B) human wild type ORF nucleic acid sequence for TYK2; C) human wild type amino acid sequence for TYK2; D) human wild type nucleic acid
15 sequence for TYK2; and E) human TYK2 nucleic acid intron involved in TYK2 translocation.

As used herein, "NPM1" refers to a gene which encodes a phosphoprotein which moves between the nucleus and the cytoplasm. The gene product is thought to be involved in several processes including regulation of the ARF/p53 pathway. An exemplary NPM1 coding
20 sequence is the human sequence described in Genbank accession number P06748. NPM1 also includes species homologs, splice variants, polymorphic variants, and conservatively modified variants of NPM1 that retain its wild type activity. Fig. 12 provides: A) human wild type cDNA sequence for NPM1; B) human wild type ORF nucleic acid sequence for NPM1; C) human wild type amino acid sequence for NPM1; D) human wild type nucleic acid
25 sequence for NPM1; and E) human NPM1 nucleic acid intron involved in NPM1 translocation.

As used herein, “NPM1-TYK2” refers to a chimeric genomic DNA, a chimeric messenger RNA, a truncated protein or a chimeric protein resulting from the fusion of at least a portion of an NPM1 gene to at least a portion of a TYK2 gene. The gene fusion need not
30 include entire genes or exons of genes. Fig. 13 provides: A) cDNA sequence for NPM1-TYK2 gene fusion; B) coding region for NPM1-TYK2 gene fusion; C) amino acid sequence for NPM1-TYK2 gene fusion product; and D) NPM1-TYK2 introns involved in translocation.

As used herein, the terms "NPM1/TYK2 informative reagent" refers to a reagent or reagents that are informative for identification of gene fusions described herein. In some embodiments, reagents are primers, probes or antibodies for detection of NPM1-TYKE2 gene fusions described herein.

5 As used herein, the term "transcriptional regulatory region" refers to the non-coding upstream regulatory sequence of a gene, also called the 5' untranslated region (5'UTR).

As used herein, the terms "detect", "detecting", or "detection" may describe either the general act of discovering or discerning or the specific observation of a detectably labeled composition.

10 As used herein, the term "inhibits at least one biological activity of a gene fusion" refers to any agent that decreases any activity of a gene fusion of the present invention (e.g., including, but not limited to, the activities described herein), via directly contacting gene fusion protein, contacting gene fusion mRNA or genomic DNA, causing conformational changes of gene fusion polypeptides, decreasing gene fusion protein levels, or interfering
15 with gene fusion interactions with signaling partners, and affecting the expression of gene fusion target genes. Inhibitors also include molecules that indirectly regulate gene fusion biological activity by intercepting upstream signaling molecules.

As used herein, the term "siRNAs" refers to small interfering RNAs. In some embodiments, siRNAs comprise a duplex, or double-stranded region, of about 18-25
20 nucleotides long; often siRNAs contain from about two to four unpaired nucleotides at the 3' end of each strand. At least one strand of the duplex or double-stranded region of a siRNA is substantially homologous to, or substantially complementary to, a target RNA molecule. The strand complementary to a target RNA molecule is the "antisense strand;" the strand homologous to the target RNA molecule is the "sense strand," and is also complementary to
25 the siRNA antisense strand. siRNAs may also contain additional sequences; non-limiting examples of such sequences include linking sequences, or loops, as well as stem and other folded structures. siRNAs appear to function as key intermediaries in triggering RNA interference in invertebrates and in vertebrates, and in triggering sequence-specific RNA degradation during posttranscriptional gene silencing in plants.

30 The term "RNA interference" or "RNAi" refers to the silencing or decreasing of gene expression by siRNAs. It is the process of sequence-specific, post-transcriptional gene silencing in animals and plants, initiated by siRNA that is homologous in its duplex region to the sequence of the silenced gene. The gene may be endogenous or exogenous to the organism, present integrated into a chromosome or present in a transfection vector that is not

integrated into the genome. The expression of the gene is either completely or partially inhibited. RNAi may also be considered to inhibit the function of a target RNA; the function of the target RNA may be complete or partial.

As used herein, the term "gene transfer system" refers to any means of delivering a composition comprising a nucleic acid sequence to a cell or tissue. For example, gene transfer systems include, but are not limited to, vectors (e.g., retroviral, adenoviral, adeno-associated viral, and other nucleic acid-based delivery systems), microinjection of naked nucleic acid, polymer-based delivery systems (e.g., liposome-based and metallic particle-based systems), biolistic injection, and the like. As used herein, the term "viral gene transfer system" refers to gene transfer systems comprising viral elements (e.g., intact viruses, modified viruses and viral components such as nucleic acids or proteins) to facilitate delivery of the sample to a desired cell or tissue. As used herein, the term "adenovirus gene transfer system" refers to gene transfer systems comprising intact or altered viruses belonging to the family Adenoviridae.

As used herein, the term "site-specific recombination target sequences" refers to nucleic acid sequences that provide recognition sequences for recombination factors and the location where recombination takes place.

As used herein, the term "nucleic acid molecule" refers to any nucleic acid containing molecule, including but not limited to, DNA or RNA. The term encompasses sequences that include any of the known base analogs of DNA and RNA including, but not limited to, 4-acetylcytosine, 8-hydroxy-N6-methyladenosine, aziridinylcytosine, pseudoisocytosine, 5-(carboxyhydroxymethyl) uracil, 5-fluorouracil, 5-bromouracil, 5-carboxymethylaminomethyl-2-thiouracil, 5-carboxymethylaminomethyluracil, dihydrouracil, inosine, N6-isopentenyladenine, 1-methyladenine, 1-methylpseudouracil, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-methyladenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarbonylmethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid, oxybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, N-uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid, pseudouracil, queosine, 2-thiocytosine, and 2,6-diaminopurine.

The term "gene" refers to a nucleic acid (e.g., DNA) sequence that comprises coding sequences necessary for the production of a polypeptide, precursor, or RNA (e.g., rRNA,

tRNA). The polypeptide can be encoded by a full length coding sequence or by any portion of the coding sequence so long as the desired activity or functional properties (e.g., enzymatic activity, ligand binding, signal transduction, immunogenicity, etc.) of the full-length or fragment is retained. The term also encompasses the coding region of a structural gene and the sequences located adjacent to the coding region on both the 5' and 3' ends for a distance of about 1 kb or more on either end such that the gene corresponds to the length of the full-length mRNA. Sequences located 5' of the coding region and present on the mRNA are referred to as 5' non-translated sequences. Sequences located 3' or downstream of the coding region and present on the mRNA are referred to as 3' non-translated sequences. The term "gene" encompasses both cDNA and genomic forms of a gene. A genomic form or clone of a gene contains the coding region interrupted with non-coding sequences termed "introns" or "intervening regions" or "intervening sequences." Introns are segments of a gene that are transcribed into nuclear RNA (hnRNA); introns may contain regulatory elements such as enhancers. Introns are removed or "spliced out" from the nuclear or primary transcript; introns therefore are absent in the messenger RNA (mRNA) transcript. The mRNA functions during translation to specify the sequence or order of amino acids in a nascent polypeptide.

As used herein, the term "heterologous gene" refers to a gene that is not in its natural environment. For example, a heterologous gene includes a gene from one species introduced into another species. A heterologous gene also includes a gene native to an organism that has been altered in some way (e.g., mutated, added in multiple copies, linked to non-native regulatory sequences, etc). Heterologous genes are distinguished from endogenous genes in that the heterologous gene sequences are typically joined to DNA sequences that are not found naturally associated with the gene sequences in the chromosome or are associated with portions of the chromosome not found in nature (e.g., genes expressed in loci where the gene is not normally expressed).

As used herein, the term "oligonucleotide," refers to a short length of single-stranded polynucleotide chain. Oligonucleotides are typically less than 200 residues long (e.g., between 15 and 100), however, as used herein, the term is also intended to encompass longer polynucleotide chains. Oligonucleotides are often referred to by their length. For example a 24 residue oligonucleotide is referred to as a "24-mer". Oligonucleotides can form secondary and tertiary structures by self-hybridizing or by hybridizing to other polynucleotides. Such structures can include, but are not limited to, duplexes, hairpins, cruciforms, bends, and triplexes.

As used herein, the terms "complementary" or "complementarity" are used in reference to polynucleotides (i.e., a sequence of nucleotides) related by the base-pairing rules. For example, the sequence "5'-A-G-T-3'," is complementary to the sequence "3'-T-C-A-5'." Complementarity may be "partial," in which only some of the nucleic acids' bases are
5 matched according to the base pairing rules. Or, there may be "complete" or "total" complementarity between the nucleic acids. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, as well as detection methods that depend upon binding between nucleic acids.

10 The term "homology" refers to a degree of complementarity. There may be partial homology or complete homology (i.e., identity). A partially complementary sequence is a nucleic acid molecule that at least partially inhibits a completely complementary nucleic acid molecule from hybridizing to a target nucleic acid is "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence
15 may be examined using a hybridization assay (Southern or Northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or probe will compete for and inhibit the binding (i.e., the hybridization) of a completely homologous nucleic acid molecule to a target under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific binding is
20 permitted; low stringency conditions require that the binding of two sequences to one another be a specific (i.e., selective) interaction. The absence of non-specific binding may be tested by the use of a second target that is substantially non-complementary (e.g., less than about 30% identity); in the absence of non-specific binding the probe will not hybridize to the second non-complementary target.

25 When used in reference to a double-stranded nucleic acid sequence such as a cDNA or genomic clone, the term "substantially homologous" refers to any probe that can hybridize to either or both strands of the double-stranded nucleic acid sequence under conditions of low stringency as described above.

A gene may produce multiple RNA species that are generated by differential splicing
30 of the primary RNA transcript. cDNAs that are splice variants of the same gene will contain regions of sequence identity or complete homology (representing the presence of the same exon or portion of the same exon on both cDNAs) and regions of complete non-identity (for example, representing the presence of exon "A" on cDNA 1 wherein cDNA 2 contains exon "B" instead). Because the two cDNAs contain regions of sequence identity they will both

hybridize to a probe derived from the entire gene or portions of the gene containing sequences found on both cDNAs; the two splice variants are therefore substantially homologous to such a probe and to each other.

When used in reference to a single-stranded nucleic acid sequence, the term "substantially homologous" refers to any probe that can hybridize (i.e., it is the complement of) the single-stranded nucleic acid sequence under conditions of low stringency as described above.

As used herein, the term "hybridization" is used in reference to the pairing of complementary nucleic acids. Hybridization and the strength of hybridization (i.e., the strength of the association between the nucleic acids) is impacted by such factors as the degree of complementary between the nucleic acids, stringency of the conditions involved, the T_m of the formed hybrid, and the G:C ratio within the nucleic acids. A single molecule that contains pairing of complementary nucleic acids within its structure is said to be "self-hybridized."

As used herein the term "stringency" is used in reference to the conditions of temperature, ionic strength, and the presence of other compounds such as organic solvents, under which nucleic acid hybridizations are conducted. Under "low stringency conditions" a nucleic acid sequence of interest will hybridize to its exact complement, sequences with single base mismatches, closely related sequences (e.g., sequences with 90% or greater homology), and sequences having only partial homology (e.g., sequences with 50-90% homology). Under "medium stringency conditions," a nucleic acid sequence of interest will hybridize only to its exact complement, sequences with single base mismatches, and closely related sequences (e.g., 90% or greater homology). Under "high stringency conditions," a nucleic acid sequence of interest will hybridize only to its exact complement, and (depending on conditions such as temperature) sequences with single base mismatches. In other words, under conditions of high stringency the temperature can be raised so as to exclude hybridization to sequences with single base mismatches.

"High stringency conditions" when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 42°C in a solution consisting of 5X SSPE (43.8 g/l NaCl, 6.9 g/l NaH_2PO_4 H_2O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.5% SDS, 5X Denhardt's reagent and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 0.1X SSPE, 1.0% SDS at 42°C when a probe of about 500 nucleotides in length is employed.

"Medium stringency conditions" when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 42°C in a solution consisting of 5X SSPE (43.8 g/l NaCl, 6.9 g/l NaH₂PO₄ H₂O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.5% SDS, 5X Denhardt's reagent and 100 µg/ml denatured salmon sperm DNA
5 followed by washing in a solution comprising 1.0X SSPE, 1.0% SDS at 42°C when a probe of about 500 nucleotides in length is employed.

"Low stringency conditions" comprise conditions equivalent to binding or hybridization at 42°C in a solution consisting of 5X SSPE (43.8 g/l NaCl, 6.9 g/l NaH₂PO₄ H₂O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.1% SDS, 5X Denhardt's reagent
10 [50X Denhardt's contains per 500 ml: 5 g Ficoll (Type 400, Pharamcia), 5 g BSA (Fraction V; Sigma)] and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 5X SSPE, 0.1% SDS at 42°C when a probe of about 500 nucleotides in length is employed.

Numerous equivalent conditions may be employed to comprise low stringency
15 conditions; factors such as the length and nature (DNA, RNA, base composition) of the probe and nature of the target (DNA, RNA, base composition, present in solution or immobilized, etc.) and the concentration of the salts and other components (e.g., the presence or absence of formamide, dextran sulfate, polyethylene glycol) are considered and the hybridization solution may be varied to generate conditions of low stringency hybridization different from,
20 but equivalent to, the above listed conditions. In addition, the conditions that promote hybridization under conditions of high stringency (e.g., increasing the temperature of the hybridization and/or wash steps, the use of formamide in the hybridization solution, etc.) (see definition above for "stringency") can be utilized.

As used herein, the term "amplification oligonucleotide" refers to an oligonucleotide
25 that hybridizes to a target nucleic acid, or its complement, and participates in a nucleic acid amplification reaction. An example of an amplification oligonucleotide is a "primer" that hybridizes to a template nucleic acid and contains a 3' OH end that is extended by a polymerase in an amplification process. Another example of an amplification oligonucleotide is an oligonucleotide that is not extended by a polymerase (e.g., because it
30 has a 3' blocked end) but participates in or facilitates amplification. Amplification oligonucleotides may optionally include modified nucleotides or analogs, or additional nucleotides that participate in an amplification reaction but are not complementary to or contained in the target nucleic acid. Amplification oligonucleotides may contain a sequence

that is not complementary to the target or template sequence. For example, the 5' region of a primer may include a promoter sequence that is non-complementary to the target nucleic acid (referred to as a "promoter-primer"). Those skilled in the art will understand that an amplification oligonucleotide that functions as a primer may be modified to include a 5' promoter sequence, and thus function as a promoter-primer. Similarly, a promoter-primer may be modified by removal of, or synthesis without, a promoter sequence and still function as a primer. A 3' blocked amplification oligonucleotide may provide a promoter sequence and serve as a template for polymerization (referred to as a "promoter-provider").

As used herein, the term "primer" refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, that is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product that is complementary to a nucleic acid strand is induced, (i.e., in the presence of nucleotides and an inducing agent such as DNA polymerase and at a suitable temperature and pH). The primer is preferably single stranded for maximum efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the inducing agent. The exact lengths of the primers will depend on many factors, including temperature, source of primer and the use of the method.

As used herein, the term "probe" refers to an oligonucleotide (i.e., a sequence of nucleotides), whether occurring naturally as in a purified restriction digest or produced synthetically, recombinantly or by PCR amplification, that is capable of hybridizing to at least a portion of another oligonucleotide of interest. A probe may be single-stranded or double-stranded. Probes are useful in the detection, identification and isolation of particular gene sequences. It is contemplated that any probe used in the present invention will be labeled with any "reporter molecule," so that is detectable in any detection system, including, but not limited to enzyme (e.g., ELISA, as well as enzyme-based histochemical assays), fluorescent, radioactive, and luminescent systems. It is not intended that the present invention be limited to any particular detection system or label.

The term "isolated" when used in relation to a nucleic acid, as in "an isolated oligonucleotide" or "isolated polynucleotide" refers to a nucleic acid sequence that is identified and separated from at least one component or contaminant with which it is ordinarily associated in its natural source. Isolated nucleic acid is such present in a form or

setting that is different from that in which it is found in nature. In contrast, non-isolated nucleic acids as nucleic acids such as DNA and RNA found in the state they exist in nature. For example, a given DNA sequence (e.g., a gene) is found on the host cell chromosome in proximity to neighboring genes; RNA sequences, such as a specific mRNA sequence
5 encoding a specific protein, are found in the cell as a mixture with numerous other mRNAs that encode a multitude of proteins. However, isolated nucleic acid encoding a given protein includes, by way of example, such nucleic acid in cells ordinarily expressing the given protein where the nucleic acid is in a chromosomal location different from that of natural cells, or is otherwise flanked by a different nucleic acid sequence than that found in nature.

10 The isolated nucleic acid, oligonucleotide, or polynucleotide may be present in single-stranded or double-stranded form. When an isolated nucleic acid, oligonucleotide or polynucleotide is to be utilized to express a protein, the oligonucleotide or polynucleotide will contain at a minimum the sense or coding strand (i.e., the oligonucleotide or polynucleotide may be single-stranded), but may contain both the sense and anti-sense
15 strands (i.e., the oligonucleotide or polynucleotide may be double-stranded).

As used herein, the term "purified" or "to purify" refers to the removal of components (e.g., contaminants) from a sample. For example, antibodies are purified by removal of contaminating non-immunoglobulin proteins; they are also purified by the removal of immunoglobulin that does not bind to the target molecule. The removal of non-
20 immunoglobulin proteins and/or the removal of immunoglobulins that do not bind to the target molecule results in an increase in the percent of target-reactive immunoglobulins in the sample. In another example, recombinant polypeptides are expressed in bacterial host cells and the polypeptides are purified by the removal of host cell proteins; the percent of recombinant polypeptides is thereby increased in the sample.

25

DETAILED DESCRIPTION OF THE INVENTION

Recurrent chromosomal translocations frequently underlie the pathogenesis of several hematopoietic malignancies and often define molecular subtypes with distinct biological behavior (see, e.g., Rowley JD, Nature reviews Cancer 2001;1:245-50; Mitelman F, et al.,
30 Nature reviews Cancer 2007;7:233-45). Frequently, these translocations target tyrosine kinases resulting in constitutive activation and promotion of oncogenesis (see, e.g., Greuber EK, et al., Nature reviews Cancer 2013;13:559-71). Cutaneous CD30-positive LPD represents a clinicopathologic spectrum including lymphomatoid papulosis (LYP) and primary cutaneous anaplastic large cell lymphoma (ALCL) (see, e.g., Swerdlow SH, Campo,

E., Harris, N.L., Jaffe, E.S., Pileri, S.A., Stein, H., Thiele, J., Vardiman, J.W., ed. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues. Lyon: IARC; 2008). Gene fusions targeting tyrosine kinases underlying the pathogenesis of CD30-positive LPD have not been described.

Experiments conducted during the course of developing embodiments for the present invention identified a novel recurrent NPM1-TYK2 gene fusion in CD30-positive LPD. Functional support for NPM1-TYK2 in mediating activation of STAT1/3/5 signaling to promote cell proliferation was provided. Functional inactivation of TYK2 was shown to significantly diminish proliferation indicating TYK2 as an oncogenic driver kinase.

Additionally, FISH studies provided evidence that TYK2 is targeted by translocation partners other than NPM1. Taken together, such results indicate TYK2 as a therapeutic target in the treatment of CD30-positive LPDs.

Accordingly, provided herein are kits, compositions and methods for cancer diagnosis, research and therapy, including but not limited to, cancer markers. In particular, the present invention relates to recurrent gene fusions (e.g., recurrent translocations involving TYK2) as diagnostic markers and clinical targets for cutaneous CD30-positive lymphoproliferative disorders (e.g., lymphomatoid papulosis; primary cutaneous anaplastic large cell lymphoma).

I. Gene Fusions

Embodiments of the present invention provide diagnostic, screening, research, and therapeutic methods for detecting, diagnosing and characterizing cancer (e.g., cutaneous CD30-positive lymphoproliferative disorders (e.g., lymphomatoid papulosis; primary cutaneous anaplastic large cell lymphoma)) based on the presence of gene fusions including TYK2 and a TYK2-gene-fusion-partner in a sample. In some embodiments, the gene fusion comprises an intact catalytic domain in TYK2, although other regions are specifically encompassed by embodiments of the present invention.

Such embodiments are not limited to a particular portion of a TYK2 gene.

In some embodiments, the portion of the TYK2 gene comprises any portion of TYK2 enables tyrosine kinase activity upon expression as a polypeptide. In some embodiments, the portion of the TYK2 gene comprises at least a portion (e.g., 1%, 10%, 25%, 50%, 75%, 80%, 95%, 99%, 100%) of the C-terminal kinase domain of TYK2. In some embodiments, the portion of the TYK2 gene comprises the entire C-terminal kinase domain of TYK2. In some embodiments, the portion of the TYK2 gene comprises at least a portion of the TYK2

pseudokinase domain. In some embodiments, the portion of the TYK2 gene comprises at least a portion (e.g., 1%, 10%, 25%, 50%, 75%, 80%, 95%, 99%, 100%) of the C-terminal kinase domain of TYK2 and/or at least a portion (e.g., 1%, 10%, 25%, 50%, 75%, 80%, 95%, 99%, 100%) of the TYK2 pseudokinase domain.

5 In some embodiments, the gene fusion is a NPM1-TYK2 gene fusion. In some embodiments, gene fusions comprise the oligomerization domain of NPM1 and an intact catalytic domain in TYK2, although other regions are specifically encompassed by embodiments of the present invention.

Such embodiments are not limited to a particular TYK2-gene-fusion-partner.

10 In some embodiments, the TYK2-gene-fusion-partner is any gene such that upon fusion with TYK2 the resulting gene fusion encodes a polypeptide capable of having constitutively activated tyrosine kinase activity when expressed in a cell.

 In some embodiments, the TYK2-gene-fusion-partner is any gene such that upon fusion with TYK2 the resulting gene fusion encodes a polypeptide that when expressed in a cell is capable of constitutively activating downstream targets of TYK2 (e.g., STAT1, STAT3, STAT5) within the cell.

 In some embodiments, the TYK2-gene-fusion-partner is any gene such that upon fusion with TYK2 the resulting gene fusion encodes a polypeptide that when expressed in a cell results in increased cellular proliferation of that cell.

20 In some embodiments, the TYK2-gene-fusion-partner is any gene such that upon fusion with TYK2 the resulting gene fusion encodes a polypeptide, wherein 1) the polypeptide is capable of having constitutively activated tyrosine kinase activity when expressed in a cell, 2) expression of the polypeptide in a cell results in constitutive activation of downstream targets of TYK2 (e.g., STAT1, STAT3, STAT5), and 3) expression of the polypeptide in a cell results in increased cellular proliferation of the cell.

 In some embodiments, the TYK2-gene-fusion-partner is any portion of NPM1.

 In some embodiments, the TYK2-gene-fusion-partner comprises at least a portion (e.g., 1%, 10%, 25%, 50%, 75%, 80%, 95%, 99%, 100%) of the NPM1 oligomerization domain. In some embodiments, the TYK2-gene-fusion-partner comprises the entire NPM1 oligomerization domain.

30 In some embodiments, the TYK2-gene-fusion-partner comprises at least a portion (e.g., 1%, 10%, 25%, 50%, 75%, 80%, 95%, 99%, 100%) of the NPM1 histone binding domains. In some embodiments, the TYK2-gene-fusion-partner comprises the entire NPM1 histone binding domains.

In some embodiments, the TYK2-gene-fusion-partner comprises at least a portion (e.g., 1%, 10%, 25%, 50%, 75%, 80%, 95%, 99%, 100%) of the distal portion of the NPM1 DNA/RNA binding domains. In some embodiments, the TYK2-gene-fusion-partner comprises the entire distal portion of the NPM1 DNA/RNA binding domains.

- 5 In some embodiments, the TYK2-gene-fusion-partner comprises the entire NPM1 oligomerization domain, the entire NPM1 histone binding domain, and the entire distal portion of the NPM1 DNA/RNA binding domain.

In some embodiments, the NPM1-TYK2 gene fusion is a fusion between chromosome 5, exon 9 of NPM1 and chromosome 19, exon 15 of TYK2 (e.g., N5;T19).

10

II. Antibodies

The gene fusion proteins of the present invention, including fragments, derivatives and analogs thereof, may be used as immunogens to produce antibodies having use in the diagnostic, research, and therapeutic methods described below. The antibodies may be polyclonal or monoclonal, chimeric, humanized, single chain or Fab fragments. Various procedures may be used for the production and labeling of such antibodies and fragments. See, e.g., Burns, ed., Immunochemical Protocols, 3rd ed., Humana Press (2005); Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory (1988); Kozbor et al., Immunology Today 4: 72 (1983); Köhler and Milstein, Nature 256: 495 (1975).

20

III. Diagnostic Applications

The present invention provides DNA, RNA and protein based diagnostic methods that either directly or indirectly detect the gene fusions. The present invention also provides compositions and kits for diagnostic purposes.

25

The diagnostic methods of the present invention may be qualitative or quantitative. Quantitative diagnostic methods may be used, for example, to discriminate between indolent and aggressive cancers via a cutoff or threshold level. Where applicable, qualitative or quantitative diagnostic methods may also include amplification of target, signal or intermediary (e.g., a universal primer).

30

An initial assay may confirm the presence of a gene fusion but not identify the specific fusion. A secondary assay is then performed to determine the identity of the particular fusion, if desired. The second assay may use a different detection technology than the initial assay.

The gene fusions of embodiments of the present invention may be detected along with other markers in a multiplex or panel format. Markers are selected for their predictive value alone or in combination with the gene fusions. Markers for other cancers, diseases, infections, and metabolic conditions are also contemplated for inclusion in a multiplex of panel format.

The diagnostic methods of embodiments of the present invention may also be modified with reference to data correlating particular gene fusions with the stage, aggressiveness or progression of the disease or the presence or risk of metastasis. Ultimately, the information provided by the methods of the present invention will assist a physician in choosing the best course of treatment for a particular patient.

A. Sample

Any patient sample suspected of containing the gene fusions may be tested according to the methods of the present invention. By way of non-limiting examples, the sample may be tissue (e.g., a biopsy sample), blood, or a fraction thereof (e.g., plasma, serum, or cells).

In some embodiments, the patient sample typically requires preliminary processing designed to isolate or enrich the sample for the gene fusions or cells that contain the gene fusions. A variety of techniques may be used for this purpose, including but not limited: centrifugation; immunocapture; cell lysis; and, nucleic acid target capture.

B. DNA and RNA Detection

The gene fusions of the present invention may be detected as chromosomal rearrangements of genomic DNA or chimeric mRNA using a variety of nucleic acid techniques, including but not limited to: nucleic acid sequencing; nucleic acid hybridization; and, nucleic acid amplification.

1. Sequencing

Illustrative non-limiting examples of nucleic acid sequencing techniques include, but are not limited to, chain terminator (Sanger) sequencing and dye terminator sequencing.

Those of ordinary skill in the art will recognize that because RNA is less stable in the cell and more prone to nuclease attack experimentally RNA is usually reverse transcribed to DNA before sequencing.

Chain terminator sequencing uses sequence-specific termination of a DNA synthesis reaction using modified nucleotide substrates. Extension is initiated at a specific site on the

template DNA by using a short radioactive, fluorescent or other labeled, oligonucleotide primer complementary to the template at that region. The oligonucleotide primer is extended using a DNA polymerase, standard four deoxynucleotide bases, and a low concentration of one chain terminating nucleotide, most commonly a di-deoxynucleotide. This reaction is repeated in four separate tubes with each of the bases taking turns as the di-deoxynucleotide. Limited incorporation of the chain terminating nucleotide by the DNA polymerase results in a series of related DNA fragments that are terminated only at positions where that particular di-deoxynucleotide is used. For each reaction tube, the fragments are size-separated by electrophoresis in a slab polyacrylamide gel or a capillary tube filled with a viscous polymer. The sequence is determined by reading which lane produces a visualized mark from the labeled primer as you scan from the top of the gel to the bottom.

Dye terminator sequencing alternatively labels the terminators. Complete sequencing can be performed in a single reaction by labeling each of the di-deoxynucleotide chain-terminators with a separate fluorescent dye, which fluoresces at a different wavelength.

Some embodiments of the present invention utilize next generation or high-throughput sequencing. A variety of nucleic acid sequencing methods are contemplated for use in the methods of the present disclosure including, for example, chain terminator (Sanger) sequencing, dye terminator sequencing, and high-throughput sequencing methods. See, e.g., Sanger et al., *Proc. Natl. Acad. Sci. USA* 74:5463-5467 (1997); Maxam et al., *Proc. Natl. Acad. Sci. USA* 74:560-564 (1977); Drmanac, et al., *Nat. Biotechnol.* 16:54-58 (1998); Kato, *Int. J. Clin. Exp. Med.* 2:193-202 (2009); Ronaghi et al., *Anal. Biochem.* 242:84-89 (1996); Margulies et al., *Nature* 437:376-380 (2005); Ruparel et al., *Proc. Natl. Acad. Sci. USA* 102:5932-5937 (2005), and Harris et al., *Science* 320:106-109 (2008); Levene et al., *Science* 299:682-686 (2003); Korlach et al., *Proc. Natl. Acad. Sci. USA* 105:1176-1181 (2008); Branton et al., *Nat. Biotechnol.* 26(10):1146-53 (2008); Eid et al., *Science* 323:133-138 (2009); each of which is herein incorporated by reference in its entirety.

In some embodiments, sequencing technology including, but not limited to, pyrosequencing, sequencing-by-ligation, single molecule sequencing, sequence-by-synthesis (SBS), massive parallel clonal, massive parallel single molecule SBS, massive parallel single molecule real-time, massive parallel single molecule real-time nanopore technology, etc. Morozova and Marra provide a review of some such technologies in *Genomics*, 92: 255 (2008), herein incorporated by reference in its entirety. Those of ordinary skill in the art will recognize that because RNA is less stable in the cell and more prone to nuclease attack experimentally RNA is usually reverse transcribed to DNA before sequencing.

A number of DNA sequencing techniques can be used including fluorescence-based sequencing methodologies (See, e.g., Birren et al., *Genome Analysis: Analyzing DNA*, 1, Cold Spring Harbor, N.Y.; herein incorporated by reference in its entirety). In some embodiments, the technology finds use in automated sequencing techniques understood in that art. In some embodiments, the present technology finds use in parallel sequencing of partitioned amplicons (PCT Publication No: WO2006084132 to Kevin McKernan et al., herein incorporated by reference in its entirety). In some embodiments, the technology finds use in DNA sequencing by parallel oligonucleotide extension (See, e.g., U.S. Pat. No. 5,750,341 to Macevicz et al., and U.S. Pat. No. 6,306,597 to Macevicz et al., both of which are herein incorporated by reference in their entireties). Additional examples of sequencing techniques in which the technology finds use include the Church polony technology (Mitra et al., 2003, *Analytical Biochemistry* 320, 55-65; Shendure et al., 2005 *Science* 309, 1728-1732; U.S. Pat. No. 6,432,360, U.S. Pat. No. 6,485,944, U.S. Pat. No. 6,511,803; herein incorporated by reference in their entireties), the 454 picotiter pyrosequencing technology (Margulies et al., 2005 *Nature* 437, 376-380; US 20050130173; herein incorporated by reference in their entireties), the Solexa single base addition technology (Bennett et al., 2005, *Pharmacogenomics*, 6, 373-382; U.S. Pat. No. 6,787,308; U.S. Pat. No. 6,833,246; herein incorporated by reference in their entireties), the Lynx massively parallel signature sequencing technology (Brenner et al. (2000). *Nat. Biotechnol.* 18:630-634; U.S. Pat. No. 5,695,934; U.S. Pat. No. 5,714,330; herein incorporated by reference in their entireties), and the Adessi PCR colony technology (Adessi et al. (2000). *Nucleic Acid Res.* 28, E87; WO 00018957; herein incorporated by reference in its entirety).

Next-generation sequencing (NGS) methods share the common feature of massively parallel, high-throughput strategies, with the goal of lower costs in comparison to older sequencing methods (see, e.g., Voelkerding et al., *Clinical Chem.*, 55: 641-658, 2009; MacLean et al., *Nature Rev. Microbiol.*, 7: 287-296; each herein incorporated by reference in their entirety). NGS methods can be broadly divided into those that typically use template amplification and those that do not. Amplification-requiring methods include pyrosequencing commercialized by Roche as the 454 technology platforms (e.g., GS 20 and GS FLX), the Solexa platform commercialized by Illumina, and the Supported Oligonucleotide Ligation and Detection (SOLiD) platform commercialized by Applied Biosystems. Non-amplification approaches, also known as single-molecule sequencing, are exemplified by the HeliScope platform commercialized by Helicos BioSciences, and emerging platforms commercialized

by VisiGen, Oxford Nanopore Technologies Ltd., Life Technologies/Ion Torrent, and Pacific Biosciences, respectively.

In pyrosequencing (Voelkerding et al., *Clinical Chem.*, 55: 641-658, 2009; MacLean et al., *Nature Rev. Microbiol.*, 7: 287-296; U.S. Pat. No. 6,210,891; U.S. Pat. No. 6,258,568; each herein incorporated by reference in its entirety), template DNA is fragmented, end-repaired, ligated to adaptors, and clonally amplified in-situ by capturing single template molecules with beads bearing oligonucleotides complementary to the adaptors. Each bead bearing a single template type is compartmentalized into a water-in-oil microvesicle, and the template is clonally amplified using a technique referred to as emulsion PCR. The emulsion is disrupted after amplification and beads are deposited into individual wells of a picotitre plate functioning as a flow cell during the sequencing reactions. Ordered, iterative introduction of each of the four dNTP reagents occurs in the flow cell in the presence of sequencing enzymes and luminescent reporter such as luciferase. In the event that an appropriate dNTP is added to the 3' end of the sequencing primer, the resulting production of ATP causes a burst of luminescence within the well, which is recorded using a CCD camera. It is possible to achieve read lengths greater than or equal to 400 bases, and 10^6 sequence reads can be achieved, resulting in up to 500 million base pairs (Mb) of sequence.

In the Solexa/Illumina platform (Voelkerding et al., *Clinical Chem.*, 55: 641-658, 2009; MacLean et al., *Nature Rev. Microbiol.*, 7: 287-296; U.S. Pat. No. 6,833,246; U.S. Pat. No. 7,115,400; U.S. Pat. No. 6,969,488; each herein incorporated by reference in its entirety), sequencing data are produced in the form of shorter-length reads. In this method, single-stranded fragmented DNA is end-repaired to generate 5'-phosphorylated blunt ends, followed by Klenow-mediated addition of a single A base to the 3' end of the fragments. A-addition facilitates addition of T-overhang adaptor oligonucleotides, which are subsequently used to capture the template-adaptor molecules on the surface of a flow cell that is studded with oligonucleotide anchors. The anchor is used as a PCR primer, but because of the length of the template and its proximity to other nearby anchor oligonucleotides, extension by PCR results in the "arching over" of the molecule to hybridize with an adjacent anchor oligonucleotide to form a bridge structure on the surface of the flow cell. These loops of DNA are denatured and cleaved. Forward strands are then sequenced with reversible dye terminators. The sequence of incorporated nucleotides is determined by detection of post-incorporation fluorescence, with each fluor and block removed prior to the next cycle of dNTP addition. Sequence read length ranges from 36 nucleotides to over 50 nucleotides, with overall output exceeding 1 billion nucleotide pairs per analytical run.

Sequencing nucleic acid molecules using SOLiD technology (Voelkerding et al., Clinical Chem., 55: 641-658, 2009; MacLean et al., Nature Rev. Microbiol., 7: 287-296; U.S. Pat. No. 5,912,148; U.S. Pat. No. 6,130,073; each herein incorporated by reference in their entirety) also involves fragmentation of the template, ligation to oligonucleotide adaptors, attachment to beads, and clonal amplification by emulsion PCR. Following this, beads bearing template are immobilized on a derivatized surface of a glass flow-cell, and a primer complementary to the adaptor oligonucleotide is annealed. However, rather than utilizing this primer for 3' extension, it is instead used to provide a 5' phosphate group for ligation to interrogation probes containing two probe-specific bases followed by 6 degenerate bases and one of four fluorescent labels. In the SOLiD system, interrogation probes have 16 possible combinations of the two bases at the 3' end of each probe, and one of four fluors at the 5' end. Fluor color, and thus identity of each probe, corresponds to specified color-space coding schemes. Multiple rounds (usually 7) of probe annealing, ligation, and fluor detection are followed by denaturation, and then a second round of sequencing using a primer that is offset by one base relative to the initial primer. In this manner, the template sequence can be computationally re-constructed, and template bases are interrogated twice, resulting in increased accuracy. Sequence read length averages 35 nucleotides, and overall output exceeds 4 billion bases per sequencing run.

In certain embodiments, the technology finds use in nanopore sequencing (see, e.g., Astier et al., J. Am. Chem. Soc. 2006 Feb 8; 128(5):1705-10, herein incorporated by reference). The theory behind nanopore sequencing has to do with what occurs when a nanopore is immersed in a conducting fluid and a potential (voltage) is applied across it. Under these conditions a slight electric current due to conduction of ions through the nanopore can be observed, and the amount of current is exceedingly sensitive to the size of the nanopore. As each base of a nucleic acid passes through the nanopore, this causes a change in the magnitude of the current through the nanopore that is distinct for each of the four bases, thereby allowing the sequence of the DNA molecule to be determined.

In certain embodiments, the technology finds use in HeliScope by Helicos BioSciences (Voelkerding et al., Clinical Chem., 55: 641-658, 2009; MacLean et al., Nature Rev. Microbiol., 7: 287-296; U.S. Pat. No. 7,169,560; U.S. Pat. No. 7,282,337; U.S. Pat. No. 7,482,120; U.S. Pat. No. 7,501,245; U.S. Pat. No. 6,818,395; U.S. Pat. No. 6,911,345; U.S. Pat. No. 7,501,245; each herein incorporated by reference in their entirety). Template DNA is fragmented and polyadenylated at the 3' end, with the final adenosine bearing a fluorescent label. Denatured polyadenylated template fragments are ligated to poly(dT) oligonucleotides

on the surface of a flow cell. Initial physical locations of captured template molecules are recorded by a CCD camera, and then label is cleaved and washed away. Sequencing is achieved by addition of polymerase and serial addition of fluorescently-labeled dNTP reagents. Incorporation events result in fluor signal corresponding to the dNTP, and signal is captured by a CCD camera before each round of dNTP addition. Sequence read length ranges from 25–50 nucleotides, with overall output exceeding 1 billion nucleotide pairs per analytical run.

