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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 August 2010 (05.08.2010)

PCT

(10) International Publication Number
WO 2010/087702 A1

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

(21) International Application Number:
PCT/NL2010/050034

(22) International Filing Date:
26 January 2010 (26.01.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09151787.0 30 January 2009 (30.01.2009) EP
61/148,477 30 January 2009 (30.01.2009) US

(71) Applicant *(for all designated States except US)*: **STICHTING KATHOLIEKE UNIVERSITEIT** [NL/NL]; Geert Grooteplein-Noord 9, NL-6525 EZ Nijmegen (NL).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **JANSEN, Jozef Henricus** [NL/NL]; Delistraat 13, NL-6524 KM Nijmegen (NL).

(74) Agent: **KETELAARS, Maarten**; J.W. Frisolaan 13, NL-2517 JS The Hague (NL).

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

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

Published:

— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

(54) **Title**: TET2 gene as a marker for diagnosing a myelodysuplastic syndrome (MDS) or an acute myeloid leukemia (AML) and determining the prognosis in a subject

Fig 1a

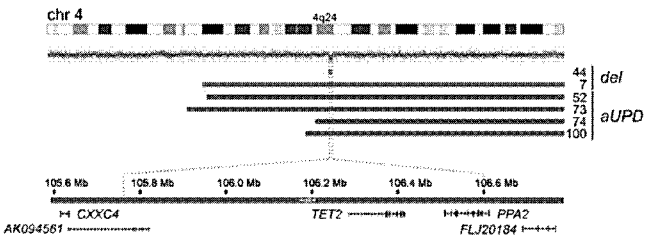
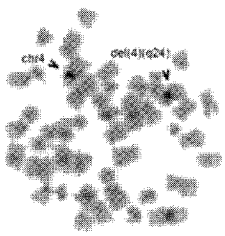


Fig 1b



(57) **Abstract**: The present invention relates to a method for diagnosing a MDS or an AML in a subject,the method comprising the step of identifying the presence of a mutation in an allele of the TET2 gene of said subject by comparing said TET2 gene to a control TET2 gene being represented by SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3or an equivalent thereof. The method further relates to a method for treating such diseases in a subject using a control TET2 gene. The invention also relates to such TET2 gene for use in such a method.

WO 2010/087702 A1

TET2 gene as a marker for diagnosing a myelodysplastic syndrome (MDS) or an acute myeloid leukemia (AML) and determining the prognosis in a subject

Field of the invention

The present invention relates to a method for diagnosing a MDS or an AML in a subject, the method comprising the step of identifying the presence of a mutation in an allele of the TET2 gene of said subject by comparing said TET2 gene to a control TET2 gene being represented by SEQ ID NO:1, SEQ IDNO:2 or SEQ ID NO:3 or an equivalent thereof (NM_001127208 / UniRef100Q6N021). The method further relates to a method for determining the prognosis of a subject with MDS or AML. In addition, the method relates to treating such diseases in a subject using a control TET2 gene. The invention also relates to such TET2 gene for use in such a method.

Background of the invention

MDS are clonal hematopoietic disorders that are characterized by dysplasia of the myeloid, megakaryocytic and/or erythroid lineages. In addition, cytopenias are observed. The incidence is approximately 3-10 per 100.000 persons annually[2, 3]. MDS mainly affects people from middle age onwards but the disease also occurs at more juvenile ages. The clinical course ranges from a smoldering disease that may last for years, to a more acute manifestation with bone marrow failure resulting in life-threatening complications. About 30% of the patients develop AML, but most patients eventually die from complications of bone marrow failure. Proper diagnosis of MDS is often difficult as various conditions (like infections, medication, toxic agents) may give rise to dysplasia without clonal disease. The international prognostic scoring system (IPSS) is a helpful tool to predict survival and progression towards AML[4, 5]. This classification is based on the number of cytopenias, the percentage of bone marrow blasts, and the type of cytogenetic aberrations. Four risk groups are defined: low-risk, intermediate-1 (int-1), intermediate-2 (int-2) and high-risk. Treatment of MDS is diverse and ranges from supportive treatment to allogeneic stem cell transplantation[6-8]. Recently, an analogue of thalidomide (Lenalidomide/Revlimid) has shown significant activity in MDS[9], particularly in patients with a 5q- chromosome abnormality.

If present, genetic markers facilitate the diagnosis of MDS. Using karyotyping and fluorescent in-situ hybridization (FISH), chromosomal aberrations can be found in approximately 60% of the cases[10]. Common aberrations are loss of (part of) chromosomes 5, 7, 17, 20 and Y, and trisomy 8[10-12]. In spite of this knowledge, information on specific genes that are affected in MDS is still scarce. Mutations in *N-RAS* have been described in 10-15% of the patients, as well as mutations in *FLT-3* (2-5%), *TP53* (5-10%) and *RUNX1/AML1* (2-10%). Recently, haploinsufficiency of the *RPS14* gene was implicated in the impaired erythropoiesis in 5q- patients[13, 14].

Lack of genetic markers hampers proper diagnosis, assessment of prognosis and response to therapy and the development of novel forms of targeted therapy. The inventors identified and characterized new chromosomal regions and subsequently a specific gene, TET2, located therein as being a specific marker for diagnosing a MDS or an AML in a subject. Mutations present in this gene account for more than 25% of the patients with MDS, with a more than 40% incidence in the IPPS low and int1 categories In addition, TET2 mutations were found in more than 15% of the AML patients.

Description of the figures

Figure 1. Deletion and uniparental disomy of chromosome 4q detected by SNP-array and FISH.

A: Using SNP-arrays, deletions (patient 7 and 44), or uniparental disomy (patients 52, 73, 74 and 100) of chromosome 4q were seen in 2 and 4 patients respectively. Patient 44 showed a small 0,8 Mb deletion, harbouring the *PPA2* and *TET2* genes. Patient 7 showed a large deletion, spanning most of the q-arm of chromosome 4. In four patients acquired UPD of (part of) the long arm of chromosome 4 was revealed. T cells that were isolated from the same patients did not show these aberrations, indicating that they represented acquired, non-constitutional aberrations.

B: Confirmation of the 4q24 deletion in patient UPN44 by FISH. FISH probes were prepared from a BAC clone library (RP11-449G2 and RP11-542F11, BACPAC Resources, Oakland, USA). The 4q24 probe was labeled in green, and cells from patient 44 were stained. To identify the two chromosomes 4, a chromosome centromere

probe was used that was labeled in red. One characteristic metaphase is depicted, showing two red signals, and only one green signal.

Figure 2: *TET2* splice variants.

Three putative isoforms of *TET2* that can be deduced from sequences in public databases. The existence of these different isoforms was confirmed by PCR and sequencing. The location of the start codons does not differ between the isoforms. The locations of the stop codons vary and are located in exon 10 (isoform 1), 3B (isoform 2) or 5 (isoform 3). The boxes represent exons, the black boxes indicate the translated region. AA = amino acid.

Figure 3: Localization of *TET2* mutations

The localization of all *TET2* mutations shown at the protein level. Missense mutations are indicated by dark gray boxes, nonsense- and frameshift mutations in light gray boxes. All but one missense mutations are located in one of the conserved protein regions. The nonsense and frameshift mutations are mainly situated before or in box 1. UPN = unique patient number (corresponding to Table 1), AA = amino acid.

Figure 4: Expression of three *TET2* isoforms in various cells and tissues

Expression of *TET2* isoforms was measured by isoform-specific quantitative PCR. Expression is depicted relative to the housekeeping gene GAPDH. Expression levels were determined in (A) different hematopoietic cell fractions, (B) various tissues and (C) the promyelocytic cell line NB4 before and after induction of granulocytic differentiation using *all-trans* retinoic acid (ATRA, 10-6 M). The mean expression levels are indicated. Expression in granulocytes was highly increased compared to other hematopoietic cell fractions and tissues (note the difference in scale between Figure 4A and 4B/C). Error bars represent the standard deviation. CD = cluster of differentiation.

Figure 5: Characteristics and prognosis of patients with *TET2* mutations

TET2 mutations occur more often in IPSS low and int-1 patients compared to IPSS int-2 and high risk patients. When studying the different components of the IPSS

classification, *TET2* mutations appear to be more frequent in patients with a normal karyotype, 0/1 cytopenias and a low number of blasts.

* indicates statistically significant differences.

Description of the invention

Methods of diagnosis

In a first aspect, the invention relates to a method for diagnosing a MDS or AML in a subject, the method comprising the step of identifying the presence of a mutation in an allele of the TET2 gene of said subject by comparing said TET2 gene to a control TET2 gene being represented by SEQ ID NO:1, SEQ ID NO:2 and/or SEQ ID NO:3 or an equivalent thereof.