The Ion Torrent technology is a method of DNA sequencing based on the detection of hydrogen ions that are released during the polymerization of DNA (see, e.g., Science 327(5970): 1190 (2010); U.S. Pat. Appl. Pub. Nos. 20090026082, 20090127589, 20100301398, 20100197507, 20100188073, and 20100137143, incorporated by reference in their entireties for all purposes). A microwell contains a template DNA strand to be sequenced. Beneath the layer of microwells is a hypersensitive ISFET ion sensor. All layers are contained within a CMOS semiconductor chip, similar to that used in the electronics industry. When a dNTP is incorporated into the growing complementary strand a hydrogen ion is released, which triggers a hypersensitive ion sensor. If homopolymer repeats are present in the template sequence, multiple dNTP molecules will be incorporated in a single cycle. This leads to a corresponding number of released hydrogens and a proportionally higher electronic signal. This technology differs from other sequencing technologies in that no modified nucleotides or optics are used. The per-base accuracy of the Ion Torrent sequencer is ~99.6% for 50 base reads, with ~100 Mb generated per run. The read-length is 100 base pairs. The accuracy for homopolymer repeats of 5 repeats in length is ~98%. The benefits of ion semiconductor sequencing are rapid sequencing speed and low upfront and operating costs.

The technology finds use in another nucleic acid sequencing approach developed by Stratos Genomics, Inc. and involves the use of Xpandomers. This sequencing process typically includes providing a daughter strand produced by a template-directed synthesis. The daughter strand generally includes a plurality of subunits coupled in a sequence corresponding to a contiguous nucleotide sequence of all or a portion of a target nucleic acid in which the individual subunits comprise a tether, at least one probe or nucleobase residue, and at least one selectively cleavable bond. The selectively cleavable bond(s) is/are cleaved to yield an Xpandomer of a length longer than the plurality of the subunits of the daughter strand. The Xpandomer typically includes the tethers and reporter elements for parsing genetic information in a sequence corresponding to the contiguous nucleotide sequence of all

or a portion of the target nucleic acid. Reporter elements of the Xpandomer are then detected. Additional details relating to Xpandomer-based approaches are described in, for example, U.S. Pat. Pub No. 20090035777, entitled "High Throughput Nucleic Acid Sequencing by Expansion," filed June 19, 2008, which is incorporated herein in its entirety.

5 Other emerging single molecule sequencing methods include real-time sequencing by synthesis using a VisiGen platform (Voelkerding et al., Clinical Chem., 55: 641–58, 2009; U.S. Pat. No. 7,329,492; U.S. Pat. App. Ser. No. 11/671956; U.S. Pat. App. Ser. No. 11/781166; each herein incorporated by reference in their entirety) in which immobilized, primed DNA template is subjected to strand extension using a fluorescently-modified
10 polymerase and fluorescent acceptor molecules, resulting in detectable fluorescence resonance energy transfer (FRET) upon nucleotide addition.

 In some embodiments, capillary electrophoresis (CE) is utilized to analyze amplification fragments. During capillary electrophoresis, nucleic acids (e.g., the products of a PCR reaction) are injected electrokinetically into capillaries filled with polymer. High
15 voltage is applied so that the fluorescent DNA fragments are separated by size and are detected by a laser/camera system. In some embodiments, CE systems from Life Technologies (Grand Island, NY) are utilized for fragment sizing (See e.g., US 6706162, US8043493, each of which is herein incorporated by reference in its entirety).

20 **2. Hybridization**

 Illustrative non-limiting examples of nucleic acid hybridization techniques include, but are not limited to, in situ hybridization (ISH), microarray, and Southern or Northern blot.

 In situ hybridization (ISH) is a type of hybridization that uses a labeled complementary DNA or RNA strand as a probe to localize a specific DNA or RNA sequence
25 in a portion or section of tissue (in situ), or, if the tissue is small enough, the entire tissue (whole mount ISH). DNA ISH can be used to determine the structure of chromosomes. RNA ISH is used to measure and localize mRNAs and other transcripts within tissue sections or whole mounts. Sample cells and tissues are usually treated to fix the target transcripts in place and to increase access of the probe. The probe hybridizes to the target sequence at
30 elevated temperature, and then the excess probe is washed away. The probe that was labeled with either radio-, fluorescent- or antigen-labeled bases is localized and quantitated in the tissue using either autoradiography, fluorescence microscopy or immunohistochemistry, respectively. ISH can also use two or more probes, labeled with radioactivity or the other non-radioactive labels, to simultaneously detect two or more transcripts.

2.1 FISH

In some embodiments, fusion sequences are detected using fluorescence in situ hybridization (FISH). The preferred FISH assays for the present invention utilize bacterial
5 artificial chromosomes (BACs). These have been used extensively in the human genome sequencing project (see Nature 409: 953-958 (2001)) and clones containing specific BACs are available through distributors that can be located through many sources, e.g., NCBI. Each BAC clone from the human genome has been given a reference name that unambiguously identifies it. These names can be used to find a corresponding GenBank
10 sequence and to order copies of the clone from a distributor. In some embodiments, the detection assay is a FISH assay utilizing a probe for TYK2 and/or a TYK2-gene-fusion-partner (e.g., NPM1).

2.2 Microarrays

15 Different kinds of biological assays are called microarrays including, but not limited to: DNA microarrays (e.g., cDNA microarrays and oligonucleotide microarrays); protein microarrays; tissue microarrays; transfection or cell microarrays; chemical compound microarrays; and, antibody microarrays. A DNA microarray, commonly known as gene chip, DNA chip, or biochip, is a collection of microscopic DNA spots attached to a solid surface
20 (e.g., glass, plastic or silicon chip) forming an array for the purpose of expression profiling or monitoring expression levels for thousands of genes simultaneously. The affixed DNA segments are known as probes, thousands of which can be used in a single DNA microarray. Microarrays can be used to identify disease genes by comparing gene expression in disease and normal cells. Microarrays can be fabricated using a variety of technologies, including
25 but not limiting: printing with fine-pointed pins onto glass slides; photolithography using pre-made masks; photolithography using dynamic micromirror devices; ink-jet printing; or, electrochemistry on microelectrode arrays.

Southern and Northern blotting is used to detect specific DNA or RNA sequences, respectively. DNA or RNA extracted from a sample is fragmented, electrophoretically
30 separated on a matrix gel, and transferred to a membrane filter. The filter bound DNA or RNA is subject to hybridization with a labeled probe complementary to the sequence of interest. Hybridized probe bound to the filter is detected. A variant of the procedure is the reverse Northern blot, in which the substrate nucleic acid that is affixed to the membrane is a

collection of isolated DNA fragments and the probe is RNA extracted from a tissue and labeled.

3. Amplification

5 Chromosomal rearrangements of genomic DNA and chimeric mRNA may be amplified prior to or simultaneous with detection. Illustrative non-limiting examples of nucleic acid amplification techniques include, but are not limited to, polymerase chain reaction (PCR), reverse transcription polymerase chain reaction (RT-PCR), transcription-mediated amplification (TMA), ligase chain reaction (LCR), strand displacement
10 amplification (SDA), and nucleic acid sequence based amplification (NASBA). Those of ordinary skill in the art will recognize that certain amplification techniques (e.g., PCR) require that RNA be reversed transcribed to DNA prior to amplification (e.g., RT-PCR), whereas other amplification techniques directly amplify RNA (e.g., TMA and NASBA).

The polymerase chain reaction (U.S. Pat. Nos. 4,683,195, 4,683,202, 4,800,159 and
15 4,965,188, each of which is herein incorporated by reference in its entirety), commonly referred to as PCR, uses multiple cycles of denaturation, annealing of primer pairs to opposite strands, and primer extension to exponentially increase copy numbers of a target nucleic acid sequence. In a variation called RT-PCR, reverse transcriptase (RT) is used to make a complementary DNA (cDNA) from mRNA, and the cDNA is then amplified by PCR to
20 produce multiple copies of DNA. For other various permutations of PCR see, e.g., U.S. Pat. Nos. 4,683,195, 4,683,202 and 4,800,159; Mullis et al., Meth. Enzymol. 155: 335 (1987); and, Murakawa et al., DNA 7: 287 (1988), each of which is herein incorporated by reference in its entirety.

Transcription mediated amplification (U.S. Pat. Nos. 5,480,784 and 5,399,491, each
25 of which is herein incorporated by reference in its entirety), commonly referred to as TMA, synthesizes multiple copies of a target nucleic acid sequence autocatalytically under conditions of substantially constant temperature, ionic strength, and pH in which multiple RNA copies of the target sequence autocatalytically generate additional copies. See, e.g., U.S. Pat. Nos. 5,399,491 and 5,824,518, each of which is herein incorporated by reference in
30 its entirety. In a variation described in U.S. Publ. No. 20060046265 (herein incorporated by reference in its entirety), TMA optionally incorporates the use of blocking moieties, terminating moieties, and other modifying moieties to improve TMA process sensitivity and accuracy.

The ligase chain reaction (Weiss, R., Science 254: 1292 (1991), herein incorporated by reference in its entirety), commonly referred to as LCR, uses two sets of complementary DNA oligonucleotides that hybridize to adjacent regions of the target nucleic acid. The DNA oligonucleotides are covalently linked by a DNA ligase in repeated cycles of thermal
5 denaturation, hybridization and ligation to produce a detectable double-stranded ligated oligonucleotide product.

Strand displacement amplification (Walker, G. et al., Proc. Natl. Acad. Sci. USA 89: 392-396 (1992); U.S. Pat. Nos. 5,270,184 and 5,455,166, each of which is herein incorporated by reference in its entirety), commonly referred to as SDA, uses cycles of
10 annealing pairs of primer sequences to opposite strands of a target sequence, primer extension in the presence of a dNTP α S to produce a duplex hemiphosphorothioated primer extension product, endonuclease-mediated nicking of a hemimodified restriction endonuclease recognition site, and polymerase-mediated primer extension from the 3' end of the nick to
15 displace an existing strand and produce a strand for the next round of primer annealing, nicking and strand displacement, resulting in geometric amplification of product. Thermophilic SDA (tSDA) uses thermophilic endonucleases and polymerases at higher temperatures in essentially the same method (EP Pat. No. 0 684 315).

Other amplification methods include, for example: nucleic acid sequence based amplification (U.S. Pat. No. 5,130,238, herein incorporated by reference in its entirety),
20 commonly referred to as NASBA; one that uses an RNA replicase to amplify the probe molecule itself (Lizardi et al., BioTechnol. 6: 1197 (1988), herein incorporated by reference in its entirety), commonly referred to as Q β replicase; a transcription based amplification method (Kwoh et al., Proc. Natl. Acad. Sci. USA 86:1173 (1989)); and, self-sustained sequence replication (Guatelli et al., Proc. Natl. Acad. Sci. USA 87: 1874 (1990), each of
25 which is herein incorporated by reference in its entirety). For further discussion of amplification methods see Persing, David H., "In Vitro Nucleic Acid Amplification Techniques" in Diagnostic Medical Microbiology: Principles and Applications (Persing et al., Eds.), pp. 51-87 (American Society for Microbiology, Washington, DC (1993)).

4. Detection Methods

30 Non-amplified or amplified gene fusion nucleic acids can be detected by any means. For example, the gene fusions can be detected by hybridization with a detectably labeled probe and measurement of the resulting hybrids. Illustrative non-limiting examples of detection methods are described below.

One illustrative detection method, the Hybridization Protection Assay (HPA) involves hybridizing a chemiluminescent oligonucleotide probe (e.g., an acridinium ester-labeled (AE) probe) to the target sequence, selectively hydrolyzing the chemiluminescent label present on unhybridized probe, and measuring the chemiluminescence produced from the remaining
5 probe in a luminometer. See, e.g., U.S. Pat. No. 5,283,174 and Norman C. Nelson et al., Nonisotopic Probing, Blotting, and Sequencing, ch. 17 (Larry J. Kricka ed., 2d ed. 1995, each of which is herein incorporated by reference in its entirety).

Another illustrative detection method provides for quantitative evaluation of the amplification process in real-time. Evaluation of an amplification process in “real-time”
10 involves determining the amount of amplicon in the reaction mixture either continuously or periodically during the amplification reaction, and using the determined values to calculate the amount of target sequence initially present in the sample. A variety of methods for determining the amount of initial target sequence present in a sample based on real-time amplification including methods disclosed in U.S. Pat. Nos. 6,303,305 and 6,541,205, each of
15 which is herein incorporated by reference in its entirety can be used. Another method for determining the quantity of target sequence initially present in a sample, but which is not based on a real-time amplification, is disclosed in U.S. Pat. No. 5,710,029, herein incorporated by reference in its entirety.

Amplification products may be detected in real-time through the use of various self-
20 hybridizing probes, most of which have a stem-loop structure. Such self-hybridizing probes are labeled so that they emit differently detectable signals, depending on whether the probes are in a self-hybridized state or an altered state through hybridization to a target sequence. By way of non-limiting example, “molecular torches” are a type of self-hybridizing probe that includes distinct regions of self-complementarity (referred to as “the target binding
25 domain” and “the target closing domain”) which are connected by a joining region (e.g., non-nucleotide linker) and which hybridize to each other under predetermined hybridization assay conditions. In a preferred embodiment, molecular torches contain single-stranded base regions in the target binding domain that are from 1 to about 20 bases in length and are accessible for hybridization to a target sequence present in an amplification reaction under
30 strand displacement conditions. Under strand displacement conditions, hybridization of the two complementary regions, which may be fully or partially complementary, of the molecular torch is favored, except in the presence of the target sequence, which will bind to the single-stranded region present in the target binding domain and displace all or a portion of the target closing domain. The target binding domain and the target closing domain of a molecular

torch include a detectable label or a pair of interacting labels (e.g., luminescent/quencher) positioned so that a different signal is produced when the molecular torch is self-hybridized than when the molecular torch is hybridized to the target sequence, thereby permitting detection of probe:target duplexes in a test sample in the presence of unhybridized molecular torches. Molecular torches and a variety of types of interacting label pairs are disclosed in U.S. Pat. No. 6,534,274, herein incorporated by reference in its entirety.

Another example of a detection probe having self-complementarity is a “molecular beacon.” Molecular beacons include nucleic acid molecules having a target complementary sequence, an affinity pair (or nucleic acid arms) holding the probe in a closed conformation in the absence of a target sequence present in an amplification reaction, and a label pair that interacts when the probe is in a closed conformation. Hybridization of the target sequence and the target complementary sequence separates the members of the affinity pair, thereby shifting the probe to an open conformation. The shift to the open conformation is detectable due to reduced interaction of the label pair, which may be, for example, a fluorophore and a quencher (e.g., DABCYL and EDANS). Molecular beacons are disclosed in U.S. Pat. Nos. 5,925,517 and 6,150,097, herein incorporated by reference in its entirety.

Other self-hybridizing are, for example, probe binding pairs having interacting labels, such as those disclosed in U.S. Pat. No. 5,928,862 (herein incorporated by reference in its entirety) might be adapted for use in the present invention. Probe systems used to detect single nucleotide polymorphisms (SNPs) might also be utilized in the present invention. Additional detection systems include “molecular switches,” as disclosed in U.S. Publ. No. 20050042638, herein incorporated by reference in its entirety. Other probes, such as those comprising intercalating dyes and/or fluorochromes, are also useful for detection of amplification products in the present invention. See, e.g., U.S. Pat. No. 5,814,447 (herein incorporated by reference in its entirety).

C. Protein Detection

The gene fusions of the present invention may be detected as truncated or chimeric proteins using a variety of protein techniques, including but not limited to: protein sequencing; and, immunoassays.

1. Sequencing

Illustrative non-limiting examples of protein sequencing techniques include, but are not limited to, mass spectrometry and Edman degradation.

Mass spectrometry can, in principle, sequence any size protein but becomes computationally more difficult as size increases. A protein is digested by an endoprotease, and the resulting solution is passed through a high pressure liquid chromatography column. At the end of this column, the solution is sprayed out of a narrow nozzle charged to a high positive potential into the mass spectrometer. The charge on the droplets causes them to fragment until only single ions remain. The peptides are then fragmented and the mass-charge ratios of the fragments measured. The mass spectrum is analyzed by computer and often compared against a database of previously sequenced proteins in order to determine the sequences of the fragments. The process is then repeated with a different digestion enzyme, and the overlaps in sequences are used to construct a sequence for the protein.

In the Edman degradation reaction, the peptide to be sequenced is adsorbed onto a solid surface (e.g., a glass fiber coated with polybrene). The Edman reagent, phenylisothiocyanate (PTC), is added to the adsorbed peptide, together with a mildly basic buffer solution of 12% trimethylamine, and reacts with the amine group of the N-terminal amino acid. The terminal amino acid derivative can then be selectively detached by the addition of anhydrous acid. The derivative isomerizes to give a substituted phenylthiohydantoin, which can be washed off and identified by chromatography, and the cycle can be repeated. The efficiency of each step is about 98%, which allows about 50 amino acids to be reliably determined.

2. Immunoassays

Illustrative non-limiting examples of immunoassays include, but are not limited to: immunoprecipitation; Western blot; ELISA; immunohistochemistry; immunocytochemistry; flow cytometry; and, immuno-PCR. Polyclonal or monoclonal antibodies detectably labeled using various techniques (e.g., colorimetric, fluorescent, chemiluminescent or radioactive) are suitable for use in the immunoassays.

Immunoprecipitation is the technique of precipitating an antigen out of solution using an antibody specific to that antigen. The process can be used to identify protein complexes present in cell extracts by targeting a protein believed to be in the complex. The complexes are brought out of solution by insoluble antibody-binding proteins isolated initially from bacteria, such as Protein A and Protein G. The antibodies can also be coupled to sepharose beads that can easily be isolated out of solution. After washing, the precipitate can be analyzed using mass spectrometry, Western blotting, or any number of other methods for identifying constituents in the complex.

A Western blot, or immunoblot, is a method to detect protein in a given sample of tissue homogenate or extract. It uses gel electrophoresis to separate denatured proteins by mass. The proteins are then transferred out of the gel and onto a membrane, typically polyvinylidene difluoride or nitrocellulose, where they are probed using antibodies specific to the protein of interest. As a result, researchers can examine the amount of protein in a given sample and compare levels between several groups.

An ELISA, short for Enzyme-Linked Immunosorbent Assay, is a biochemical technique to detect the presence of an antibody or an antigen in a sample. It utilizes a minimum of two antibodies, one of which is specific to the antigen and the other of which is coupled to an enzyme. The second antibody will cause a chromogenic or fluorogenic substrate to produce a signal. Variations of ELISA include sandwich ELISA, competitive ELISA, and ELISPOT. Because the ELISA can be performed to evaluate either the presence of antigen or the presence of antibody in a sample, it is a useful tool both for determining serum antibody concentrations and also for detecting the presence of antigen.

Immunohistochemistry and immunocytochemistry refer to the process of localizing proteins in a tissue section or cell, respectively, via the principle of antigens in tissue or cells binding to their respective antibodies. Visualization is enabled by tagging the antibody with color producing or fluorescent tags. Typical examples of color tags include, but are not limited to, horseradish peroxidase and alkaline phosphatase. Typical examples of fluorophore tags include, but are not limited to, fluorescein isothiocyanate (FITC) or phycoerythrin (PE).

Flow cytometry is a technique for counting, examining and sorting microscopic particles suspended in a stream of fluid. It allows simultaneous multiparametric analysis of the physical and/or chemical characteristics of single cells flowing through an optical/electronic detection apparatus. A beam of light (e.g., a laser) of a single frequency or color is directed onto a hydrodynamically focused stream of fluid. A number of detectors are aimed at the point where the stream passes through the light beam; one in line with the light beam (Forward Scatter or FSC) and several perpendicular to it (Side Scatter (SSC) and one or more fluorescent detectors). Each suspended particle passing through the beam scatters the light in some way, and fluorescent chemicals in the particle may be excited into emitting light at a lower frequency than the light source. The combination of scattered and fluorescent light is picked up by the detectors, and by analyzing fluctuations in brightness at each detector, one for each fluorescent emission peak, it is possible to deduce various facts about the physical and chemical structure of each individual particle. FSC correlates with the cell volume and

SSC correlates with the density or inner complexity of the particle (e.g., shape of the nucleus, the amount and type of cytoplasmic granules or the membrane roughness).

Immuno-polymerase chain reaction (IPCR) utilizes nucleic acid amplification techniques to increase signal generation in antibody-based immunoassays. Because no protein equivalence of PCR exists, that is, proteins cannot be replicated in the same manner that nucleic acid is replicated during PCR, the only way to increase detection sensitivity is by signal amplification. The target proteins are bound to antibodies which are directly or indirectly conjugated to oligonucleotides. Unbound antibodies are washed away and the remaining bound antibodies have their oligonucleotides amplified. Protein detection occurs via detection of amplified oligonucleotides using standard nucleic acid detection methods, including real-time methods.

D. Data Analysis

In some embodiments, a computer-based analysis program is used to translate the raw data generated by the detection assay (e.g., the presence, absence, or amount of a given marker or markers) into data of predictive value for a clinician. The clinician can access the predictive data using any suitable means. Thus, in some preferred embodiments, the present invention provides the further benefit that the clinician, who is not likely to be trained in genetics or molecular biology, need not understand the raw data. The data is presented directly to the clinician in its most useful form. The clinician is then able to immediately utilize the information in order to optimize the care of the subject.

The present invention contemplates any method capable of receiving, processing, and transmitting the information to and from laboratories conducting the assays, information provides, medical personal, and subjects. For example, in some embodiments of the present invention, a sample (e.g., a biopsy or a serum sample) is obtained from a subject and submitted to a profiling service (e.g., clinical lab at a medical facility, genomic profiling business, etc.), located in any part of the world (e.g., in a country different than the country where the subject resides or where the information is ultimately used) to generate raw data. Where the sample comprises a tissue or other biological sample, the subject may visit a medical center to have the sample obtained and sent to the profiling center, or subjects may collect the sample themselves and directly send it to a profiling center. Where the sample comprises previously determined biological information, the information may be directly sent to the profiling service by the subject (e.g., an information card containing the information may be scanned by a computer and the data transmitted to a computer of the profiling center

using an electronic communication systems). Once received by the profiling service, the sample is processed and a profile is produced (i.e., expression data), specific for the diagnostic or prognostic information desired for the subject.

The profile data is then prepared in a format suitable for interpretation by a treating clinician. For example, rather than providing raw expression data, the prepared format may represent a diagnosis or risk assessment (e.g., likelihood of cancer being present) for the subject, along with recommendations for particular treatment options. The data may be displayed to the clinician by any suitable method. For example, in some embodiments, the profiling service generates a report that can be printed for the clinician (e.g., at the point of care) or displayed to the clinician on a computer monitor.

In some embodiments, the information is first analyzed at the point of care or at a regional facility. The raw data is then sent to a central processing facility for further analysis and/or to convert the raw data to information useful for a clinician or patient. The central processing facility provides the advantage of privacy (all data is stored in a central facility with uniform security protocols), speed, and uniformity of data analysis. The central processing facility can then control the fate of the data following treatment of the subject. For example, using an electronic communication system, the central facility can provide data to the clinician, the subject, or researchers.

In some embodiments, the subject is able to directly access the data using the electronic communication system. The subject may chose further intervention or counseling based on the results. In some embodiments, the data is used for research use. For example, the data may be used to further optimize the inclusion or elimination of markers as useful indicators of a particular condition or stage of disease.

E. In vivo Imaging

The gene fusions of the present invention may also be detected using in vivo imaging techniques, including but not limited to: radionuclide imaging; positron emission tomography (PET); computerized axial tomography, X-ray or magnetic resonance imaging method, fluorescence detection, and chemiluminescent detection. In some embodiments, in vivo imaging techniques are used to visualize the presence of or expression of cancer markers in an animal (e.g., a human or non-human mammal). For example, in some embodiments, cancer marker mRNA or protein is labeled using a labeled antibody specific for the cancer marker. A specifically bound and labeled antibody can be detected in an individual using an in vivo imaging method, including, but not limited to, radionuclide imaging, positron

emission tomography, computerized axial tomography, X-ray or magnetic resonance imaging method, fluorescence detection, and chemiluminescent detection. Methods for generating antibodies to the cancer markers of the present invention are described below.

The in vivo imaging methods of the present invention are useful in the diagnosis of cancers that express the cancer markers of the present invention. In vivo imaging is used to visualize the presence of a marker indicative of the cancer. Such techniques allow for diagnosis without the use of an unpleasant biopsy. The in vivo imaging methods of the present invention are also useful for providing prognoses to cancer patients. For example, the presence of a marker indicative of cancers likely to metastasize can be detected. The in vivo imaging methods of the present invention can further be used to detect metastatic cancers in other parts of the body.

In some embodiments, reagents (e.g., antibodies) specific for the cancer markers of the present invention are fluorescently labeled. The labeled antibodies are introduced into a subject (e.g., orally or parenterally). Fluorescently labeled antibodies are detected using any suitable method (e.g., using the apparatus described in U.S. Pat. No. 6,198,107, herein incorporated by reference).

In other embodiments, antibodies are radioactively labeled. Sumerdon et al., (Nucl. Med. Biol 17:247-254 [1990]) have described an optimized antibody-chelator for the radioimmunoscinotgraphic imaging of tumors using Indium-111 as the label. Griffin et al., (J Clin Onc 9:631-640 [1991]) have described the use of this agent in detecting tumors in patients suspected of having recurrent colorectal cancer. Agents with paramagnetic ions as labels for magnetic resonance imaging can be utilized (Lauffer, Magnetic Resonance in Medicine 22:339-342 [1991]). The label used will depend on the imaging modality chosen. Radioactive labels such as Indium-111, Technetium-99m, or Iodine-131 can be used for planar scans or single photon emission computed tomography (SPECT). Positron emitting labels such as Fluorine-19 can also be used for positron emission tomography (PET). For MRI, paramagnetic ions such as Gadolinium (III) or Manganese (II) can be used.

Radioactive metals with half-lives ranging from 1 hour to 3.5 days are available for conjugation to antibodies, such as scandium-47 (3.5 days) gallium-67 (2.8 days), gallium-68 (68 minutes), technetium-99m (6 hours), and indium-111 (3.2 days), of which gallium-67, technetium-99m, and indium-111 are preferable for gamma camera imaging, gallium-68 is preferable for positron emission tomography.

A useful method of labeling antibodies with such radiometals is by means of a bifunctional chelating agent, such as diethylenetriaminepentaacetic acid (DTPA), as

described, for example, by Khaw et al. (Science 209:295 [1980]) for In-111 and Tc-99m, and by Scheinberg et al. (Science 215:1511 [1982]). Other chelating agents may also be used, but the 1-(p-carboxymethoxybenzyl)EDTA and the carboxycarbonic anhydride of DTPA are advantageous because their use permits conjugation without affecting the antibody's immunoreactivity substantially.

Another method for coupling DTPA to proteins is by use of the cyclic anhydride of DTPA, as described by Hnatowich et al. (Int. J. Appl. Radiat. Isot. 33:327 [1982]) for labeling of albumin with In-111, but which can be adapted for labeling of antibodies. A suitable method of labeling antibodies with Tc-99m which does not use chelation with DTPA is the pretinning method of Crockford et al., (U.S. Pat. No. 4,323,546, herein incorporated by reference).

A preferred method of labeling immunoglobulins with Tc-99m is that described by Wong et al. (Int. J. Appl. Radiat. Isot., 29:251 [1978]) for plasma protein, and recently applied successfully by Wong et al. (J. Nucl. Med., 23:229 [1981]) for labeling antibodies.

In the case of the radiometals conjugated to the specific antibody, it is likewise desirable to introduce as high a proportion of the radiolabel as possible into the antibody molecule without destroying its immunospecificity. A further improvement may be achieved by effecting radiolabeling in the presence of the specific cancer marker of the present invention, to insure that the antigen binding site on the antibody will be protected. The antigen is separated after labeling.

In still further embodiments, in vivo biophotonic imaging (Xenogen, Alameda, CA) is utilized for in vivo imaging. This real-time in vivo imaging utilizes luciferase. The luciferase gene is incorporated into cells, microorganisms, and animals (e.g., as a fusion protein with a cancer marker of the present invention). When active, it leads to a reaction that emits light. A CCD camera and software is used to capture the image and analyze it.

F. Compositions & Kits

Compositions for use in the diagnostic methods of the present invention include, but are not limited to, probes, amplification oligonucleotides, and antibodies. Particularly preferred compositions detect a product only when a gene fusion of TYK2 and a TYK2-gene-fusion-partner (e.g., NPM1) is present in a sample (e.g., NPM1-TYK2). In some embodiments, the compositions include a single labeled probe comprising a sequence that hybridizes to the junction at which a 5' portion from a TYK2-gene-fusion-partner (e.g., NPM1) fuses to a 3' portion from a TYK2 gene (i.e., spans the gene fusion junction). In some

embodiments, the compositions include a pair of amplification oligonucleotides wherein the first amplification oligonucleotide comprises a sequence that hybridizes to a TYK2-gene-fusion-partner (e.g., NPM1) and the second amplification oligonucleotide comprises a sequence that hybridizes to TYK2 gene. In some embodiments, the compositions include an antibody to a gene fusion between TYK2 and a TYK2-gene-fusion-partner (e.g., NPM1).

Other useful compositions, however, include: a pair of labeled probes wherein the first labeled probe comprises a sequence that hybridizes to a TYK2-gene-fusion-partner (e.g., NPM1) and a second labeled probe comprises a sequence that hybridizes to an TYK2 gene.

Any of these compositions, alone or in combination with other compositions of the present invention, may be provided in the form of a kit. For example, the single labeled probe and pair of amplification oligonucleotides may be provided in a kit for the amplification and detection of gene fusions of the present invention. Kits may further comprise appropriate controls and/or detection reagents.

The probe and antibody compositions of the present invention may also be provided in the form of an array.

IV. Drug Screening Applications

In some embodiments, the present invention provides drug screening assays (e.g., to screen for anticancer drugs). The screening methods of the present invention utilize cancer markers identified using the methods of the present invention (e.g., including but not limited to gene fusions between TYK2 and a TYK2-gene-fusion-partner (e.g., NPM1)). For example, in some embodiments, the present invention provides methods of screening for compounds that alter (e.g., decrease) the expression of cancer marker genes. The compounds or agents may interfere with transcription, by interacting, for example, with the promoter region. The compounds or agents may interfere with mRNA produced from the fusion (e.g., by RNA interference, antisense technologies, etc.). The compounds or agents may interfere with pathways that are upstream or downstream of the biological activity of the fusion. In some embodiments, candidate compounds are antisense or interfering RNA agents (e.g., oligonucleotides) directed against cancer markers. In other embodiments, candidate compounds are antibodies or small molecules that specifically bind to a cancer marker regulator or expression products of the present invention and inhibit its biological function.

In one screening method, candidate compounds are evaluated for their ability to alter cancer marker expression by contacting a compound with a cell expressing a cancer marker and then assaying for the effect of the candidate compounds on expression. In some

embodiments, the effect of candidate compounds on expression of a cancer marker gene is assayed for by detecting the level of cancer marker mRNA expressed by the cell. mRNA expression can be detected by any suitable method. In other embodiments, the effect of candidate compounds on expression of cancer marker genes is assayed by measuring the level of polypeptide encoded by the cancer markers. The level of polypeptide expressed can be measured using any suitable method, including but not limited to, those disclosed herein.

Specifically, the present invention provides screening methods for identifying modulators, i.e., candidate or test compounds or agents (e.g., proteins, peptides, peptidomimetics, peptoids, small molecules or other drugs) which bind to cancer markers of the present invention, have an inhibitory (or stimulatory) effect on, for example, cancer marker expression or cancer marker activity, or have a stimulatory or inhibitory effect on, for example, the expression or activity of a cancer marker substrate. Compounds thus identified can be used to modulate the activity of target gene products (e.g., cancer marker genes) either directly or indirectly in a therapeutic protocol, to elaborate the biological function of the target gene product, or to identify compounds that disrupt normal target gene interactions. Compounds that inhibit the activity or expression of cancer markers are useful in the treatment of proliferative disorders, e.g., cancer, particularly cutaneous CD30-positive lymphoproliferative disorders (e.g., lymphomatoid papulosis; primary cutaneous anaplastic large cell lymphoma).

In one embodiment, the invention provides assays for screening candidate or test compounds that are substrates of a cancer marker protein or polypeptide or a biologically active portion thereof. In another embodiment, the invention provides assays for screening candidate or test compounds that bind to or modulate the activity of a cancer marker protein or polypeptide or a biologically active portion thereof.

The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods, including biological libraries; peptoid libraries (libraries of molecules having the functionalities of peptides, but with a novel, non-peptide backbone, which are resistant to enzymatic degradation but which nevertheless remain bioactive; see, e.g., Zuckermann et al., J. Med. Chem. 37: 2678-85 [1994]); spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library and peptoid library approaches are preferred for use with peptide libraries, while the other four

approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam (1997) *Anticancer Drug Des.* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt et al., *Proc. Natl. Acad. Sci. U.S.A.* 90:6909 [1993]; Erb et al., *Proc. Nad. Acad. Sci. USA* 91:11422 [1994]; Zuckermann et al., *J. Med. Chem.* 37:2678 [1994];
5 Cho et al., *Science* 261:1303 [1993]; Carrell et al., *Angew. Chem. Int. Ed. Engl.* 33:2059 [1994]; Carell et al., *Angew. Chem. Int. Ed. Engl.* 33:2061 [1994]; and Gallop et al., *J. Med. Chem.* 37:1233 [1994].

Libraries of compounds may be presented in solution (e.g., Houghten, *Biotechniques*
10 13:412-421 [1992]), or on beads (Lam, *Nature* 354:82-84 [1991]), chips (Fodor, *Nature* 364:555-556 [1993]), bacteria or spores (U.S. Pat. No. 5,223,409; herein incorporated by reference), plasmids (Cull et al., *Proc. Nad. Acad. Sci. USA* 89:18651869 [1992]) or on phage (Scott and Smith, *Science* 249:386-390 [1990]; Devlin *Science* 249:404-406 [1990]; Cwirla et al., *Proc. Natl. Acad. Sci.* 87:6378-6382 [1990]; Felici, *J. Mol. Biol.* 222:301
15 [1991]).

VI. Therapeutic Applications

In some embodiments, the present invention provides therapies for cancer (e.g., cutaneous CD30-positive lymphoproliferative disorders (e.g., lymphomatoid papulosis;
20 primary cutaneous anaplastic large cell lymphoma)). In some embodiments, therapies directly or indirectly target cancer markers (e.g., including but not limited to, gene fusions between TYK2 and a TYK2-gene-fusion-partner (e.g., NPM1).

A. RNA Interference and Antisense Therapies

In some embodiments, the present invention targets the expression of cancer markers. For example, in some embodiments, the present invention employs compositions comprising oligomeric antisense or RNAi compounds, particularly oligonucleotides (e.g., those identified in the drug screening methods described above), for use in modulating the function of nucleic acid molecules encoding cancer markers of the present invention, ultimately modulating the
30 amount of cancer marker expressed.

1. RNA Interference (RNAi)

In some embodiments, RNAi is utilized to inhibit fusion protein function. RNAi represents an evolutionary conserved cellular defense for controlling the expression of

foreign genes in most eukaryotes, including humans. RNAi is typically triggered by double-stranded RNA (dsRNA) and causes sequence-specific mRNA degradation of single-stranded target RNAs homologous in response to dsRNA. The mediators of mRNA degradation are small interfering RNA duplexes (siRNAs), which are normally produced from long dsRNA by enzymatic cleavage in the cell. siRNAs are generally approximately twenty-one nucleotides in length (e.g. 21-23 nucleotides in length), and have a base-paired structure characterized by two nucleotide 3'-overhangs. Following the introduction of a small RNA, or RNAi, into the cell, it is believed the sequence is delivered to an enzyme complex called RISC (RNA-induced silencing complex). RISC recognizes the target and cleaves it with an endonuclease. It is noted that if larger RNA sequences are delivered to a cell, RNase III enzyme (Dicer) converts longer dsRNA into 21-23 nt ds siRNA fragments. In some embodiments, RNAi oligonucleotides are designed to target the junction region of fusion proteins.

Chemically synthesized siRNAs have become powerful reagents for genome-wide analysis of mammalian gene function in cultured somatic cells. Beyond their value for validation of gene function, siRNAs also hold great potential as gene-specific therapeutic agents (Tuschl and Borkhardt, *Molecular Intervent.* 2002; 2(3):158-67, herein incorporated by reference).

The transfection of siRNAs into animal cells results in the potent, long-lasting post-transcriptional silencing of specific genes (Caplen et al, *Proc Natl Acad Sci U.S.A.* 2001; 98: 9742-7; Elbashir et al., *Nature.* 2001; 411:494-8; Elbashir et al., *Genes Dev.* 2001;15: 188-200; and Elbashir et al., *EMBO J.* 2001; 20: 6877-88, all of which are herein incorporated by reference). Methods and compositions for performing RNAi with siRNAs are described, for example, in U.S. Pat. 6,506,559, herein incorporated by reference.

siRNAs are extraordinarily effective at lowering the amounts of targeted RNA, and by extension proteins, frequently to undetectable levels. The silencing effect can last several months, and is extraordinarily specific, because one nucleotide mismatch between the target RNA and the central region of the siRNA is frequently sufficient to prevent silencing (Brummelkamp et al, *Science* 2002; 296:550-3; and Holen et al, *Nucleic Acids Res.* 2002; 30:1757-66, both of which are herein incorporated by reference).