In the context of the invention, MDS are clonal hematopoietic disorders that are characterized by dysplasia of the myeloid, megakaryocytic and/or erythroid lineages, ineffective haematopoiesis and increased risk of development of acute myeloid leukaemia. (Swerdlow SH, Campo E, Harris NL, Jaffe ES, Pileri SA, Stein H, Thiele J, Vardiman JW (Eds) WHO classification of tumors of haematopoietic and lymphoid tissues, Chapter 5, Brunning RD et al, page 88-107, WHO press 2008, ISBN 978-92-832-2431-0). In the context of the invention, AML is a clonal myeloid disease characterized by an excess of more than 20% immature blast cells in the bone marrow (Swerdlow SH, Campo E, Harris NL, Jaffe ES, Pileri SA, Stein H, Thiele J, Vardiman JW (Eds) WHO classification of tumors of haematopoietic and lymphoid tissues, Chapter 6 DA Arber et al, page 110-147, WHO press 2008, ISBN 978-92-832-2431-0). The presence of such types of cells can be detected by techniques known to the skilled person. In the context of the invention and as known to the skilled person, MDS is a syndrome which is distinct from MPN (myeloid proliferative neoplasm) as exemplified in Swerdlow et al (SH, Campo E, Harris NL, Jaffe ES, Pileri SA, Stein H, Thiele J, Vardiman JW (Eds) WHO classification of tumors of haematopoietic and lymphoid tissues, Chapter 5, Brunning RD et al, page 88-107, WHO press 2008, ISBN 978-92-832-2431-0). The fact that a marker is characteristic of MPN does not have any implications for its role as a marker of MDS, and vice versa. The same holds between MPN and AML.

In the context of the invention, diagnosis means the assessment of the presence of a clonal malignant AML or MDS disease in a subject. Such diagnostic method is attractive since it can be reached in an early stage of said disease. The availability of a genetic marker may help to establish that a clonal disease is present, excluding non-malignant, polyclonal hematopoietic aberrations that may be seen for instance during infections. Clearly these non-malignant aberrations have a very different prognosis compared to MDS or AML.

In the context of the invention, a subject may be an animal or a human being. Preferably, a subject is a human being.

A method of the invention comprises the step of identifying the presence of a mutation in an allele of the TET2 gene of said subject by comparing said TET2 gene to a control TET2 gene being represented by SEQ ID NO:1 or SEQ ID NO:2 or SEQ ID NO:3 or an equivalent thereof. Transcription of the TET2 gene leads to the production of three distinct cDNAs obtained by distinct splicing mechanisms, each of these cDNAs being represented by SEQ ID NO:1, SEQ ID NO:2 or SEQ ID NO:3 respectively. The translation of each of these cDNAs leads to the production of three TET2 protein isoforms, each being respectively represented by SEQ ID NO:4, SEQ ID NO5 and SEQ ID NO6. Since the identification of a mutation in a gene may be difficult to be measured in a subject, a sample from a subject is preferably used. According to another preferred embodiment, a mutation is determined or identified *ex vivo* in a sample obtained from a subject. A sample preferably comprises a blood and/or bone marrow sample from a subject. In a preferred embodiment, a method is applied on a myeloid cell isolated and optionally purified from bone marrow or blood of said subject. The isolation and optionally purification from myeloid cells can be achieved by density gradient centrifugation or ammonium chloride lysis of red cells using standard technology known to the skilled person.

A control TET2 gene is preferably represented by SEQ ID NO:1 or SEQ ID NO:2 or SEQ ID NO:3 or by an equivalent thereof. A control TET2 gene encodes a control or wild type or functional TET2 protein or an equivalent thereof. A preferred control

TET2 protein is represented by SEQ ID NO:4, SEQ ID NO:5 or SEQ ID NO:6 respectively. An equivalent of a control TET2 gene or protein preferably means a nucleotide respectively an amino acid sequence that is represented by SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3, respectively SEQ ID NO:4, SEQ ID NO:5, or SEQ ID NO:6, which does not have any of the mutations as disclosed in table 1 or figure 3 but which may have any of the polymorphisms as disclosed in table 4. Table 4 discloses some SNP as identified in the TET2 gene of healthy subjects. Each of the mutations identified in table 1 or 4 are identified by reference to isoform 1 of the TET2 protein, which is represented by SEQ ID NO:4. The skilled person understands that each of these mutations may also be present in isoform 2 and/or 3 of the TET2 protein, represented by SEQ ID NO:5, SEQ ID NO:6 respectively.

The identification of the presence of a mutation in an allele of the TET2 gene may be carried out by any method known to the skilled person. A mutation may be determined as being present in the gene TET2 (DNA level), or as present in the messenger RNA coding for TET2 (mRNA level), and/or as present in the protein TET2 obtained from a subject. This determination or identification may be carried out using a direct and/or an indirect method. The type of method used and means used in said method are not critical for the invention as long as these methods and means are able to identify a mutation. Preferred means include SNP arrays, Southern blotting, PCR using primers, FISH, nucleic acid probes Westen-blot analysis or immunofluorescence microscopy or FACS using antibodies . Each of these techniques is known to the skilled person (CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, AUSBEL FM ET AL (ed) WILEY AND SONS Inc, ISBN 0-471-50338-X).

The type and the number of mutations identified in the TET2 gene are also not critical for the method of the invention. As soon as a mutation has been identified in the TET2 gene by comparison to a TET2 control gene as defined herein, one can speak of a subject having been diagnosed as a MDS or an AML patient.

Within the context of the invention, a mutation in an allele of the TET2 gene is preferably an acquired mutation. Accordingly, in a preferred embodiment, a method of the invention comprises the identification of a given mutation in the TET2 gene in a first sample from a subject, preferably in a blood and/or bone marrow sample, while a second sample of a same subject comprises cells that do not belong to the clonal

haematopoietic disease, preferably buccal swap or other non-haematopoietic cells, that do not have a mutation in an allele of the TET2 gene as identified in said first sample.

Preferred mutations in an allele of the TET2 gene is a deletion, an insertion, an amplification, a copy neutral loss of heterozygosity, a mono- or a bi-allelic nonsense, a missense and/or a frameshift.

It is also encompassed by the present invention that at least one mutation is found in each allele of the TET2 gene.

Preferred mutations have been described in table 1 and in Figure 3:

One preferred mutation is a deletion or amplification at 4q24.

A mutation may lead to a TET2 protein which is altered compared to the TET2 protein represented by SEQ ID NO:4, SEQ ID NO:5 or SEQ ID NO:6. An insertion, deletion and/or non-sense mutation leading to a truncated TET2 protein may be expected in any part of a TET2 protein, whereas a missense mutation is preferably clustered in boxes 1 and/or 2 as identified in figure 3. An altered TET2 protein may have an amino acid deletion and/or substitution and/or insertion and/or may be a truncated TET2 protein by comparison with a TET2 protein as represented by SEQ ID NO:4, SEQ ID NO:5 or SEQ ID NO:6.

Nucleic acid construct

In a further aspect, the invention relates to a nucleic acid construct. A nucleic acid construct comprises all or a part of a nucleotide sequence that encodes a TET2 protein that comprises an amino acid sequence that is encoded by a nucleotide sequence selected from:

(a) a nucleotide sequence that has at least 80, 85, 90, 95, 98 or 99% identity with SEQ ID NO:1, SEQ ID NO:2 or SEQ ID NO:3 or which is an equivalent of SEQ ID NO:1, SEQ ID NO:2 or SEQ ID NO:3; and/or,

(b) a nucleotide sequence that encodes an amino acid sequence that has at least 80, 85, 90, 95, 98 or 99% amino acid identity with an amino acid sequence encoded by SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3.

Preferably, a nucleotide sequence is operably linked to a promoter that is capable of driving expression of the nucleotide sequence in a hematopoietic cell, more preferably a human hematopoietic cell.

Said nucleic acid construct may be an expression construct or an inactivating construct. An expression construct is intended to be used for expressing a control or a wild type TET2 protein or an equivalent thereof. An inactivating construct is intended to be used for down-regulating the expression of a mutated TET2 gene and/or an altered TET2 protein. Each of these features have already been defined earlier herein.

An inactivating nucleic acid construct of the invention may comprises or consists of a nucleotide sequence that encodes an RNAi agent, i.e. an RNA molecule that is capable of RNA interference or that is part of an RNA molecule that is capable of RNA interference. Such RNA molecules are referred to as siRNA (short interfering RNA, including e.g. a short hairpin RNA). A nucleotide sequence that encodes a RNAi agent preferably has sufficient complementarity with a cellular nucleotide sequence to be capable of inhibiting the expression of an altered TET2 protein wherein optionally the nucleotide sequence encoding the RNAi agent is operably linked to a promoter that is capable of driving expression of the nucleotide sequence in a hematopoietic cell.

In a nucleic acid construct of the invention, a promoter which may be present is preferably a promoter that is specific for a hematopoietic cell. More preferably, a promoter chosen is specific for and functional in a human hematopoietic cell. A promoter that is specific for a hematopoietic cell is a promoter with a transcription rate that is higher in a hematopoietic cell than in other types of cells. Preferably the promoter's transcription rate in a hematopoietic cell is at least 1.1, 1.5, 2.0 or 5.0 times higher than in a non-hematopoietic cell as measured by PCR of the construct in the hematopoietic as compared to a non-hematopoietic cell.

A suitable promoter for use in a nucleic acid construct of the invention and that is capable of driving expression in a hematopoietic cell includes a promoter of a TET2 gene. Other suitable promoters for use in a nucleic acid construct of the invention and that is capable of driving expression in a hematopoietic cell include A promoter for use in a DNA construct of the invention is preferably of mammalian origin, more preferably of human origin.

In a preferred embodiment a nucleic acid construct is a viral gene therapy vector selected from gene therapy vectors based on an adenovirus, an adeno-associated virus (AAV), a herpes virus, a pox virus and a retrovirus. A preferred viral gene therapy vector is an AAV or Lentiviral vector. Such vectors are further described herein below.

Method for preventing and/or treating a MDS or an AML

There are currently various medicaments that may be used in a method for treating a MDS or an AML that is used in patients. Standard treatment in AML comprises high dose chemotherapy and, if applicable bone marrow transplantation. In MDS, the only generally accepted curative option is allogeneic bone marrow transplantation. In spite of these treatment options, in AML overall approximately 60% of the patients die within 5 years after diagnosis. In MDS, the course of the disease may be somewhat less aggressive, but most patients die of bone marrow failure within 10 years after diagnosis.

Accordingly, in a further aspect, the invention provides a method for treating a MDS or an AML in a subject, the method comprising pharmacologically decreasing the activity or the steady-state level of a TET2 protein produced from a mutated TET2 gene and/or expressing a wild type TET2 gene, preferably being a control or wild type TET2 gene or an equivalent thereof.

All the features of this preferred method have already been described herein.

In a preferred method of the invention, an activity or steady-state level of a mutated TET2 gene or altered TET2 protein is altered in order to mimic its physiological level in a healthy subject.

An activity or steady-state level of a TET2 protein may be altered at the level of the protein itself, e.g. by providing said wild type or non-altered protein to a subject, preferably to a hematopoietic cell of a subject, said protein being from an exogenous source, or by adding an antagonist or inhibitor of said altered protein to a subject, preferably to a hematopoietic cell, such as e.g. an antibody against the altered protein, preferably a neutralizing antibody. For provision of a protein from an exogenous source, a protein may conveniently be produced by expression of a nucleic acid encoding a protein in a suitable host cell as described below. An antibody against a TET2 protein may be obtained as described below.

Preferably, however, an activity or steady-state level of a TET2 protein is altered by regulating the expression level of a nucleotide sequence encoding a TET2 protein. Preferably, the expression level of a nucleotide sequence is regulated in a hematopoietic cell. The expression level of a non-altered or wild type TET2 protein may be increased by introduction of an expression construct (or vector) into a

hematopoietic cell, whereby an expression vector comprises a nucleotide sequence encoding a said polypeptide, and whereby a nucleotide sequence is under control of a promoter capable of driving expression of a nucleotide sequence in a hematopoietic cell.

Alternatively or in combination with previous embodiment of increasing the expression level of a non-mutated TET2 gene or corresponding protein, if so required for preventing and/or treating a MDS or an AML, the expression level of an altered TET2 protein may be decreased by providing an antisense molecule to a hematopoietic cell, whereby an antisense molecule is capable of inhibiting the biosynthesis (usually the translation) of a nucleotide sequence encoding said protein. Decreasing gene expression by providing antisense or interfering RNA molecules is described below herein and is e.g. reviewed by Famulok et al. (2002, Trends Biotechnol., 20(11): 462-466). An antisense molecule may be provided to a cells as such or it may be provided by introducing an expression construct into a hematopoietic cell, whereby an expression construct comprises an antisense nucleotide sequence that is capable of inhibiting the expression of a nucleotide sequence encoding said altered TET2 protein, and whereby an antisense nucleotide sequence is under control of a promoter capable of driving transcription of an antisense nucleotide sequence in a hematopoietic cell. The expression level of an altered TET2 protein may also be decreased by introducing an expression construct into a hematopoietic cell, whereby an expression construct comprises a nucleotide sequence encoding a factor capable of trans-repression of an endogenous nucleotide sequence encoding said altered TET2 protein. An antisense or interfering nucleic acid molecule may be introduced into a cell directly "as such", optionally in a suitable formulation, or it may be produce *in situ* in a cell by introducing into a cell an expression construct comprising a (antisense or interfering) nucleotide sequence that is capable of inhibiting the expression of a nucleotide sequence encoding said altered TET2 protein, whereby, optionally, an antisense or interfering nucleotide sequence is under control of a promoter capable of driving expression of an nucleotide sequence in a hematopoietic cell.

A method of the invention preferably comprises the step of administering to a subject a therapeutically effective amount of a pharmaceutical composition comprising a nucleic acid construct for modulating the activity or steady state level of an altered TET2 protein and/or a neutralizing antibody and/or providing a non-altered TET2 gene

or protein as defined herein. A nucleic acid construct may be an expression construct as further specified herein. Preferably, an expression construct is a viral gene therapy vector selected from gene therapy vectors based on an adenovirus, an adeno-associated virus (AAV), a herpes virus, a pox virus and a retrovirus. A preferred viral gene therapy vector is an AAV or Lentiviral vector. Alternatively, a nucleic acid construct may be for inhibiting expression of an altered TET2 protein such as an antisense molecule or an RNA molecule capable of RNA interference (see below).

In a method of the invention, a hematopoietic cell is preferably a hematopoietic cell from a subject shown to have MDS or an AML. In a method, a hematopoietic cell chosen to be treated is preferably isolated from the subject they belong to (ex vivo method). Cells are subsequently treated by altering an activity or the steady state level of an altered TET2 protein. This treatment is preferably performed by infecting them with a non-altered TET2 protein and/or a nucleic acid construct of the invention comprising a nucleic acid or a gene encoding for a non-altered TET2 protein and/or a neutralizing antibody as earlier defined herein. Finally, treated cells are placed back into the subject they belong to. Alternatively or in combination with other preferred methods, in a method of the invention, a nucleic acid construct and/or a neutralizing antibody and/or a non-altered TET2 protein is preferably administered into a subject intravenously where treatment is needed.

In another treating method, the invention mentioned herein may be combined with standard treatments of a MDS or an AML such as any of those as earlier identified herein.

TET2 gene or protein as a medicament for treating MDS or AML

Accordingly in a further aspect, the invention relates to a TET2 protein, a TET2 gene or a nucleic acid construct comprising said gene for use as a medicament for treating a MDS or an AML, preferably wherein the TET2 gene is represented by SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3 or an equivalent thereof and/or wherein the TET2 protein is represented by SEQ ID NO:4, SEQ ID NO:5 or SEQ ID NO:6 or an equivalent thereof.

Use of a nucleic acid construct

In a further aspect the invention relates to a use of a nucleic acid construct as defined for modulating the activity or steady state level of an altered TET2 protein as defined herein, for the manufacture of a medicament for treating a MDS or an AML in a subject, preferably in a method of the invention as defined herein above.

Sequence identity

"Sequence identity" is herein defined as a relationship between two or more amino acid (polypeptide or protein) sequences or two or more nucleic acid (nucleotide, polynucleotide, gene) sequences, as determined by comparing the sequences. Preferably identity is assessed on the whole length of a given SEQ ID NO as identified herein. In the art, "identity" also means the degree of sequence relatedness between amino acid or nucleic acid sequences, as the case may be, as determined by the match between strings of such sequences. "Similarity" between two amino acid sequences is determined by comparing the amino acid sequence and its conserved amino acid substitutes of one polypeptide to the sequence of a second polypeptide. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in (Computational Molecular Biology, Lesk, A. M., ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D. W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part I, Griffin, A. M., and Griffin, H. G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heine, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., SIAM J. Applied Math., 48:1073 (1988).

Preferred methods to determine identity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Preferred computer program methods to determine identity and similarity between two sequences include e.g. the GCG program package (Devereux, J., et al., Nucleic Acids Research 12 (1): 387 (1984)), BestFit, BLASTP, BLASTN, and FASTA (Altschul, S. F. et al., J. Mol. Biol. 215:403-410 (1990). The BLAST X program is publicly available from NCBI and other sources (BLAST Manual, Altschul, S., et al., NCBI NLM NIH Bethesda, MD 20894; Altschul,

S., et al., J. Mol. Biol. 215:403-410 (1990). The well-known Smith Waterman algorithm may also be used to determine identity.