An important factor in the design of siRNAs is the presence of accessible sites for siRNA binding. Bahoia et al., (*J. Biol. Chem.*, 2003; 278: 15991-15997; herein incorporated by reference) describe the use of a type of DNA array called a scanning array to find accessible sites in mRNAs for designing effective siRNAs. These arrays comprise oligonucleotides

ranging in size from monomers to a certain maximum, usually Comers, synthesized using a physical barrier (mask) by stepwise addition of each base in the sequence. Thus the arrays represent a full oligonucleotide complement of a region of the target gene. Hybridization of the target mRNA to these arrays provides an exhaustive accessibility profile of this region of the target mRNA. Such data are useful in the design of antisense oligonucleotides (ranging from 7mers to 25mers), where it is important to achieve a compromise between oligonucleotide length and binding affinity, to retain efficacy and target specificity (Sohail et al, Nucleic Acids Res., 2001; 29(10): 2041- 2045). Additional methods and concerns for selecting siRNAs are described for example, in WO 05054270, WO05038054A1, WO03070966A2, J Mol Biol. 2005 May 13;348(4):883-93, J Mol Biol. 2005 May 13;348(4):871-81, and Nucleic Acids Res. 2003 Aug 1;31(15):4417-24, each of which is herein incorporated by reference in its entirety. In addition, software (e.g., the MWG online siMAX siRNA design tool) is commercially or publicly available for use in the selection of siRNAs.

2. Antisense

In other embodiments, fusion protein expression is modulated using antisense compounds that specifically hybridize with one or more nucleic acids encoding cancer markers of the present invention. The specific hybridization of an oligomeric compound with its target nucleic acid interferes with the normal function of the nucleic acid. This modulation of function of a target nucleic acid by compounds that specifically hybridize to it is generally referred to as "antisense." The functions of DNA to be interfered with include replication and transcription. The functions of RNA to be interfered with include all vital functions such as, for example, translocation of the RNA to the site of protein translation, translation of protein from the RNA, splicing of the RNA to yield one or more mRNA species, and catalytic activity that may be engaged in or facilitated by the RNA. The overall effect of such interference with target nucleic acid function is modulation of the expression of cancer markers of the present invention. In the context of the present invention, "modulation" means either an increase (stimulation) or a decrease (inhibition) in the expression of a gene. For example, expression may be inhibited to potentially prevent tumor proliferation.

It is preferred to target specific nucleic acids for antisense. "Targeting" an antisense compound to a particular nucleic acid, in the context of the present invention, is a multistep process. The process usually begins with the identification of a nucleic acid sequence whose

function is to be modulated. This may be, for example, a cellular gene (or mRNA transcribed from the gene) whose expression is associated with a particular disorder or disease state, or a nucleic acid molecule from an infectious agent. In the present invention, the target is a nucleic acid molecule encoding a cancer marker of the present invention. The targeting
5 process also includes determination of a site or sites within this gene for the antisense interaction to occur such that the desired effect, e.g., detection or modulation of expression of the protein, will result.

Chimeric antisense compounds of the present invention may be formed as composite structures of two or more oligonucleotides, modified oligonucleotides, oligonucleosides
10 and/or oligonucleotide mimetics as described above.

The present invention also includes pharmaceutical compositions and formulations that include the antisense compounds of the present invention as described below.

B. Gene Therapy

The present invention contemplates the use of any genetic manipulation for use in modulating the expression of cancer markers of the present invention. Examples of genetic manipulation include, but are not limited to, gene knockout (e.g., removing the fusion gene from the chromosome using, for example, recombination), expression of antisense constructs with or without inducible promoters, and the like. Delivery of nucleic acid construct to cells
20 in vitro or in vivo may be conducted using any suitable method. A suitable method is one that introduces the nucleic acid construct into the cell such that the desired event occurs (e.g., expression of an antisense construct). Genetic therapy may also be used to deliver siRNA or other interfering molecules that are expressed in vivo (e.g., upon stimulation by an inducible promoter).

Introduction of molecules carrying genetic information into cells is achieved by any of various methods including, but not limited to, directed injection of naked DNA constructs, bombardment with gold particles loaded with said constructs, and macromolecule mediated gene transfer using, for example, liposomes, biopolymers, and the like. Preferred methods use gene delivery vehicles derived from viruses, including, but not limited to, adenoviruses, retroviruses, vaccinia viruses, and adeno-associated viruses. Because of the higher efficiency
30 as compared to retroviruses, vectors derived from adenoviruses are the preferred gene delivery vehicles for transferring nucleic acid molecules into host cells in vivo. Adenoviral vectors have been shown to provide very efficient in vivo gene transfer into a variety of solid tumors in animal models and into human solid tumor xenografts in immune-deficient mice.

Examples of adenoviral vectors and methods for gene transfer are described in PCT publications WO 00/12738 and WO 00/09675 and U.S. Pat. Appl. Nos. 6,033,908, 6,019,978, 6,001,557, 5,994,132, 5,994,128, 5,994,106, 5,981,225, 5,885,808, 5,872,154, 5,830,730, and 5,824,544, each of which is herein incorporated by reference in its entirety.

5 Vectors may be administered to subject in a variety of ways. For example, in some embodiments of the present invention, vectors are administered into tumors or tissue associated with tumors using direct injection. In other embodiments, administration is via the blood or lymphatic circulation (See e.g., PCT publication 99/02685 herein incorporated by reference in its entirety). Exemplary dose levels of adenoviral vector are preferably 10^8 to
10 10^{11} vector particles added to the perfusate.

C. Antibody Therapy

In some embodiments, the present invention provides antibodies that target cells that express a cancer marker of the present invention (e.g., a gene fusion between TYK2 and a
15 TYK2-gene-fusion-partner (e.g., NPM1)). Any suitable antibody (e.g., monoclonal, polyclonal, or synthetic) may be utilized in the therapeutic methods disclosed herein. In preferred embodiments, the antibodies used for cancer therapy are humanized antibodies. Methods for humanizing antibodies can be utilized (See e.g., U.S. Pat. Nos. 6,180,370, 5,585,089, 6,054,297, and 5,565,332; each of which is herein incorporated by reference).

20 In some embodiments, the therapeutic antibodies comprise an antibody generated against a cancer marker of the present invention (e.g., a gene fusion between TYK2 and a TYK2-gene-fusion-partner (e.g., NPM1)), wherein the antibody is conjugated to a cytotoxic agent. In such embodiments, a tumor specific therapeutic agent is generated that does not target normal cells, thus reducing many of the detrimental side effects of traditional
25 chemotherapy. For certain applications, it is envisioned that the therapeutic agents will be pharmacologic agents that will serve as useful agents for attachment to antibodies, particularly cytotoxic or otherwise anticellular agents having the ability to kill or suppress the growth or cell division of endothelial cells. The present invention contemplates the use of any pharmacologic agent that can be conjugated to an antibody, and delivered in active form.
30 Exemplary anticellular agents include chemotherapeutic agents, radioisotopes, and cytotoxins. The therapeutic antibodies of the present invention may include a variety of cytotoxic moieties, including but not limited to, radioactive isotopes (e.g., iodine-131, iodine-123, technetium-99m, indium-111, rhenium-188, rhenium-186, gallium-67, copper-67, yttrium-90, iodine-125 or astatine-211), hormones such as a steroid, antimetabolites such as

cytosines (e.g., arabinoside, fluorouracil, methotrexate or aminopterin; an anthracycline; mitomycin C), vinca alkaloids (e.g., demecolcine; etoposide; mithramycin), and antitumor alkylating agent such as chlorambucil or melphalan. Other embodiments may include agents such as a coagulant, a cytokine, growth factor, bacterial endotoxin or the lipid A moiety of bacterial endotoxin. For example, in some embodiments, therapeutic agents will include plant-, fungus- or bacteria-derived toxin, such as an A chain toxins, a ribosome inactivating protein, α -sarcin, aspergillin, restrictocin, a ribonuclease, diphtheria toxin or pseudomonas exotoxin, to mention just a few examples. In some preferred embodiments, deglycosylated ricin A chain is utilized.

In any event, it is proposed that agents such as these may, if desired, be successfully conjugated to an antibody, in a manner that will allow their targeting, internalization, release or presentation to blood components at the site of the targeted tumor cells as required using conjugation technology (See, e.g., Ghose et al., *Methods Enzymol.*, 93:280 [1983]).

For example, in some embodiments the present invention provides immunotoxins targeted to a cancer marker of the present invention (e.g., a gene fusion between TYK2 and a TYK2-gene-fusion-partner (e.g., NPM1)). Immunotoxins are conjugates of a specific targeting agent typically a tumor-directed antibody or fragment, with a cytotoxic agent, such as a toxin moiety. The targeting agent directs the toxin to, and thereby selectively kills, cells carrying the targeted antigen. In some embodiments, therapeutic antibodies employ crosslinkers that provide high in vivo stability (Thorpe et al., *Cancer Res.*, 48:6396 [1988]).

In other embodiments, particularly those involving treatment of solid tumors, antibodies are designed to have a cytotoxic or otherwise anticellular effect against the tumor vasculature, by suppressing the growth or cell division of the vascular endothelial cells. This attack is intended to lead to a tumor-localized vascular collapse, depriving the tumor cells, particularly those tumor cells distal of the vasculature, of oxygen and nutrients, ultimately leading to cell death and tumor necrosis.

In preferred embodiments, antibody based therapeutics are formulated as pharmaceutical compositions as described below. In preferred embodiments, administration of an antibody composition of the present invention results in a measurable decrease in cancer (e.g., decrease or elimination of tumor).

D. Pharmaceutical Compositions

The present invention further provides pharmaceutical compositions (e.g., comprising pharmaceutical agents that modulate the expression or activity of gene fusions of the present

invention (e.g., a gene fusion between TYK2 and a TYK2-gene-fusion-partner (e.g., NPM1))). The pharmaceutical compositions of the present invention may be administered in a number of ways depending upon whether local or systemic treatment is desired and upon the area to be treated. Administration may be topical (including ophthalmic and to mucous
5 membranes including vaginal and rectal delivery), pulmonary (e.g., by inhalation or insufflation of powders or aerosols, including by nebulizer; intratracheal, intranasal, epidermal and transdermal), oral or parenteral. Parenteral administration includes intravenous, intraarterial, subcutaneous, intraperitoneal or intramuscular injection or infusion; or intracranial, e.g., intrathecal or intraventricular, administration.

10 Pharmaceutical compositions and formulations for topical administration may include transdermal patches, ointments, lotions, creams, gels, drops, suppositories, sprays, liquids and powders. Pharmaceutical carriers, aqueous, powder or oily bases, thickeners and the like may be necessary or desirable.

 Compositions and formulations for oral administration include powders or granules,
15 suspensions or solutions in water or non-aqueous media, capsules, sachets or tablets. Thickeners, flavoring agents, diluents, emulsifiers, dispersing aids or binders may be desirable.

 Compositions and formulations for parenteral, intrathecal or intraventricular administration may include sterile aqueous solutions that may also contain buffers, diluents
20 and other suitable additives such as, but not limited to, penetration enhancers, carrier compounds and other pharmaceutically acceptable carriers or excipients.

 Pharmaceutical compositions of the present invention include, but are not limited to, solutions, emulsions, and liposome-containing formulations. These compositions may be generated from a variety of components that include, but are not limited to, preformed
25 liquids, self-emulsifying solids and self-emulsifying semisolids.

 The pharmaceutical formulations of the present invention, which may conveniently be presented in unit dosage form, may be prepared according to techniques such as, for example, bringing into association the active ingredients with the pharmaceutical carrier(s) or excipient(s). In general the formulations are prepared by uniformly and intimately bringing
30 into association the active ingredients with liquid carriers or finely divided solid carriers or both, and then, if necessary, shaping the product.

 The compositions of the present invention may be formulated into any of many possible dosage forms such as, but not limited to, tablets, capsules, liquid syrups, soft gels, suppositories, and enemas. The compositions of the present invention may also be

formulated as suspensions in aqueous, non-aqueous or mixed media. Aqueous suspensions may further contain substances that increase the viscosity of the suspension including, for example, sodium carboxymethylcellulose, sorbitol and/or dextran. The suspension may also contain stabilizers.

5 In one embodiment of the present invention the pharmaceutical compositions may be formulated and used as foams. Pharmaceutical foams include formulations such as, but not limited to, emulsions, microemulsions, creams, jellies and liposomes. While basically similar in nature these formulations vary in the components and the consistency of the final product.

10 Agents that enhance uptake of oligonucleotides at the cellular level may also be added to the pharmaceutical and other compositions of the present invention. For example, cationic lipids, such as lipofectin (U.S. Pat. No. 5,705,188), cationic glycerol derivatives, and polycationic molecules, such as polylysine (WO 97/30731), also enhance the cellular uptake of oligonucleotides.

15 The compositions of the present invention may additionally contain other adjunct components found in pharmaceutical compositions. Thus, for example, the compositions may contain additional, compatible, pharmaceutically-active materials such as, for example, antipruritics, astringents, local anesthetics or anti-inflammatory agents, or may contain additional materials useful in physically formulating various dosage forms of the compositions of the present invention, such as dyes, flavoring agents, preservatives,
20 antioxidants, opacifiers, thickening agents and stabilizers. However, such materials, when added, should not unduly interfere with the biological activities of the components of the compositions of the present invention. The formulations can be sterilized and, if desired, mixed with auxiliary agents, e.g., lubricants, preservatives, stabilizers, wetting agents, emulsifiers, salts for influencing osmotic pressure, buffers, colorings, flavorings and/or
25 aromatic substances and the like which do not deleteriously interact with the nucleic acid(s) of the formulation.

 Certain embodiments of the invention provide pharmaceutical compositions containing (a) one or more antisense compounds and (b) one or more other chemotherapeutic agents that function by a non-antisense mechanism. Examples of such chemotherapeutic
30 agents include, but are not limited to, anticancer drugs such as daunorubicin, dactinomycin, doxorubicin, bleomycin, mitomycin, nitrogen mustard, chlorambucil, melphalan, cyclophosphamide, 6-mercaptopurine, 6-thioguanine, cytarabine (CA), 5-fluorouracil (5-FU), floxuridine (5-FUdR), methotrexate (MTX), colchicine, vincristine, vinblastine, etoposide, teniposide, cisplatin and diethylstilbestrol (DES). Anti-inflammatory drugs, including but not

limited to nonsteroidal anti-inflammatory drugs and corticosteroids, and antiviral drugs, including but not limited to ribivirin, vidarabine, acyclovir and ganciclovir, may also be combined in compositions of the invention. Other non-antisense chemotherapeutic agents are also within the scope of this invention. Two or more combined compounds may be used
5 together or sequentially.

Dosing is dependent on severity and responsiveness of the disease state to be treated, with the course of treatment lasting from several days to several months, or until a cure is effected or a diminution of the disease state is achieved. Optimal dosing schedules can be calculated from measurements of drug accumulation in the body of the patient. The
10 administering physician can easily determine optimum dosages, dosing methodologies and repetition rates. Optimum dosages may vary depending on the relative potency of individual oligonucleotides, and can generally be estimated based on EC₅₀s found to be effective in in vitro and in vivo animal models or based on the examples described herein. In general, dosage is from 0.01 µg to 100 g per kg of body weight, and may be given once or more daily,
15 weekly, monthly or yearly. The treating physician can estimate repetition rates for dosing based on measured residence times and concentrations of the drug in bodily fluids or tissues. Following successful treatment, it may be desirable to have the subject undergo maintenance therapy to prevent the recurrence of the disease state, wherein the oligonucleotide is administered in maintenance doses, ranging from 0.01 µg to 100 g per kg of body weight,
20 once or more daily, to once every 20 years.

EXPERIMENTAL

The following examples are illustrative, but not limiting, of the compounds, compositions, and methods of the present invention. Other suitable modifications and
25 adaptations of the variety of conditions and parameters normally encountered in clinical therapy and which are obvious to those skilled in the art are within the spirit and scope of the invention.

Example 1.

30 This example describes the methods for Examples 2 and 3.

RNAseq RNA was isolated from the MyLa cell lines using RNeasy kit from Qiagen (Valencia, CA, USA) and quantitated using the Agilent 2100 bio-analyzer (Agilent

Biotechnologies, Santa Clara, CA, USA). Samples with RNA integrity score (RIN) >6 were used for preparing cDNA libraries for transcriptome sequencing. Paired-end libraries for sequencing with Illumina Genome Analyzer II were prepared according to the protocol provided by Illumina, with minor modifications, with the mRNA-seq sample prep kit (Illumina). Sequence analysis was carried out using a combination of custom designed bioinformatics tools and Chimerascan (see, e.g., Iyer MK, Bioinformatics 2011;27:2903-4).

Clinical samples Clinical samples of primary cutaneous CD30-positive T-cell lymphoproliferative disorders (n = 47); mycosis fungoides (n = 44); systemic ALK-negative anaplastic large cell lymphoma (n = 44); ALK-positive anaplastic large cell lymphoma (n = 22); peripheral T-cell lymphoma, not otherwise specified (n = 24); angioimmunoblastic T-cell lymphoma (n = 7); extranodal NK/T-cell lymphoma (n = 5); enteropathy-associated T-cell lymphoma (n = 3); hepatosplenic T-cell lymphoma (n = 1); cutaneous gamma/delta T-cell lymphoma (n = 1) were obtained with IRB approval from the pathology archives of the University of Michigan, the University of Texas MD Anderson Cancer Center, Henry Ford Health System, and Seoul National University Hospital. Diagnoses were established by using 2008 World Health Organization criteria (see, e.g., Swerdlow SH, et al., WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues. Lyon: IARC; 2008). Patients with CD30-positive T-cell lymphoproliferative disorders were further subclassified as having either lymphomatoid papulosis or primary cutaneous anaplastic large cell lymphoma based on the combination of histologic and clinical features (see, e.g., Bekkenk MW, et al., Blood 2000;95:3653-61). 1mm diameter cores were taken in triplicate from tumor-rich areas of each formalin-fixed, paraffin-embedded specimen for tissue microarray construction and subsequent immunohistochemical and fluorescence in situ hybridization analysis.

Clinical information Clinical information was available for four of the seven patients demonstrating TYK2 translocations, including 1 patient with NPM1-TYK2 fusion (Fig. 1). The patient with NPM1-TYK2 fusion presented at the age of 71 with localized disease diagnosed as LYP. Phototherapy resulted in 6 years of remission. The patient's disease recurred and was controlled with methotrexate. Another patient presented at the age of 40 with localized disease diagnosed as LYP. After initial treatment with radiation therapy, he developed more widespread involvement that was currently controlled with methotrexate and topical steroids. A third patient presented at the age of 49 with disease diagnosed as LYP. Her disease was currently controlled with methotrexate. The fourth patient was diagnosed

with LYP at the age of 31 which progressed to cutaneous ALCL. The patient received radiation and methotrexate with stable disease. Of note, no difference in clinical behavior or presentation was discernable between cases with and without a TYK2 rearrangement.

5 Fluorescence in situ hybridization (FISH) Interphase FISH was performed to assess for TYK2 rearrangements and for specific NPM1-TYK2 fusions. Fluorescently labeled BAC clones were purchased from Empire Genomics, Buffalo, NY, USA. Tissue microarrays containing clinical samples were sectioned at 4µm and deparaffinized. The deparaffinized tissue microarray and cytospin slides of MyLa cells were hybridized with the probes
10 overnight and assessed by fluorescent microscopy. BAC clones used for FISH included TYK2 probes: RP11-177J4 labeled with ROX (5' of TYK2), RP11-347E20, labeled with fluorescein (3' of TYK2) and NPM1 probe RP11-1072I20, labeled with ROX (5' of NPM1).

15 Quantitative real time polymerase chain reaction (qRT-PCR) validation Reverse transcription. 1µg of total RNA was used for preparation of cDNA using SuperScript II reverse transcription kit (Life Technologies, Grand Island, NY, USA). The final product was diluted to 100µl in nuclease free water. 1µl of this diluted cDNA was used for quantitative real-time polymerase chain reaction (qRT-PCR) analyses.

20 SYBR green assay. qRT-PCR was performed using Power SYBR Green Mastermix (Applied Biosystems, Carlsbad, CA, USA) on an Applied Biosystems StepOne Plus Real-Time PCR System. All oligonucleotide primers were obtained from Integrated DNA Technologies (Coralville, IA). Control primers were used to amplify the GAPDH housekeeping gene. All assays were performed and repeated twice and results were plotted as
25 average fold change relative to GAPDH.

 Primers used for validation of NPM1-TYK2 fusion transcripts in cell lines by SYBR green assay:

 NPM-TYK2 Q F1 5'ACTCAAACCATCATCAACACCA3' (SEQ ID NO: 1)

 NPM-TYK2 Q R1 5'GTTCCGGCCACACACATTAC3' (SEQ ID NO: 2)

30 Sanger sequencing of NPM1-TYK2 fusion chimera NPM1-TYK2 fusion chimera was amplified from MyLa T-cell line using two different primer sets for NPM1 and TYK2. The primers were designed to amplify across the fusion chimera generating two different amplicons (175 bp and 468 bp) using Phusion DNA polymerase (New England Biolabs,

Ipswich, MA, USA) followed by conventional Sanger sequencing technology using BigDye version 3.1 chemistry run on an Applied Biosystems 3730xl DNA Sequencer. The PCR products were separated on a 1% agarose gel, bands excised and purified for Sanger sequencing.

5

Genomic PCR for genomic breakpoint mapping To identify the genomic break point in the index NPM1-TYK2 fusion event, a series of forward primers were designed that spanned from last exon of NPM1 present in NPM1-TYK2 fusion to the end of the following intron. Similarly, reverse primers were designed to span the region between the first exon of TYK2 in NPM1-TYK2 through the entire intron proximal to the exon. Genomic DNA was isolated from MyLa and HH cells using QIAamp DNA extraction kit (Valencia, CA, USA). 40ng of genomic DNA was used as template for performing genomic PCR using High Fidelity Phusion DNA Polymerase.

10

Primers used for amplifying across the juxtaposed NPM1 and TYK2 introns:

15

NPMg F 5'ACAGAGGGCGTATGAGACTTC3' (SEQ ID NO: 3)

TYK2g R 5'CCATAGGGACAGCCAATGTGAAAA3' (SEQ ID NO: 4)

Molecular cloning of NPM1-TYK2 full length fusion gene and transient expression in HEK293FT cells The full length wild-type NPM1-TYK2 fusion gene was amplified from

20

MyLa cell line using high fidelity Phusion DNA polymerase. The amplified gene product was then cloned into pLentilox IRES GFP mammalian expression Lenti virus vector (University of Michigan Vector Core, MI, USA) between Xho1 and Xba1 restriction sites. FLAG sequences were artificially introduced at the N-terminal end of the gene by PCR. NPM1-TYK2 K462R kinase dead mutant (see, e.g., Gauzzi MC, et al., The Journal of biological chemistry 1996;271:20494-500) was generated by performing site directed mutagenesis PCR using mutant primers and wild type NPM1-TYK2 gene as template. Sequence verified clones were transfected into HEK293FT cells and checked for expression of fusion protein after 48hrs post transfection using PolyJet (SignaGen, Rockville, MD, USA).

25

30

Cell culture The hematological cell lines used for such experiments were stocked and HEK293FT cells were obtained from American Type Culture Collection (ATCC, Manassas, VA, USA). T-cell derived cell lines were maintained in RPMI-1640 medium supplemented with 20% FBS, L-glutamine and penicillin/streptomycin (Thermo Scientific, Waltham, MA, USA). HEK293FT cells were maintained in Dulbecco's modified Eagle's medium (DMEM)

supplemented with 10%FBS, glutamine and penicillin/streptomycin. Stable TYK2 knockdown MyLa cell lines were generated by transducing the lentivirus supernatant containing shRNAs for TYK2 and selecting them with puromycin (1µg/mL) for 48hrs post transduction.

5

Immunoblotting The levels of pSTAT1, pSTAT3, pSTAT5 and pTYK2 were analysed in cell lines after treating them with 1mM activated Na₃V04 for 1hr prior harvesting. The lysates used in the immunoblotting experiments were prepared from the cell lines using RIPA buffer containing 50mM Tris HCl, pH 7.4; 1% NP-40; 0.25% Na-deoxycholate; 150mM NaCl; 1mM PMSF; 1mM EDTA; 1mM Na₃V0₄; 1mM NaF; and 0.1% SDS. The samples were incubated on ice for 30 min followed by centrifugation at 10,000x for 15 min. The protein content of the supernatants was estimated using a protein BCA assay kit (Pierce protein research products, Rockford, IL, USA). 20µg of protein extract was separated on a high resolution SDS PAGE using MES SDS running buffer, and analyzed for expression of proteins by immunoblotting using antibodies specific for monoclonal ANTI-FLAG M2 (clone M2) antibody (Sigma, St. Louis, MO, USA). pSTAT1 (Y701), STAT1, pSTAT3 (Y705), STAT3, pSTAT5 (Y694), STAT5, pTYK2 (Y1054/55) and TYK2 antibodies were purchased from Cell Signaling Technologies (Beverly, MA, USA) and used in our study.

20

Transcriptional activation assay DNA constructs containing vector and NPM1-TYK2 wild type and K462R mutant were co-transfected along with luciferase reporter system plasmids sensitive to appropriate STAT proteins (pGL4.52 luc2P/STAT5RE; pGL4.45 luc2P/ISRE; Promega; Madison, WI, Cignal STAT3 reporter, SA Biosciences, Valencia, CA) to HEK293FT cells. After 48hrs post transfection, the cells were lysed and transcriptional activity of STAT proteins were measured by using the Dual Luciferase kit (Promega) and a specially configured luminometer (Berthold Technologies; Germany). Western blotting was performed using the extracts to ensure equal expression of different constructs (data not shown).

30

Stable knockdown of TYK2 expression in MyLa Lenti-viral mediated gene transduction was performed as described earlier (see, e.g., Sahasrabudhe AA, et al., Oncogene 2014). Briefly, 293FT cells were co-transfected with lentiviral construct and packaging vectors (Invitrogen). Virus-producing supernatant was harvested 48 h after

transfection and supplemented with polybrene and used to infect MyLa cells. Lentiviral mission shRNA constructs targeting different region of TYK2 (TRCN0000003123-5'CGTGAGCCTAACCATGATCTT3' (SEQ ID NO: 5) and TRCN0000320620-5'GAGTGCCTGAAGGAGTATAAG3' (SEQ ID NO: 6)) and Vector control (SHC201) were purchased from Sigma (St.Louis, MO, USA).

Cell proliferation assay Cell proliferation assay was performed using water soluble tetrazolium-1 (WST-1) procured from Roche Applied Sciences, Indianapolis, IN, USA, based on manufacturer's protocol (see, e.g., McDonnell SR, et al., Oncogene 2011). Briefly, equal numbers of cells were plated in 96-well plates with growth media and 10µl of WST-1 reagent was added at different time intervals (0, 12, 36, 48 hours). After addition of WST-1 reagent, the plates were incubated at 37°C in an incubator with 5% CO₂ for 2 hours. Based on the cleavage of the tetrazolium salt WST-1 by mitochondrial dehydrogenases in viable cells, the intensities of the color developed in each well was determined by measuring the absorbance at 440nm. The assay was performed in triplicate for all the samples at each time point.

Screening for somatic mutations in TYK2 by Sanger resequencing PCR fragments targeting exon 14 to 16 and 20 to 23, were amplified from genomic DNA of CD30 positive LPD samples (n=25). Genomic DNA was extracted using both the QIAGEN BioRobot EZ1 and QIAamp FFPE DNA extraction kits (QIAGEN). For all Sanger sequencing reactions, PCR amplification was performed using Phusion DNA polymerase (New England Biolabs) followed by conventional Sanger sequencing technology using BigDye version 3.1 chemistry run on an Applied Biosystems 3730xl DNA Sequencer at the University of Michigan DNA sequencing Core. All sequencing reactions were performed using nested sequencing primers. Sequencing trace analysis was performed using Mutation Surveyor software. All mutations were verified in at least two independent PCR amplification and sequencing reactions.

Immunostaining for NPM1 and pSTAT5 Immunohistochemical staining was performed on formalin-fixed, paraffin-embedded tissue microarrays containing patient samples using an automated immunostainer (Autostainer Link, Dako, Carpinteria, CA). Following deparafinization and heat-induced epitope retrieval, the slides were incubated with either anti-NPM1 (1:800 dilution, clone EP1848Y, Abcam, Cambridge, MA) or anti-pSTAT5 (1:400 dilution, clone AX1, Advantex BioReagents LLC, Houston, TX) for 60 min. Staining was visualized with Dako Envision detection system and diaminobenzadine chromagen.

Immunofluorescence analysis Cultured suspension cells from HH and MyLa cell lines were deposited onto glass slides by cytocentrifugation. Cells were fixed and permeabilized with methanol/glacial acetic acid (3:1). After fixation, cells were first incubated with the following primary antibodies (Cell Signaling Technology): Y701 phospho-STAT1 (clone 58D6, rabbit monoclonal IgG), Y705 phospho-SATA3 (mouse monoclonal IgG1, clone M9C6) or Y694/Y699 phospho-STAT5 monoclonal antibody (rabbit monoclonal IgG, clone D47E7). The secondary antibodies were Alexa-conjugated donkey anti-rabbit IgG or Alexa-conjugated goat anti-mouse IgG (Alexa594, Life Technologies). Coverslips were mounted on standard slides with mounting media supplemented with 4', 6'-diamidino-2-phenylindole (DAPI). The images were captured and recorded using an Olympus BX-51 upright light microscope equipped with an Olympus DP-70 camera.

Example 2.

Patient biopsy samples were obtained with institutional review board approval. RNA was subjected to chimera analysis by producing paired end libraries sequenced on the Illumina Genome Analyzer II. Sequencing data was analyzed using custom bioinformatics tools and Chimerascan software (see, e.g., Iyer MK, et al., Bioinformatics 2011;27:2903-4). Sequencing confirmation of NPM1-TYK2 fusion transcripts was achieved using SYBR Green- based quantitative real-time polymerase chain reaction (qRT-PCR) assays and Sanger sequencing of amplicons by reverse transcription (RT)-PCR. Fluorescence In Situ Hybridization (FISH) was performed on tissue microarrays of primary patient samples using standard methods. TYK2 break-apart FISH and NPM1-TYK2 fusion FISH assays were designed to detect TYK2 rearrangements and NPM1-TYK2 fusions, respectively. Immunohistochemistry for pSTAT5 (AX1, Advantex BioReagents LLC, Houston, TX) and NPM1 (EP1848Y, Abcam, Cambridge, MA) was performed on primary patient samples. Mutations in TYK2 were assessed by targeted Sanger sequencing of the genomic DNA. The NPM1-TYK2 fusion gene was amplified from MyLa cells and cloned into a mammalian expression vector for functional studies. Stable MyLa cell lines depleted of TYK2 were generated using lenti-viral mediated gene transduction (see, e.g., Sahasrabudde AA, et al., Oncogene 2014) of shRNAs and used in cell proliferation assays.

Example 3.

To discover novel gene fusions that may be involved in the pathogenesis of cutaneous T-cell lymphoproliferative disorders, RNAseq was performed on the cutaneous T-cell lymphoma-derived cell line MyLa. Interestingly, RNA sequencing analysis of MyLa revealed 803 total fragments representing a chimeric fusion between the N-terminal region of NPM1 and the C-terminal region of TYK2 (Fig. 2A, upper panel; Fig. 3). RT-PCR analysis confirmed high levels and exclusive expression of the NPM1-TYK2 RNA chimera in MyLa, but not in 11 other hematological cell lines (Fig. 4; Fig. 5). The fusion occurred at NPM1 NM_002520 c.1016 (exon 9) and TYK2 NM_003331 c.2554 (exon 16) and was confirmed by Sanger sequencing (Fig. 2A, lower panel). The fusion gene was predicted to yield a product comprising the entire oligomerization, histone binding domains and the distal portion of the DNA/RNA binding domains of NPM1 and a small portion of the TYK2 pseudokinase domain and the entire C-terminal kinase domain (Fig. 2B).

To precisely map the genomic breakpoints, a conventional PCR-based assay using a series of forward primers starting from the C-terminal part of exon 9 through intron 10 of NPM1 (positive strand) and reverse primers located from exon 16 through intron 16 of TYK2 (negative strand) was designed. Bi-directional Sanger sequencing of the PCR product revealed break points at position chr5:170,832,813 within the NPM1 gene and chr19:10,469,817 in the TYK2 gene (Fig. 1C; Fig. 1D; Fig. 6). TYK2 break-apart and NPM1-TYK2 fusion FISH assays confirmed the presence of the NPM1-TYK2 rearrangement in MyLa (Fig. 2E).

Next the TYK2 break-apart FISH assay was employed to screen for TYK2 gene rearrangements in a large cohort of primary mature T-cell LPD/lymphomas (n=198). TYK2 rearrangements were identified in 7/47 (15%) primary CD30-positive LPD including 3/15 (20.0%) cases of LYP and 4/32 (12.5%) cases of primary cutaneous ALCL. Each of the 7 cases that were identified by TYK2 break-apart FISH was also evaluated with the NPM1-TYK2 fusion FISH assay. Two cases (1 LYP, 1 cutaneous ALCL) harbored the NPM1-TYK2 rearrangement (Fig. 2E; Fig. 2F). These results indicate that TYK2 is targeted by multiple different translocation partners in CD30-positive LPD. This is not unexpected since other prominent oncogenes are known to utilize multiple translocation partners (see, e.g., Mitelman F, et al., *Nature reviews Cancer* 2007;7:233-45; Chen YW, et al., *Blood* 2003;102:1931-2; author reply 2; Harper DP, et al., *Cancer research* 2008;68:10024-7; Chiarle R, et al., *Nature reviews Cancer* 2008;8:11-23). Of note, TYK2 translocations were absent in all (n = 151) cases of other mature T-cell lymphomas including mycosis fungoides (n = 44); systemic ALK-negative ALCL (n = 44) systemic ALK-positive ALCL (n = 22);

peripheral T-cell lymphoma, not otherwise specified (n = 24); angioimmunoblastic T-cell lymphoma (n = 7); extranodal NK/T-cell lymphoma (n = 5); enteropathy-associated T-cell lymphoma (n = 3); hepatosplenic T-cell lymphoma (n = 1) and cutaneous gamma/delta T-cell lymphoma (n = 1). To investigate whether point mutations may represent a mechanism for
5 activation of TYK2, targeted sequencing of TYK2 pseudokinase and kinase domains including previously reported mutations in T-lymphoblastic lymphoma cell lines was performed (see, e.g., Sanda T, et al., Cancer discovery 2013;3:564-77). No mutations were identified in TYK2 in 25 cases of CD30-positive LPD.

In the NPM1-TYK2 fusion reported herein, the 5' partner of the fusion protein NPM1
10 retained the oligomerization domain (Fig. 2B) which was hypothesized to lead to constitutive activation of TYK2 through a mechanism similar to that observed with NPM1-ALK (see, e.g., Pearson JD, et al., Journal of signal transduction 2012;2012:123253; Bischof D, et al., Molecular and cellular biology 1997;17:2312-25). The pTYK2 (Y1054/1055) levels in the NPM1-TYK2-positive MyLa cell line and in other T-cell lines were evaluated, and confirmed
15 high levels of pTYK2 in only MyLa indicating constitutive activation of TYK2 kinase (Fig. 7A). Notably, western blot analysis for NPM1 and total TYK2 demonstrates the presence of a fusion protein (81kD) supporting the expression of the NPM1-TYK2 fusion protein. Further, immunostaining for NPM1 in NPM1-TYK2 fusion positive cases revealed cytoplasmic reactivity and mislocalization of NPM1, a predominantly nuclear protein, indicating
20 abnormal subcellular localization (Fig. 8). To investigate whether the NPM1-TYK2 fusion conferred constitutive activation of STAT signaling, western blot analysis and immunofluorescence microscopy was performed which revealed that STAT-family members (pSTAT1, pSTAT3 and pSTAT5) downstream of TYK2 are constitutively activated in the MyLa cell line (Fig. 7B; Fig. 9). Importantly, constitutively activated STAT signaling was
25 confirmed in primary NPM1-TYK2 fusion positive CD30 positive LPD tissue samples (n= 2), where nuclear expression of pSTAT5 was observed compared to TYK2 negative cases (Fig. 10).

FLAG-tagged NPM1-TYK2 and NPM1-TYK2 (K462R) kinase-defective mutant proteins were ectopically expressed in HEK293FT cells. Expression of NPM1-TYK2 protein
30 demonstrated constitutive TYK2 autophosphorylation whereas the kinase-defective NPM1-TYK2 (K462R) mutant completely abolished its phosphorylation (Fig. 7C). Similarly, pSTAT1, pSTAT3 and pSTAT5 were induced by the kinase-active NPM1-TYK2 while the kinase-defective NPM1-TYK2 (K462R) mutant completely abrogated their phosphorylation. Activation of STAT1/3/5 specifically by NPM1-TYK2 was further corroborated by

transcriptional activation (reporter) assays (Fig. 7D). Altogether, these results indicate that expression of the NPM1-TYK2 fusion protein results in constitutive activation of the STAT signaling pathway.

The oncogenic potential of NPM1-TYK2 was further assessed by examining the effect of TYK2 knockdown on STAT1/3/5 activation and cell proliferation. Lentivirus-based shRNA-mediated knockdown of TYK2 in MyLa resulted in decreased pSTAT1, pSTAT3 and pSTAT5 without affecting their total protein levels (Fig. 7E). Importantly, a significant decrease (2.2 fold) in cell proliferation was observed compared to vector control ($P < 0.01$) (Fig. 7F).

Example 4.

In some embodiments, the following primers are used for cloning in pCAGGS:

NPM 1 exon 1 ecor1 F

AAgaattcAGAAAGGAGTGGGGTTGAAAAGCGCTTG (SEQ ID NO: 7)

TYK2 exon 23 xho1 R

AActcgagTCCAGCAGAGAAAACATGAGTTTATTACCAGATGG (SEQ ID NO: 8).

In some embodiments, the following primers are used for cloning in plentilox IRES GFP:

NPM1 xho1 flag F

AActcgagATGGACTACAAGGACGACGATGACAAGGAAGATTTCGATGGACATGGACATGAGCCC (SEQ ID NO: 9)

TYK2 xba1 stop R

AAtctagaTCAGCACACGCTGAACACTGAAGGGGC (SEQ ID NO: 10).

In some embodiments, the following PCR primers are used for PCR amplification for sequencing a NPM1-TYK2 gene fusion:

NPM1 PCR

AAGCGCCAGTGAAGAAATCT (SEQ ID NO: 11)

TYK2 PCR

CCTCTCCGTCAAAGCAGATC (SEQ ID NO: 12).