Preferred parameters for polypeptide sequence comparison include the following: Algorithm: Needleman and Wunsch, J. Mol. Biol. 48:443-453 (1970); Comparison matrix: BLOSSUM62 from Hentikoff and Hentikoff, Proc. Natl. Acad. Sci. USA. 89:10915-10919 (1992); Gap Penalty: 12; and Gap Length Penalty: 4. A program useful with these parameters is publicly available as the "Ogap" program from Genetics Computer Group, located in Madison, WI. The aforementioned parameters are the default parameters for amino acid comparisons (along with no penalty for end gaps).

Preferred parameters for nucleic acid comparison include the following: Algorithm: Needleman and Wunsch, J. Mol. Biol. 48:443-453 (1970); Comparison matrix: matches=+10, mismatch=0; Gap Penalty: 50; Gap Length Penalty: 3. Available as the Gap program from Genetics Computer Group, located in Madison, Wis. Given above are the default parameters for nucleic acid comparisons.

Optionally, in determining the degree of amino acid similarity, the skilled person may also take into account so-called "conservative" amino acid substitutions, as will be clear to the skilled person. Conservative amino acid substitutions refer to the interchangeability of residues having similar side chains. For example, a group of amino acids having aliphatic side chains is glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains is serine and threonine; a group of amino acids having amide-containing side chains is asparagine and glutamine; a group of amino acids having aromatic side chains is phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains is lysine, arginine, and histidine; and a group of amino acids having sulphur-containing side chains is cysteine and methionine. Preferred conservative amino acids substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, and asparagine-glutamine. Substitutional variants of the amino acid sequence disclosed herein are those in which at least one residue in the disclosed sequences has been removed and a different residue inserted in its place. Preferably, the amino acid change is conservative. Preferred conservative substitutions for each of the naturally occurring amino acids are as follows: Ala to ser; Arg to lys; Asn to gln or his; Asp to glu; Cys to ser or ala; Gln to asn; Glu to asp; Gly to pro; His to asn or gln; Ile to leu or

val; Leu to ile or val; Lys to arg; gln or glu; Met to leu or ile; Phe to met, leu or tyr; Ser to thr; Thr to ser; Trp to tyr; Tyr to trp or phe; and, Val to ile or leu.

Recombinant techniques and methods for recombinant production of a TET2 protein

A TET2 protein for use in the present invention can be prepared using recombinant techniques, in which a nucleotide sequence encoding protein is expressed in a suitable host cell. The present invention thus also concerns the use of a vector or nucleic acid construct comprising a nucleic acid molecule or nucleotide sequence as defined above. Preferably, a vector is a replicative vector comprising an origin of replication (or autonomously replication sequence) that ensures multiplication of a vector in a suitable host for said vector. Alternatively a vector is capable of integrating into a host cell's genome, e.g. through homologous recombination or otherwise. A particularly preferred vector is an expression vector wherein a nucleotide sequence encoding said protein, is operably linked to a promoter capable of directing expression of a nucleotide sequence (i.e. a coding sequence) in a host cell for the vector.

As used herein, the term "promoter" refers to a nucleic acid fragment that functions to control the transcription of one or more genes (or coding sequence), located upstream with respect to the direction of transcription of the transcription initiation site of the gene, and is structurally identified by the presence of a binding site for DNA-dependent RNA polymerase, transcription initiation sites and any other DNA sequences, including, but not limited to transcription factor binding sites, repressor and activator protein binding sites, and any other sequences of nucleotides known to one of skill in the art to act directly or indirectly to regulate the amount of transcription from the promoter. A "constitutive" promoter is a promoter that is active under most physiological and developmental conditions. An "inducible" promoter is a promoter that is regulated depending on physiological or developmental conditions. A "tissue specific" promoter is only active in specific types of differentiated cells/tissues, such as preferably a hematopoietic cell or tissue derived therefrom.

Expression vectors allow a TET2 protein as defined above to be prepared using recombinant techniques in which a nucleotide sequence encoding said protein is expressed in a suitable cell, e.g. cultured cells or cells of a multicellular organism, such as described in Ausubel et al., "Current Protocols in Molecular Biology", Greene Publishing and Wiley-Interscience, New York (1987) and in Sambrook and Russell

(2001, *supra*); both of which are incorporated herein by reference in their entirety. Also see, Kunkel (1985) Proc. Natl. Acad. Sci. 82:488 (describing site directed mutagenesis) and Roberts et al. (1987) Nature 328:731-734 or Wells, J.A., et al. (1985) Gene 34: 315 (describing cassette mutagenesis).

Typically, a nucleic acid or nucleotide sequence encoding a TET2 protein is used in an expression vector. The phrase "expression vector" generally refers to a nucleotide sequence that is capable of effecting expression of a gene in a host compatible with such sequences. These expression vectors typically include at least suitable promoter sequences and optionally, transcription termination signals. An additional factor necessary or helpful in effecting expression can also be used as described herein. A nucleic acid or DNA or nucleotide sequence encoding a TET2 protein is incorporated into a DNA construct capable of introduction into and expression in an *in vitro* cell culture. Specifically, a DNA construct is suitable for replication in a prokaryotic host, such as bacteria, *e.g.*, *E. coli*, or can be introduced into a cultured mammalian, plant, insect, e.g., Sf9, yeast, fungi or other eukaryotic cell lines.

A DNA construct prepared for introduction into a particular host typically include a replication system recognized by the host, an intended DNA segment encoding a desired polypeptide, and transcriptional and translational initiation and termination regulatory sequences operably linked to the polypeptide-encoding segment. A DNA segment is “operably linked” when it is placed into a functional relationship with another DNA segment. For example, a promoter or enhancer is operably linked to a coding sequence if it stimulates the transcription of the sequence. DNA for a signal sequence is operably linked to DNA encoding a polypeptide if it is expressed as a preprotein that participates in the secretion of a polypeptide. Generally, a DNA sequence that is operably linked are contiguous, and, in the case of a signal sequence, both contiguous and in reading phase. However, enhancers need not be contiguous with a coding sequence whose transcription they control. Linking is accomplished by ligation at convenient restriction sites or at adapters or linkers inserted in lieu thereof.

The selection of an appropriate promoter sequence generally depends upon the host cell selected for the expression of a DNA segment. Examples of suitable promoter sequences include prokaryotic, and eukaryotic promoters well known in the art (see, e.g. Sambrook and Russell, 2001, *supra*). A transcriptional regulatory sequence typically includes a heterologous enhancer or promoter that is recognised by the host.

The selection of an appropriate promoter depends upon the host, but promoters such as the trp, lac and phage promoters, tRNA promoters and glycolytic enzyme promoters are known and available (see, e.g. Sambrook and Russell, 2001, *supra*). An expression vector includes the replication system and transcriptional and translational regulatory sequences together with the insertion site for the polypeptide encoding segment can be employed. Examples of workable combinations of cell lines and expression vectors are described in Sambrook and Russell (2001, *supra*) and in Metzger et al. (1988) Nature 334: 31-36. For example, suitable expression vectors can be expressed in, yeast, *e.g. S. cerevisiae*, *e.g.*, insect cells, *e.g.,* Sf9 cells, mammalian cells, *e.g.*, CHO cells and bacterial cells, *e.g.*, *E. coli*. A host cell may thus be a prokaryotic or eukarotic host cell. A host cell may be a host cell that is suitable for culture in liquid or on solid media. A host cell is preferably used in a method for producing a TET2 protein as defined above. A method comprises the step of culturing a host cell under conditions conducive to the expression of a TET2 protein. Optionally the method may comprise recovery said protein. A protein may e.g. be recovered from the culture medium by standard protein purification techniques, including a variety of chromatography methods known in the art per se.

Alternatively, a host cell is a cell that is part of a multicellular organism such as a transgenic plant or animal, preferably a non-human animal. A transgenic plant comprises in at least a part of its cells a vector as defined above. Methods for generating transgenic plants are e.g. described in U.S. 6,359,196 and in the references cited therein. Such transgenic plant or animal may be used in a method for producing a polypeptide of the invention as defined above. For transgenic plant, a method comprises the step of recovering a part of a transgenic plant comprising in its cells the vector or a part of a descendant of such transgenic plant, whereby the plant part contains a TET2 protein, and, optionally recovery of a TET2 protein from the plant part. Such methods are also described in U.S. 6,359,196 and in the references cited therein. Similarly, a transgenic animal comprises in its somatic and germ cells a vector as defined above. A transgenic animal preferably is a non-human animal. Methods for generating transgenic animals are e.g. described in WO 01/57079 and in the references cited therein. Such transgenic animals may be used in a method for producing a TET2 protein as defined above, the method comprising the step of recovering a body fluid from a transgenic animal comprising the vector or a female descendant thereof, wherein

the body fluid contains a TET2 protein, and, optionally recovery of said protein from the body fluid. Such methods are also described in WO 01/57079 and in the references cited therein. A body fluid containing a TET2 protein is preferably blood or more preferably milk.