In some embodiments, the following primers are used for NPM1-TYK2 Q-PCR:

NPM-TYK2 Q F1

ACTCAAAACCATCATCAACACCA (SEQ ID NO: 1)

5 NPM-TYK2 Q R1

GTTCCGGCCACACACATTAC (SEQ ID NO: 2)

NPM-TYK2 Q F2

CACCAAAAGGACCTAGTTCTGT (SEQ ID NO: 13)

NPM-TYK2 Q R2

10 CTCAGCTTGATGAAGGGGCT (SEQ ID NO: 14)

NPM-TYK2 Q F3

ACCATCATCAACACCAAGATCA (SEQ ID NO: 15)

NPM-TYK2 Q R3

GCCAGGATCACTCAGCTTGA (SEQ ID NO: 16).

15 In some embodiments, NPM1-TYK2 Q-PCR yields the following sequencing result:

GCTGGAAAGTAGATGCTGGCTGTTGCTTGGTATTATAGGGAGGGCCAGGAGCCC
 AGGAGTTCGAGACCAGCCTGGGCAACAGGGCGAGACCCCATCTTTTTTGTGTTGTT
 TTGTTTTAGGTGGAGTTTCGCCCTGTACCCAGGCTGGAGTGCAATGGCATGATC
 20 TCGGCTCACTGCAACCTTCGCCTCCCGGGTTCAAATGATTCTCCCGCCTCAGCCTC
 CCAAGTAGCTGGGATTATAGGCTCCTGCCATCACGCCAGCTAATTTTTGTATTTT
 TAGTAGAGATGGGGTTTCACCATGTTGGTCAGGCTGGTCTCGAACTTCTGACCTC
 GTGATCCATCCGCCTCAGCCTCCCAAAGTGCTGGGATTCCAGGCGTGAGCCACCA
 CGCCCAGCTGGCCCCATCTTTTAAAAATAAACAAATAGCCAGGTGTGGTGGCTCA
 25 TACCTGTAATCCCAGCAGTTTGGGAGGCCAAGGCGGGTGGATCACCTGAGGTCA
 GGAGTTTGAGACCAGCCTGGCCAGTATGGCAAAACCTTGCCTCTACTAAAAATAC
 AAAAAAATTAGCCAATGTGGTAACCTGCACCTGTAGGCCGAGGTACTCAGGAA
 GTTGAGGTGGGAGGATCACCTGAGCTCAGGAATTTGAGGCCGTAGTAAGCTATG
 ATCACACCACTGCACCCCAGCCTGGGCAGCAGCATAGCTAGACCCCATCTCCACC
 30 AAAAATTTAAGAATCAGCTAGGCTGTGGTGATGTGCACCTGTAATTCTCGCTACT
 TGGAAGGCTTGAGCCCAGGAGTTTGAAGCTGCAGTGAGCTATGTTTCGTGCCACTA
 TACTCCAACCTGAAAGACAGAGTGAGACCCTGTCTT (SEQ ID NO: 17).

In some embodiments, the NPM1-TYK2 gene fusion break point on sequencing is provided:

Gtgagtaaagttatcttaaaaaaactttgtctccccctcaaattgcacgtgtctggttgcatagacttgaatgttctgtttttgtttgttt
 ggtttaataacttgccctgggtcggtatgaatttttcaaaaatttctataaaacatttataatcgtgtctgtggtgattttgcatatgcaaaa
 ttaaatatgcctattttccattatgaaggaacgtagtcactgggtgcaagataacattctgacctccatgttaaaatagatcagtga
 accctttgcctattctgggtgaagatatgtagagaaccaacagagggcggtatgagacttcattaaaattacaacagctggaaaagta
 5 gatgtcggctgttgcctgtggtattatagggagggccaggagcccaggagttcgagaccagcctgggcaacagggcgagaccccatctt
 tttgtttgtttgttttaggtggagtttcgacctgtcaccaggctggagtgaatggcatgatctcggtcactgcaaccttcgctcccg
 gggtcaaatgattctccgcctcagcctcccaagtagctgggattataggtcctgccatcacgcccagctaattttgtatttttagtagag
 atgggggttcaccatgttggtcaggctgggtcgaactctgacctcgtgatccatccgctcagcctcccaagtgctgggattccagg
 cgtgagccaccacgcccagctggccccatctttaaaaataaacaatatagccaggtgtggtggctcatacctgtaatcccagcagttg
 10 ggaggccaaggcgggtggatcacctgaggtcaggagttgagaccagcctggccagtatggcaaaccttccttactaaaaatac
 aaaaaaattagccaatgtggttaactgcacctgtaggccgaggtactcaggaagttgaggtgggaggatcacctgagctcaggaatt
 tgaggccgtagtaagctatgatcacaccactgcaccccagcctgggcagcagcatagctagaccccatctccacaaaaatttaagaa
 tcagctaggctgtggtgatgtgcacctgtaattctcgctacttgaaggcttgagcccaggagtttgaagctgcagtgcagctatgttcgtg
 ccactatactccaacctgagagacagagtgcacctgtcttaaaaaaaaaaaaaaaaaaaaaaacaagaaaaaataacccccca
 15 aaaaacaaaacagaaagaacccagtgatagtcacagttgtcctttcacattggctgtccctatggggaccagggttgggggtggcgtc
 tgtgcctctcctgagtggccacacccctctcctgccacctcag (SEQ ID NO: 18).

In some embodiments, the following primers are provided:

NPMg F1 exonic primer

CACCAAAAGGACCTAGTTCTGT (SEQ ID NO: 13)

20 NPMg F2

cttgcctgggttcgtggtatg (SEQ ID NO: 19)

NPMg F3

aaccaacagagggcggtatga (SEQ ID NO: 20)

NPMg F4

25 tctcatctgcttgactgggg (SEQ ID NO: 21)

NPMg F5

gcaaatgtgggtggtgaaa (SEQ ID NO: 22)

NPMg F6

ttttagatgcccccccc (SEQ ID NO: 23)

30 NPMg F7

ttctgatttgccacccatgc (SEQ ID NO: 24)

NPMg F8

ttctgcaaatgccccttgga (SEQ ID NO: 25)

NPMg F9

accgagttgccatgtttgtt (SEQ ID NO: 26)

NPMg F10

tctaaaggtatctctctcggtgt (SEQ ID NO: 27)

TYK2g R1

5 ggatggatcacgaggtcaga (SEQ ID NO: 28)

TYK2g exonic R2

GTTCCGGCCACACACATTAC (SEQ ID NO: 2)

NPMg F1 exonic primer long	GAAAAAACTCCTAAAACACCAA AAGGACCTAGTTCTGTAGAAG (SEQ ID NO: 29)
TYK2g exonic R2 long	GTAATGTGTGTGGCCGGAAC ATCCTGCTGGCCCG (SEQ ID NO: 30)

10 TYK2 K930R F

ACTGGCGAGATGGTGGCGGTGAGAGCCCTCAAGGCAGACTGCGGC (SEQ ID NO:
31)

TYK2 K930R R

GCCGCAGTCTGCCTTGAGGGCTCTCACCGCCACCATCTCGCCAGT (SEQ ID NO:
15 32)

TYK2 Y1054Fe F

GCCGTGCCCCGAAGGCCACGAGTTCTACCGCGTGCGCGAGGATGGGG (SEQ ID NO:
33)

TYK2 Y1054Fe R

20 CCCCATCCTCGCGCACGCGGTAGAACTCGTGGCCTTCGGGCACGGC (SEQ ID NO:
34)

TYK2 Y1054/55Fe F

GCCGTGCCCCGAAGGCCACGAGTTCTTCCGCGTGCGCGAGGATGGGG (SEQ ID NO:
35)

25 TYK2 Y1054/55Fe R

CCCATCCTCGCGCACGCGGAAGAACTCGTGGCCTTCGGGCACGGC (SEQ ID NO:
36).

All publications, patents, patent applications and accession numbers mentioned in the above specification are herein incorporated by reference in their entirety. Although the invention has been described in connection with specific embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific
5 embodiments. Indeed, various modifications and variations of the described compositions and methods of the invention will be apparent to those of ordinary skill in the art and are intended to be within the scope of the following claims.

CLAIMS

We claim:

1. A method for diagnosing a neoplastic disorder or susceptibility to a neoplastic
5 disorder comprising detecting the presence of a gene fusion between TYK2 and a TYK2-
gene-fusion-partner in a nucleic acid sample from a subject, and diagnosing the subject as
having or being susceptible to a neoplastic disorder when the gene fusion is present.
2. The method of Claim 1, wherein the neoplastic disorder is a cutaneous CD30-positive
10 lymphoproliferative disorder.
3. The method of Claim 2, wherein the cutaneous CD30-positive lymphoproliferative
disorder is lymphomatoid papulosis.
- 15 4. The method of Claim 2, wherein the cutaneous CD30-positive lymphoproliferative
disorder is primary cutaneous anaplastic large cell lymphoma.
5. The method of Claim 1, wherein the gene fusion comprises at least a portion of the
C-terminal kinase domain of TYK2.
20
6. The method of Claim 1, wherein the wherein the TYK2-gene-fusion-partner is NPM1.
7. The method of Claim 6, wherein the gene fusion is an NPM1-TYK2 gene fusion.
- 25 8. The method of Claim 7, wherein the NPM1-TYK2 gene fusion is a fusion between at
least a portion of exon 9 of NPM1 and at least a portion of exon 15 of TYK2.
9. The method of Claim 7, wherein the NPM1-TYK2 gene fusion comprises the
oligomerization domain of NPM1.
30
10. The method of Claim 7, wherein the NPM1-TYK2 gene fusion comprises the histone
binding domain of NPM1.

11. The method of Claim 7, wherein the NPM1-TYK2 gene fusion comprises the DNA/RNA binding domain of NPM1.
12. The method of Claim 7, wherein the NPM1-TYK2 gene fusion comprises the
5 pseudokinase domain of TYK2.
13. The method of Claim 7, wherein the NPM1-TYK2 gene fusion comprises the kinase domain of TYK2.
- 10 14. The method of Claim 7, wherein the NPM1-TYK2 gene fusion comprises one or more of the following domains: the oligomerization domain of NPM1, the histone binding domain of NPM1, the DNA/RNA binding domain of NPM1, the the pseudokinase domain of TYK2, and the kinase domain of TYK2.
- 15 15. The method of Claim 7, wherein the NPM1-TYK2 gene fusion is N5;T19.
16. The method of Claim 1, wherein the sample is selected from the group consisting of: plasma, serum, and tissue.
- 20 17. The method of claim 1, wherein the detecting comprises amplifying SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, and/or a fragment thereof.
18. The method of claim 1, wherein the detecting comprises PCR, RT-PCR, or nucleic acid hybridization.
25
19. The method of claim 1, wherein the detecting comprises reverse transcriptase PCR.
20. The method of claim 1, wherein the RT-PCR employs a detectably labeled primer.
- 30 21. The method of claim 1, wherein the detecting is accomplished with electrophoresis.
22. The method of claim 1, wherein the detecting is accomplished using a real-time PCR detection system.

23. A kit for detecting a NPM1-TYK2 gene fusion in a sample comprising at least one oligonucleotide comprising the sequence of SEQ ID NOs: 1, 2, 3, 4, or complements thereof.

24. An antibody that specifically binds to the polypeptide encoded by a NPM1-TYK2 gene fusion.

25. A kit for detecting gene fusions associated with cutaneous CD30-positive lymphoproliferative disorders in a subject, comprising: at least a first gene fusion informative reagent for identification of a gene fusion between TYK2 and a TYK2-gene-fusion-partner.

26. The kit of claim 25, wherein said reagent is a probe that specifically hybridizes to the fusion junction of a gene fusion involving TYK2 and a TYK2-gene-fusion-partner.

27. The kit of claim 25, wherein said reagent is a pair of primers that amplify a fusion junction of a gene fusion between TYK2 and a TYK2-gene-fusion-partner.

28. The kit of claim 27, wherein said pair of primers comprise a first primer that hybridizes to a TYK2 nucleic acid and second primer that hybridizes to nucleic acid from the TYK2-gene-fusion-partner.

29. The kit of claim 25, wherein said reagent is an antibody that binds to the fusion junction of a fusion polypeptide encoded by a gene fusion between TYK2 and a TYK2-gene-fusion-partner.

30. The kit of claim 25, wherein the reagent is a sequencing primer that binds to a gene fusion between TYK2 and a TYK2-gene-fusion-partner and generates an extension product that spans the fusion junction of said gene fusion.

31. The kit of Claim 30, wherein the gene fusion is NPM1-TYK2.

32. The kit of claim 31, wherein said reagent comprises a pair of probes wherein said first probe hybridizes to NPM1 and said second probe hybridizes to TYK2.

33. The kit of claim 32, wherein the first probe comprises SEQ ID NO: 1 or a fragment thereof, and the second probe comprises SEQ ID NO: 2, or a fragment thereof.
34. The kit of claim 32, wherein the first probe comprises SEQ ID NO: 3 or a fragment thereof, and the second probe comprises SEQ ID NO: 4, or a fragment thereof.
- 5

FIG. 1

FISH Results and Clinical Data for Cutaneous CD30-Positive Lymphoproliferative Disorders

Case	Diagnosis	FISH Results			Age (years)	Sex	Site(s)	Clinical Data	
		TYK2 Break-Apart	Percent Positive	NPM1-TYK2 Fusion				Treatment Received	Last Known Status
1	PC-ALCL	Negative	0%	N/P	47	Male	Lower extremity, regional lymph node	Radiation, cyclophosphamide, adriamycin, vincristine, prednisone	Remission
2	PC-ALCL	Negative	0%	N/P	28	Female	Lower extremity	Radiation, cyclophosphamide, adriamycin, vincristine, prednisone	Remission
3	LyP	Negative	0%	N/P	65	Female	Trunk, upper and lower extremities	Radiation, topical steroids	Persistent disease
4	LyP	Negative	0%	N/P	63	Male	Upper extremity	Methotrexate	Persistent disease
5	LyP	Negative	0%	N/P	12	Female	Upper and lower extremities	N/A	Persistent disease
6	LyP	Negative	0%	N/P	50	Female	Trunk, upper and lower extremities, vulva	Methotrexate, topical steroids, phototherapy	Persistent disease
7	LyP	Negative	0%	N/P	59	Male	Trunk, upper and lower extremities	Radiation, methotrexate, bexarotene	Deceased, renal failure
8	LyP	Negative	0%	N/P	37	Female	Face, neck, upper and lower extremities	Topical steroids	Persistent disease
9	LyP	Negative	0%	N/P	55	Male	Trunk, upper and lower extremities, scalp	Topical steroids	Persistent disease
10	LyP	Positive	75%	Negative	40	Male	Trunk, upper and lower extremities	Radiation, methotrexate, topical steroids	Persistent disease
11	LyP	Negative	0%	N/P	56	Male	Upper extremity	Radiation, methotrexate, topical steroids	Persistent disease
12	PC-ALCL	Negative	0%	N/P	55	Male	Trunk, upper and lower extremities	Radiation, topical steroids	Persistent disease
13	PC-ALCL	Negative	0%	N/P	56	Female	Trunk, upper and lower extremities	Radiation, topical steroids	Persistent disease
14	LyP	Positive	85%	Positive	71	Male	Trunk, upper and lower extremities	Radiation, methotrexate, phototherapy	Persistent disease

FIG. 1 (Cont.)

Case	Diagnosis	FISH Results			Clinical Data					Last Known Status
		TYK2 Break-Apart	Percent Positive	NPM1-TYK2 Fusion	Percent Positive	Age (years)	Sex	Site(s)	Treatment Received	
15	LyP	Negative	0%	N/P	0%	62	Male	Lower extremity	Radiation, topical steroids	Persistent disease
16	LyP	Negative	0%	N/P	0%	46	Female	Upper and lower extremities, buttocks	Methotrexate	Persistent disease
17	PC-ALCL	Negative	0%	N/P	0%	74	Male	Upper back	Radiation	Remission
18	PC-ALCL	Negative	0%	N/P	0%	80	Male	Upper extremity, regional lymph node	Radiation, cyclophosphamide, adriamycin, vincristine, prednisone, etoposide, carboplatin, palatrexate	Persistent disease
19	LyP	Positive	30%	Negative	0%	49	Female	Trunk, upper and lower extremities	Methotrexate	Persistent disease
20	LyP	Negative	0%	N/P	0%	81	Female	Upper and lower extremities, buttocks	Topical steroids, phototherapy	Persistent disease
21	LyP	Negative	0%	N/P	0%	45	Male	Lower extremities, buttocks, axilla	Topical steroids, phototherapy	Persistent disease
22	PC-ALCL	Negative	0%	N/P	0%	56	Female	N/A	N/A	N/A
23	PC-ALCL	Negative	0%	N/P	0%	87	Female	N/A	N/A	N/A
24	PC-ALCL	Positive	20%	Negative	0%	53	Female	N/A	N/A	N/A
25	PC-ALCL	Negative	0%	N/P	0%	25	Female	N/A	N/A	N/A
26	PC-ALCL	Negative	0%	N/P	0%	56	Male	N/A	N/A	N/A
27	PC-ALCL	Negative	0%	N/P	0%	78	Female	N/A	N/A	N/A
28	PC-ALCL	Positive	80%	Negative	0%	31	Male	Thigh	Radiation, methotrexate	N/A
29	PC-ALCL	Negative	0%	N/P	0%	43	Male	N/A	N/A	N/A
30	PC-ALCL	Positive	90%	Positive	90%	N/A	N/A	N/A	N/A	N/A

FIG. 1 (Cont.)

Case	Diagnosis	FISH Results				Clinical Data				
		TYK2 Break-Apart	Percent Positive	NPM1-TYK2 Fusion	Percent Positive	Age (years)	Sex	Site(s)	Treatment Received	Last Known Status
1	PC-ALCL	Negative	0%	Negative	0%	N/A	N/A	N/A	N/A	N/A
32	PC-ALCL	Positive	90%	Negative	0%	N/A	N/A	N/A	N/A	N/A
33	PC-ALCL	Negative	0%	Negative	0%	N/A	N/A	N/A	N/A	N/A
34	PC-ALCL	Negative	0%	Negative	0%	N/A	N/A	N/A	N/A	N/A
35	PC-ALCL	Negative	0%	Negative	0%	N/A	N/A	N/A	N/A	N/A
36	PC-ALCL	Negative	0%	Negative	0%	N/A	N/A	N/A	N/A	N/A
37	PC-ALCL	Negative	0%	Negative	0%	N/A	N/A	N/A	N/A	N/A
38	PC-ALCL	Negative	0%	Negative	0%	N/A	N/A	N/A	N/A	N/A
39	PC-ALCL	Negative	0%	Negative	0%	N/A	N/A	N/A	N/A	N/A
40	PC-ALCL	Negative	0%	Negative	0%	N/A	N/A	N/A	N/A	N/A
41	PC-ALCL	Negative	0%	Negative	0%	N/A	N/A	N/A	N/A	N/A
42	PC-ALCL	Negative	0%	Negative	0%	N/A	N/A	N/A	N/A	N/A
43	PC-ALCL	Negative	0%	Negative	0%	N/A	N/A	N/A	N/A	N/A
44	PC-ALCL	Negative	0%	Negative	0%	N/A	N/A	N/A	N/A	N/A
45	PC-ALCL	Negative	0%	Negative	0%	N/A	N/A	N/A	N/A	N/A
46	LYP	Negative	0%	Negative	0%	58	Male	Upper back	Topical steroids	Remission
47	LYP	Negative	0%	Negative	0%	49	Female	Lower extremities, scalp	Methotrexate, topical steroids	Persistent disease

FISH, fluorescence in-situ hybridization; PC-ALCL, Primary cutaneous anaplastic large cell lymphoma; LYP, Lymphomatoid Papulosis; N/P, not performed; N/A, information not available

FIG. 2A

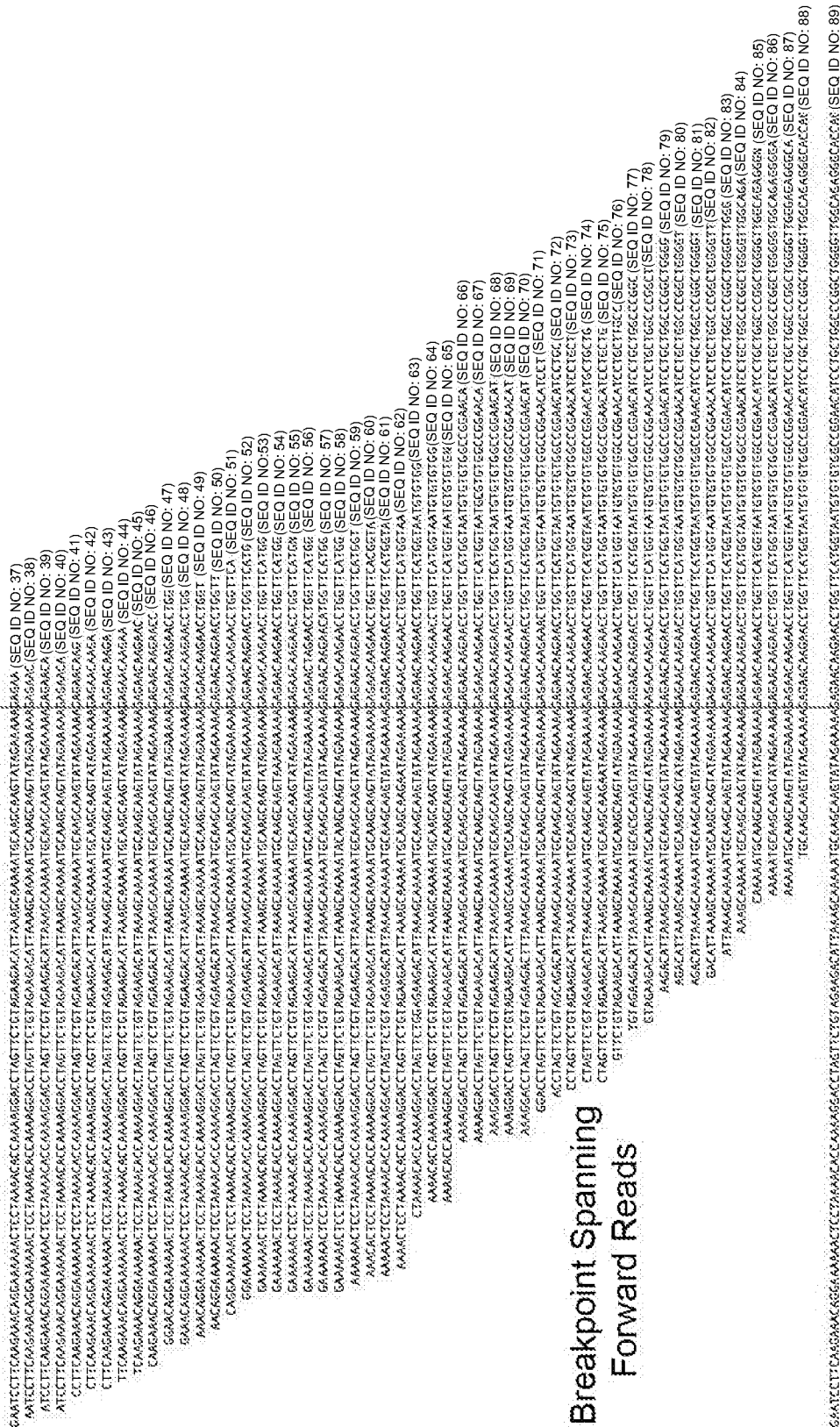


FIG. 2B

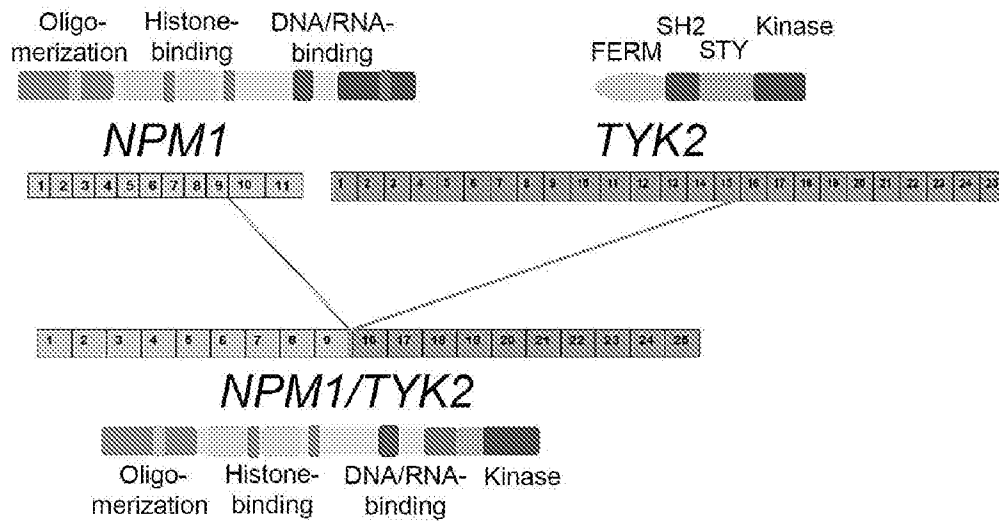


FIG. 2C

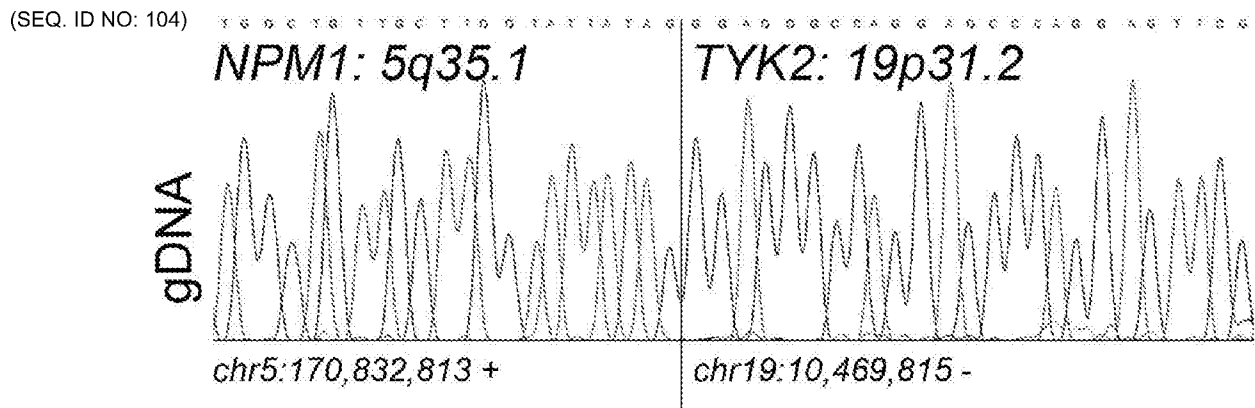


FIG. 2D

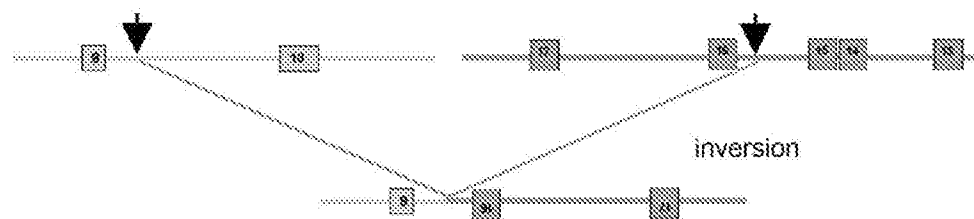


FIG. 2E

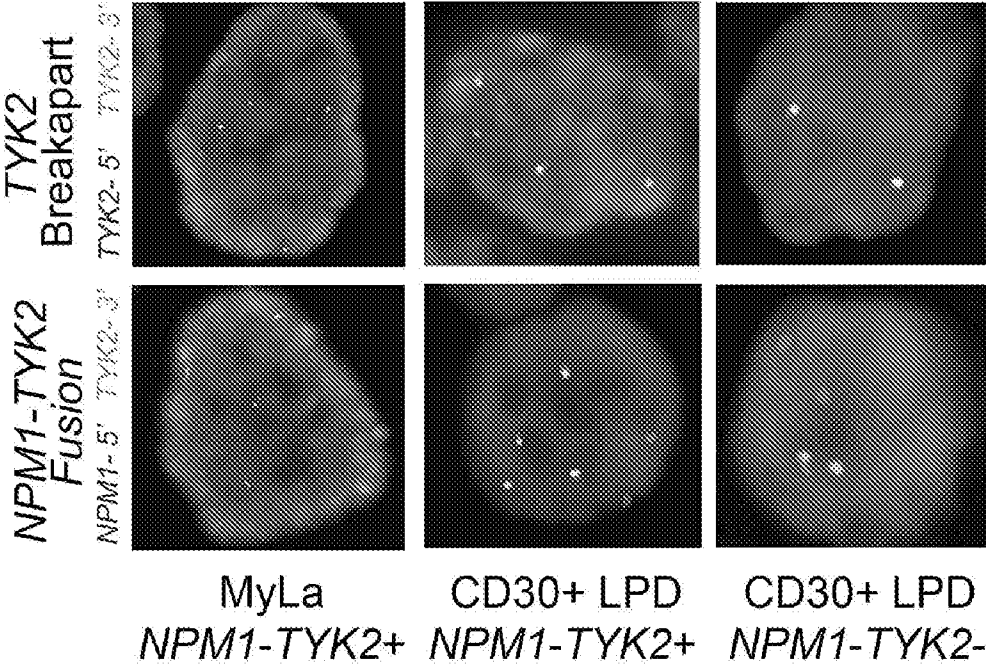


FIG. 2F

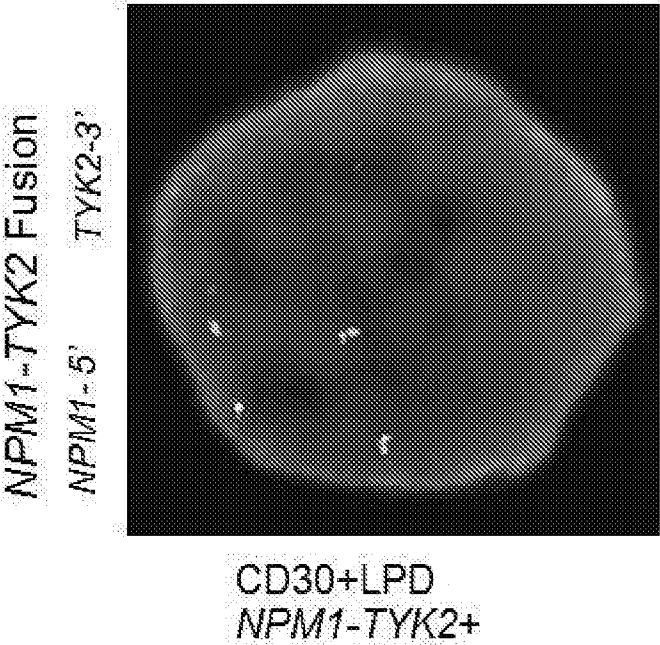


FIG. 3

Summary of RNAseq data for Myla

3'										5'									
Chr	Start	End	Strand	Gene	Chr	Start	End	Strand	Gene	Type	Total Reads	Breakpoint Spanning Reads							
chr5	170,814,707	170,832,406	+	NPM1	chr19	10,461,203	10,468,813	-	TYK2	Interchromosomal	803	766							
chr6	29,910,246	29,911,319	+	HLA-A	chr6	31,323,093	31,324,144	-	HLA-B	Intrachromosomal_Converging	84	0							
chr1	35,652,601	35,658,742	-	SFPQ	chr1	35,641,980	35,642,999	-	ALB1889	Read_Through	36	1							
chr13	41,885,340	41,885,717	+	NAA16	chr16	70,404,125	70,407,263	+	DDX194	Interchromosomal	36	33							
chr12	4,829,751	4,874,711	+	GALNT8	chr12	4,959,503	4,960,277	+	KCNAB	Read_Through	19	12							
chr10	17,270,257	17,279,591	+	VIM	chr1	115,520,177	115,537,989	+	SYCP1	Interchromosomal	24	0							
chr1	27,022,521	27,100,207	+	ARID1A	chr1	27,117,249	27,124,893	+	PGV	Read_Through	15	14							
chr5	70,196,890	70,196,975	+	SEPF1	chr5	69,322,233	69,328,524	+	SEPF1B	Intrachromosomal_Complex	14	14							
chr5	69,321,472	69,321,557	+	SEPF1	chr5	69,322,233	69,328,524	+	SEPF1B	Overlapping_Complex	14	14							
chr5	70,196,890	70,196,975	+	SEPF1	chr5	70,197,651	70,203,943	+	SEPF1B	Overlapping_Complex	14	14							
chr5	69,321,472	69,321,557	+	SEPF1	chr5	70,197,651	70,203,943	+	SEPF1B	Intrachromosomal	14	14							
chr14	24,761,389	24,768,665	-	DHRS1	chr14	24,734,743	24,740,742	-	RABGGTA	Read_Through	16	0							
chr19	17,940,916	17,958,840	-	JAK3	chr19	17,927,321	17,927,867	-	RSL3	Read_Through	10	8							
chr2	88,081,642	88,083,286	-	RGP02	chr2	109,111,746	109,125,853	+	GCC2	Intrachromosomal_Diverging	14	7							
chr2	108,453,905	108,459,385	+	RGP01	chr2	109,111,746	109,125,853	+	GCC2	Intrachromosomal	14	7							
chr2	87,208,275	87,214,931	+	RGP02	chr2	109,111,746	109,125,853	+	GCC2	Intrachromosomal	14	7							
chr6	29,591,116	29,591,703	+	HLA-F	chr6	31,321,648	31,324,449	-	HLA-B	Intrachromosomal_Converging	11	0							
chr2	133,015,101	133,015,541	-	ANKRD308L	chr6	46,188,466	46,459,059	-	RCA2	Intrachromosomal	10	0							
chr7	64,541,128	64,541,657	-	BCO44608	chr7	56,078,743	56,088,923	-	PSPH	Intrachromosomal	9	0							
chr14	21,205,020	21,205,283	+	TCRA, TCR	chr14	21,975,833	23,019,607	+	TCRA, TRA@ TBAC	Intrachromosomal	8	0							
chr11	61,560,108	61,560,499	+	FEN1	chr11	61,605,249	61,634,824	+	FADS2	Read_Through	8	8							
chr19	10,713,120	10,713,236	+	SLC44A2	chr19	17,448,816	17,453,839	+	GTPBP2	Intrachromosomal	6	6							
chr19	5,785,393	5,791,246	-	DUG3L	chr19	5,782,970	5,784,449	-	PAR22	Read_Through	6	5							
chr6	81,323,943	81,324,449	-	HLA-B	chr6	29,511,044	29,513,660	+	HLA-A	Intrachromosomal_Converging	9	0							
chr8	67,834,848	67,837,776	-	SMHG6	chr8	67,834,708	67,834,783	-	SNORD87	Overlapping_Complex	6	0							
chr7	130,630,478	130,791,788	-	FLJ43663	chr7	130,557,250	130,569,154	-	LOC648329	Read_Through	8	7							
chr12	57,127,930	57,146,145	-	PRIM1	chr12	57,105,210	57,108,470	-	NACA	Read_Through	7	7							
chr5	2,752,261	2,755,510	+	C5orf38	chr5	99,614,625	99,757,881	-	TGSA10	Interchromosomal	12	0							
chr6	34,704,576	34,708,691	+	HMGA1	chr5	180,374,511	180,377,905	+	BTNL8	Intrachromosomal	6	0							
chr5	180,325,076	180,334,754	+	BTNL8	chr6	34,212,607	34,214,007	+	HMGA1	Intrachromosomal	5	0							
chr19	1,269,266	1,274,438	+	CREP	chr19	1,277,181	1,279,242	+	C19orf24	Read_Through	4	4							
chr5	156,893,290	156,714,116	+	CYBP2	chr15	22,928,128	22,933,154	+	CYBP1	Interchromosomal	6	3							
chr19	17,516,551	17,517,104	+	AK311380, AK055623	chr19	17,531,121	17,532,021	+	FAM125A	Read_Through	4	3							

FIG. 3 (Cont.)

chr6	155,153,580	155,154,582	+	TAM2	chr6	155,228,863	155,229,999	+	AK124484	Overlapping_Same	4	3
chr1	173,834,608	173,837,124	-	GAS5	chr1	173,834,487	173,834,567	-	SWORD79	Overlapping_Complex	6	0
chr19	3,623,685	3,625,812	-	C19orf29	chr16	732,863	733,767	+	STUR1	Interchromosomal	5	0
chr7	45,798,872	45,808,616	-	SEP77P2	chr7	55,840,717	55,841,375	+	CO9	Interchromosomal_Diverging	5	0
chr6	144,329,226	144,339,540	-	PLAGL1	chr2	47,530,805	47,710,359	+	MIR2	Interchromosomal	4	0
chr17	16,343,349	16,343,419	+	SNORD49A	chr17	16,343,424	16,345,339	+	C17orf75-AS1	Overlapping_Complex	4	0
chr6	143,380,489	143,634,581	+	AK1	chr3	183,547,372	183,584,352	-	PATL	Interchromosomal	4	0
chr1	153,934,695	153,935,797	-	SLC39A1	chr1	153,920,147	153,927,641	-	CRTC2	Read_Through	6	6
chr19	34,919,267	34,937,918	+	UBA2	chr19	34,981,280	34,992,084	+	WNP	Read_Through	4	4
chr1	42,921,172	42,923,020	+	PPI5	chr1	42,938,792	43,120,334	+	CCOC30	Read_Through	4	4
chr1	143,915,747	143,936,307	+	SRGAP2P2	chr1	206,557,668	206,528,045	+	SRGAP2	Interchromosomal	4	4
chr1	27,022,521	27,100,207	+	ARID1A	chr1	27,158,937	27,182,210	+	ZDHHC18	Interchromosomal	4	4
chr2	181,988,563	182,227,644	+	AK129001	chr2	182,339,866	182,402,767	+	PIGA4	Read_Through	3	3
chr12	123,237,370	123,238,353	+	OEVR	chr12	123,261,937	123,311,976	+	CCOC52	Read_Through	3	3
chr1	26,146,896	26,152,316	+	FAM54B	chr1	25,810,806	25,826,897	+	TNFRSF57	Interchromosomal_Complex	3	3
chr12	56,708,634	56,710,127	-	CNPY2	chr12	56,665,482	56,680,428	-	CS	Read_Through	3	3
chr1	27,114,453	27,121,724	+	PIGV	chr1	27,158,837	27,181,210	+	ZDHHC18	Read_Through	4	2
chr7	99,156,044	99,158,317	+	ZNRF55	chr7	99,197,827	99,208,455	+	LOC100269187	Interchromosomal	4	2
chrX	3,778,830	3,781,614	-	AK000470	chrX	3,735,375	3,736,743	-	LOC338906	Read_Through	3	2
chrX	3,817,256	3,820,040	-	AK000470	chrX	3,735,375	3,736,743	-	LOC338905	Interchromosomal	3	2
chrX	3,746,346	3,761,934	-	LOC338906	chrX	3,771,864	3,779,036	-	AK000470	Adjacent_Complex	3	1
chrX	3,746,546	3,761,934	-	LOC338905	chrX	3,803,488	3,811,462	-	AK000470	Interchromosomal_Complex	3	1
chr9	37,423,680	37,424,971	+	GRHR	chr17	43,340,487	43,364,191	-	MARSK14	Interchromosomal	6	0
chr1	38,471,028	38,471,186	+	FHL3	chr19	47,931,278	47,975,306	-	SLC8A2	Interchromosomal	6	0
chr17	5,404,718	5,487,831	-	MLRPI	chr17	5,402,746	5,403,269	-	LOC728382	Read_Through	5	0
chr7	5,821,171	5,821,350	-	RNF216	chr1	110,527,366	110,566,363	+	AHCY11	Interchromosomal	5	0
chr17	7,482,804	7,485,428	+	CD88	chr17	7,486,864	7,491,526	+	MPOU1	Read_Through	5	0
chr4	120,375,997	120,376,187	+	LOC45513	chr7	55,840,871	55,841,375	+	CO9	Interchromosomal	4	0
chr17	18,761,380	18,761,573	+	PRPSA2	chr19	12,810,807	12,812,856	-	TNPO2	Interchromosomal	4	0
chr17	7,123,152	7,128,584	+	ACADVL	chr10	134,325,129	134,325,664	-	AF086154	Interchromosomal	3	0
chr9	111,754,688	111,754,848	-	S_RS_PXA	chr14	100,705,101	100,744,803	+	YY1	Interchromosomal	3	0
chr22	29,702,956	29,703,142	+	GAS2L1	chr17	48,211,100	48,227,352	-	PPP1R9B	Interchromosomal	3	0
chr1	75,928,908	75,938,110	-	GCF2	chr12	49,566,845	49,567,808	-	TUBA1C	Interchromosomal	3	0
chr6	30,573,976	30,576,949	+	GTF2H4	chr19	12,335,500	12,336,093	-	ZNRF44	Interchromosomal	3	0
chr12	120,933,625	120,933,748	-	LOC	chr8	80,831,094	80,915,414	-	MIRP528	Interchromosomal	3	0
chr6	144,329,276	144,329,540	-	PLAGL1	chr17	48,190,034	48,207,166	-	SAMD14	Interchromosomal	3	0
chr3	52,719,935	52,724,719	+	GNL3	chr3	53,724,753	53,724,845	+	SNORD198	Read_Through	3	0
chrM	14,856	15,887	+	Cytchrome_b	chr15	90,805,237	90,815,442	+	TILL13	Interchromosomal	35	35
chr5	69,321,071	69,322,241	+	SERF18	chr5	70,234,865	70,248,841	+	SMN2-5MIR1	Interchromosomal	5	4
chr5	70,196,489	70,197,759	+	SERF18	chr5	69,359,241	69,373,421	+	SMN2-5MIR1	Interchromosomal_Complex	5	4
chr5	70,196,489	70,197,759	+	SERF18	chr5	70,234,865	70,248,841	+	SMN2-5MIR1	Read_Through	5	4
chr5	69,321,071	69,322,241	+	SERF18	chr5	69,359,241	69,373,421	+	SMN2-5MIR1	Read_Through	5	4
chr17	43,597,980	43,707,513	+	M2C57345	chr17	43,723,266	43,723,594	+	C17orf69	Read_Through	4	4
chr1	151,150,476	151,162,639	-	VP572	chr1	151,142,462	151,142,578	-	TMD04	Read_Through	4	4

FIG. 3 (Cont.)

chr11	119,252,487	119,252,617	+	LOC	100499227	chr11	117,975,405	117,990,555	+	TMPS3A	Intrachromosomal_Complex	3	3
chr19	5,921,211	5,934,472	-	PANBP3		chr19	5,993,174	5,004,508	-	RFX2	Adjacent_Complex	3	3
chr16	29,674,579	29,674,856	+	SNW		chr16	29,708,316	29,708,318	+	QPR7	Read_Through	3	3
chr16	21,868,579	21,902,952	-	LOC23117		chr16	18,451,950	18,466,710	-	MRP1	Intrachromosomal	3	3
chr16	21,436,299	21,471,255	-	LOC	100271836	chr16	18,451,950	18,466,710	-	MRP1	Intrachromosomal	3	3
chr7	150,076,405	150,085,588	+	ZNF775		chr7	150,107,247	150,109,557	+	LOC728743	Read_Through	3	3
chr2	54,093,221	54,197,976	-	PSME4		chr2	54,080,049	54,082,001	-	GPRT5	Read_Through	3	3
chr16	81,118,067	81,129,979	-	GSH		chr16	81,087,101	81,087,698	-	C16orf46	Read_Through	3	3
chr9	56,431,910	56,432,751	-	GAP1		chr9	56,437,794	56,457,793	+	LOC399765	Intrachromosomal_Diverging	3	3
chr16	21,436,299	21,471,255	-	LOC	100271836	chr16	16,429,717	16,444,464	+	MRP1	Intrachromosomal_Converging	3	3
chr16	21,868,579	21,902,952	-	LOC23117		chr16	16,429,717	16,444,464	+	MRP1	Intrachromosomal_Converging	3	3
chr17	56,083,151	56,084,706	-	SRSF1		chr17	56,066,404	56,067,156	-	AK126318	Read_Through	3	3
chr17	44,171,925	44,270,165	-	KIAA1267		chr17	44,194,067	44,194,234	-	ARL17A	Intrachromosomal_Complex	3	2
chr17	44,171,925	44,270,165	-	KIAA1267		chr17	44,376,499	44,430,295	-	ARL17B,ARL17A	Intrachromosomal_Complex	3	2
chr17	29,421,944	29,670,152	+	NF1		chr17	65,613,950	65,697,217	+	AK4	Intrachromosomal	3	2
chr19	1,863,039	1,863,563	-	KLF16		chr19	1,815,244	1,823,528	-	REXO1	Read_Through	2	2
chr1	131,260,008	131,261,256	+	EMBP1		chr2	96,142,714	96,144,604	-	TBM43B	Intrachromosomal	2	2
chr8	23,261,466	23,261,721	-	LOC12		chr8	23,243,441	23,243,557	-	ENTPD4	Overlapping_Complex	2	2
chr10	5,734,565	5,751,825	+	C10orf18		chr10	5,754,785	5,805,782	+	FAM208B	Overlapping_Complex	2	2
chr15	68,521,839	68,522,079	-	C15orf6		chr15	68,483,042	68,489,965	-	CALML4	Read_Through	2	2
chr5	69,697,279	69,697,494	+	SMN4		chr5	68,860,916	68,876,220	+	GTF2H3C	Intrachromosomal_Complex	2	2
chr1	49,201,910	49,242,546	-	BRNDS		chr1	49,998,525	49,128,912	-	AGBL4	Overlapping_Same	2	2
chr5	69,697,279	69,697,494	+	SMN4		chr5	69,716,273	69,746,188	+	GTF2H2B	Overlapping_Same	2	2
chr11	107,573,726	107,729,913	-	SLC35F2		chr11	107,578,100	107,578,730	-	SLN	Read_Through	2	2
chr1	117,131,315	117,210,313	-	GSRF3		chr22	73,003,820	73,024,971	+	G6H1	Intrachromosomal	2	2
chr9	86,431,910	86,432,751	-	GAP1		chr9	86,451,614	86,530,592	-	KIF27	Adjacent_Complex	2	2
chr1	131,260,008	131,261,256	+	EMBP1		chr2	96,263,627	96,263,468	+	TBM43	Intrachromosomal	2	2
chr11	65,029,431	65,063,460	+	POL42		chr11	65,088,014	65,089,899	+	CDC42EP2	Read_Through	2	2
chr5	69,697,279	69,697,494	+	SMN4		chr5	70,330,950	70,358,621	-	GTF2H2	Intrachromosomal_Converging	2	2
chr3	136,581,049	136,647,088	+	MOX1		chr3	136,699,307	136,729,925	+	L20RB	Read_Through	3	1
chr20	31,172,371	31,172,874	-	LOC28404		chr20	31,052,883	31,115,715	-	C20orf112	Read_Through	3	1
chr19	41,082,756	41,086,841	+	S-KRP1		chr19	41,066,054	41,066,657	+	SPTBN4	Adjacent_Complex	2	1
chr1	173,445,770	173,446,293	-	LOC	100506023	chr1	173,387,097	173,387,760	-	PC138808	Overlapping_Complex	2	1
chr6	5,004,049	5,021,706	+	A0094934		chr6	5,042,739	5,043,172	+	BC133670	Overlapping_Same	2	1
chr10	111,762,710	111,768,056	+	A0093		chr7	1,570,867	1,582,678	+	KAPK	Intrachromosomal	8	0
chr9	36,249,188	36,258,495	-	GVE		chr9	99,212,413	99,253,617	+	HABP4	Intrachromosomal_Diverging	5	0
chr17	58,054,016	58,096,385	-	TBC1D3P1		chr17	58,029,722	58,040,644	-	RMFT1	Read_Through	4	0
chr16	70,020,283	70,076,905	-	PDXDC2P		chr16	22,535,140	22,547,840	+	LOC100132247	Intrachromosomal_Converging	4	0

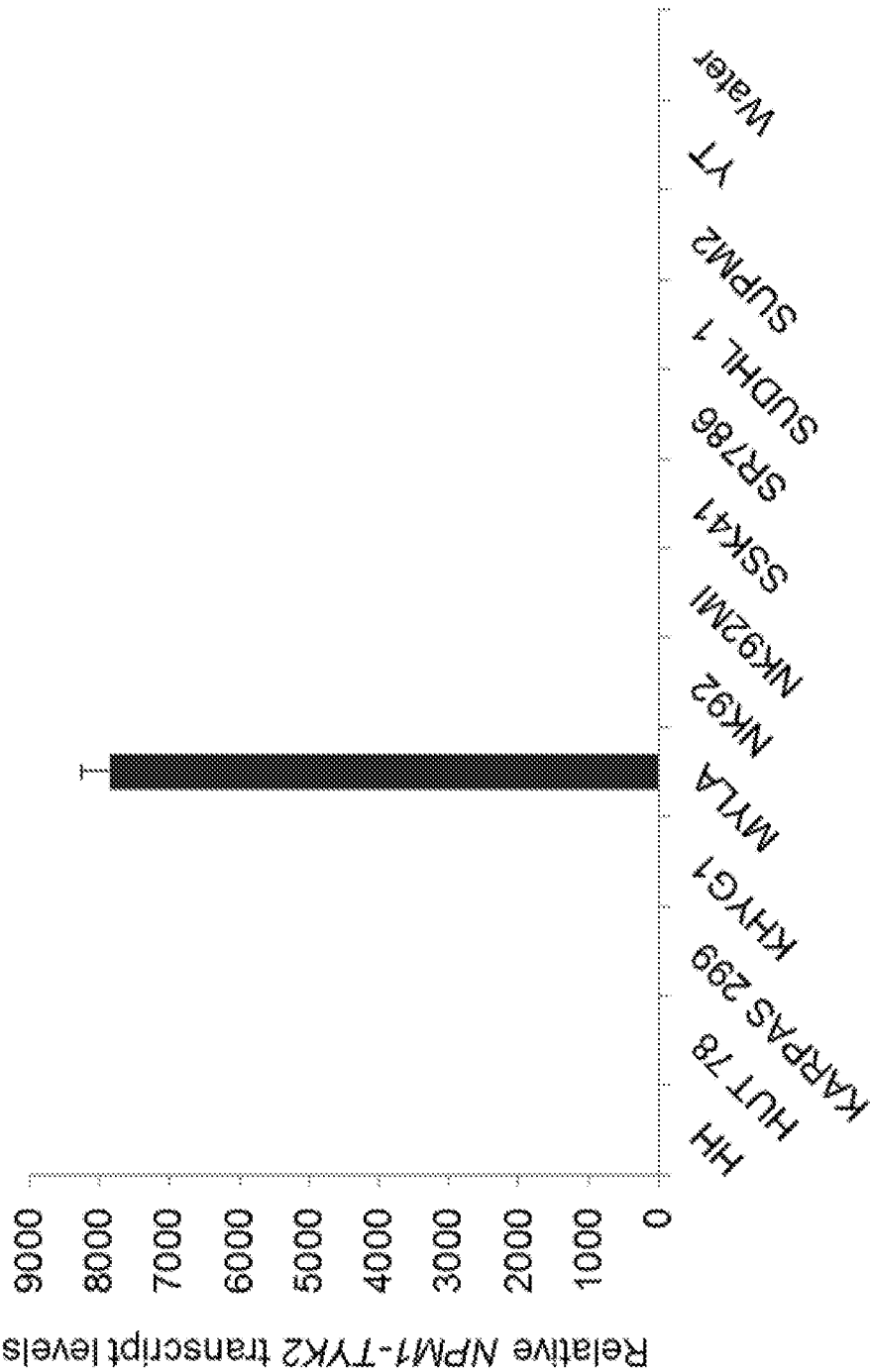
FIG. 3 (Cont.)

chr18	70,040,283	70,075,905	-	PDX022P	chr16	30,234,349	30,246,899	-	LQ0813037	intrachromosomal	4	0
chr6	136,589,299	136,610,968	-	BCLA1	chr1	36,767,153	36,770,956	+	THRAP3	interchromosomal	4	0
chr16	70,020,283	70,076,905	-	PDX022P	chr16	21,415,197	21,426,405	-	NPPL3	intrachromosomal	4	0
chr7	100,071,896	100,076,901	-	TSC22D4	chr7	100,054,237	100,061,893	-	C7orf51	Read_Through	4	0
chr17	57,642,885	57,676,311	+	DMK40	chr17	58,039,732	58,040,844	-	RPR1	intrachromosomal_Converging	4	0
chr1	223,280,321	223,289,539	+	SGP1	chr9	137,318,315	137,332,430	+	NRX1	interchromosomal	4	0
chr1	147,931,733	147,931,979	-	FLJ39739	chr1	145,004,780	145,005,286	+	BC063231	intrachromosomal_Converging	3	0
chr2	219,524,378	219,526,675	+	BCS1L	chr12	9,013,730	9,028,376	+	A2ML1	interchromosomal	3	0
chr7	96,338,999	96,339,202	-	SPFM1	chr7	96,330,968	96,375,247	-	LOC100506186	Read_Through	3	0
chr1	241,803,183	241,803,700	-	OPM3	chr1	241,792,166	241,799,231	-	C-MIL	Overlapping_Complex	3	0
chr1	149,576,452	149,576,699	+	BC110321	chr1	145,004,780	145,005,286	+	BC063231	intrachromosomal_Complex	3	0
chr2	92,082,423	92,096,341	+	OTUD6	chr2	92,114,846	92,231,463	+	LRR59	Read_Through	3	0
chr2	132,249,448	132,249,658	-	MET3A	chr14	103,398,715	103,440,485	-	CD44BP8	interchromosomal	3	0
chr15	74,244,204	74,249,419	-	BC043527	chr3	190,936,535	190,967,909	+	OST1	interchromosomal	3	0
chr2	48,667,907	48,738,606	+	PPP1R21	chr2	48,757,063	49,003,655	+	STON1	Read_Through	3	0
chr1	144,340,525	144,340,772	-	LOC728855	chr1	145,004,780	145,005,286	+	BC063231	intrachromosomal_Diverging	3	0
chr1	143,744,200	143,744,585	-	LOC	chr1	145,004,780	145,005,286	+	BC063231	intrachromosomal_Diverging	3	0
chr11	62,344,126	62,359,108	-	TUT1	chr11	62,327,072	62,342,400	-	EFF1G	Read_Through	3	0
chr11	133,828,585	133,826,879	-	SGSF8	chr14	35,515,710	35,532,588	+	FAM177A1	interchromosomal	3	0
chr17	36,955,518	36,956,157	-	P94K2B	chr19	47,567,446	47,615,845	-	ZC3H4	interchromosomal	3	0
chr1	39,538,962	39,539,049	-	MYCBP	chr1	39,531,478	39,577,105	-	RHBO12	intrachromosomal_Complex	3	0
chr8	144,661,866	144,672,250	-	EEF1D	chr8	144,656,954	144,659,908	-	NAPRT1	Read_Through	3	0
chr12	132,589,111	132,594,376	+	EP420NL	chr12	132,604,964	132,613,428	+	EP400ML	Read_Through	3	0
chr12	132,589,111	132,594,376	+	EP420NL	chr12	132,472,249	132,475,258	+	EP400	intrachromosomal_Complex	3	0
chr1	133,309,826	133,321,021	-	PGLYBP4	chr1	133,270,337	133,283,193	-	PGLYBP3	Read_Through	3	0
chr9	111,754,698	111,754,848	-	SLS_RNA	chr1	22,140,890	22,151,713	+	LDLRAD2	interchromosomal	3	0
chr1	144,530,773	144,531,803	-	LOC728875	chr1	145,004,780	145,005,286	+	BC063231	intrachromosomal_Diverging	3	0
chr9	38,543,235	38,577,760	-	ANKRD18A	chr17	34,493,080	34,499,312	-	TBC1D38	interchromosomal	2	0
chr12	49,666,045	49,667,808	-	TUBA1L	chr3	45,265,935	45,267,813	-	TBMX158	interchromosomal	2	0
chr9	38,543,235	38,577,760	-	ANKRD18A	chr17	60,345,505	60,348,386	-	TBC1D3P2	interchromosomal	2	0
chr17	36,955,518	36,956,157	-	P94K2B	chr1	62,626,131	62,657,483	-	PIERH35	interchromosomal	2	0
chr9	140,083,053	140,084,821	+	SSNA1	chr1	62,146,718	62,159,455	-	TM1D1	interchromosomal	2	0
chr9	36,258,317	36,258,495	-	GME	chr1	54,892,194	54,872,091	-	SSBP3	interchromosomal	2	0
chr10	48,281,846	49,338,995	-	ANK209P2	chr7	42,948,871	42,967,057	-	C7orf25	interchromosomal	2	0
chr11	117,049,938	117,066,716	+	SID2	chr11	117,073,717	117,075,507	+	TAC1N	Read_Through	2	0
chr3	52,321,835	52,325,126	+	GLYCN	chr3	52,350,334	52,434,512	+	DNAH1	Read_Through	2	0
chr16	17,584,290	17,584,797	-	WNT1	chr19	49,671,113	49,715,067	+	TBR14	interchromosomal	2	0
chr9	38,543,235	38,577,760	-	ANKRD18A	chr17	36,337,527	36,343,994	-	TBC1D3	interchromosomal	2	0
chr1	25,943,958	25,944,837	+	MAN1C1	chr4	25,749,046	25,864,608	-	SL113	interchromosomal	2	0
chr19	36,266,416	36,266,565	+	ARHGAP33	chr3	63,498,263	63,582,294	+	ATXN7	interchromosomal	2	0

FIG. 3 (Cont.)

chr8	145,538,246	145,530,566	-	DCAF1	chr16	29,848,042	29,859,359	+	MVP	interchromosomal	2	0
chr9	38,543,235	38,577,260	-	ANKRD18A	chr17	34,581,084	34,587,366	-	TBC1D3H, TBC1D3C, TBC1D3G	interchromosomal	2	0
chr16	72,166,831	72,106,063	-	PMF8P1	chr8	52,255,095	52,265,010	-	TIR9	interchromosomal	2	0
chr2	97,760,457	97,760,581	-	FAHD2B	chr20	48,697,660	48,770,334	-	TMEM189, TMEM189- UBE2V1	interchromosomal	2	0
chr2	133,015,101	133,015,541	-	ANKRD30B	chr2	180,206,710	180,726,231	-	ZNF385B	intrachromosomal_Complex	2	0
chr9	38,543,235	38,577,260	-	ANKRD18A	chr17	36,288,650	36,295,097	+	TBC1D3, TBC1D3F	interchromosomal	2	0
chr19	54,694,137	54,697,504	+	TSEB4	chr19	54,705,009	54,711,514	+	RPS9	Read_Through	2	0
chr19	10,982,252	11,018,820	+	CARM1	chr8	130,397,777	130,435,442	-	PK3R4	interchromosomal	2	0
chr1	155,190,699	155,292,920	+	RUSC1	chr12	102,091,418	102,122,845	+	CHMT1	interchromosomal	2	0
chr9	123,452,203	123,476,764	-	MEGF9	chr9	123,519,253	123,555,663	-	FBXW2	Adjacent_Complex	2	0
chr10	51,606,987	51,823,335	-	TMM23	chr10	51,224,887	51,247,973	-	AGAP8	intrachromosomal	2	0
chr22	40,742,503	40,761,059	+	AOL	chr22	40,766,594	40,806,292	+	SGSM3	Read_Through	2	0
chr7	123,673,258	124,028,580	+	E2F3B17	chr17	70,842,084	70,845,942	-	SIC3B411	interchromosomal	2	0
chr6	7,910,746	7,911,046	-	TRNDC5	chr10	17,271,274	17,279,591	+	VM	interchromosomal	2	0
chr9	100,818,958	100,819,221	+	NAN5	chr1	38,463,441	38,471,186	-	FHL3	interchromosomal	2	0
chr10	11,653,411	11,653,678	-	USP6NL	chr7	26,538,338	26,538,593	+	LOC411204	interchromosomal	2	0
chr6	144,607,889	144,607,706	+	UTRN	chr3	50,650,278	50,686,727	+	MAPKAPK3	interchromosomal	2	0
chr16	30,064,410	30,078,685	+	ALDO4	chrM	10,059	10,404	+	AD_1	interchromosomal	2	0
chr17	72,920,869	72,921,039	+	OTOF2	chr20	26,167,654	26,189,868	-	LOC284901	interchromosomal	2	0
chr7	56,088,765	56,119,267	-	PSPH	chr7	55,840,871	55,841,375	+	CO9	intrachromosomal_Converging	2	0
chr1	222,885,895	222,904,857	+	BRDX	chr1	222,906,202	222,908,532	+	AKO25140	Read_Through	2	0
chr7	45,796,672	45,808,615	-	SEPT72	chr7	56,078,743	56,079,561	-	PSPH	intrachromosomal_Complex	2	0
chr10	51,371,394	51,387,762	+	TMM23B	chr10	51,224,687	51,247,973	-	AGAP8	intrachromosomal_Diverging	2	0
chr9	2,621,792	2,654,484	+	VLDR	chr2	112,973,438	112,996,104	-	ZC3H8	interchromosomal	2	0
chr10	46,584,431	46,620,057	-	PTPN20A	chr7	41,948,871	42,967,057	-	C7orf25	interchromosomal	2	0
chr2	96,676,298	96,676,479	+	LOC729734	chr20	48,697,660	48,770,334	-	TMEM189, TMEM189- UBE2V1	interchromosomal	2	0
chr5	180,256,830	180,258,598	-	AKO26101	chr8	38,845,454	38,846,180	+	HTRA4	interchromosomal	2	0
chr19	46,520,055	46,566,547	-	SYMPK	chr12	122,267,971	122,370,561	+	SETD1B	interchromosomal	2	0
chr16	30,418,234	30,418,984	+	ZNF771	chr20	33,705,827	33,823,993	-	EDEM2	interchromosomal	2	0
chr16	1,607,935	1,636,275	-	IFT140	chr16	1,603,655	1,604,130	-	BC114455	Overlapping_Complex	2	0