Another method for preparing a TET2 protein is to employ an *in vitro* transcription/translation system. A DNA encoding a TET2 protein is cloned into an expression vector as described *supra*. An expression vector is then transcribed and translated *in vitro*. A translation product can be used directly or first purified. A TET2 protein resulting from *in vitro* translation typically do not contain the post-translation modifications present on said protein synthesised *in vivo*, although due to the inherent presence of microsomes some post-translational modification may occur. A method for synthesis of said protein by *in vitro* translation is described by, for example, Berger & Kimmel, Methods in Enzymology, Volume 152, Guide to Molecular Cloning Techniques, Academic Press, Inc., San Diego, CA, 1987.

Gene therapy

Some aspects of the invention concern the use of a nucleic acid construct or expression vector comprising a nucleotide sequence as defined above, wherein the vector is a vector that is suitable for gene therapy. Vectors that are suitable for gene therapy are described in Anderson 1998, Nature 392: 25-30; Walther and Stein, 2000, Drugs 60: 249-71; Kay et al., 2001, Nat. Med. 7: 33-40; Russell, 2000, J. Gen. Virol. 81: 2573-604; Amado and Chen, 1999, Science 285: 674-6; Federico, 1999, Curr. Opin. Biotechnol.10: 448-53; Vigna and Naldini, 2000, J. Gene Med. 2: 308-16; Marin et al., 1997, Mol. Med. Today 3: 396-403; Peng and Russell, 1999, Curr. Opin. Biotechnol. 10: 454-7; Sommerfelt, 1999, J. Gen. Virol. 80: 3049-64; Reiser, 2000, Gene Ther. 7: 910-3; and references cited therein.

A particularly suitable gene therapy vector includes an Adenoviral and Adeno-associated virus (AAV) vector. These vectors infect a wide number of dividing and non-dividing cell types including neuronal cells. In addition adenoviral vectors are capable of high levels of transgene expression. However, because of the episomal nature of the adenoviral and AAV vectors after cell entry, these viral vectors are most suited for therapeutic applications requiring only transient expression of the transgene (Russell, 2000, J. Gen. Virol. 81: 2573-2604; Goncalves, 2005, Virol J. 2(1):43) as

indicated above. Preferred adenoviral vectors are modified to reduce the host response as reviewed by Russell (2000, *supra*). Method for neuronal gene therapy using AAV vectors are described by Wang et al., 2005, J Gene Med. March 9 (Epub ahead of print), Mandel et al., 2004, Curr Opin Mol Ther. 6(5):482-90, and Martin et al., 2004, Eye 18(11):1049-55. For gene transfer into a hematopoietic cell, a AAV serotype 2 is an effective vector and therefore a preferred AAV serotype.

A preferred retroviral vector for application in the present invention is a lentiviral based expression construct. Lentiviral vectors have the unique ability to infect non-dividing cells (Amado and Chen, 1999 Science 285: 674-6). Methods for the construction and use of lentiviral based expression constructs are described in U.S. Patent No.'s 6,165,782, 6,207,455, 6,218,181, 6,277,633 and 6,323,031 and in Federico (1999, Curr Opin Biotechnol 10: 448-53) and Vigna et al. (2000, J Gene Med 2000; 2: 308-16).

Generally, gene therapy vectors will be as the expression vectors described above in the sense that they comprise a nucleotide sequence encoding a TET2 protein to be expressed, whereby a nucleotide sequence is operably linked to the appropriate regulatory sequences as indicated above. Such regulatory sequence will at least comprise a promoter sequence. Suitable promoters for expression of a nucleotide sequence encoding a polypeptide from gene therapy vectors include e.g. cytomegalovirus (CMV) intermediate early promoter, viral long terminal repeat promoters (LTRs), such as those from murine moloney leukaemia virus (MMLV) rous sarcoma virus, or HTLV-1 , the simian virus 40 (SV 40) early promoter and the herpes simplex virus thymidine kinase promoter. Suitable promoters are described below.

Several inducible promoter systems have been described that may be induced by the administration of small organic or inorganic compounds. Such inducible promoters include those controlled by heavy metals, such as the metallothionine promoter (Brinster et al. 1982 Nature 296: 39-42; Mayo et al. 1982 Cell 29: 99-108), RU-486 (a progesterone antagonist) (Wang et al. 1994 Proc. Natl. Acad. Sci. USA 91: 8180-8184), steroids (Mader and White, 1993 Proc. Natl. Acad. Sci. USA 90: 5603-5607), tetracycline (Gossen and Bujard 1992 Proc. Natl. Acad. Sci. USA 89: 5547-5551; U.S. Pat. No. 5,464,758; Furth et al. 1994 Proc. Natl. Acad. Sci. USA 91: 9302-9306; Howe et al. 1995 J. Biol. Chem. 270: 14168-14174; Resnitzky et al. 1994 Mol. Cell. Biol. 14: 1669-1679; Shockett et al. 1995 Proc. Natl. Acad. Sci. USA 92: 6522-6526) and the

tTAER system that is based on the multi-chimeric transactivator composed of a tetR polypeptide, as activation domain of VP16, and a ligand binding domain of an estrogen receptor (Yee et al., 2002, US 6,432,705).

Suitable promoters for nucleotide sequences encoding small RNAs for knock down of specific genes by RNA interference (see below) include, in addition to the above mentioned polymerase II promoters, polymerase III promoters. The RNA polymerase III (pol III) is responsible for the synthesis of a large variety of small nuclear and cytoplasmic non-coding RNAs including 5S, U6, adenovirus VA1, Vault, telomerase RNA, and tRNAs. The promoter structures of a large number of genes encoding these RNAs have been determined and it has been found that RNA pol III promoters fall into three types of structures (for a review see Geiduschek and Tocchini-Valentini, 1988 Annu. Rev. Biochem. 57: 873-914; Willis, 1993 Eur. J. Biochem. 212: 1-11; Hernandez, 2001, J. Biol. Chem. 276: 26733-36). Particularly suitable for expression of siRNAs are the type 3 of the RNA pol III promoters, whereby transcription is driven by cis-acting elements found only in the 5'-flanking region, i.e. upstream of the transcription start site. Upstream sequence elements include a traditional TATA box (Mattaj et al., 1988 Cell 55, 435-442), proximal sequence element and a distal sequence element (DSE; Gupta and Reddy, 1991 Nucleic Acids Res. 19, 2073-2075). Examples of genes under the control of the type 3 pol III promoter are U6 small nuclear RNA (U6 snRNA), 7SK, Y, MRP, H1 and telomerase RNA genes (see e.g. Myslinski et al., 2001, Nucl. Acids Res. 21: 2502-09).

A gene therapy vector may optionally comprise a second or one or more further nucleotide sequence coding for a second or further polypeptide. A second or further polypeptide may be a (selectable) marker polypeptide that allows for the identification, selection and/or screening for cells containing the expression construct. Suitable marker proteins for this purpose are e.g. the fluorescent protein GFP, and the selectable marker genes HSV thymidine kinase (for selection on HAT medium), bacterial hygromycin B phosphotransferase (for selection on hygromycin B), Tn5 aminoglycoside phosphotransferase (for selection on G418), and dihydrofolate reductase (DHFR) (for selection on methotrexate), CD20, the low affinity nerve growth factor gene. Sources for obtaining these marker genes and methods for their use are provided in Sambrook and Russel (2001) "Molecular Cloning: A Laboratory Manual (3rd edition), Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, New York.

Alternatively, a second or further nucleotide sequence may encode a polypeptide that provides for fail-safe mechanism that allows to cure a subject from the transgenic cells, if deemed necessary. Such a nucleotide sequence, often referred to as a suicide gene, encodes a polypeptide that is capable of converting a prodrug into a toxic substance that is capable of killing the transgenic cells in which the polypeptide is expressed. Suitable examples of such suicide genes include e.g. the *E.coli* cytosine deaminase gene or one of the thymidine kinase genes from Herpes Simplex Virus, Cytomegalovirus and Varicella-Zoster virus, in which case ganciclovir may be used as prodrug to kill the IL-10 transgenic cells in the subject (see e.g. Clair et al., 1987, Antimicrob. Agents Chemother. 31: 844-849).

A gene therapy vector is preferably formulated in a pharmaceutical composition comprising a suitable pharmaceutical carrier as defined below.

RNA interference

For knock down of expression of an altered TET2 protein of the invention, a gene therapy vector or other expression construct is used for the expression of a desired nucleotide sequence that preferably encodes an RNAi agent, i.e. an RNA molecule that is capable of RNA interference or that is part of an RNA molecule that is capable of RNA interference. Such RNA molecules are referred to as siRNA (short interfering RNA, including e.g. a short hairpin RNA). Alternatively, a siRNA molecule may directly, e.g. in a pharmaceutical composition that is administered within or in the neighborhood of a hematopoietic cell.