FIG. 4



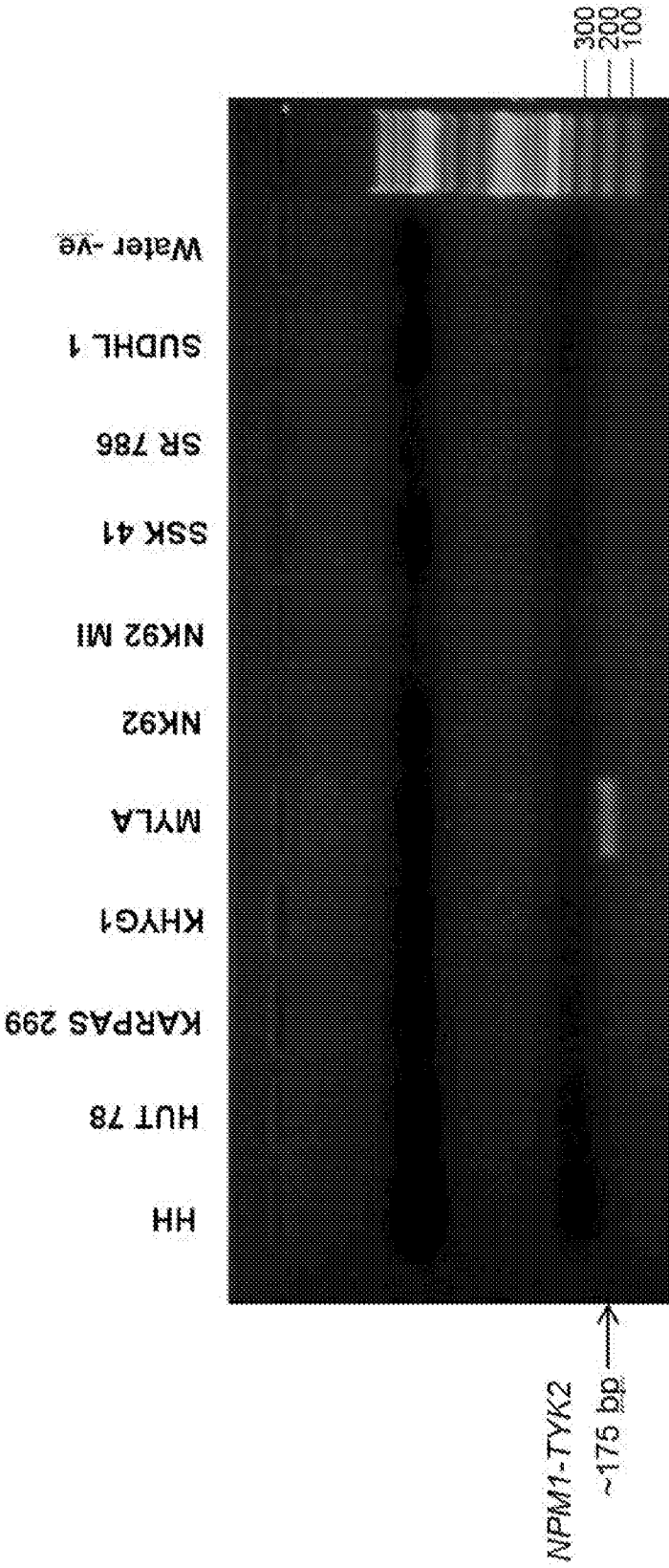


FIG. 5

FIG. 6

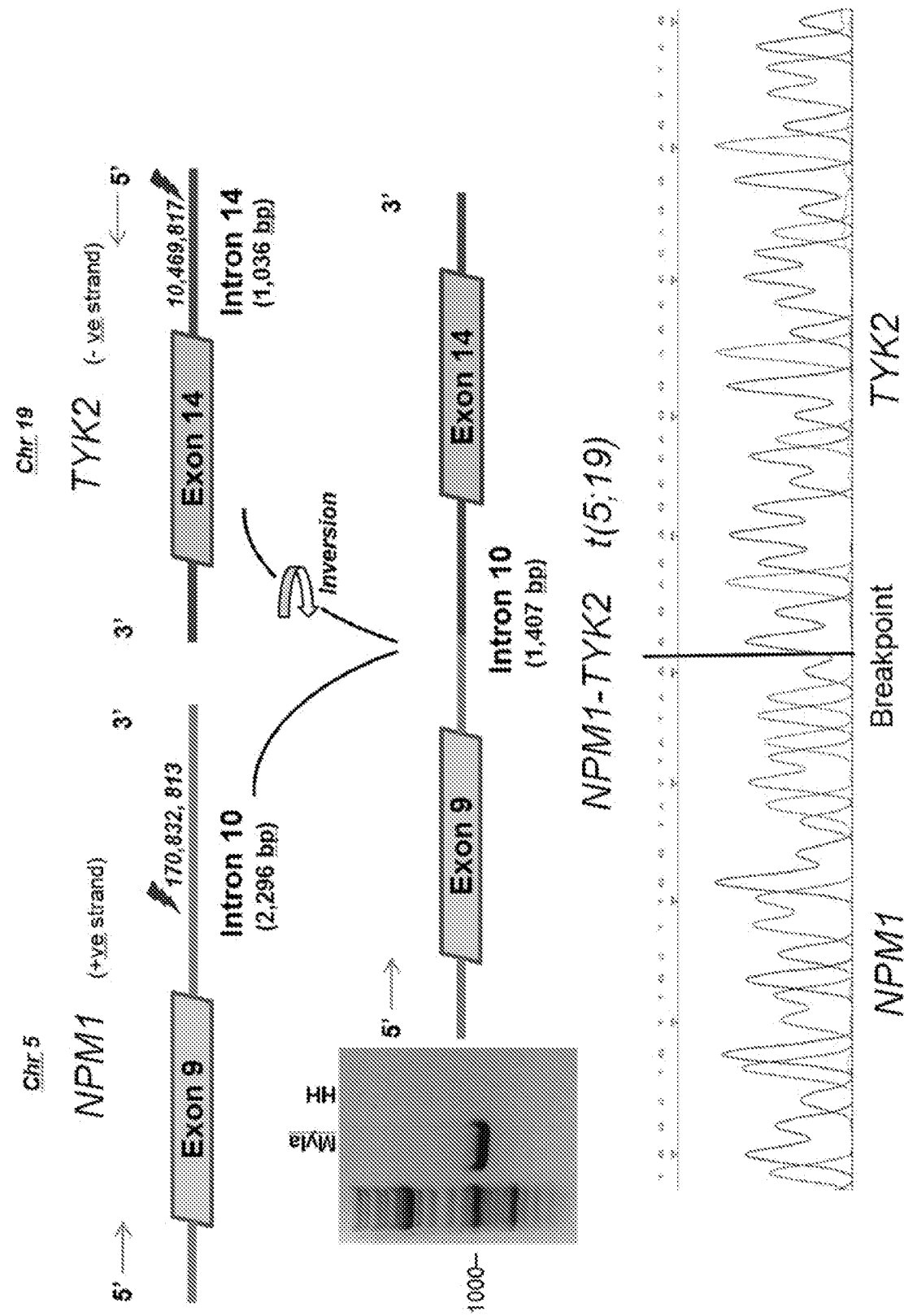


FIG. 7A

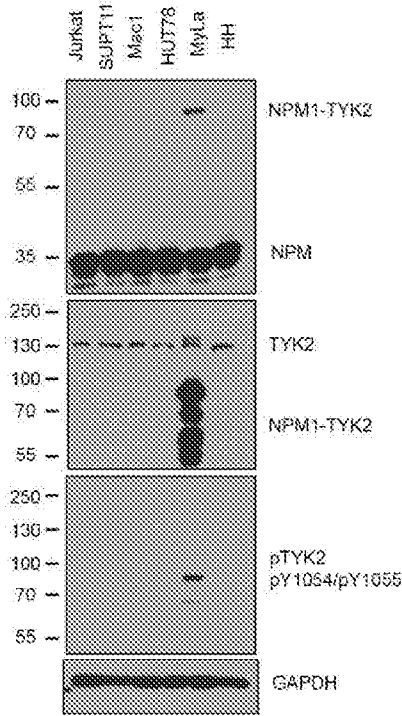


FIG. 7B

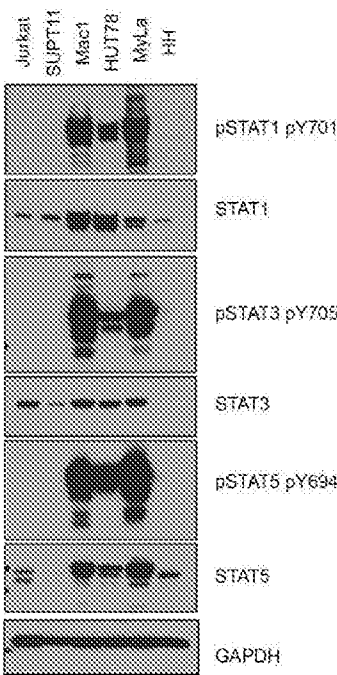


FIG. 7C

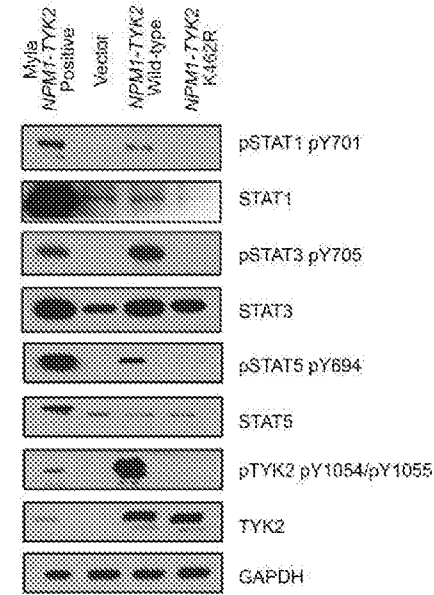


FIG. 7D

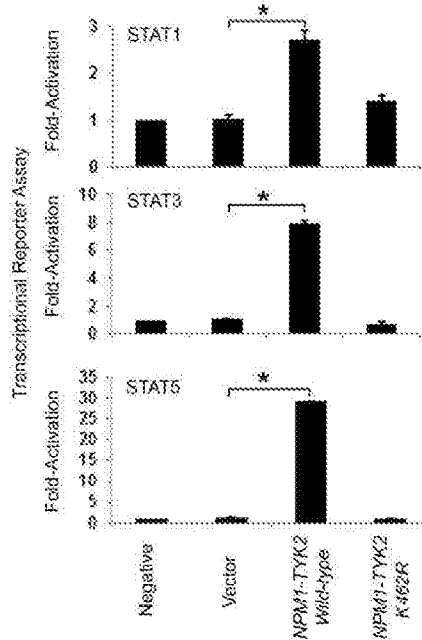


FIG. 7E

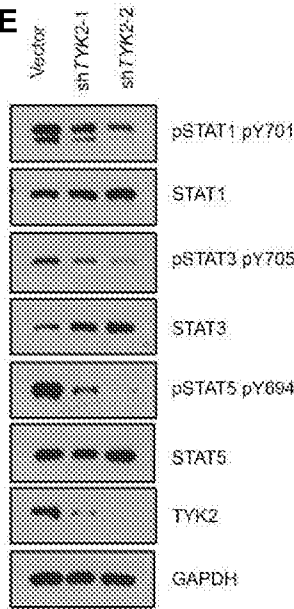


FIG. 7F

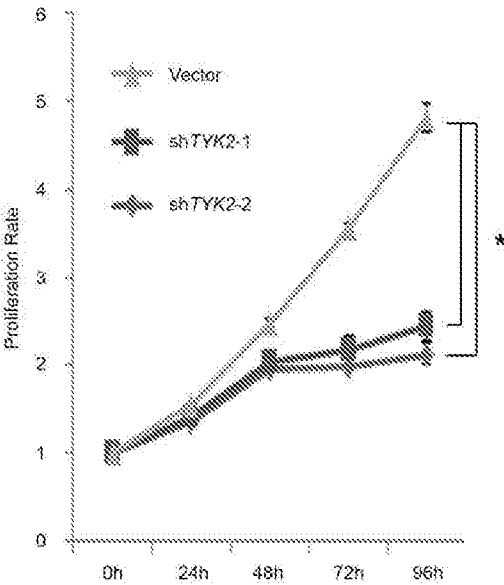


FIG. 8

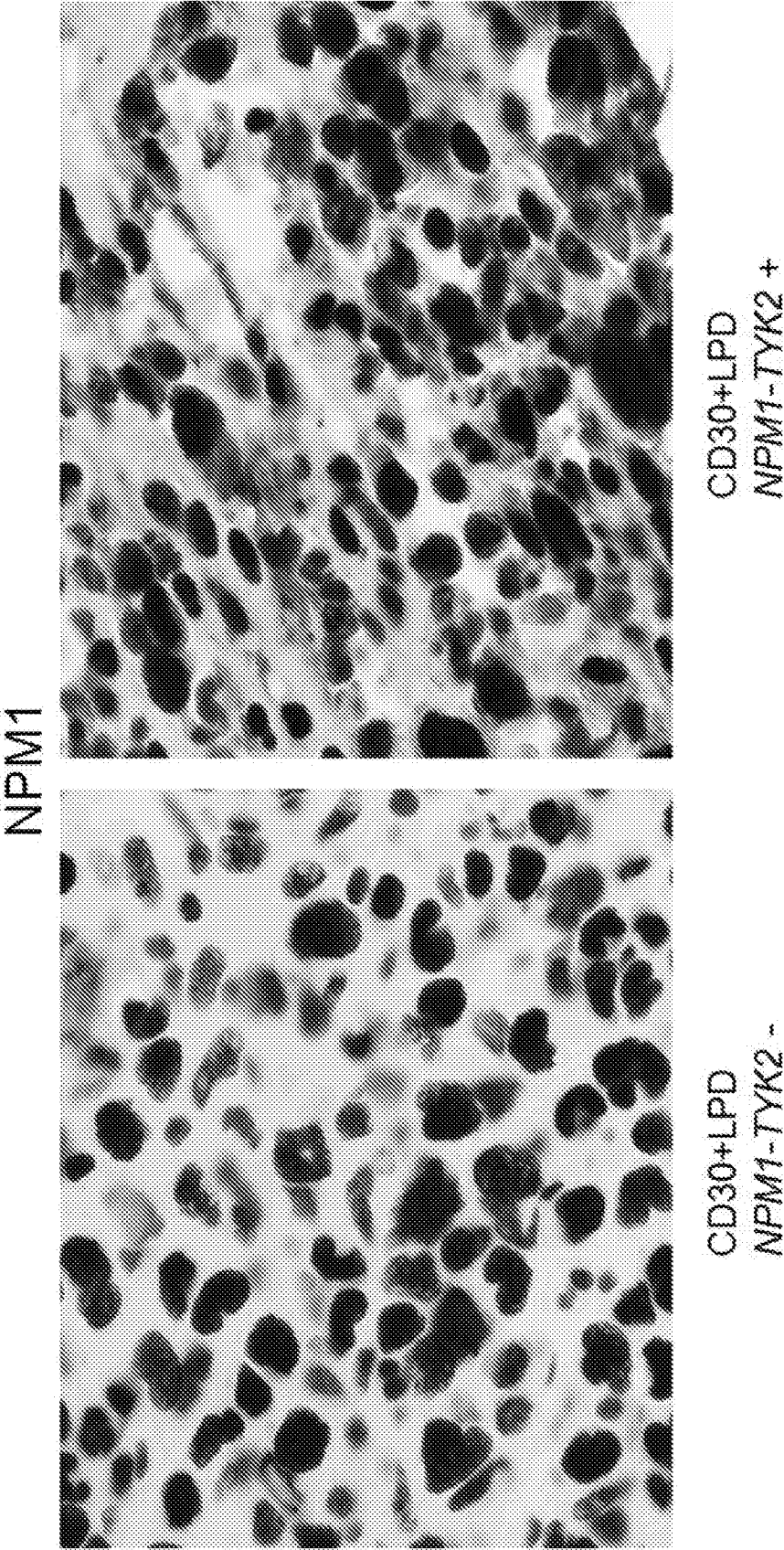


FIG. 9

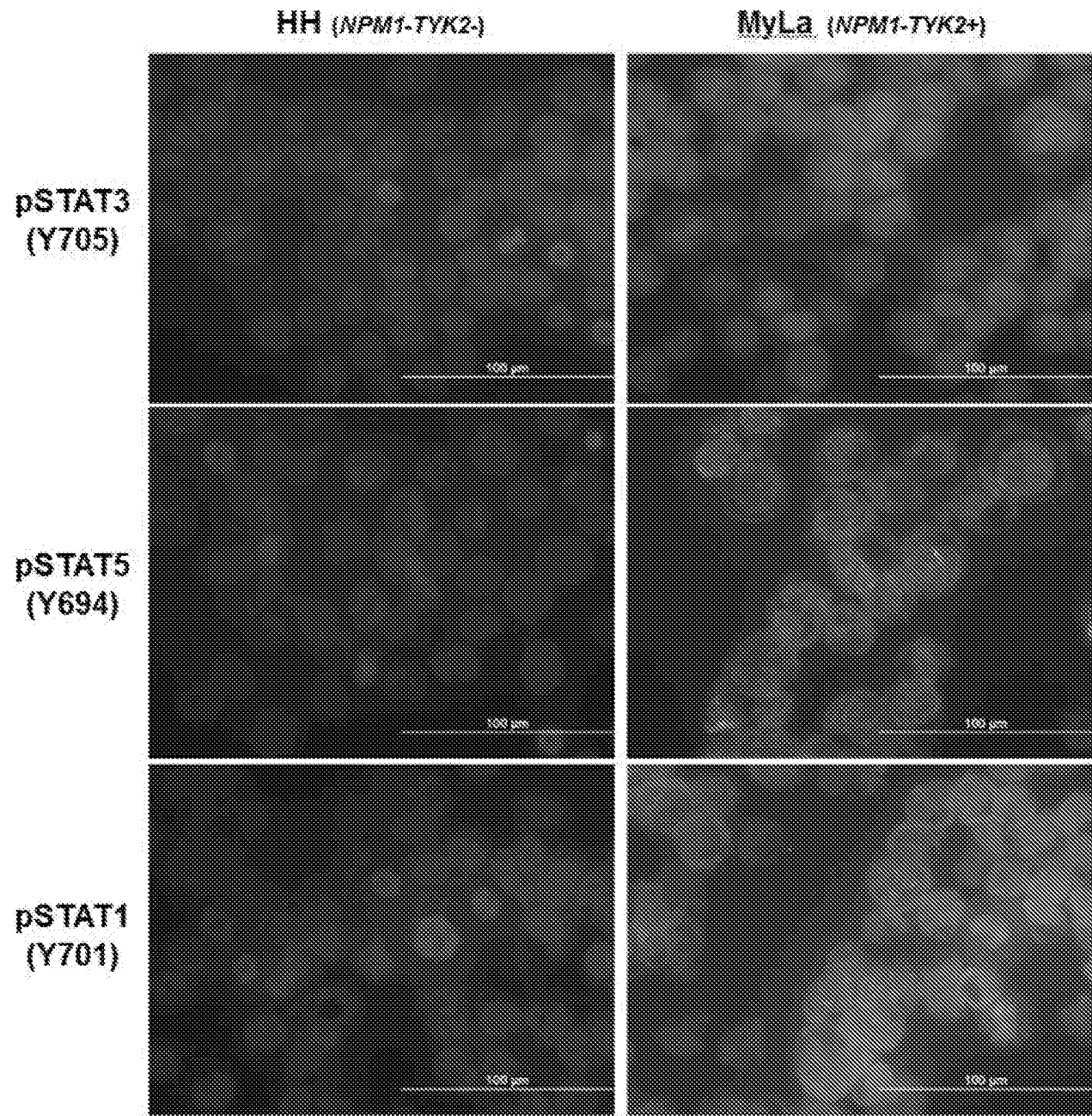
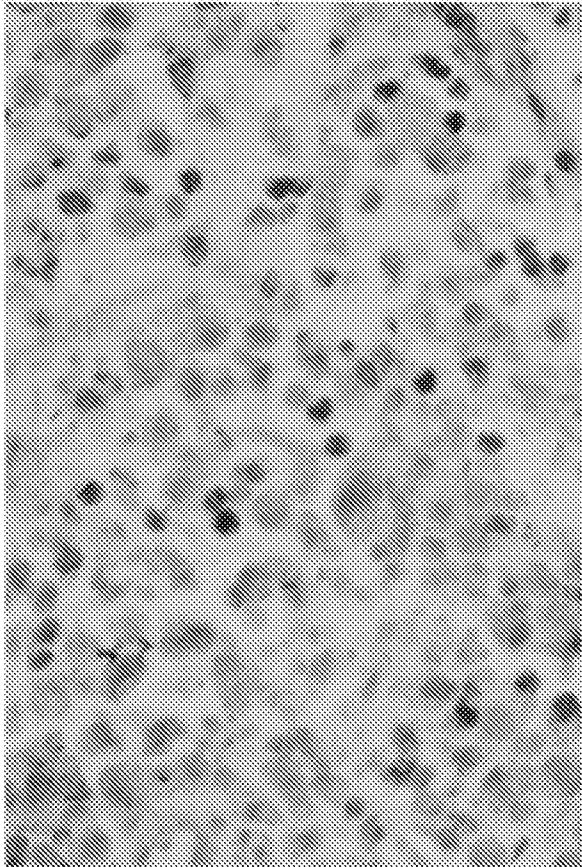
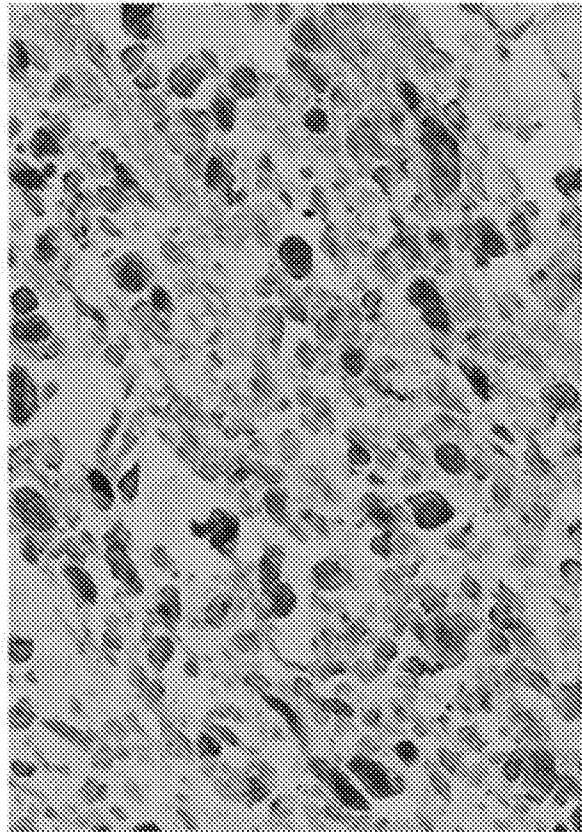


FIG. 10

pSTAT5



CD30+LPD
NPM1-TYK2 -



CD30+LPD
NPM1-TYK2 +

FIG. 11A

TYK2 cDNA 4262 bp

AAGCAGTAGCTACCCGCGGGAGCGGGGAGGGGTCCGGGTTCGAGCTTGTGTTCCCCCGGAAGGGTGAGTC
TGGACGCGGGCGCGGAAGGAGCGCGGCCGAGGTCTCAGGAAGAAGCCGCGGGGACTGGCTGCGCTTGA
CAGGCTGCACTTGGATGGGAGCACCTGGTGCCTCGGGACTGCTCCGATGCCCCGGGTCTGTGCTGAATGTG
TAATATGCGGAACTATATTGAAACATTACAACCATCTTTTGATGGCAACACCCTGAGGACCTCCCTTTTC
CAGATGGGGAAACTGAGGCCCAGAATTGCTAAGTGGCTTGCTTGAGTTGACACAGGGAGCTCCAGGACTC
ACCCTCAGCTGAGCCACCTGCCGGGAGCATGCCTCTGCGCCACTGGGGGATGGCCAGGGGCAGTAAGCCC
GTTGGGGATGGAGCCCAGCCCATGGCTGCCATGGGAGGCCTGAAGGTGCTTCTGCACTGGGCTGGTCCAG
GCGGCGGGGAGCCCTGGGTCACTTTCAGTGAGTCATCGCTGACAGCTGAGGAAGTCTGCATCCACATTGC
ACATAAAGTTGGTATCACTCCTCCTTGCTTCAATCTCTTTGCCCTCTTCGATGCTCAGGCCCAAGTCTGG
TTGCCCCCAAACCACATCCTAGAGATCCCCAGAGATGCAAGCCTGATGCTATATTTCCGCATAAGGTTTT
ATTTCCGGAACCTGGCATGGCATGAATCCTCGGGAACCGGCTGTGTACCGTTGTGGGCCCCCAGGAACCGA
GGCATCCTCAGATCAGACAGCACAGGGGATGCAACTCCTGGACCCAGCCTCATTTGAGTACCTCTTTGAG
CAGGGCAAGCATGAGTTTGTGAATGACGTGGCATCACTGTGGGAGCTGTGACCCGAGGAGGAGATCCACC
ACTTTAAGAATGAGAGCCTGGGCATGGCCTTCTGACCTCTGTACCTCGCTCTCCGCCATGGCATCCC
CCTGGAGGAGGTGGCCAAGAAGACCAGCTTCAAGGACTGCATCCCGCGCTCCTTCCGCCGGCATATCCGG
CAGCACAGCGCCCTGACCCGGCTGCGCCTTCGGAACGTCTTCCGCAGGTTCTGCGGGACTTCCAGCCGG
GCCGACTCTCCCAGCAGATGGTCATGGTCAAATACCTAGCCACACTCGAGCGGCTGGCACCCCGCTTCGG
CACAGAGCGTGTGCCCCGTGTGCCACCTGAGGCTGCTGGCCCAGGCCGAGGGGGAGCCCTGCTACATCCGG
GACAGTGGGGTGGCCCCCTACAGACCCTGGCCCTGAGTCTGCTGCTGGGCCCCCAACCCACGAGGTGCTGG
TGACAGGCACTGGTGGCATCCAGTGGTGGCCAGTAGAGGAGGAGGTGAACAAGGAGGAGGGTTCTAGTGG
CAGCAGTGGCAGGAACCCCCAAGCCAGCCTGTTTGGGAAGAAGGCCAAGGCTCACAAGGCAGTCGGCCAG
CCGGCAGACAGGCCGCGGGAGCCACTGTGGGCCTACTTCTGTGACTTCCGGGACATCACCCACGTGGTGC
TGAAAGAGCACTGTGTACAGATCCACCGGCAGGACAACAAGTGCTGGAGCTGAGCTTGCCCTCCCGGGC
TGCGGCGCTGTCTTTCGTGTGCTGGTGGACGGCTATTTCCGCCTGACGGCCGACTCCAGCCACTACCTG
TGCCACGAGGTGGCTCCCCACGGCTGGTGTATGAGCATCCGGGATGGGATCCACGGACCCCTGCTGGAGC
CATTTGTGCAGGCCAAGCTGCGGCCCGAGGACGGCCTGTACCTCATTCAGTGGAGACCAGCCACCCCTA
CCGCCTGATCCTCACAGTGGCCCAGCGTAGCCAGGCACCAGACGGCATGCAGAGCTTGCGGCTCCGAAAG
TTCCCCATTGAGCAGCAGGACGGGGCCTTCGTGCTGGAGGGCTGGGGCCGGTCTTCCCCAGCGTTCGGG
AACTTGGGGCTGCCTTGCAGGGCTGCTTGCTGAGGGCCGGGGATGACTGCTTCTCTCTGCGTCGCTGTTG
CCTGCCCCAACCAGGAGAAACCTCCAATCTCATCATCATGCGGGGGGCTCGGGCCAGCCCCAGGACACTC
AACCTCAGCCAGCTCAGCTTCCACCGGGTTGACCAGAAGGAGATCACCCAGCTGTCCCACTTGGGCCAGG
GCACAAGGACCAACGTGTATGAGGGCCGCTGCGAGTGGAGGGCAGCGGGGACCCTGAGGAGGGCAAGAT
GGATGACGAGGACCCCTCGTGCCTGGCAGGGACCGTGGGCAGGAGCTACGAGTGGTGTCAAAGTGCTG
GACCCTAGTCACCATGACATCGCCCTGGCCTTCTACGAGACAGCCAGCCTCATGAGCCAGGTCTCCCA
CGCACCTGGCCTTCGTGCATGGCGTCTGTGTGCGCGGCCCTGAAAATATCATGGTGACAGAGTACGTGGA
GCACGGACCCCTGGATGTGTGGCTGCGGAGGGAGCGGGGCCATGTGCCCATGGCTTGGAAGATGGTGGTG
GCCAGCAGCTGGCCAGCGCCCTCAGCTACCTGGAGAACAAGAACCTGGTTTCATGGTAATGTGTGTGGCC
GGAACATCCTGCTGGCCCGGCTGGGGTTGGCAGAGGGCACCAGCCCCTTCATCAAGCTGAGTGATCCTGG
CGTGGGCCTGGGCGCCCTCTCCAGGGAGGAGCGGGTGGAGAGGATCCCCTGGCTGGCCCCCGAATGCCTA
CCAGGTGGGGCCAACAGCCTAAGCACCGCCATGGACAAGTGGGGGTTTGGCGCCACCCTCCTGGAGATCT
GCTTTGACGGAGAGGCCCTCTGCAGAGCCGCAGTCCCTCCGAGAAGGAGCATTTCTACCAGAGGCAGCA

FIG. 11A (Cont.)

CCGGCTGCCCCGAGCCCTCCTGCCCACAGCTGGCCACACTCACCAGCCAGTGTCTGACCTATGAGCCAACC
CAGAGGCCATCATTCCGCACCATCCTGCGTGACCTCACCCGGCTGCAGCCCCACAATCTTGCTGACGTCT
TGA CTGTGAACCCGGACTCACCGGCGTCGGACCCTACGGTTTTCCACAAGCGCTATTTGAAAAAGATCCG
AGATCTGGGCGAGGGTCACTTCGGCAAGGTCAGCTTGTACTGCTACGATCCGACCAACGACGGCACTGGC
GAGATGGTGGCGGTGAAAGCCCTCAAGGCAGACTGCGGCCCCAGCACCGCTCGGGCTGGAAGCAGGAGA
TTGACATTCTGCGCACGCTCTACCACGAGCACATCATCAAGTACAAGGGCTGCTGCGAGGACCAAGGCGA
GAAGTCGCTGCAGCTGGTCATGGAGTACGTGCCCCCTGGGCAGCCTCCGAGACTACCTGCCCCGGCACAGC
ATCGGGCTGGCCCAGCTGCTGCTCTTCGCCCAGCAGATCTGCGAGGGCATGGCCTATCTGCACGCGCAGC
ACTACATCCACCGAGACCTAGCCGCGCGCAACGTGCTGCTGGACAACGACAGGCTGGTCAAGATCGGGGA
CTTTGGCCTAGCCAAGGCCGTGCCCCGAAGGCCACGAGTACTACCGCGTGCGCGAGGATGGGGACAGCCCC
GTGTTCTGGTATGCCCCAGAGTGCCCTGAAGGAGTATAAGTTCTACTATGCGTCAGATGTCTGGTCCTTCG
GGGTGACCCTGTATGAGCTGCTGACGCACTGTGACTCCAGCCAGAGCCCCCCCCACGAAATTCCTTGAGCT
CATAGGCATTGCTCAGGGTCAGATGACAGTTCTGAGACTCACTGAGTTGCTGGAACGAGGGGAGAGGCTG
CCACGGCCCCGACAAATGTCCCTGTGAGGTCTATCATCTCATGAAGAACTGCTGGGAGACAGAGGCGTCCT
TTCGCCCCAACCTTCGAGAACCTCATACCCATTTCTGAAGACAGTCCATGAGAAGTACCAAGGCCAGGCCCC
TTCAGTGTTACAGCGTGCTGAGGCACAATGGCAGCCCTGCCTGGGAGGACTGGACCAGGCAGTGGCTGC
AGAGGGAGCCTCCTGCTCCCTGCTCCAGGATGAAACCAAGAGGGGGATGTCAGCCTCACCCACACCGTGT
GCCTTACTCCTGTCTAGAGACCCACCTCTGTGAACTTATTTTTCTTTCTTGGCCGTGAGCCTAACCATG
ATCTTGAGGGACCCAACATTTGTAGGGGCACTAATCCAGCCCTTAAATCCCCAGCTTCCAAACTTGAGG
CCCACCATCTCCACCATCTGGTAATAAACTCATGTTTTCTCTGCTGGAAAAAAAAAAAAAAAA (SEQ ID
NO: 90)

FIG. 11B

TYK2 ORF 3564 bp

ATGCCTCTGCGCCACTGGGGGATGGCCAGGGGCAGTAAGCCCGTTGGGGATGGAGCCCAGCCCATGGCTG
CCATGGGAGGCCTGAAGGTGCTTCTGCACTGGGCTGGTCCAGGCGGCGGGAGCCCTGGGTCACTTTCAG
TGAGTCATCGCTGACAGCTGAGGAAGTCTGCATCCACATTGCACATAAAGTTGGTATCACTCCTCCTTGC
TTCAATCTCTTTGCCCTCTTCGATGCTCAGGCCCAAGTCTGGTTGCCCCCAAACCACATCCTAGAGATCC
CCAGAGATGCAAGCCTGATGCTATATTTCCGCATAAGGTTTTATTTCCGGAACCTGGCATGGCATGAATCC
TCGGGAACCGGCTGTGTACCGTTGTGGGCCCCCAGGAACCGAGGCATCCTCAGATCAGACAGCACAGGGG
ATGCAACTCCTGGACCCAGCCTCATTGAGTACCTCTTTGAGCAGGGCAAGCATGAGTTTGTGAATGACG
TGGCATCACTGTGGGAGCTGTGACCGAGGAGGAGATCCACCACCTTTAAGAATGAGAGCCTGGGCATGGC
CTTTCTGCACCTCTGTACCTCGCTCTCCGCCATGGCATCCCCCTGGAGGAGGTGGCCAAGAAGACCAGC
TTCAAGGACTGCATCCCGCGCTCCTTCCGCCGGCATATCCGGCAGCACAGCGCCCTGACCCGGCTGCGCC
TTCGGAACGTCTTCCGCAGGTTCCCTGCGGGACTTCCAGCCGGGCGACTCTCCCAGCAGATGGTCATGGT
CAAATACCTAGCCACACTCGAGCGGCTGGCACCCCGCTTCGGCACAGAGCGTGTGCCCGTGTGCCACCTG
AGGCTGTGGCCAGGCCGAGGGGGAGCCCTGCTACATCCGGGACAGTGGGGTGGCCCTACAGACCCTG
GCCCTGAGTCTGCTGCTGGGCCCCCAACCCACGAGGTGCTGGTGACAGGCACTGGTGGCATCCAGTGGTG
GCCAGTAGAGGAGGAGGTGAACAAGGAGGAGGGTTCTAGTGGCAGCAGTGGCAGGAACCCCCAAGCCAGC
CTGTTTGGGAAGAAGGCCAAGGCTCACAAGGCAGTCGGCCAGCCGGCAGACAGGCCGCGGGAGCCACTGT
GGGCTACTTCTGTGACTTCCGGGACATCACCCACGTGGTGCTGAAAGAGCACTGTGTGAGCATCCACCG
GCAGGACAACAAGTGCCTGGAGCTGAGCTTGCTTCCCGGGCTGCGGCGCTGTCTTCGTGTGCTGGTG
GACGGCTATTTCCGCTGACGGCCGACTCCAGCCACTACCTGTGCCACGAGGTGGCTCCCCACGGCTGG
TGATGAGCATCCGGGATGGGATCCACGGACCCCTGCTGGAGCCATTTGTGACAGCCAAGCTGCGGCCCCGA
GGACGGCCTGTACCTCATTCACTGGAGCACCAGCCACCCCTACCGCCTGATCCTCACAGTGGCCCAGCGT
AGCCAGGCACCAGACGGCATGCAGAGCTTGC GGCTCCGAAAGTTCCCCATTGAGCAGCAGGACGGGGCCT
TCGTGCTGGAGGGCTGGGGCCGGTCTTCCCCAGCGTTCCGGAACTTGGGGCTGCCTTGAGGGCTGCTT
GCTGAGGGCCGGGATGACTGCTTCTCTGCGTCTGCTGTTGCCTGCCCCAACCCAGGAGAAACCTCCAAT
CTCATCATCATGCGGGGGGCTCGGGCCAGCCCCAGGACACTCAACCTCAGCCAGCTCAGCTTCCACCGGG
TTGACCAGAAGGAGATCACCCAGCTGTCCCACTTGGGCCAGGGCACAAGGACCAACGTGTATGAGGGCCG
CCTGCGAGTGGAGGGCAGCGGGGACCTGAGGAGGGCAAGATGGATGACGAGGACCCCTCGTGCCTGGC
AGGGACCGTGGGCAGGAGCTACGAGTGGTGCTCAAAGTGCTGGACCCTAGTCACCATGACATCGCCCTGG
CCTTCTACGAGACAGCCAGCCTCATGAGCCAGGTCTCCCACACGCACCTGGCCTTCGTGCATGGCGTCTG
TGTGCGCGGCCCTGAAAATATCATGGTGACAGAGTACGTGGAGCACGGACCCCTGGATGTGTGGCTGCGG
AGGGAGCGGGGCCATGTGCCATGGCTTGGAAGATGGTGGTGGCCAGCAGCTGGCCAGCGCCCTCAGCT
ACCTGGAGAACAAGAACCTGGTTTCATGGTAATGTGTGTGGCCGGAACATCCTGCTGGCCCGGCTGGGGTT
GGCAGAGGGCACCAGCCCCTTCATCAAGCTGAGTGATCCTGGCGTGGGCCTGGGCGCCCTCTCCAGGGAG
GAGCGGGTGGAGAGGATCCCCTGGCTGGCCCCGAATGCCTACCAGGTGGGGCCAACAGCCTAAGCACCG
CCATGGACAAGTGGGGGTTTGGCGCCACCCTCCTGGAGATCTGCTTTGACGGAGAGGCCCTCTGCAGAG
CCGCAGTCCCTCCGAGAAGGAGCATTTCTACCAGAGGCAGCACCGGCTGCCCCAGCCCTCCTGCCACAG
CTGGCCCACTCACCAGCCAGTGTCTGACCTATGAGCCAACCCAGAGGCCATCATTCCGCACCATCCTGC
GTGACCTACCCGGCTGCAGCCCCACAATCTTGCTGACGTCTTGACTGTGAACCCGGACTCACCGGCGTC
GGACCCTACGGTTTTCCACAAGCGCTATTTGAAAAAGATCCGAGATCTGGGCGAGGGTCACTTCGGCAAG
GTCAGCTTGTACTGCTACGATCCGACCAACGACGGCACTGGCGAGATGGTGGCGGTGAAAGCCCTCAAGG
CAGACTGCGGCCCCCAGCACCGCTCGGGCTGGAAGCAGGAGATTGACATTCTGCGCACGCTCTACCACGA

FIG. 11B (Cont.)

GCACATCATCAAGTACAAGGGCTGCTGCGAGGACCAAGGCGAGAAGTCGCTGCAGCTGGTCATGGAGTAC
GTGCCCCCTGGGCAGCCTCCGAGACTACCTGCCCCGGCACAGCATCGGGCTGGCCCAGCTGCTGCTCTTCG
CCCAGCAGATCTGCGAGGGCATGGCCTATCTGCACGCGCAGCACTACATCCACCGAGACCTAGCCGCGCG
CAACGTGCTGCTGGACAACGACAGGCTGGTCAAGATCGGGGACTTTGGCCTAGCCAAGGCCGTGCCCCGAA
GGCCACGAGTACTACCGCGTGCGCGAGGATGGGGACAGCCCCGTGTTCTGGTATGCCCCAGAGTGCCTGA
AGGAGTATAAGTTCTACTATGCGTCAGATGTCTGGTCCTTCGGGGTGACCCTGTATGAGCTGCTGACGCA
CTGTGACTCCAGCCAGAGCCCCCCCCACGAAATTCCTTGAGCTCATAGGCATTGCTCAGGGTCAGATGACA
GTTCTGAGACTCACTGAGTTGCTGGAACGAGGGGAGAGGCTGCCACGGCCCGACAAATGTCCCTGTGAGG
TCTATCATCTCATGAAGAACTGCTGGGAGACAGAGGCGTCCTTTCGCCCCAACCTTCGAGAACCTCATACC
CATTCTGAAGACAGTCCATGAGAAGTACCAAGGCCAGGCCCTTCAGTGTTTCAGCGTGTGCTGA (SEQ
ID NO: 91)

FIG. 11C

TYK2 protein (~134 kD)

MPLRHWGMARGSKPVGDGAQPMAMGGLKVLLHWAGPGGGEPWVTFSESSLTAEVCIHIAHKVGITPPC
FNLFALEFDAQAQVWLPPNHILEIPRDASLMLYFRIRFYFRNWHGMNPREPAVYRCGPPGTEASSDQTAQG
MQLLDPASFAYLFEQKHEFVNDVASLWELSTEEIHHFKNESLGMAFLHLCHLALRHGIPLEEVAKKTS
FKDCIPRSFRRHIRQHSALTRLRLRNVFRFLRDFQPGRLSQQMVMVKYLATLERLAPRFGTERVPVCHL
RLLAQAEGEPCYIRDSGVAPTDPGPESAAGPPTHEVLVTGTGGIQWWPVEEEVNKEEGSSGSSGRNPQAS
LFGKKAKAHKAVGQPADRPREPLWAYFCDFRDITHVVLKEHCVSIHRQDNKCLELSLPSRAAALS FVSLV
DGYFRLTADSSHYLCHEVAPPRLVMSIRDGIHGPLEPFVQAKLRPEDGLYLIHWSTSHPYRLILTVAQR
SQAPDGMQSLRLRKFPFIEQQDGA FVLEGWGRSFPSVRELGAALQGCLLRAGDDCFSLRRCLPQPGETSN
LIIMRGARASPRTLNLSQLSFHRVDQKEITQLSHLGQGTRTNVYEGRLRVEGSGDPEEGKMDDDEDPLVPG
RDRGQELRVVLKVLDPSHHDIALAFYETASLMSQVSHTHLAFVHGVCVRGPENIMVTEYVEHGPLDVWLR
RERGHVPMAMKMMVVAQQLASALSYLENKNLVHGNVCGRNILLARLGLAEGTSPFIKLSDPGVGLGALSRE
ERVERIPWLAPECLPGGANSLSTAMDKWGFATLLEICFDGEAPLQSRSPSEKEHFYQRQHRLPEPSCPQ
LATLTSQCLTYEPTQRPSFRTILRDLTRLQPHNLADVLTVPNDSPASDPTVFHKRYLKKIRDLGEGHFGK
VSLYCYDPTNDGTGEMVAVKALKADCGPQHRSGWKQEIDILRTLYHEHI IKYKGCCEDQGEKSLQLVMEY
VPLGSLRDYLPRHSIGLAQLLLFAQQICEGMAYLHAQHYIHRDLAARNVLLDNDRLVKIGDFGLAKAVPE
GHEYYRVREDGDS PVFWYAPECLKEYKFYYASDVWSFGVTLYELLTHCDSSQSPPTKFLLELIGIAQQQMT
VLRLTELLERGERLPRPDKCPCEVYHLMKNCWETEASFRPTFENLIPILKTVHEKYQGQAPS VFSVC &
(SEQ ID NO: 92)

FIG. 11D (Cont.)

caacatagggagaccttgtctcaacaaaataatTTTTtagaaattagccagggcatgggtggtacatgacctgtagtccca
gcaacttgagagcctgaggcaagaggatcgcttgagcccaggagttcaagcctgcaatgaggtatggtttcgtcact
gcactccagcctgggtgacagagtgagaccagtggtcattaaaaaaagaaagtcaggccggggcgcggtgggtcatgc
ctgtaatcccagcactttgggaggtgaggtggtggatcacttgaggtcagaagtttgagagcagcctggccaaca
tggcaaaactccagctctactaaaaatacaaaaactagccgggctggggtgcgcacctgtaattccagctactcag
gaggtgagggcaggagaatagcttgaacctgggacatggaggttgagtgagccgagatcacagcactgcactccag
gctgggtgacagagctagactccgtttcaaaaataaaaaaaggccaggtgcgggtgggtcatgacctgtaatcccaaca
ctttcagaggctgaggcgggcagatcacctgagatcaggagttcaagaccagcctggccaatgtagtgaaaccgcat
ctctactaaaaatacaaaaatgtagccagggcatgggtggcaggtgcctgcaatcccagctactcagaaggctgaggcag
gagaatagcttgaacctgggacatggaggttgagtgagccgagatcacagcactgcactccaggtgggtgacaga
gctagactccgtttcaaaaataaaaaaaggccaggtgcgggtgggtcatgacctgtaatcccaacactttcagaggctg
aggcgggcagatcacctgagatcaggagttcaagaccagcctggccaatgtagtgaaaccgcatctctactaaaaat
acaaaatgtagccaggcatgggtggcaggtgcctgcaatcctagctactcaggagggtgaggcaggagaattgcttga
acttgggaggtggaggtgcagtgagctgagatcggtccattgtaactccagcctgggcaacaagagcgaaactccat
ctcaaagaaaaaaaaaaaaaaaaaagaaagaaaggctttccctcctcctgcacaggaaattgcacttccaaatatat
ctactacagtctttctttttttcttttttttgagatggaggtctcgtctgtcccagggtgaagtgcaatggcgcaat
ctcagctaaactgcaacctccgctcccggttcaagcgattctccggcctcagcctcccgagtagctgggactacag
gtgtgtgccaccatgcccggtagattttttttttttttttttgtatttttagtagaggtggggtttcccggtgttagc
caggatggctcttgatctcctgacctcgtgatctgcgcgcctcggccctccaaagtgcctgggattacagggtgtgagcc
acctgcccggcccaattttcttttttttaatgacataggggtctcactctgtcaccagggtggagtgagtggtgtg
acctcagctcactgcaacctccatgtcccagggttcaagccattctcctgcctcagccgcctgagtagctggaattac
aggtgcacaccccacgcccggctagtttttatatttttagtagagctggagttttgcoatgttgggccagggtggtct
cgaacttctgacctcaagtgatccacctgcctcagcctcccaaagtgcctgggattacaggcgtgagccaccatgcc
aaccaaggaaaggcctttctaaggcggctttgggtacatataccttccctccagaaactcctcttacctccccattc
ccagctttccatgactggatgggggcattgtaatcaatttttataaaccttccacttccccattgatggctgagtc
ttttttttttttctttcgtagagacagctctctctcttgcagggtgaaatgcagtggtgcaatcatgacgcattg
cagcctctaactgccagggtcaagtgattttcctgcctcagctctccaaatagctggggctacaggccacatgtcac
cacaccggataatttttgattttttgcatggtgcctcatgctgggtctcaaaactcctgggctcaagccatccaccc
gccttggttcccaaagtgttgggagggcatgagccatcgcgccctggccatgactgatttttttttttttttgaaac
gaagtctcgtcagtcctcaaagtggagtgcaatggcacaatcttggctcactgcaacctccgctcccggttct
agcgatcctcctgcctcagcctcctgagtagctgggattacaggcgcctgccaccacgcccagctaaattttgtatt
tttagtagagacggggtttccacatggttggtcaggctggtcttgaacatgcatgctgttccccctgcctcggcctct
caaagtgcctgggtagacagggtgtgagccaccgtgccggcttttttttttttttttttttttgagacggagttctgc
tctgtcaccagggtggagtgagtggtgtgatctcggtcactgcaacctccgctcccggttcaagtgattctc
ctgcctcagcctcctgagtagctgggattacaggcgcaccaccacacctggctaatttttgtattttcagtagag
acgggggtttccacatggttgccagggtggtctcgaactcctgacctcaagtgatccacctgcctcagactccaaa
gtgttgggattacaagtgtgcacctcgagcctggccatggctgattctttgtttgtttgtgtgtgttgaga
cagagtctcgtctgtcgcagggtggagtgagtgggcgatctcagctcactgcaagctccgctcccggttc
acgccattctcctgcctcagcctcccgaggagctgggactataggcaccaccacacgcctggctaattttttgta
tttttagtagagacagggtttcactgtgctagccaggatgggtctcgatctcctgacctcgtgatccgcccacctcag
ctcccaaagtgcctgggattacagggtgtgagccactgtgcccgcctccatgactgattattaaaggattcagacagg
tgtcggggtgatgtgaaacacgctcagcagagggaacagcatgcatggaggcccgagagcttgatctgtttgagga
agcgaaagaccttcagggtgactgggtatcaattttgagagtttggagggaagtaacaattgaaattggggcat
agcaggcaaaagccagggtgggggtactcagaatcaagggtcaagggttgggcattctccgggagcaatggggagc
catggaaggtttaagccagagagtgaagcagttctgactagttcttttttttttttttttttttactgggtctt
taagaagttgtctcaagccaggcacagtgggtcacgcctgtaatcccagcactttgggaggtgagggcgggcagatc
acttgaggttaggagtttgagaccagcatggccaacatgggtgaaaccccgctctctactaaaaataaaaaattagct
ggcggtggtgacgcacatccagctacttgggaggtgagggcaggagaatcacttaaacctgggagggcggaggttga
gtgagccgagatcgccactgcactccagcctgggtgacagagactcgggtctcaaaaaacaaaaatatgaaacatt

FIG. 11D (Cont.)

ggccaggtgccgtagctcacacttgtaatccttgcgcttagggaggccaaggcgggtggatcacttgaggtcaggag
ttggagaccagtctggtcaacatgatgaaaccccatctctactgaaaatacaaaaaattagccgggctggtggcgg
gcgcctgtaatcccagctactcaggaggtgaggcaggagaatcgcttgaaccagggggtggaggttacagtgagc
cgagatcacgccactgcactccagcctgggcaacaagagcaaaaactctatccaaaaaaaaaaaaatacaaaaaat
tagctgggcatggtggcacatgcctgtagtcccagctactcaggaggtgagacaggagaatgacttgaacctgaaa
ggcagaagttgcggtgaaccaagattgcaacattgcactccagctctgggtgactccatctaaaaaaaaatatattg
tttcacagggaggtatcctcccttcacagaggggcagtataattccatggtgtgactttggagagtgatttgataac
acagtcaaagctggatgcagtggtcgcgcctgcaatcccagcagttttggaggctaagcgtggaggatcacttgag
cccaggagttcaagatcagcctgggcaacatcgcgaaaccccgctctactaaaaatacaaaaaattagccagggtgt
gctgtgtgctggtgtgcacctgtagtcccagctactcaggaggtgaggtgggaggatcacttgagcccgggaggtt
gaagctacagtgatcatgccactgcccttcagcctaggtgacagagcaagatgctgtctcaaaaaaaaaaaaaaaa
aaaaatatatatatatatatatatatatagctgtgtgtggtggcgtgttctgtatttcccagctactcaggag
actgaggcaggaggactgcttgagccaggggaggtcatggtgcactgagccatgattgcaccactgtattccagcct
atatgtcagagcaagatcctgcttcaaaaaaaaaaaaaaacccaggcgggcgcgtggtcgcgcctgtaatccca
gcactttgggaggtgaggtgggcgaatcatgaggtcaggagattgcgaccgtcctggctaacacggtgaaaccccg
tctctactaaaaatacaaaaaattagctgggcatggtggcgggcgcctgtagtcccagctactcgggaggtga
ggcaggagaatggcgtgaaccaggaggcagagcttgcaagtgcagcggagattgagccaccacactccagcctgggcg
acagatcgagactccgtctcaaaaaaaaaaaaaaacactcagaacatcttccctgagggcatatagttccactcct
ggacagtgcagtcactctataaacagtcagctcatgtgctcccagggaacttatgagaatgttttagagccgcctt
gtttatcatatggaaaaggggagaccattgcaaagtcctcaataagagactgcctaagtaattgcaatctggttg
aatactggactactgcgcagccgggagagaaggaatttaaccagaattactcatgtcattctgcagtgaagctttataac
ctctcgtagcctgaaaaagcacattgcagagggacgtctgcagtataataccacacatttacagttcataacttttt
tttttttgagacagagtttctgctcctgttgcccagggttgagtgcaatagcacgatctcagctcaccgcaacctcca
cctcccagggttcaatttctcctgcctcagcctcccagtagctgggattacaggcatgcgccaccacgcccagcta
tttgatttttagtagagacggggtttctccatttttggtcagggtggtctcgaaactccgacctcagggtgatccgcc
tcctcagcctcccaaagtgcgtgggattacaggcgtgagccactgtgccagccacagttcataacatttaaggcta
gattgtagaggatatgaaaattataaacactacttcagggcagagcacactggcttatgctgtaatcccagcactt
tgagggtgagtagctcagcctcccagctactcaggaggtgaggtgggaggatcacttgagcctaggagtttgaga
ccagcctggacaacatggtgagaccccttctctacaaaaactacaaaaatcagctgggcatggtggcgtgcgcctgt
agtcacagctccttgggaggtgaggcaggaggattgctttagcccagtaagtcgaggggtgcagtgagccatgatca
caccactgcactccagccacacctggctaattttattttatttttagtttttttagagatgaggggtctcactatggt
gcttaggctggtcttgaactcctaggttcaagtgattctcccgccttagcctcccaaagtgcgtgggattacagatgt
gagccaccatacccagcttttagataaaaattattttattttcattttatttttttagacagagcctcactcttt
caccaggtggagcgcagtggtgatctcggtcactgcaacctctgcctccagggttcaagcaattctcttgcct
cagcctgctgagtagctgcgattacaggtgctcaccaccacaccagctaattttttgtatttttagtagagacgagg
ttttgctctggtggccaggctggtctcaaacctcctgcctcaagtgcttcacctgcctcagctccgaaagtgcgtgg
gactataggcatgggcccactgcacccagccagataccatttttttttttttttgagacagctcttgcgtgtgcgc
aggctggagtgagtggtggcacaatctcggtcactgcacctccacctccgggttcaagcgattgtcctgcgtcaac
ctccggagtagctgggactaaaggtgcacccaccatggccagccaatttttttttttcagatctcactctgcgcgc
caggctggagctcagtggtcacgatcttggctcactgcaacctccgcctccgggttcaagcagtcctcctgcctcag
cctcccaagtagctggaattacaggcacatgccaccagggtccagctatttttttttttaatatagagatgggatttcac
catgctggccacgtggtctcgaaactcctgacctcaagtgatccacctgcctcagactcccaaagtgttgggattac
aagtggtgcacctcgagcctggccatggctgattctttgtttgtttgtttgtttgtttgagacagagctctgcctc
tgtcggccaggctggagtgagtggtggcgaatctcggtcactgcaagttccacctcctgaattcacaccattctctt
gcctcagcctcccaaagtagctgggactacaggcatccgccaccatgcctggctaattttttgtatttttttagtagag
acggggtttcaccatggttagccaagatggtcttgatctcctgacctcgtggtctgcccacctcagcctcccaaagtg
ctgggattataggcgtgagccactgcgcccggccgacaggtacaatttttaaatgacatcagtagaagaaagatacc
ctgggtagctggtgtgggaaattacagcaagtctggagactggattgggggttgaggtttgcatttgaaggtggtgga
tgagagccccaggacatggaggggtgcaggaagagcaggtcttgggtgaccgatctgttttgagcaggtgggagaggtt

FIG. 11D (Cont.)

gaagataaaggaccaagtctctgggttgaagaactgggcagaaagccacacacccctggctcatagccaataatcccatca
 actccagagattgagggtgggaggatcgcttgagccctggagggtgagggtgagctgagctaggattgcaccaactgca
 ctgcaacctgggtgggcaaaagaaagacccctga cccttaaaaaaaaagggtcagaggaaaggaagggttgagtttgcca
 accacagataagat atgttg tttggaacaagccat atctaagggggggctgtgatgtccagggggagacagccagg
 aggtgactggatgt tctttt gggtctggagt tcaggagagacgcccagacaattgcagcagtttgtagagttatgtggg
 tcaggctaggattagttttgctgcaggaataagcaatcccccaaatcttaatacaagggtttaccta aaatcggccga
 gcacagtggtcatgctgt aatctcagcacctttgggggggccaagggtgggcagatcacctgagggtcaggagt ttgag
 accagcctggccaa catggcgaaacctcgtctcttctaaaaat acaaaa actagccgggttgggtggcgccacacctg
 taatccagcaactcaggagggtgagacaggagaatcacttga acccaggaggcggaagt tgcagcgagccaaagatc
 gtgtcactgctctc cagcctgggcaa cagca agactccatctcaaaaa aaaaaaaagt tttacttaaaatcaagt
 acaggttgatgggttcacctctgtcttacagcttcacatctgtgcacaa gaagccttcacataacctcgggtggtga
 aggtggcaagagaa tttatt tccaggcttggaatgt acccagccctttggccagaaccagcagtgcaattggctc
 catcctagctgcaaggggaggctggga aatgt agtctcctgtgaggcca cagca tgggcaacatacgacactgtctc
 tgccaggggcaacctgaga ctgaagctgcaaggtggggggggggggcctcagaggagtggtgaaggcagacatgggt
 ctgggggggtcttttgggggtgacggtagcaaa tgacgtgactcctgacctcgtttctcaactgactccagGTATCACTC
 CTCCTTGCCTCAATCTCTTTGCCCTCTTCGATGCTCAGGCCCAAGTCTGCTTGCCTCCCAACCACATCTAGAGATC
 CCCAGAGATGCAAGCCTGATGCTATATTCCGCATAAGGtgggtgggagacctttgcaaaagctcgtc cctcctctgtgc
 tgaagctggtctgactctgt gctaaagccccagctggttcctccttctcagGTTTATTTCCGGAAGTGGCATGG
 CATGAATCCTCGGGAACCGGCTGTGTACCGT TGTGGGCCCCAGGAACC GAGGCATCCTCAGATCA GACAGCACAGG
 GGATGCAACTCCTGGAACCCAGCCTCA TTTGAGTACCTCTTTGAGCTAGgt atgagcagggttgggtggcaagactat
 ttgtgggagacttaggggca gttgaggagccccatttcctcctctgat tcaata tagctaatagggtttcaactcat
 gctatctggggatctttttt tttttt ttgagaaggagtcctgtctgtctgcccaggctga agtgcaagtgtgtgac
 tcgggtcactgcaagctcggcctccagggttcctgtcattttcctgctcagcctcccaagtagctgggactacagg
 cgcgcgcaccatgcccagctaat tttgtatttttagtggaacgggttttcacagtgttagctaggatggtctc
 catctcctgacctcgtgatctgcccgcctcggcctcccaagtgctgggattacaggcgtgagcca cgcctccagc
 cgatctttttttt tgaatggagtcttgctctgtggcccaggctggagtgcagtgggctgatctcagctcactgca
 acttcacactcccaaggttga agcgtatcttcacactcagcctcccaagt aactgagatta cagaagcccgccaccac
 cccaggctaat ttttctatt tttagt aaaga cgggggtttcaccatgttgccagggtggtctcaaa cctcctgacctc
 aagtgttcaactg cctcagcctcctaaattgtggga tttatagggtgtgagccaccatgccagcca agagttcagggt
 tgaaggggatagagtgtctgt gatgtagttagt aatgtctgcccaggccacaaagt tgacaa gaggaggggtgtatgt
 gtgtatttggcctt tccagt aaatgggtggtctctcctcctcctcaagggtgtaagacgcaggaat tagagttctctg
 gctagagt aagctt gtccaa cccctggcctg cggactgcatgtgcccaggacagctttgaatgaagcccaacaa
 attcataaaactatcttaaaa tttctgagat tttgtgtgatt tttttt ttttt ttgaga cagagt ttcattctgtg
 tgctcaggctggagtgc aatgtcatgatctcagctcactgcaa cctctg cctccagggtt caagtgtatctcctgct
 tcagcctccggagt agctgggtatcacagaca tgcccaccaca cctggctaattttgtat ttttagtataggcgagt
 tttctcca tgttggtcagggtgggtct caaactcctga cctcagggtgacc cgcgcgccttggtcctccaaagtgtgtg
 gattataggcgtgagccaccatgcccggcttgogata tttttt tttttt ttttt ttttttagctcat cagctatcatt
 agtggttagtgtatt ttatgt agggcccaaga cacttcttcagtggttccagggttagccagaagattggatacctc
 tgggcttagagaggaacgcattctcacacctc cctttctgtc aatttctctgttccagGGCAAGCATGAGTTTGTGA
 ATGAGTGGCATCACTGTGGGAGCTGTGACCGAGGAGGASATCCACCACTTTAAGAAATGAGAGCCTGGGCATGCCC
 TTTCTGCACCTCTGTACCTCGCTCTCCGCCATGGCATCCCCCTGGAGGAGGTGGCCAAGAAGACCAAGgtgtctggg
 aatgggggtggggaacgtgtgttaggca gggggaaatgt cctgagcatctgggagccagggt agactttgcataattgcc
 gtgcctctactttt tttttt tctgagccggagtctctctctgtcacct aggtggagta cagtggcgcaatcttgg
 ctcaactgcaacctctgctccaggt tcaagagattctccttctcagcctcctgagtagctggga ttacagggtgcc
 caccatcaagcccagctaat ttttgtatttttagtagagacagactgagtttca tcatgt tggccagggtgtctctg
 aactcctgacctca agcgtatcttccgccttggcttcccaagtagctgagactacaggt caagtcaccacacccgg
 ctagtgtgctgttttccagatgaggcagctgaggct caaaga ggttaagccacttgccctgtgtcaccagctggg
 cctgcaaccaggt cctgacagcctcctca ctgtgt tccccccagCTTCAGGACTGCATCCCGCCTCTCCG
 CCGCATATCCGGCAGCACAGCGCCCTGACC CGGCTGCGCCTT CGGAACGTCTTCCGCAGSTTCTGCGGSACTTCC

Ex 5

Ex 6

Ex 7

Ex 4

FIG. 11D (Cont.)

cgggaaggagtttgggggttgaggat aaatt cctacctaaggt gattct caggagatgtgggtgggaccctctgctt
 tgaggagaggcctaggtgccaggaggt cttaa tagagcggagtagggcgggcctggggtat tcogaaaggatctgctt
 ggaggagtggcccgggaaat cgggggttggggagtggtgcagggattggggagggtggagctgaccctcgtgcgctcc
 cgaccagGGTCACTTCGGCAAGSTCAGCTTG TACTGCTACGATCCGACCAACGACGGCAC TGGCGAGATGGTGGCGG
 TGAAGCCCTCAAGGCAGACTGCGGCCCCAGCACCGCTCGGGCTGGAAGCAGGAGATTGACATTCTGCGCACGCTC
 TACCACGAGCACATCATCAAGTACAA GGGCTGCTGCSAGGACCAAGgtggcggggaccccgcgagtcctagagactt
 gaggggggcgtggttagagcgtgggtgtggccttgtggcgggggccagtcctggcctagccaattagagactctcaacc
 ccaggctgcagtagcggttcttgccatctgggtgagcgtggctgggctgcctccttcccatctgggggggtgctacc
 tgctggggccaggatggagcgggagccaccagcagctccctcaa gtgagaatggggttctggatgggtccatccacc
 tgagaactgggtctagtgtgcgcggttcttgcccttccctcccgaccttggaggggctctgctgggctcaaggtagg
 cggtecccgggccgctctatcgteccca cggggcgccctctctcgggcagGCGAGAGTTCGCTGCACTGG
 TCATGGAGTACGTGCCCCTGGGCGAGCTCCGAGACTA CCTGCCCGGGCACAGCA TCGGGCTGGCCCCAGCTGCTGCTC
 TTCGCCAGCAGATCTGCGAGgttggtcggc cccgcccctgctcttgga gcttgcccttctctt cagcttgggt
 ggcttgagcgatccggtgca gctgctctca cgtccgcccctgctgcttgccagggtgcttgcttgccttgcaactg
 acctccgaactgtttggctt gcttggccaca cccctcctcccaaggga ccacgcacagcgggtcggccctccca
 actcgccctcctgcccagct ttgctcgtctagcccg tccctgcttcttggtcagcctccttttgcttggtgccac
 gtgccatccggccctctcca gggcggggtagccaccggcctgacctga ctgtcccccagccctcctggctgctcag
 gtccctggcgccgcccggcgccgaacc cggcccactgaaactcagcagccctgcccgtccccagGGCATGGCCTAT
 CTGCACGCGCAGCACTACAT CCACCGAGACCTAGCCGCGCGCAACGTCTGCTGACAAACGACAGGCTGGTCAAGAT
 CGGGGACTTTGGCTTAGCCAAGGCCGTGCCC GAAGGCCACGAGTACTACCGGTGCGCGAGGATGGGGACAGCCCCG
 TGTTCIGgtgaccaggcggtgagctctgaggggggtgctgggggtcacttggaacaggcagggtggagtagacggat
 gctgggtatcatgatagggcctgacccagtgaaagcgcggggggcttagcggtgccccgggtggggcctatggaggc
 aaggggcggtgtatagagtgcggcttggcctggtggcggggttaacccggcctagccaa tgagcgccctagccaat
 gagcatatccccgctcaggtcggtcgtcttaagggtgggtcccaactgagcccagagggtccttcagtcctcaggtg
 agggcttcagcctggtctga tccccaagccctcagtgagccccccgtggcctgagtgcccccctcccccagGTAT
 GCCCCAGAGTGCCCTGAAGGAGTATAA STTCTACTATGCGTCAGATGTCTGGTCTTCGGGCTGACCCTGTATGAGCT
 GCTGACGCACTGTGACTCCAGCCAGAGCCCC CCGGgtgagagccaggcccgagccccacggggagtttgctaga
 gcaattagaaaggcataggctggggccaggcgcggtggctcaagcctgtaatcccagcactttgggaggccgaggcg
 gggatcacgaggtcaggagatcgagaccaa cctggctaaacaggtgaaacccgtctctactaaaaatacaaaaaaa
 attagccaggcgtggtggcgggcgccctgtagtccagctactcaggaga ctgaggcaggagaatggcatgaa cccgg
 gaggcgagccttgcaagtgcggaga tctccgactgcactccagctgggtgacagagcgagactcgtctaaagaca
 ggacaggacggagagagagaga gaaagaaaaga aaggaaagaaaggcatagccggggcggggtgggtgacacctgtaat
 cgcagcactttgggaggccgaggcgggcagatcacgaggtcaggagatcgagaccagcctggccaa catggtgaaac
 ccgctctctactaaaaatacaaaaaat tagccggggcgtggtggcaggtgcctgtaatcccagctactcaggcggtga
 ggcaggggagaattgcttgaa cccggtaggcgaaaggttgcagtgagccagatcatgccactgcactccagcctgggt
 gaaagagcgagagactgtctcaaaaaaaaattctggacggtagaacaatctgctgtgaggcttaataataacggc
 aaacacctgactca cgttactctatgaata ttaggttaacgttcacagtaatcccaggtggcaga tgttat tacc
 tattttacggatgagaaaacaggcccagaga agtgaaactcaaa ttcaaagtacacagctgagaagtgttaagaggtg
 ggattggaacccatgcagtcgatttcatgaagtgcagagtga tttctt tttct tttttgagatggagtttcaact
 tgtatcccaggctggagtgcaatggcacaatctcggtcactgcaacctccgctcccaggttcagacgactcctct
 gectcagcctcccaagtagctgggat taaggcgccctgccattatgccagctaat ttttgtattt ttagtagagat
 ggggtttcaccatgttgggcaggtggtctcgaactcctgacctcaggtgatccaccacctcagccttccaaagt
 ctgggattacaggcatgagccacgctgacctgctttcattgctcttgaa aaagagcttgtcttggttataactgtagA
 AATTCCCTTGAGCTCATAGGCATTGCTCAGGGTCAGATGACASTTCTGAGACTCACTGAGT TGCTGGAAACAGGGGAG
 AGGCTGCCACGGCCCGACAAATGTCCCTGTGAGgtgagacttctttgtctttccctgacccactatcctgctgtg
 gagagggcagccccaccagccctgtttttctttctagSTCTATCATCTCATGAAGAACTTGCTGGGAGACAGAGGC
 GTCCCTTCGCCAACTTCGAGAACCTCATA CCCATTCTGAAGACAGTTCATGAGAAAGTCCAAGGCCAGGCCCTT
 CAGTGTTCAGCGTGTGCTGAAGGCACAATGGCAGCCCTGCCTGGGAGGACTGGAC CAGGCA GTGGCTGCAGAGGGAGC
 CTCCTGCTCCCTGCTCCAGGATGAAACCAAGAGGGGGATGTCAAGCTCA CCCACACCGTGTGCTTACTCTGTCTA

Ex 20

Ex 21

Ex 22

Ex 23

Ex 24

Ex 25

FIG. 11D (Cont.)

GAGACCCACCTCTGTGAACTTATTTTCTTTCTTGGCCGTGAGCCTAACCATGATCTTGAGGGACCCAACATTTGT
AGGGGCACTAATCCAGCCCTTAAATCCCCAGCTTCCAACTTGAGGCCCACCATCTCCACCATCTGGTAATAAACT
CATGTTTTCTCTGCTGGA (SEQ ID NO: 93)

FIG. 11E

Intron involved in TYK2 translocation

gtgtgtggcctgtgtgtgtggggcctgggtcggtcagggagggccaggagcccaggagttcgagaccagcctgggcaac
agggcgagaccccatcttttttgtttgttttgttttaggtggagtttcgcccgtgtcaccaggtggagtgcaatgg
catgatctcgggtcactgcaaccttcgctcccggttcaaagtattctccgcctcagcctcccaagtagctggga
ttataggctcctgccatcacgcccagctaatttttgtatttttagtagagatgggggtttaccatgttggtcaggct
ggtctcgaacttctgacctcgtgatccatccgcctcagcctcccaaagtgtgtgggattccaggcgtgagccaccacg
cccagctggcccatcttttaaaaataaacaatatagccaggtgtggtgggtcatacctgtaatcccagcagtttggg
aggccaaggcggtggatcacctgaggtcaggagtttgagaccagcctggccagtatggcaaaccttgctctact
aaaaatacaaaaaaattagccaatgtggttaacttgacactgtaggccgaggtactcaggaagttgaggtgggagga
tcacctgagctcaggaatttgaggccgtagtaagctatgatcacaccactgcaccccagcctgggcagcagcatagc
tagaccccatctccacaaaaatttaagaatcagctaggtgtggtgatgtgcacctgtaattctcgctacttgga
ggcttgagcccaggagtttgaagctgcagtgagctatgttcgtgccactatactccaacctgagagacagagtgaga
ccctgtcttaaaaaaaaaaaaaaaaaaaaaaaaaacaagaaaaaataaccccccaaaaaacagaaagaaaccc
agtgatagtcacagttgtccttttcacattggctgtccctatggggaccaggtttgggggttggcgtctgtgcctctc
ctgagtggccacacccctctcctgccacctcag (SEQ ID NO: 94)

FIG. 12A

NPM 1 cDNA 1449 bp

AGAAAGGAGTGGGGTTGAAAAGCGCTTGCGCAGGACGGCTACGGTACGGGGGTGGGAGGGCTTCGGAGCA
CGCGCGCGGAGGCGGGACTTGGGAAGCGCTCGCGAGATCTTCAGGGTCTATATATAAGCGCGGGGAGCCT
GCGTCCTTTCCCTGGTGTGATTCCGTCCTGCGCGGTTGTTCTCTGGAGCAGCGTTCTTTTATCTCCGTCC
GCCTTCTCTCCTACCTAAGTGCGTGCCGCCACCCGATGGAAGATTTCGATGGACATGGACATGAGCCCCCT
GAGGCCCCAGAACTATCTTTTCGGTTGTGAACTAAAGGCCGACAAAGATTATCACTTTAAGGTGGATAAT
GATGAAAATGAGCACCAGTTATCTTTAAGAACGGTCAGTTTAGGGGCTGGTGCAAAGGATGAGTTGCACA
TTGTTGAAGCAGAGGCAATGAATTACGAAGGCAGTCCAATTAAAGTAACACTGGCAACTTTGAAAATGTC
TGTACAGCCAACGGTTTCCCTTGGGGGGCTTTGAAATAACACCACCAGTGGTCTTAAGGTTGAAGTGTGGT
TCAGGGCCAGTGCATATTAGTGGACAGCACTTAGTAGCTGTGGAGGAAGATGCAGAGTCAGAAGATGAAG
AGGAGGAGGATGTGAAACTCTTAAGTATATCTGGAAAGCGGTCTGCCCCTGGAGGTGGTAGCAAGGTTCC
ACAGAAAAAAGTAAACTTGCTGCTGATGAAGATGATGACGATGATGATGAAGAGGATGATGATGAAGAT
GATGATGATGATGATTTTGATGATGAGGAAGCTGAAGAAAAAGCGCCAGTGAAGAAATCTATACGAGATA
CTCCAGCCAAAAATGCACAAAAGTCAAATCAGAATGGAAAAGACTCAAACCATCATCAACACCAAGATC
AAAAGGACAAGAATCCTTCAAGAAACAGGAAAAAACTCCTAAAACACCAAAGGACCTAGTTCTGTAGAA
GACATTAAAGCAAAAATGCAAGCAAGTATAGAAAAAGGTGGTTCTCTTCCCAAAGTGAAGCCAAATTCA
TCAATTATGTGAAGAATTGCTTCCGGATGACTGACCAAGAGGCTATTCAAGATCTCTGGCAGTGGAGGAA
GTCTCTTTAAGAAAAATAGTTTAAACAATTTGTTAAAAAATTTTCCGTCTTATTTCAATTTCTGTAACAGTT
GATATCTGGCTGTCCTTTTTTATAATGCAGAGTGAGAACTTTCCCTACCGTGTTTGATAAATGTTGTCCAG
GTTCTATTGCCAAGAATGTGTTGTCCAAAATGCCTGTTTAGTTTTTAAAGATGGAAGTCCACCCTTTGCT
TGGTTTTAAGTATGTATGGAATGTTATGATAGGACATAGTAGTAGCGGTGGTCAGACATGGAAATGGTGG
GGAGACAAAAATATACATGTGAAATAAACTCAGTATTTTAATAAAGTA (SEQ ID NO: 95)

ATGGAAGATTTCGATGGACATGGACATGAGCCCCCTGAGGCCCCAGAACTATCTTTTCGGTTGTGAACATA
AGGCCGACAAAGATTATCACTTTAAGGTGGATAATGATGAAAATGAGCACCAAGTTATCTTTAAGAACGGT
CAGTTTAGGGGCTGGTGCAAAGGATGAGTTGCACATTGTTGAAGCAGAGGCAATGAATTACGAAGGCAGT
CCAATTAAGTAACACTGGCAACTTTGAAAATGTCTGTACAGCCAACGGTTTCCCTTGGGGGCTTTGAAA
TAACACCACCAGTGGTCTTAAGGTTGAAGTGTGGTTCAGGGCCAGTGCATATTAGTGGACAGCACTTAGT
AGCTGTGGAGGAAGATGCAGAGTCAGAAGATGAAGAGGAGGAGGATGTGAAACTCTTAAGTATATCTGGA
AAGCGGTCTGCCCCCTGGAGGTGGTAGCAAGGTTCCACAGAAAAAAGTAAAACTTGCTGCTGATGAAGATG
ATGACGATGATGATGAAGAGGATGATGATGAAGATGATGATGATGATGATTTTGATGATGAGGAAGCTGA
AGAAAAAGCGCCAGTGAAGAAATCTATACGAGATACTCCAGCCAAAAATGCACAAAAGTCAAATCAGAAT
GGAAAAGACTCAAAACCATCATCAACACCAAGATCAAAAGGACAAGAATCCTTCAAGAAACAGGAAAAAA
CTCCTAAAAACCCAAAAGGACCTAGTTCTGTAGAAGACATTAAAGCAAAAATGCAAGCAAGTATAGAAAA
AGGTGGTTCTCTTCCCAAAGTGGAAGCCAAATTCATCAATTATGTGAAGAATTGCTTCCGGATGACTGAC
CAAGAGGCTATTCAAGATCTCTGGCAGTGGAGGAAGTCTCTTTAA (SEQ ID NO: 96)

FIG. 12C

NPM protein (~35kD)

MEDSMDMDMSPLRPQNYLFGCELKADKDYHFKVDNDENEHQLSLRTVSLGAGAKDELHIVEAEAMNYEGS
PIKVTLATLKMSVQPTVSLGGFEITPPVVLRLKCGSGPVHISGQHLVAVEEDAEESEDEEEEDVKLLSISG
KRSAPGGGSKVPQKKVKLADEDDDDDDDEDDDEDDDDDDDFDDEEAEEKAPVKKSIRDTPAKNAQKSNQN
GKDSKPSSTPRSKGQESFKKQEKTPKTPKGPSSVEDIKAKMQASIEKGGSLPKVEAKFINYVKNCFRMTD
QEAIQDLWQWRKSL (SEQ ID NO: 97)

FIG. 12D

Homo sapiens nucleophosmin (nucleolar phosphoprotein B23, numatrin) (NPM1), transcript variant 1, mRNA

NCBI Reference Sequence: NM_002520.6

AGAAAGGAGTGGGGTTGAAAAGCGCTTGCGCAGGACGGCTACGGTACGGGGGTGGGAGG
GCTTCGGAGCACGCGCGCGGAGGCGGGACTTGGGAAGCGCTCGCGAGATCTTCAGGGTC
TATATATAAGCGCGGGGAGCCTGCGTCCTTTCCCTGGTGTGATTCCGTCCTGCGCGGTT
GTTCTCTGGAGCAGCGTTCTTTTATCTCCGTCCGCCTTCTCTCCTACCTAAGTGCGTG
CGCCACCCGATGGAAGATTGATGGACATGGACATGAGCCCCCTGAGGCCCCAGAACTA
TCTTTTCGgtaactgctggggggagctggagcgaggccgagcggggcctggtggcggtg
aggggtgggggtgaggggcggaatccggctgcaaccgggtctggtggagaccgccagac
cgacgggaggcctgagcggtgggaagagctgcctgagccctggacctggagggatggg
agccgcggggggagccgggtggcggtgaagggggaggagacgagcggcctgaagcgctctgg
ggcggtgagcggtccgagggccggggcctgaggtaaagtggaaaccggccgcctggaaagt
cggcctttgcccgggatacgtggcgagctccaggcggtgtttattttctaacggcggt
ggttcataggcttctgatgccaggccccgggttggttggtcctggaggctgggttcaccg
ggaagcatgggcctgcttgttgagcgggtagatgcacagcctggcgttattctggcct
cagagcctggagctgccggacgtagacttacctcgctgcttcggccagttactggggg
tggggaggcgcggcgcggcggtgaggaggccgcaaccgctgggagcacgtggttgccacg
tggttgggggagggagggggggtgtggggcgctacatccgggactcacccggcggttgcca
ctcgtgagccagggatgctgtggctctgttgaagtgggaaatgttaagaggctggtagc
actaggggttctggagggtctttgttcagtgttacagcatgtactgtgatggcagctctc
atTTTTTTgagcccttttctgtctcactgcggtaggtggagctagccccactttgcag
atgaggatactggtgcccagagataaaactcacttaacgaagggttgtaacttagtaaaggc
aataagagactaaatctaaaccaagacgcttgactccgtagctgctgcttcgtgaatat
taatgaggcttcggttttagggggacgtccctcgcatgccgccatttgtgcggcattgacc
aaactgcttttctaaatacaagaaacacgtagagtgttcggtttgagtcagatcaagac
ctaagactattcgcagaattttgggagcaccacccccctcgacttgcggtagtccctgcc
ccaaaaaggagagtcccacttggggttttgatctcagtaggattaattgcagaagccta
atctttgtactattcagactttgaagtaaattgcaaatcctaccgtgaattttatcttgt
tcattatccgtgctgagatttgttaattaaacgtctgttcacatttaagcatgtatatg
tattggtaacgggttagatgatagttaaatggaatcatttaaaatttttattctgacagt
gttttgagaattttgcccattcatggagtggagtagacatatgtattcaaaacatta
aggctttaattcctagttatTTTggatagctgagaattctggtaaatttatatcagtac
TTTTtgataactatcggaagaacccctccttgctcttaaacataatttctgaatatagtt
tgaggcgatgtcactttctgttgctttaagagtgagggttgatccctgtagcaaac
ttcggttcttttacttttgtcatttttcttttgtaatagtaagctattcatgtttgtaa
ccgttctcatttcgaaattgccatttggaatttaggtaggtaaatacagctgttttaa
tagtgggaaacttaacattggacattacatttggtttcagataattgcttcagggttgag
atttactaaaccattgagcttggttatgcagaataacagtacctatcagcgtagttaga
gatgaaatatggtatgatgagagaatggaattatagctttttatgggtgtctgtagagg
tttctgttcaagaaatctaggaattgatcattctgctttgccctctggtagctaaaata

Ex 1

FIG. 12D (Cont.)

gtgaaaaactagttcaaaagtttagacctgaccttttttggttaccacacttaagtttca
gtgttatttttctccttggttagagttgcttttttcttcatttacagGTTGTGAACATAAA
GGCCGACAAAGATTATCACTTTAAGGTGGATAATGATGAAAATGAGCACCAGTTATCTT
TAAGAACGgtacttaaactttcaaaaataaactacttaacccctacttgatttcagccctt
tagtttctattcatgttggttgagactttttttcctttgctgactgcttataaaatact
atctcttacacctgggtattgtgtgtacctcactgtctgtacatctttgaagaacttgg
catctataccactcatttgggtaaccatgtaaattccagtcacactgttcaacttgcctt
atctcttcagttgaattacaaagcccttgtaaaaggcatcgataatctttcatgtcta
ccagagactagccttctgctcaacttagaagttgctcaataaatatgttaggtggtcta
tattgtgtgtatgcataatgatgcatttcttaaattttctgaatatgagaaactgattt
gccagatcacaaaacctgaggacaacattgcacaaatttgttttccagggttaatagtg
agaggtttatacttggttttttaaaaaaatcacagtcattgtgcctcatttgaactgctgg
gtcaacagatggatcgcatatacaatggtgggtcccataagattattataccttattttt
aatatacctttcctgttttttttttttttcttggaggcagttctcgctctgtcaccaggct
ggagtgcagtggtgtaattctcggtcactgcaacctctgatttccgagttcaagtaatt
ccctgcctcacctcccgagtagctgggattacaggcatgtgccaccacacctggctaa
tttttggtgttttttagtagagacgaggttggttaccatgttggtccaggctgggtctcaaact
cctgacctcatgatctgcccgcctcagcctccctaagtgttgggattacaggcatgagc
caccaagcccggccctttcctatgcttagatgcacaaatactgtgtttcgggttgcttac
agtattcagtcacagtaacactgtacaggtttgtagcctaggtgtgtgttagtaggctata
ccatctaggtttgtataagcacattcttatgattgtacaaagatgaaattgtctaaaca
cacatttctcagaacctagccctgtggttaagtgaacgcatggctgcataaacatttag
tgggggggtgtaaaataggtggaactcaaaagttgaagtagtatttttttttgttcac
agGTCAGTTTAGGGCTGGTGCAAAGGATGAGTTGCACATTTGTTGAAGCAGAGGCAATG
AATTACGAAGGCAGTCCAATTAAAGTAACACTGGCAACTTTGAAAATGTCTGTACAGCC
AACGgtaagggcacttacatactttggatgttggttcaaggtttaattctgttttaagg
taggttttggtgtcatttagttgtgccaggagatagaaagtgggtcttcttctctgtc
actggagttcgatggtcaactcttgaacatgggggcttctgctgctacttttatcagag
gtggaaaaacaggttcaactggtttgttgatttggcttatgtgtttgcctgtaattgttta
ttgttcaattttcttcacatgttttagtgatgaaaaatttctccctcttagGTTTCCCTTG
GGGCTTTGAAATAACACCAACAGTGGTCTTAAGGTTGAAGTGTGGTTCAGGGCCAGTG
CATATTAGTGGACAGCACTTAGTAGgtatgttatttttatataataactacttagttt
gtcctctttagtgagttgcttggttccagtttggacttaaaagcatgggtatagtact
actgtctttttaataggttccaatgtgagtcctagaaattggagaggacaaataaatttt
tggggcgggggggagaggaaatcttgctgtcaccaggctagagtacagtggtcacgac
ttggctcactggaacctctccgggattcaagcgatcctcctgtctcagcctcccagta
gctggggccacagacgtgcacccaagtcagtttgcgtttcatagttatagttagagaa
gggttttcgtgatgttggtccaggccgatcttgaactcctggccttaacgtgatctgcac
gccctggcctctcaaaagtgttagtattacacgtgtcagccactgtgcctggcctaaaaa
ttattttttaataaagacagttctcattataacggctggagtgcagtgatgtgatcatag
cttgctatatcctcgaactgtactgggttcacctcagcctctggaaatagctagaacta
caggcacactccacgcctggctaattttttttgtatatgtgcagatgggggtctcagtat
gttggccagattggactcttggcctcaagtgggtccgccttggcctcccaaaagttagat
tacaggcatgagccacctcccaggcttcttgcatttaaaacctggcagtgaaacatta

Ex 2

Ex 3

Ex 4

FIG. 12D (Cont.)

ggcctcaaaaataactttttgttataaaaagttccttttcccatgtgctccttttttttttttttt
tttaaatagaaatagaagtcctcagtttttagagtatttactatcagtggtccttttttttttt
ctgacttccttgctgcttgagttttataatgtctaataaattgtatttttagCTGTGGAG
GAAGATGCAGAGTCAGAAGATGAAGAGGAGGAGGATGTTGAAACTCTTAAGTATATCTG
GAAAGCGGTCTGCCCTGGAGGTGGTAGCAAGGTTCCACAGgtagagatggcaatttt
attatagggttttgattatagccttttagtttggtgatagaacagctcctgttcattgagt
acgtatccttttctttttaaagAAAAAGTAAACTTGCTGCTGATGAAGATGATGACG
ATGATGATGAAGAGGATGATGATGAAGAGtaagtatgatttttagaaacttgatatact
tccggaatccttgacaaaaaaaggaatttgacatagttatatgcatgaggggttttataa
aagtcatttacaaaaagccatcctatgtaaatagtttataataaaaagggcaggtggtcat
ctgttgctcagtttaagttaaatgagctgagttgaaaggataattgggtctgtgagccttt
acaatgctgtgacttggtgactcctcagaagggttagactatagtggttggaagtttgat
tatgtccctttgttctgaagatttagtggtatggttataccatcaagcctgggtatggt
ttatgggtgagcagttaatgagattgggttgaaagaaaaatgattaaatagctctatat
tcattttacaagttgttactcaagggttggttattccctaaaaggattttgtcttatgggt
tttatgtagatatattatgacaaaaataagattctaaaagggatattaagattttcttg
ggatttaaaaatatggttggaacaatatattgatgactttatattaaactagatcaaact
attgttacaaacagtttaatacgcacactgggtataaagtactgtttataattgggtcttat
gtgtgccagtagcagtaattgcattgaatactttgatttggtctcagcttggtccttca
gttctgaggttggtccatatgcatttattgaaaacaaatataagaacatgcactttaaa
agagaacctgcatgaaagatcaaattgggagtttaggttttaagctggtggttcttcaa
aatctttgagcatgacgatgaaggcagaaaacaggaaaaaggccgaaagagccgaaagc
ttaaaaaattcaaagtatgaccaggcgcagtggttcacacccgtaatcctaacaaccagc
caagatgagaggatctcttgaggccagcccggtcaaaaatagcaagaccaatctcaaaa
gaaaaataaaaaattcaaagtacttgcatgtcaaactttataaatgaacgcaaacttaag
gtagtatttgactgttgctgtttttgttttcttttttttttttttttttttttgagatgg
agtctccagttgctcaggctggagtgcggtggcgccatctcggtcaccacaaactctg
cctcctgggttcaagcgattgtcctgcctcagcctcccggttagctgggactgcaggct
cggggcaccatgcctgctaatttttgatttagtagagacagggttcagtatgttgg
tccggctggtcctgaaactcctaacctcgtgatctacctgcctcggtctcccaaagtgt
gggattgcagatgtgaaccactgcgcccagccaaatttttgatttttaatatagacatgg
ggtttcaccacgttggttaggctattctcaaaactcctaacctcgggcgatctggtcacc
tcattcctcccaaagtgtggtgattacaggcatgagccaacctgcccggccactttttt
cttgagatggtgttttaccatgctactcagactggacttgaatctctgggcacaaagcta
tcctccctctcagactatgaagcaggtgggtttacaggagcataaccaagccagctt
gtttggttatcacttttaagaatatctctcgttagtaagaattgaaatacattccaaga
gaagaatgggaaacaggctaaaaaacacaaattagaaatagggatgggtatggttcggatt
ggttttagtctgattttgagttacctttgtacaagtttataaaaataagtggttaatatagca
ttcaccgaggctcggggacaagcaatcccttcagaaaggctttggagtaggacctgat
tgtagtattgacctgttggtgggttttggaagatttcttttttaaaaattgatataatt
aggccaggcatgggtgggtcaccgctgtaatcccagcactttgggaggctgaggtgggtg
gatcacttgaggtcaggagttcaagaccagcctggccaacatgggtgaaacccggtatct
actaaaaatacaaaaattagctgggtatcgtgggtgcctgcctattgggtcccagttactt
aggaggttgagtcaggagaattgcttgaaacctgggaggtggaggttgagcaagccaag

Ex 5

Ex 6

FIG. 12D (Cont.)