A desired nucleotide sequence comprises an antisense code DNA coding for the antisense RNA directed against a region of the target gene mRNA, and/or a sense code DNA coding for the sense RNA directed against the same region of the target gene mRNA. In a DNA construct of the invention, an antisense and sense code DNAs are operably linked to one or more promoters as herein defined above that are capable of expressing an antisense and sense RNAs, respectively. "siRNA" preferably means a small interfering RNA that is a short-length double-stranded RNA that are not toxic in mammalian cells (Elbashir et al., 2001, Nature 411: 494-98; Caplen et al., 2001, Proc. Natl. Acad. Sci. USA 98: 9742-47). The length is not necessarily limited to 21 to 23 nucleotides. There is no particular limitation in the length of siRNA as long as it does not show toxicity. "siRNAs" can be, e.g. at least 15, 18 or 21 nucleotides and up to 25,

30, 35 or 49 nucleotides long. Alternatively, the double-stranded RNA portion of a final transcription product of siRNA to be expressed can be, e.g. at least 15, 18 or 21 nucleotides and up to 25, 30, 35 or 49 nucleotides long.

"Antisense RNA" is preferably an RNA strand having a sequence complementary to a target gene mRNA, and thought to induce RNAi by binding to the target gene mRNA. "Sense RNA" has a sequence complementary to the antisense RNA, and annealed to its complementary antisense RNA to form siRNA. The term "target gene" in this context preferably refers to a gene whose expression is to be silenced due to siRNA to be expressed by the present system, and can be arbitrarily selected. As this target gene, for example, genes whose sequences are known but whose functions remain to be elucidated, and genes whose expressions are thought to be causative of diseases are preferably selected. A target gene may be one whose genome sequence has not been fully elucidated, as long as a partial sequence of mRNA of the gene having at least 15 nucleotides or more, which is a length capable of binding to one of the strands (antisense RNA strand) of siRNA, has been determined. Therefore, genes, expressed sequence tags (ESTs) and portions of mRNA, of which some sequence (preferably at least 15 nucleotides) has been elucidated, may be selected as the "target gene" even if their full length sequences have not been determined.

The double-stranded RNA portions of siRNAs in which two RNA strands pair up are not limited to the completely paired ones, and may contain nonpairing portions due to mismatch (the corresponding nucleotides are not complementary), bulge (lacking in the corresponding complementary nucleotide on one strand), and the like. A non-pairing portions can be contained to the extent that they do not interfere with siRNA formation. The "bulge" used herein preferably comprise 1 to 2 non-pairing nucleotides, and the double-stranded RNA region of siRNAs in which two RNA strands pair up contains preferably 1 to 7, more preferably 1 to 5 bulges. In addition, the "mismatch" used herein is preferably contained in the double-stranded RNA region of siRNAs in which two RNA strands pair up, preferably 1 to 7, more preferably 1 to 5, in number. In a preferable mismatch, one of the nucleotides is guanine, and the other is uracil. Such a mismatch is due to a mutation from C to T, G to A, or mixtures thereof in DNA coding for sense RNA, but not particularly limited to them. Furthermore, in the present invention, a double-stranded RNA region of siRNAs in which two RNA strands pair up may contain both bulge and mismatched, which sum up to, preferably 1 to 7, more

preferably 1 to 5 in number. Such non-pairing portions (mismatches or bulges, etc.) can suppress the below-described recombination between antisense and sense code DNAs and make the siRNA expression system as described below stable. Furthermore, although it is difficult to sequence stem loop DNA containing no non-pairing portion in the double-stranded RNA region of siRNAs in which two RNA strands pair up, the sequencing is enabled by introducing mismatches or bulges as described above. Moreover, siRNAs containing mismatches or bulges in the pairing double-stranded RNA region have the advantage of being stable in *E. coli* or animal cells.

The terminal structure of siRNA may be either blunt or cohesive (overhanging) as long as siRNA enables to silence the target gene expression due to its RNAi effect. The cohesive (overhanging) end structure is not limited only to the 3' overhang, and the 5' overhanging structure may be included as long as it is capable of inducing the RNAi effect. In addition, the number of overhanging nucleotide is not limited to the already reported 2 or 3, but can be any numbers as long as the overhang is capable of inducing the RNAi effect. For example, the overhang consists of 1 to 8, preferably 2 to 4 nucleotides. Herein, the total length of siRNA having cohesive end structure is expressed as the sum of the length of the paired double-stranded portion and that of a pair comprising overhanging single-strands at both ends. For example, in the case of 19 bp double-stranded RNA portion with 4 nucleotide overhangs at both ends, the total length is expressed as 23 bp. Furthermore, since this overhanging sequence has low specificity to a target gene, it is not necessarily complementary (antisense) or identical (sense) to the target gene sequence. Furthermore, as long as siRNA is able to maintain its gene silencing effect on the target gene, siRNA may contain a low molecular weight RNA (which may be a natural RNA molecule such as tRNA, rRNA or viral RNA, or an artificial RNA molecule), for example, in the overhanging portion at its one end.

In addition, the terminal structure of the "siRNA" is necessarily the cut off structure at both ends as described above, and may have a stem-loop structure in which ends of one side of double-stranded RNA are connected by a linker RNA (a "shRNA"). The length of the double-stranded RNA region (stem-loop portion) can be, e.g. at least 15, 18 or 21 nucleotides and up to 25, 30, 35 or 49 nucleotides long. Alternatively, the length of the double-stranded RNA region that is a final transcription product of siRNAs to be expressed is, e.g. at least 15, 18 or 21 nucleotides and up to 25, 30, 35 or 49 nucleotides long. Furthermore, there is no particular limitation in the length of the

linker as long as it has a length so as not to hinder the pairing of the stem portion. For example, for stable pairing of the stem portion and suppression of the recombination between DNAs coding for the portion, the linker portion may have a clover-leaf tRNA structure. Even though the linker has a length that hinders pairing of the stem portion, it is possible, for example, to construct the linker portion to include introns so that the introns are excised during processing of precursor RNA into mature RNA, thereby allowing pairing of the stem portion. In the case of a stem-loop siRNA, either end (head or tail) of RNA with no loop structure may have a low molecular weight RNA. As described above, this low molecular weight RNA may be a natural RNA molecule such as tRNA, rRNA, snRNA or viral RNA, or an artificial RNA molecule.

To express antisense and sense RNAs from the antisense and sense code DNAs respectively, a DNA construct of the present invention comprise a promoter as defined above. The number and the location of the promoter in the construct can in principle be arbitrarily selected as long as it is capable of expressing antisense and sense code DNAs. As a simple example of a DNA construct of the invention, a tandem expression system can be formed, in which a promoter is located upstream of both antisense and sense code DNAs. This tandem expression system is capable of producing siRNAs having the aforementioned cut off structure on both ends. In the stem-loop siRNA expression system (stem expression system), antisense and sense code DNAs are arranged in the opposite direction, and these DNAs are connected via a linker DNA to construct a unit. A promoter is linked to one side of this unit to construct a stem-loop siRNA expression system. Herein, there is no particular limitation in the length and sequence of the linker DNA, which may have any length and sequence as long as its sequence is not the termination sequence, and its length and sequence do not hinder the stem portion pairing during the mature RNA production as described above. As an example, DNA coding for the above-mentioned tRNA and such can be used as a linker DNA.

In both cases of tandem and stem-loop expression systems, the 5' end may be have a sequence capable of promoting the transcription from the promoter. More specifically, in the case of tandem siRNA, the efficiency of siRNA production may be improved by adding a sequence capable of promoting the transcription from the promoters at the 5' ends of antisense and sense code DNAs. In the case of stem-loop siRNA, such a sequence can be added at the 5' end of the above-described unit. A

transcript from such a sequence may be used in a state of being attached to siRNA as long as the target gene silencing by siRNA is not hindered. If this state hinders the gene silencing, it is preferable to perform trimming of the transcript using a trimming means (for example, ribozyme as are known in the art). It will be clear to the skilled person that an antisense and sense RNAs may be expressed in the same vector or in different vectors. To avoid the addition of excess sequences downstream of the sense and antisense RNAs, it is preferred to place a terminator of transcription at the 3' ends of the respective strands (strands coding for antisense and sense RNAs). The terminator may be a sequence of four or more consecutive adenine (A) nucleotides.