aatgcgccaccgtgttccatcctgggcaacagagggagactcccatctcagaaaaatgg
gtataaattcatgatgtaaccacaatgtaattttgtttgtctttaagttgggcattgat
aggaatgaaaagtgtagatatcaaggtccaaatcagtacctgggtttttttgtgggatt
tttttttttcggcaagtctcgctcttgtgccccaggtggtgtgcaatggtatgatct
ctgctcgctgcaacctctgccttccaggttcaagtgattctcctgcctcggcctccaga
gtagcttggattacaggcacttggcaccacgcctggctaagttttgtattttagtaga
gacgggggtttcaccatgttggccaagctggtctcgaaactccaaagtgctgggattacag
gcgtgagccaccatgcaccgcccgaagttttcatataagttgaagaaagtgtactaagg
tctgcatagtagtaaaggatgccttgagggaacaaatattaatagaaacttcagtgg
gagatggcaagggcccagcatagatagatggcaatgaaaatgcaaagaggtgcatgaag
gttcggttatagttacttagaaaatctaaccttttgaacacagatcaaaggggaatttgg
tcttttttgaggatggaatgggtatatggtgtgggctcagatgactcttgatttaagca
agaaaggctatgtaatgtgcatagtgtgatgtatactatacatagatgtatgtaatac
gttgatagtatgtggcccttaaatgtcttttttaatttttggggaatttcttaagtaaa
gctgaattttttttttttttttttgggagacagtctctgttgccccggctgaagtgcattg
ttggctaggctggagtgcagtgggtgcaatcatggctcactgtaacttctgcctcccggg
ttcaagtgattctcctgcctcagcctcccaagtagctgggattacaggcatgtgccagc
atgccaggctaattttttgtatttttaatagagatgggggtttcgccatgttggtcaggc
tagtcttgaactcctgacctcagggtgatctgcccaccccagcctcccagagtgtgaga
ttacagggtgtgagccaccacaccctgccagtaaagctgttttgatagtagttttgatag
ctattttgatagtagtttaatagacttgttttaacaaataagaaaaaatgtttaaaaaa
agcattctcatcttgttttctagcacaggggaggcacctgcagggattgggttctaattgc
cagaaacttgtactgacaaaatcactgttaaaaagccacttgaagggctatttgtgaca
gctttttaagttatgatactttttctaaatacagcaaataatttcttcgggtattggaaag
atagggtgtttcttacatgaagttgctgtattgggatttagagaccagtagcttcagttg
ttgaattaaacgtgaaccccttggtatttgctaataagagacttctgcctgacttgcct
ccagtgactcgatttgattactccctccattgtttacctaataacagttcacacctgt
aatcccagcactttgggtggctcgatgcgggcagatcacctgaggtcaggagttgaagac
cagcctggccaatatggtgaaaccccgctctctgctaatacagctgggtgtgatggg
ggcgcttgcaatcccagctgctggggaggctgaggcaggagaattgcttgaacccggga
ggcggagggtgcagtgagccgagatcgaccattgcattccagattgggcaacaagagt
aaaactcaaaaaaaaaatagtgtgcagttcagtggttttagtatgcatagttgtgtagcc
atcaccataatcaatttttagaacatttcatcacctcaatgagaaatcgtactctatagg
tattacccctcatgctcttcagctctagtcaaccacgaatgaactttgtctatagggtt
cctgtccctcatattttgcacgaatggacttctgtgactgttttcttagcacagtgtt
tcaagtttcatccatgttatagcattatcagtactccattttatttatatgggttgaatac
attgtatgggtatgtttgggtattcatcagttgggtgagcatttgagttgcttctacttg
ttgactgctatgaacgctttagacatgtcattttacttggatgtacacctaaagcaga
gtggctcaccattgtattcccactagcagtgatggggcttctgatttccactccaacc
tactgatttccactacaccctcacttggcattatctgactctaatacctgggtggatgaag
tgctgattgtgggttttgattgctctccctgtggactaataatgagcatcttttcatgt
attcattggccatatatctttggagaaatgtttttgcccacttaaatccatttacttct
ctttgtctttactaagttgtaagaattcttttttcttttttctttttgatatgatttt
gtatttttagtagagaaggggttttgccatgttggccaggctgttcttaaacctcctgtc

FIG. 12D (Cont.)

ttcagtgatccacctccctctgctcccaactgctgagattacagctgtgagccactg
tgccctggcaagaattcttttttttttttttttgacaagtctctgttgccaggcttgag
tgcagtgatgtgatcttggctcaactgcaacctcctcctcccttggttcaagcaattcttc
tgccctctcccaagtagttgggattacaggcacccccccaccacgcccggctaattttc
tgtatttttaaatagagaccgggtttcgccatggtggccaggctgggtcttgaactcctga
ggtcaggcccgcctcgccctcccacaagtgctgggattacaggcgtgagccactgcact
gggcccagaattctttatgtattctgcaaacaagtcccttatcagacacaagatttac
agattatcttccaccattccgtgaatttgttacttaactttcatgatgggtggcctttgaa
caagtttttaatttttatgacgtccgatttcttttgttgcttgtgcctttgggtgtcaat
cctaagaaatcatttgctaaattcaagttgtgaaaatttgcccccttaattctgagttt
tgtcctttatatattagggctttgttccattttgagtaaacttatgtgtatgggtgtgaat
ggaggggtctaaatatagactttggatgtagattagcagttgccccacatgatttgtt
caaaaactatttttccccattgaatgatcttggcaacctttaaagaggtcacttatgg
ctgggtgtgggtggctcacacctgtaatcccggcactttgagaggctgaggcaggcggt
catgaccaggagattgagactatcctggccaaatttggtgaaacctgtctctactaaaa
atacaaaattttagctgggcatggtagcgtgtgcctgtagtcaccagctactcaggaggct
gaggctggagaattcgcttgaacctgggaggcagagggttgcagtgatctgagattgcac
cactgcactccagcctggcgacagagcgagactccgcctcaaaaaacaaaaaatcact
gtttctggactgttctgttgatgtgtctatcccttaacatgtagaacagttacaacactg
ttcctctgtggtttttatttatttatttttttgaaatgagttcactgtcaccaggct
gtagcacagtggcgtgatcttggctcactgcaacctccgcctcctgggttcctcctgctt
cagcttcctgagtagctgggattacaggcgcccgccaccatgcccggctaatttttgta
tttttagtagagactgggtttccccatttagccaggctgggtctcaaacctcctgaacctcag
gtgatctggccgtctcgccctcctaaaaagtggcgggattacaagcgtgagccacgggtg
ccttggccccctttgtaggtcttatagctgattttttgaaatcaagtgtgagttttctag
ctttcttcccttccagattgcatttggcctctttgggtcccttaagtgtcttttttggt
tttggtttttgagatagggtcttgctctgtcaccaggctggagtgagtgccagaatt
acacagttacagttcactgtctcaagcagatcctcctgcttagcttttcaaatagcta
ggactacaggcgccacaccacgcctgttaattaaaaattttttttagtagagggtt
ctcattatgttgcttaggtgtgtcttgaactaaaaatcctccacctcagcctccca
aagtgttaggattataggtgtgagacactgtgccagttcttgggtttgtttataaattt
taggatctgtttctacagagaaaacagcttggggagtcttctaacgattgtatttaacc
tgtatatccggggagtgtttccatcttggcaatatctgaacatggaatgtttctattta
ggctctttaatttttttttaacacaaattcactctgagtgtaggtttaatagaacta
tacattacttaaatatacatacattatacagtggtatagtggtatttatgcaaaacta
ctcattgagcaacttcatttcccagccccctagcgatcacaattataccttctatttctt
tgagtttgactacttttagatacttaatgaatagaatcatacagtggtttgttcttttgt
ggctggcttatttcactttgttcaagggttcagtcatgtagcatgcaagaggattttcat
tgtgtgtacatatacattttcatttagtcatttaacacttgggttgctttcacctcttg
gctgtttgtgaacaatgcttcgataaaacatgggtgcacaaataatcacttcaaggctcctg
ctttcaattcttgtgtctactccaaattttgaaagtgttaatgtcttgacatttcat
ttgtagTGATGATGATGATGATTTTGATGATGAGGAAGCTGAAGAAAAAGCGCCAGTGA
AGAAAgtagtagatacaatgctacaagggttggttaactaaacaatagaaatgggtgattt
tttagtgctatttgccttggtttgttagttaagggaagctgggtgtgggagatcatctcata

[Ex 7]

FIG. 12D (Cont.)

ctgaaaattagtcctgaggaggattacagaaaacttaagagtgagggaatgggtctgtttt
 ctttatccatgtggccctccaccagtttggttagtcttggtgtaaaccttttggtccaagtg
 gttgctgcttttctctcttttttttttttttttttgagacggaggtctcaatgtcacccg
 ggctggagtacagtggcgcgatctcagctcactgcaacttctgcctcccggttgaagc
 gattctcctgcctcagcctcctgagtagctggaattacagactcatgcccacacacca
 gctaattttttgcatttttagtagaccaccatgtcggccaggctgggtctcgaaactcctga
 cctcgtgatccacctgcctcggcctcccaaagtcctgggattacaggcatgagccacca
 cgccaagcctgtggttgctgcttgctttacatggcttggacagctttgtttgcactgtt
 gttggggtcagggacagtgaattaagataaattttctaattgcagTCTATACGAGATACTC
 CAGCCAAAAATGCACAAAAGTCAAAATCAGAATGGAAAAGACTCAAAACCATCATCAACA
 CCAAGATCAAAAgttaagtggctacatttacacgtgggtctcattgatctagttggggaa
 aaagattctactgttggaagaatctagtgtgtctgaaatttgataggcctttatagaacc
 cctgtaattgctgttttaaaagttaaaatcagcttgctgcagccaggctcagtggctcac
 tctgtaatccagcactttgggagggaggccaagggtgggtgggatcacctgagggtcag
 gagtttgagactagcccggccaaacatcgtgaaacctgtctttactaaaaacacaaaaa
 ttageccaggcatggtggtatgtgcctgtaattccacctactcaggaggtggagacagga
 gaattgcttgaacctgggaggtggagtgcagtgcagattgcaccactgcactccagcttg
 ggcaaaagagcgagactccgcctccaaaaaaaataatcagcttggtgtgtgtgttaggt
 acacacacacacacaaacataactaaatgttaaggtgggtggggcggggggagccaatagac
 ttttttgaaaagagatggaccctcactttgtcactcagctctgggttagagttgcgtgatc
 tcggcttactgcaaccttcaacttccgggttcaagcgattctcctgcctcagcttccca
 agtagctgggactacatgcgcgtgccaccaggcccagcgtatttttggtatttttgagta
 gagatgggggtttcactctatatgttggtggccaggctgggtctcaaaccctgactgcagggtg
 atccgccctcctgggcctctcaaagtgtgtgagccaccacgcctggcctgaattttttg
 tgattgtgaaagtcaatagttgttttctgttaaggaatctttgttgaaagggtatgtctgca
 tagagtagaagttctcaaccttggtgcatgttagcttgaaagtagaacatttgggccta
 gacaggggtgtattgatctcattcctgacctaccacttccctgcagaaaatggaattt
 aaaggaaaaaaagtttaagggaacccagtaacttggtgtctctacttatggagattgtcat
 aaaaatggtctgggggtgggcgcaggtcatcagactacagtgaactttcttgctttattct
 tagggattttctgatttcttttttttttttttttaacctgatgaaatagttttatttac
 aatttaaacgaataaatttcatattctacaacacataataatctattgactaaatgtaac
 taatgatgaacctccatacaaaacacccactaaaaagtgaggcaaaaggacataaacaga
 cacttttcaaaagacatacaagtggccaaaaagcatatgaaaaacagctcaatatcact
 gattagagaaatgcaaatoaaaaacacaaatgagataccatctcacaccagtcagaatag
 ctattattaaaaaactcaaaaaacaaacagctgctggcgaagggtgcagaaaaaagggaatg
 cttacacactcttggtgggagtgtaaattagttcaaccattgtggaaagcagagtggtg
 attcctcaaagagctaaaaacagaactaccatttgaccagcaatcccattacattccc
 aaagggaatataaatcattctaccataaagatacatgcacacaaatattcactgcagcac
 tattcacaatagcaaaagacatggaatcaacctaatacctgatattggtctggctttgtgt
 ccccacccacatctcatctcaaaattgtaatccccacgtgttggaagaggggctgggtgg
 gaggttaattggatcatggggttagatttccctcttgctgttctgttgatagtgaaattctc
 atgagatctggttggttgaaagtgtgtagcacctccccctttgtgctctctctctctg
 ctgccatgtaagatgtgccttgcttccctttgecttctgccatgattgtaaatttct
 gtggcctcctagccatgcttctatacagcctacagcactgtaagtcaattaaacctct

Ex 8

FIG. 12D (Cont.)

gccctgggttcgtggtatgaattttttcaaaaatttcttataaaaacatttataaatcgtgtc
 tgtggtgatttttgcataatgcaaaattaaatatgccttattttccattatgcaagggaac
 gtagtgcactgggttgcaagataaacattctgaccttccatgttaaaatagatcagtga
 accctttgcctattctggttgtaagatatgctagagaaccaacagaggggcgtatgagac
 ttcattaaaaattacaaacagctgggaaaagttagatgctggctgttgcttgggtattatagt
 taatattcatgattgaccctagtcagaagtgattttcagcagattgagacatttctctt
 tggccttttacaccaagttgtcaccagttaatacatgtagtaatagctagatttcgtgaa
 tgactaataggcttaaatctagttttccttttgtcttaaataatggaaatgtgggaagag
 ggttaggtctctcatctgcttgactgggggatttgagtaggatatggaaataaatgggtt
 aaatgccttggaagcacatttgagagaagctttaattctatgaataacttttagact
 ccaactattgagagcaaatgtgggtggtgaaaaattgccaactcatttcccatgttt
 ttcttgtaatctatagatttttcgtggcttcagtagattccttgtagctgataagtcc
 actggaaaaagaattttagaactggaaacatatctcttttagtccactgtaaaccctt
 agccttccctggttatcactgatttttgtcccttggatacaatgctaaaatgacatctt
 tagatgccccctccccctccattttaatatgggtcctatctccttggtttaattgcaggcgc
 attgaacagtcctggggcactacatgtaaatgaagcccaaagatggggagaaaggaaaag
 gagagacaaaatatagtcatactgagtgatcaacaatccagactgaagtcttctatt
 ttaatctcaatccccctttctgatttgccaccatgcctcttcaggctggaaacaatct
 cttggttccctaaagcactttctctgactgctgtgattcagtgaaccttgcctttgc
 tttctattacttgtgcatttgcctcacctgacaatgttttaaategcctttgtatctcc
 ttagctgctcaataaataatttgatgcatacaattagaatgtatgtgacaataattttg
 aaatggaaatgttgaggtttattttacttaggttatggaggcagttaagtttcttaatg
 ctatacctgattctgccaagtccttggacatttgaaaaccttttcaatcttttaaaa
 atgcataaaaagatacttagagggaaagggtattaattaaatttagtaaaaacaaatatat
 tcaatgcttttaaatccaaagtagtgataaaatacaaacatttcaagatatgtgcaatag
 taatgttttgaaattttggttacaaagtcacagggtcttgccaatactgttgtagttct
 tgcctatatccgtaatttgaaaggaaatggtgagagtgattagagaagtgtaattactgt
 aattttttccctattgtagtctcttgctgtatccataatttgaaaggaaatggtaga
 gtgattagtgaatgttaattactgttaatttttccccattcaactttatatatctttaa
 ctgatgaaccagatcattgttgttctgaaccagtttgtggtcagcaagtggtttgtggg
 ttttgtttgtttgtttttaagaaacagtttgggtcacttgacatggttctccaaaggga
 tgttatgggttgatattgtgttctgggtgataaccgacttgttagataatttagataagc
 aaccgagttgcatgtttgtttgttcgaatctcaagtgtagcttatattttatgttcta
 gagagggtgtcaggggaagatttgaccccttggcaaatctgtttgaatagagatactagc
 catgctgccaatagggctttctggccctgaaaaagatacggagtattcttggaaggt
 gaagggaaaaaagaataaactgatccatgtagtagcatgcagattattgaggaattttct
 aaaggtatctctctcggtgtattttctctacttaacctgtataaatgcttttgtcttaata
 gGGTGGTTCTCTTCCCAAAGTGGAAGCCAAATTCATCAATTATGTGAAGAATTGCTTCC
 GGATGACTGACCAAGAGgtaactggattttctggggacatgattaaatccaagtttttt
 tgtgattttattatgattctgcctttaccctttttaagtgttggctctttttaaaaagtt
 cctaaccacagataaatatttcatttaattgaataacctgaaaactcatgtatttaatgaa
 atgcatgaagaatacagttggcagttctgagatgataaactttagtatattttccagat
 ttccctatagttgggtaaattatttcttgggtatttgaattcagttttacttttttgggtact
 atgttgggttaaggaaattgctgtggttttcccttgctgttccatttgactgcttggggca

[Ex 10]

FIG. 12D (Cont.)

gtctgggtcttggccttttttgggagttgggtagagatcttttttttttttttcttagatct
 tatgggaaatacattatgccccaaatcatcctaacaatctggaagcaagacactaa
 attgtggctactaccagatacataagccttgagttagctgttagatttttacagttagtac
 atgtgaaagtcttagtcaacataaaagattatgcatttattttttgccttattttacagtt
 gactcataaagtgttaaggacaatgtaaaagtgggttaaataagggatgttaggtatcgtgtg
 gagaacgggatgctgcatttaattggtctttaggtttcctgtgggtcagcttttttcaaaaa
 attaggaaaaatttggcaattagctggttttttttctgccatagtattttgcacttcaat
 tatttaattgcacttcaatttttagttttcgggttcgtaaaaccctgataatggttaatt
 tttgagtttagctttgtgaaagttgtagtgatgggctataagatgatcaggttcagttct
 ttgggtgctgttctgttttttctctacatttgaatcctaaagtttaaagtgtgagtttt
 gggagtttatttggaataactaaacattgattaatcataaatttattttgaaactccgcgt
 tatgcaactgtttattttaatattgaaaatagtggaagtcagacataaataaccagcat
 gtactaacagctaagcagcatagagatgcatcttcaactactgccaggaaagtgtat
 gataaatttaattgattaaagttgcagtagtcttctgtactttataacaaaggccttgtagc
 tacaaggtcttgtagctattaaattactgaaccagcacttctacaagaagtattttctt
 tttataaaaaagtatttttggccaggttggtcttgaactcctggcctcaagcagtcctcc
 catcttttctctccaaagtgtctaggattacaggcagcgagccaccttgcccagcctacca
 gaagtatttttttgcctctagaattgaaagtgttctcaagtgtcagaatgaaatttctgg
 aataaatagtaaatgttttaataaacttgaacttgaattgctgttacagagattatcaaaa
 ctaaagtcttctgtttttgacatagagatgcatgatgcaagcaacgacaatgtatat
 tgtttgaacctgatgataagatctctgggtgggctggtagttcatgcctataatccca
 gcactttgggaggcaacatggcaaaactctgtctctactaaaaataactaaatttagctg
 ggcatgggtggcatgtacctgtaatcccacctactcaagaggcggaggcaggaggatcgc
 ttgaacctgggaggcagaggttgcagtgagcgccaactgcactccagttctgggtggcaga
 gtgagagttctatctcagaaagaaagatttctatgatatttaatttggcattcgggtatt
 cagttacctgtacctgagaacccattggctgtgaaacagtgacagctgagagaattcct
 gagtcatctcatcttagttcttgggtgaacttctggacttttcttcagaaccaccttgc
 catgttggccaggctggtcttgaactcctgacctctcaggtgatccaacaccttggcct
 cttaaagtgtctgggattacaggcatgagccacctgacctggccagctgttttttttgggt
 ggtttgttttttgggttaccatctgtagtgtgatcttggctcactgcaacctctg
 cctcttgggtcaggcagtcctcccacctcagcctcctgagtagctgggcctcctgtag
 ttgcacaccaaccaagcctggctaattttttgcattttttagtagacagggtttcaccatgt
 tgcccaggctggtctcaaattcctgagctgaagtgtatctgcccgcctcagttctccaaa
 gtgtagggtattacaggcgtgagccacctgacctagcctcagcatatagttttttctaaa
 tgtacacatgcccaggcacaacatgcacaggcaattcagaataagtttctgggtgttatg
 taacttttatttgccaaatctggccaactctaaagctgatctcgggagatgaagttggaa
 gtaacattggccatatgggtctctgttcttctgttgatttcttaagtaataatgct
 aaactattaaataattatttagtatattgttcacatttttatgactgattaaagtgtttg
 gaattaaattacatctgagtataaattttcttggagtcatatctttatctagagttaac
 tctctgggtggtagaatgaaaaatagatgttgaactatgcaaagagacatttaatttatt
 gatgtctatgaagtgttgggttctttaaaccacatttcttttttttttttttccagGCTA
 TTCAAGATCTCTGGCAGTGGAGGAAGTCTCTTTAAGAAAATAGTTTAAACAATTTGTTA
 AAAAAATTTTCGCTCTTATTTCAATTTCTGTAAACAGTTGATATCTGGCTGTCCTTTTATA
 ATGCAGAGTGAGAACTTTCCCTACCGTGTTTGATAAATGTTGTCCAGGTTCTATTGCCA

[Ex 11]

FIG. 12D (Cont.)

AGAATGTGTTGTCCAAAATGCCTGTTTAGTTTTTAAAGATGGAACTCCACCCTTTGCTT
GGTTTTAAGTATGTATGGAATGTTATGATAGGACATAGTAGTAGCGGTGGTCAGACATG
GAAATGGTGGGGAGACAAAAATATACATGTGAAATAAACTCAGTATTTTAATAAAGTA
(SEQ ID NO: 98)

FIG. 12E

Intron involved in NPM1 translocation

gtgagtaaagttatcttaaaaaaactttgtctccccctcaaattgcacgtgtctgggt
tgcatagacttgaatgtttcttggttttttgggtttaataacttgcctgg
ttcgtgggtatgaattttttcaaaaatttcttataaaacatttataatcgtgtctgtgg
gatttttgcataatgcaaaaattaaatatgccttattttccattatgcaaggaaacgtagtg
cactgggttgcaagataacattctgaccttccatgttaaaatagatcagtgaaccctt
tgcctattctgggttgtaagatatgctagagaaaccaacagagggcgatgagacttcatt
aaaattacaaacagctggaaaagtagatgctggctgttgcttgggtattatagttaatat
tcatgattgaccctagtcagaagtgttttcagcagattgagacatttctctttgacct
ttacaccaagttgtcaccagtttaatacatgtagtaatactagatttctgtgaatgacta
ataggcttaaatctagtttcttttgtcttaataatggaaatgtgggaagaggggttag
gtctctcatctgcttgactgggggatttgagtaggatatggaataaatgggttaaatgc
cttgggaagcacatttgagagaagctttaaattctatgaataatacttttagactccaact
attgagagcaaagtgtgggggtgggtgaaaaattgccaaactcatttccccatgtttttcttg
taatctatagatttttctgtggcttcagtagattccttgtacagtgataagtccactgga
aaaagaattttagaactggaacataatttcttttagtcacactgtaaaccctttagcctt
cctgggttatcactgatttttgtcccttggatacaatgctaaaatgacatcttttagatg
cccctccccctccatttttaatatgggtcctatctccttgggtttaattgcaggcgcatgaa
cagtcctgggcactacatgtaaattaagcccaaagatggggagaaaaggaaaaggagaga
caaatatagtcatactgagtgatcaacaatccagactgaagtcttctatttttaatc
tcaatcccccttttctgatttgccacccatgcctcttcaggctggaaacaatctcttgggt
tccctaaagcacttttcttctgactgctgtgattcagtgaaccttgccctttgctttcta
ttacttgtgcattttgcctcacctgacaatgttttaaatcgccctttgtatctccttagct
gctcaataaatatttgaatgcatcaattaagaatgtatgtgacaataattttgaaatgg
aaattgtgaggtttattttactaggttatttgaggagcagtttaagtttcttaatgctatac
ctgattctgccaaagtccttggacatttgaaaaccttttcaatcttttaaaaatgcat
aaaagatacttagagggaaaggtattaattaaatttagtaaaaacaaatatattcaatg
ctttaaattccaaagtagtgataaatacaaacatttcaagatatttgcaatagtaatgt
tttgaaattttgggttacaaagtcacaggtcttgccaatactgttgtagtttcttgacct
tatccgtaatttgaaggaaatgggtgagagtgattagagaagtgttaattactgtaatttt
ttccccctattgtagtttcttgacctgtatccataatttgaaggaaatggtaagagtgatt
agtgaatgtaattactgtaattttttccccattcaactttatatatctttaactgatg
accagatcattgttggtctgaaccagtttgtgggtcagcaagtgttttgtgggggtttgt
ttgtttgttttttaagaacagtttgggtcacttgacatgggttctccaaagggtatgttat
gggttgatattgggtctgggtgataaccgacttggttagataatttagataagcaaccga
gttgccatgtttgtttgtcgaatctcaagtgtagcttataattttatgttcctagagagg
ttgtcagggaagatttgacccttggcaaactctgtttgaatagagatactagccatgct
gccaatagggtcttctggccctgaaaaagatacggagatttcttggaaggttgaaggg
aaaaagaataaactgatccatgtagtagcatgcagattattgaggaattttctaaaggt
atctctctcggtgtatttctctacttacctgtaataatgcttttgtcttaatag (SEQ
ID NO: 99)

FIG. 13A

NPM1/TYK2 cDNA 2711 bp

AGAAAGGAGTGGGGTTGAAAAGCGCTTGCGCAGGACGGCTACGGTACGGGGGTGGGAGGGCTTCGGAGCA
CGCGCGCGGAGGCGGGACTTGGAAGCGCTCGCGAGATCTTCAGGGTCTATATATAAGCGCGGGGAGCCT
GCGTCCTTTCCCTGGTGTGATTCCGTCCTGCGCGGTTGTTCTCTGGAGCAGCGTTCCTTTATCTCCGTCC
GCCTTCTCTCCTACCTAAGTGCGTGCCGCCACCCGATGGAAGATTCGATGGACATGGACATGAGCCCCCT
GAGGCCCCAGA ACTATCTTTTCGGTTGTGA ACTAAAGGCCGACAAAGATTATCACTTTAAGGTGGATAAT
GATGAAAATGAGCACCAGTTATCTTTAAGAACGGTCAGTTTAGGGGCTGGTGCAAAGGATGAGTTGCACA
TTGTTGAAGCAGAGGCAATGAATTACGAAGGCAGTCCAATTAAAGTAACACTGGCAACTTTGAAAATGTC
TGTACAGCCAACGGTTTCCCTTGGGGGCTTTGAAATAACACCACCAGTGGTCTTAAGGTTGAAGTGTGGT
TCAGGGCCAGTG CATATTAGTGACAGCACTTAGTAGCTGTGGAGGAAGATGCAGAGTCAGAAGATGAAG
AGGAGGAGGATGTGAAACTCTTAAGTATATCTGGAAAGCGGTCTGCCCCCTGGAGGTGGTAGCAAGGTTCC
ACAGAAAAAAGTAAACTTGCTGCTGATGAAGATGATGACGATGATGATGAAGAGGATGATGATGAAGAT
GATGATGATGATGATTTTGATGATGAGGAAGCTGAAGAAAAAGCGCCAGTGAAGAAATCTATACGAGATA
CTCCAGCCAAAAATGCACAAAAGTCAAATCAGAATGGAAGAGACTCAAACCATCATCAACACCAAGATC
AAAAGGACAAGAATCCTTCAAGAAACAGGAAAAAACTCCTAAACACCAAAAGGACCTAGTTCTGTAGAA
GACATTAAAGCAAAAATGCAAGCAAGTATAGAAAAAGAGAACAAGAACCTGGTTCATGGTAATGTGTGTG
GCCGGAACATCCTGCTGGCCCCGGCTGGGGTTGGCAGAGGGCACCAGCCCCCTTCATCAAGCTGAGTGATCC
TGGCGTGGGCCTGGGCGCCCTCTCCAGGGAGGAGCGGGTGGAGAGGATCCCCCTGGCTGGCCCCCGAATGC
CTACCAGGTGGGGCCAACAGCCTAAGCACCGCCATGGACAAGTGGGGGTTTGGCGCCACCCCTCCTGGAGA
TCTGCTTTGACGGAGAGGCCCCCTCTGCAGAGCCG CAGTCCCTCCGAGAAGGAGCATTTCTACCAGAGGCA
GCACCGGCTGCCCCGAGCCCTCCTGCCCACAGCTGGCCACACTCACCAGCCAGTGTCTGACCTATGAGCCA
ACCCAGAGGCCATCATTCCGCACCATCCTGCGTGACCTCACCCGGCTGCAGCCCCACAATCTTGCTGACG
TCTTGACTGTGAACCCGGACTCACCGGCGTCGGACCCTACGGTTTTCCACAAGCGCTATTTGAAAAGAT
CCGAGATCTGGGCGAGGGTCACTTCGGCAAGGT CAGCTTGTACTGCTACGATCCGACCAACGACGGCACT
GGCGAGATGGTGGCGGTGAAAGCCCTCAAGGCAGACTGCGGCCCCCAGCACCGCTCGGGCTGGAAGCAGG
AGATTGACATTCTGCGCACGCTCTACCACGAGCACATCATCAAGTACAAGGGCTGCTGCGAGGACCAAGG
CGAGAAGTCGCTGCAGCTGGTCATGGAGTACGTGCCCCCTGGGCAGCCTCCGAGACTACCTGCCCCGGCAC
AGCATCGGGCTGGCCCAGCTGCTGCTCTTCGCCCAGCAGATCTGCGAGGGCATGGCCTATCTGCACGCGC
AGCACTACATCCACCGAGACCTAGCCGCGCGCAACGTGCTGCTGGACAACGACAGGCTGGTCAAGATCGG
GGACTTTGGCCTAGCCAAGGCCGTGCCGAAGGCCACGAGTACTACCGCGTGCGCGAGGATGGGGACAGC
CCCGTGTTCTGGTATGCCCCAGAGTGCTGAAGGAGTATAAGTTCTACTATGCGTCAGATGTCTGGTCCT
TCGGGGTGACCCTGTATGAGCTGCTGACGCACTGTGACTCCAGCCAGAGCCCCCCCCACGAAATTCCTTGA
GCTCATAGGCATTGCTCAGGGTCAGATGACAGTTCTGAGACTCACTGAGTTGCTGGAACGAGGGGAGAGG
CTGCCACGGCCCCGACAAATGTCCCTGTGAGGTCTATCATCTCATGAAGAACTGCTGGGAGACAGAGGCGT
CCTTTTCGCCCAACCTTCGAGAACCTCATACCCATTCTGAAGACAGTCCATGAGAAGTACCAAGGCCAGGC
CCCTTCAGTGTTACGCGTGTGCTGAGGCACAATGGCAGCCCTGCCCTGGGAGGACTGGACCAGGCAGTGGC
TGCAGAGGGAGCCTCCTGCTCCCTGCTCCAGGATGAAACCAAGAGGGGGATGTCAGCCTCACCCACACCG
TGTGCCTTACTCCTGTCTAGAGACCCACCTCTGTGA ACTTATTTTTCTTTCTTGCCGTGAGCCTAACC
ATGATCTTGAGGGACCCAACATTTGTAGGGGCACTAATCCAGCCCTTAAATCCCCCAGCTTCCAACTTG
AGGCCCACCATCTCCACCATCTGGTAATAAACTCATGTTTTCTCTGCTGGA (SEQ ID NO: 100)

FIG. 13C

NPM-TYK2 protein (~81 kD)

MEDSMDMDMSPLRPQNYLFGCELKADKDYHFKVDNDENEHQLSLRTVSLGAGAKDELHIVEAEAMNYEGSPIKVTLA
TLKMSVQPTVSLGGFEITPPVVLRLKCGSGPVHISGQHLVAVEEDAEESEDEEEEDVKLLSISGKRSAPGGGSKVPQK
KVKLAADEDDEDDDEDDDDDDDFDDEEAEEKAPVKKSIRDTPAKNAQKSNQNGKDSKPSSTPRSKGQESFKKQ
EKTPKTPKGPSSVEDIKAKMQASIEKENKNLVHGNVCGRNILLARLGLAEGTSPFIKLSDPGVGLGALSREERVERI
PWLAPCECLPGGANSLSTAMDKWGFATLLEICFDGEAPLQSRSPSEKEHFYQRQHRLPEPSCPQLATLTSQCLTYEP
TQRPSFRTILRDLTRLQPHNLADVLTVNPDSPASDPTVFKRYLKKIRDLDGEGHFGKVSPLYCYDPTNDGTGEMVAVK
ALKADCGPQHRSGWKQEIDILRTLYHEHIKYKGCCEDQGEKSLQVMYEVPLGSLRDYLPRH SIGLAQLLLFAQQI
CEGMAYLHAQHYIHRDLAARNVLLDNDRLVKIGDFGLAKAVPEGHEYRVREDGDS PVFWYAPCECLKEYKFYYASDV
WSFGVTLYELLTHCDSSQSPPTKFLELIGIAQGQMTVLRLTELLERGERLPRPDKCPCEVYHLMKNCWETEASFRPT
FENLIPILKTVHEKYQGQAPSVFVVC (SEQ ID NO: 102)

FIG. 13D

NPM-TYK2 introns involved in translocation (3332 bp)

Gtgagtaaagttatcttaaaaaaacttttgtctccccctcaaattgcacgtgtctgggt
tgcatagacttgaatgtttcttggttttttggtttggttttaataatacttgccctgg
ttcgtgggtatgaattttttcaaaaatttcttataaaacatttataatcgtgtctgtgg
gatttttgcataatgcaaaaattaaatatgccttattttccattatgcaaggaacgtagtg
cactgggtgcaagataacattctgaccttccatgttaaaatagatcagtgaaaaccctt
tgcttattctgggttgtaagatatgctagagaaaccaacagagggcgtagagacttcatt
aaaattacaaacagctggaaaagtagatgctggctgttgcttggtattata

gttaatatcatgattgaccctagtcagaagtgattttcagcagattgagacatttctc
tttgccctttacaccaagttgtcaccagtttaatacatgtagtaatagctagatttcgtg
aatgactaataggcttaaatctagtttcccttttgtcttaataatggaaatgtgggaag
agggttaggtctctcatctgcttgactgggggatttgagtaggatatggaataaatggt
ttaaatgccttggaagcacatttgagagaagctttaattctatgaataatacttttaga
ctccaactattgagagcaaatgtgggggtggtgaaaaattgccaactcatttccccatgt
ttttcttgtaatctatagatttttctgtggcttcagtagattccttgtagagataagt
ccactggaaaaagaatttttagaactggaacataatttcttttagtcacactgtaaacct
ttagccttccctgggttatcactgattttttgtcccttggaatacaatgctaaaatgacatct
tttagatgccccctccccctccattttaatatggctcctatctccttggtttaattgcaggc
gcattgaacagtcctgggcactacatgtaaattaagcccaaagatggggagaaaggaaa
aggagagacaaatatagtcatactgagtgatcaacaatccagactgaagtcttcta
ttttaatctcaatcccccttttctgatttgccaccatgcctcttcaggctggaaacaat
ctcttggttccctaaagcactttctctgactgctgtgattcagtgaaccttgcccttt
gctttctattacttggtgcatttgccctcacctgacaatgttttaaatcgcccttgtagct
ccttagctgctcaataaatatttgatgcatcaattaagaatgtatgtgacaataattt
tgaaatggaaattgtgaggtttattttactaggttatttgaggcagtttaagtttcttaa
tgctataacctgattctgccaagtccttggaacatttgaaaaccttttcaatcttttaa
aatgcataaaagatacttagaggggaaaggtattaattaaatttagtaaaaacaaatat
attcaatgcttttaaattccaaagtagtgataaaatacaaacatttcaagatatttgcaat
agtaatgttttgaaattttgggtacaaagtcacaggtcttgccaatactgttgtagttt
cttgccctatatccgtaatttgaaaggaaatgggtgagagtgattagagaagtgttaattact
gtaattttttccccctattgtagtttcttgccctgtatccataatttgaaaggaaatggtaa
gagtgatttagtgaaatgtaattactgtaattttttccccattcaactttatatatcttt
aactgatgaccagatcattgttggtctgaaccagtttggtggtcagcaagtgttttggtg
gggttttggttggttggttttaaaagaacagtttggtgacttgacatgggttctccaaagg
gatgttatgggttgtagtttggttctgggtgataaccgacttggttagataatttagataa
gcaaccgagttgccaatgtttgtttgtcgaatctcaagtgtagcttatattttatgttcc
tagagaggttgctcaggggaagatttgaccctttggcaaatctgtttgaatagagatacta
gccatgctgccaatagggtcttctggccctgaaaaagatacggagatttcttggaag
ttgaagggaagaaagaataaactgatccatgtagtagcatgcagattattgaggaatttt
ctaaaggtagtctctctcggtgtatttctctacttacctgtaataatgcttttgcttaa

FIG. 13D (Cont.)

taggtgtgtggcctgtgtgtggggcctgggtcgggtcagggagggccaggagcccaggag
ttcgagaccagcctgggcaacagggcgagaccccatcttttttgtttgttttgtttttag
gtggagtttcgccctgtcaccaggttgagtgcaatggcatgatctcggctcactgca
accttcgcctcccggttcaaatagattctccgcctcagcctcccaagtagctgggatt
ataggctcctgccatcacgcccagctaatttttgatatttttagtagagatggggtttca
ccatgttggtcaggctgggtctcgaacttctgacctcgtgatccatccgcctcagcctcc
caaagtgtgtgggattccaggcgtgagccaccacgcccagctggccccatcttttaaaaa
taaacaaatagccaggtgtggtgggtcatacctgtaatcccagcagtttgggaggccaa
ggcgggtggatcacctgaggtcaggagtttgagaccagcctggccagtatggcaaaacc
ttgcctctactaaaaatacaaaaaaaaaattagccaatgtggtaacttgcacctgtaggcc
gaggtactcaggaagttgaggtgggaggatcacctgagctcaggaatttgaggccgtag
taagctatgatcacaccactgcaccccagcctgggacagcagcatagctagaccccatct
ccaccaaaaatttaagaatcagctaggctgtggtgatgtgcacctgtaattctcgctac
ttggaaggcttgagcccaggagtttgaagctgcagtgagctatgttcgtgccactatac
tccaacctgagagacagagtgagaccctgtcttaaaaaaaaaaaaaaaaaaaaaaaaaaca
agaaaaaaaaatacccccccaaaaaacaaaacagaaagaaacccagtgatagtcacagttg
tcctttttcacattggctgtccctatggggaccagggttgggggttggcgtctgtgcctct
cctgagtgggccacaccccctctcctgcccacctcag (SEQ ID NO: 103)