Antibodies

Some aspects of the invention concern the use of an antibody or antibody-fragment that specifically binds to an altered TET2 protein or a TET2 protein encoded by a mutated TET2 gene as defined above. Methods for generating an antibody or antibody-fragment that specifically binds to a given polypeptide are described in e.g. Harlow and Lane (1988, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY) and WO 91/19818; WO 91/18989; WO 92/01047; WO 92/06204; WO 92/18619; and US 6,420,113 and references cited therein. The term "specific binding," as used herein, includes both low and high affinity specific binding. Specific binding can be exhibited, e.g., by a low affinity antibody or antibody-fragment having a Kd of at least about 10^{-4} M. Specific binding also can be exhibited by a high affinity antibody or antibody-fragment, for example, an antibody or antibody-fragment having a Kd of at least about of 10^{-7} M, at least about 10^{-8} M, at least about 10^{-9} M, at least about 10^{-10} M, or can have a Kd of at least about 10^{-11} M or 10^{-12} M or greater. A preferred embodiment relates to an antibody directed to an altered TET2 protein or to a TET2 protein encoded by a mutated TET2 gene as defined herein, more preferably a human antibody, even more preferably a neutralizing anti-human antibody. A neutralizing antibody is preferably an antibody which is able to bind and to inactivate the action of an altered TET2 protein or a TET2 protein encoded by a mutated TET2 gene to at least some extent in a given assay.

Peptidomimetics

A peptide-like molecule (referred to as peptidomimetics) or non-peptide molecule that specifically binds to an altered TET2 protein or to a TET2 protein encoded by a mutated TET2 gene and that may be applied in a method of the invention as defined herein (for altering the activity or steady state level of said protein) as an agonist or antagonist of an altered TET2 protein or of a TET2 protein encoded by a mutated TET2 gene and may be identified using a method known in the art per se, as e.g. described in detail in US 6,180,084 which incorporated herein by reference. Such a methods includes e.g. screening libraries of peptidomimetics, peptides, DNA or cDNA expression libraries, combinatorial chemistry and, particularly useful, phage display libraries. These libraries may be screened for an agonists and/or an antagonist of an altered/non-altered TET2 protein by contacting the libraries with a substantially purified polypeptide of the invention, fragments thereof or structural analogues thereof.

Pharmaceutical compositions

The invention further relates to a pharmaceutical preparation or composition comprising as active ingredient an ingredient selected from the group consisting of: a TET2 protein, a non-altered TET2 protein, a control, wild type or functional TET2 protein, a nucleic acid, a nucleic acid construct, a gene therapy vector and an antibody. All these ingredients were already defined herein. A composition preferably at least comprises a pharmaceutically acceptable carrier in addition to the active ingredient.
In some methods, a non-altered TET2 protein as purified from mammalian, insect or microbial cell cultures, from milk of transgenic mammals or other source is administered in purified form together with a pharmaceutical carrier as a pharmaceutical composition. Methods of producing a pharmaceutical composition comprising a polypeptide are described in US Patents No.'s 5,789,543 and 6,207,718. The preferred form depends on the intended mode of administration and therapeutic application.

The pharmaceutical carrier can be any compatible, non-toxic substance suitable to deliver a protein, antibody or gene therapy vector to a patient. Sterile water, alcohol, fats, waxes, and inert solids may be used as the carrier. Pharmaceutically acceptable adjuvants, buffering agents, dispersing agents, and the like, may also be incorporated into a pharmaceutical composition.

The concentration of a protein or antibody of the invention in a pharmaceutical composition can vary widely, i.e., from less than about 0.1% by weight, usually being at least about 1% by weight to as much as 20% by weight or more.

For oral administration, an active ingredient can be administered in solid dosage forms, such as capsules, tablets, and powders, or in liquid dosage forms, such as elixirs, syrups, and suspensions. Active component(s) can be encapsulated in gelatin capsules together with inactive ingredients and powdered carriers, such as glucose, lactose, sucrose, mannitol, starch, cellulose or cellulose derivatives, magnesium stearate, stearic acid, sodium saccharin, talcum, magnesium carbonate and the like. Examples of additional inactive ingredients that may be added to provide desirable colour, taste, stability, buffering capacity, dispersion or other known desirable features are red iron oxide, silica gel, sodium lauryl sulfate, titanium dioxide, edible white ink and the like. Similar diluents can be used to make compressed tablets. Both tablets and capsules can be manufactured as sustained release products to provide for continuous release of medication over a period of hours. Compressed tablets can be sugar coated or film coated to mask any unpleasant taste and protect the tablet from the atmosphere, or enteric-coated for selective disintegration in the gastrointestinal tract. Liquid dosage forms for oral administration can contain colouring and flavouring to increase patient acceptance.

A protein, antibody or nucleic acid construct or gene therapy vector is preferably administered parentally or systemically. A protein, antibody, nucleic acid construct or vector for preparations must be sterile. Sterilisation is readily accomplished by filtration through sterile filtration membranes, prior to or following lyophilisation and reconstitution. One preferred route of administration is systemic, more preferably orally. Another preferred route is a parental route for administration of a protein, antibody nucleic acid construct or vector is in accord with known methods, e.g. injection or infusion by subcutaneous, intravenous, intraperitoneal, intramuscular, intraarterial, intralesional, intracranial, intrathecal, transdermal, nasal, buccal, rectal, or vaginal routes. More preferably, a route for administration is intravenous or subcutaneous. A protein, antibody nucleic acid construct or vector is administered continuously by infusion or by bolus injection. A typical composition for intravenous infusion could be made up to contain 10 to 50 ml of sterile 0.9% NaCl or 5% glucose optionally supplemented with a 20% albumin solution and 1 to 50 μg of the protein,

antibody nucleic acid construct or vector. A typical pharmaceutical composition for intramuscular injection would be made up to contain, for example, 1 - 10 ml of sterile buffered water and 1 to 100 μg of a polypeptide, antibody, nucleic acid construct or vector of the invention. Methods for preparing parenterally administrable compositions are well known in the art and described in more detail in various sources, including, for example, Remington's Pharmaceutical Science (15th ed., Mack Publishing, Easton, PA, 1980) (incorporated by reference in its entirety for all purposes).

For a therapeutic application, a pharmaceutical composition is administered to a subject suffering from a MDS or an AML in an amount sufficient to reduce the severity of symptoms and/or prevent or arrest further development of symptoms. An amount adequate to accomplish this is defined as a "therapeutically-" or "prophylactically-effective dose". Such effective dosages will depend on the severity of the condition and on the general state of the subject's health. In general, a therapeutically- or prophylactically-effective dose preferably is a dose, which is sufficient to reverse a symptoms, i.e. to restore or stimulate normal hematopoiesis and/or effectively decrease the number of malignant cells in the bone marrow and/or blood, thereby stimulating the generation of functional, mature blood cells in the peripheral blood to an acceptable level, preferably (close) to the average levels found in normal unaffected healthy subjects.

In a present method, a protein or antibody is usually administered at a dosage of about 1 μg/kg subject body weight or more per week to a subject. Often dosages are greater than 10 μg/kg per week. Dosage regimes can range from 10 μg/kg per week to at least 1 mg/kg per week. Typically dosage regimes are 10 μg/kg per week, 20 μg/kg per week, 30 μg/kg per week, 40 μg/kg week, 60 μg/kg week, 80 μg/kg per week and 120 μg/kg per week. In preferred regimes 10 μg/kg, 20 μg/kg or 40 μg/kg is administered once, twice or three times weekly. Treatment is preferably administered by parenteral route.

Microarrays

Another aspect of the invention relates to microarrays (or other high throughput screening devices) comprising a nucleic acid, protein or antibody as defined above. A microarray is a solid support or carrier containing one or more immobilised nucleic acid or protein fragments for analysing nucleic acid or amino acid sequences or

mixtures thereof (see e.g. WO 97/27317, WO 97/22720, WO 97/43450, EP 0 799 897, EP 0 785 280, WO 97/31256, WO 97/27317, WO 98/08083 and Zhu and Snyder, 2001, Curr. Opin. Chem. Biol. 5: 40-45). Microarrays comprising a nucleic acid may be applied e.g. in methods for analysing genotypes or expression patterns as indicated above. Microarrays comprising a protein may be used for detection of suitable candidates of substrates, ligands or other molecules interacting with a protein. Microarrays comprising an antibody may be used for in methods for analysing expression patterns of a polypeptide as indicated above.

General

In this document and in its claims, the verb "to comprise" and its conjugations is used in its non-limiting sense to mean that items following the word are included, but items not specifically mentioned are not excluded. In addition the verb “to consist” may be replaced by “to consist essentially of” meaning that a nucleotide sequence, a TET2 protein, a nucleic acid construct or a pharmaceutical composition as defined herein may comprise additional component(s) than the ones specifically identified, said additional component(s) not altering the unique characteristic of the invention. In addition, reference to an element by the indefinite article "a" or "an" does not exclude the possibility that more than one of the element is present, unless the context clearly requires that there be one and only one of the elements. The indefinite article "a" or "an" thus usually means "at least one".

Each of the embodiment as identified herein may be combined together unless otherwise indicated.

All patent and literature references cited in the present specification are hereby incorporated by reference in their entirety.

The following examples are offered for illustrative purposes only, and are not intended to limit the scope of the present invention in any way.

Examples

To better characterize the chromosomal regions that are implicated in MDS, several SNP-array studies have been performed. These have shown that this technology

correctly identifies the known cytogenetic regions, and provides additional information on potentially important chromosomal regions[15-18].

To identify genes that are affected in MDS we performed high resolution SNP-array analysis. DNA from neoplastic hematopoietic cells of 100 MDS patients was isolated and hybridized to SNP-arrays. As control, DNA from 231 healthy donors was used. Profiles of the patients and controls were analyzed for amplifications and deletions as well as copy neutral loss of heterozygosity (LOH) caused by mitotic recombination resulting in uniparental disomy (UPD). The overview of these results will be reported elsewhere (S. Langemeijer, manuscript in preparation). Six MDS patients showed aberrations on the long arm of chromosome 4. In two patients (UPN 7 and 44) this was caused by a deletion, and in 4 patients (UPN 52, 73,74 and 100) by copy-neutral LOH (Figure 1), which was not seen in any of the 231 controls. One patient (UPN 44) showed a 0.8 Mb deletion at 4q24 defining a minimal region of overlap between the affected patients. This region contained two genes, *PPA2* and *TET2*, a gene encoding a (hypothetical) protein with no known function, indicated in databases as *KIAA1546*, *TET2* or *FLJ20032*. Genomic sequencing of the coding region revealed no mutations in *PPA2*, but showed a nonsense mutation introducing a premature STOP codon in the remaining copy of the *TET2* gene of the same patient. Both the loss of the 4q24 allele (shown by SNP-array), as well as the nonsense mutation of the remaining allele (shown by genomic sequencing) were absent in non-neoplastic T-lymphocytes of the same patient, showing that these were acquired mutations. Therefore, we set out to investigate the occurrence of *TET2* mutations in MDS in more depth.

As ambiguity existed in databases with respect to splice variants of *TET2*, we analyzed the presence of various reported transcripts by reverse transcriptase-PCR and sequencing using RNA from NB4 cells (for primers see Supplemental Table 1). We confirmed that various splice variants are expressed potentially predicting three protein isoforms, the longest coding for a 2002 amino acid protein (Figure 2). Based on these transcripts, we designed primers (Supplemental Table 2) for genomic sequencing of the entire coding region of *TET2*. In addition splice donor-acceptor sites were analyzed. In 104 healthy controls, we found several SNPs that were predicted by various databases (HAPMAP, Ensembl), as well as several SNPs that were not reported previously (Supplemental Table 3).

In addition to the index patient (UPN44), the second patient (UPN7) that showed loss of one allele in the SNP-array analysis had a mutation of the remaining *TET2* allele (P1962L, Table 1). Furthermore, all four patients with uniparental disomy on the long arm of chromosome 4 carried a homozygous mutation of *TET2*. Thus, in all six patients with aberrations at 4q24 on SNP array analysis, defects in *TET2* were bi-allelic. Further sequencing of the complete cohort of MDS patients revealed mutations in *TET2* in 27/102 patients (Table 1). T cells were isolated to test whether the mutations were acquired. In all cases where T cells were available (10/27 patients), the *TET2* mutations that were found in the hematopoietic fraction, were not detectable in T cells. In addition to the six patients who showed aberrations at 4q24 on the SNP-array, there were nine who showed two different *TET2* mutations. Aditionally, in one patient (UPN 46), three different mutations were observed. As this patient did not show amplification of the 4q24 locus or trisomy of chromosome 4 in the SNP array analysis or on karyotype analysis, we hypothesize that in this patient, an initial clone carrying one *TET2* mutation might have arisen, from which later two subclones evolved carrying two different extra *TET2* mutations.

Nonsense and frameshift mutations leading to premature STOP codons were scattered across the *TET2* coding sequence, with preference for the N-terminal and middle part of the protein. Remarkably, all the missense mutations leading to amino acid substitutions and the three-base pair deletion leading to loss of one amino acid clustered in two distinct more C-terminal regions (Figure 3). These regions proved to represent highly conserved regions, when *TET2* orthologues from different species were aligned (Supplemental Figure 1). In addition, these two regions are conserved in the two human *TET2* paralogues *TET1* and *TET3* (Supplemental Figure 2), suggesting that they represent biologically important domains. The specific clustering pattern of the mutations suggests that truncation of the Tet2 protein or destruction of one of the two conserved domains is critical for the pathogenic effect of *TET2* in MDS.

The biological function of *TET2* remains obscure. The predicted proteins are large (224, 130 and 134 KD for isoforms 1, 2 and 3 respectively) but do not contain clear domains that allow functional classification. Interestingly, the *TET2* homologue *TET1* has been described as a fusion partner of *MLL* in a translocation identified in a case of AML, implicating this gene in oncogenesis[19]. However, also for *TET1*, no clear biological function has been described. Since in all patients carrying *TET2* mutations

expression of a C-terminally truncated *Tet2* protein and/or a mutated form of *Tet2* carrying a single amino acid substitution in one of the conserved regions can be predicted, this could be compatible with a dominant-negative or gain-of-function mutation. However, as in all patients at least one, and in many patients two alleles were affected, this might indicate that loss-of-function is the most probable mechanism by which these mutations contribute to the malignant transformation of the cells. To gain insight into the function of *TET2* we investigated the expression in various tissues and hematological cells (Figure 3A and 3B). High mRNA levels were present in hematological cells, particularly granulocytes (note that the scale in figures 3A and 3B is different). To further investigate whether *TET2* is differentially regulated during myelopoieisis, *TET2* expression was analyzed before and after induction of granulocytic differentiation by *all-trans* retinoic acid of the promyelocytic cell line NB4. *TET2* mRNA expression was enhanced (Figure 3C) during differentiation suggesting a role for *TET2* during hematopoietic development.

TET2 mutations occurred in all IPSS subtypes, but were significantly more frequent in the low (41%) and int-1 categories (27%) compared to the int-2 (13%) and high risk (14%) groups (Figure 5C). As the IPSS score is built up from the number of cytopenias, blast count and the cytogenetic subcategory, we analyzed which of these factors correlated with *TET2* mutation status. *TET2* mutations were significantly less common in patients with multiple cytopenias or high blast counts. The high incidence in the IPSS low and int-1 risk group is of particular interest, as patients with normal cytogenetics, normal blast counts and dysplasia in only one lineage are difficult to diagnose. Identification of *TET2* mutations in these patients may help to distinguish these patients from non-clonal hematological disorders.

METHODS

Patients and healthy controls

Bone marrow and blood from 102 MDS patients was collected after informed consent. Patients from all different WHO and IPSS categories and cytogenetic subgroups were included. Patient characteristics are shown in supplemental Table 4. DNA from FACS-sorted CD13+/33+ myeloid bone marrow cells, or Ficoll density gradient isolated mononuclear cells from peripheral blood was prepared. In addition, in a subset of patients DNA from T cells, that normally do not belong to the neoplastic clone in MDS were isolated by FACS-sorting for CD3+/CD19-/CD45+ cells. In case limited amounts of cells were available and the percentage of T cells was too low to allow the isolation of sufficient T cells by direct FACS, in vitro expansion of polyclonal T cells was performed. Cells were cultured in IMDM and 10% HSA for 7-14 days in medium containing IL-2 (100 IU/ml) and beads coated with anti-CD3 and CD28 (dynabeads, Invitrogen), providing vigorous stimulation of T cell proliferation by cross-linking of the T cell receptor. After culture, the purity of the T cells was measured using CD3 and CD45 surface markers. To ensure high purity FACS-sorting of the CD3 positive cells was performed. For the SNP arrays, DNA from 231 healthy donors was used, for the sequencing a different cohort of 104 healthy donors was used after informed consent. DNA was isolated using Qiagen spin-columns, the quality was checked on gel and by using a nanodrop ND-1000 spectrophotometer.

SNP array

250k SNP array hybridization was performed. 250 μg of total genomic DNA was digested using the Nsp I restriction enzyme and ligated to adaptors. A single primer that recognizes the adaptor sequence was used to amplify adaptor ligated DNA fragments. The amplified DNA was purified, digested with DNase I and labeled with biotin. Subsequently, samples were hybridized to the GeneChip Human mapping 250k array (Affymetrix). The arrays were washed and stained by streptavidin-phycoerythrin conjugates using the Affymetrix GeneChip Fluidics Station 450. Subsequently, the array was scanned with the Affymetrix GeneChip Scanner 3000 to obtain the fluorescence intensity data. Analysis of data, including quality control, was performed

with the Affymetrix GeneChip® Genotyping Analysis Software (GTYPE), which was designed to give highly accurate, automated SNP allele calls for the GeneChip Mapping Arrays. CNAG2.0 software (Copy Number Analyzer for Affymetrix GeneChip Mapping 100k arrays, Version 2.0 by Nannya et al, 2005) was used to calculate copy number variations by comparing the obtained SNP hybridization signal intensities in the patient samples with the SNP signal intensities in the reference set of healthy controls. Copy number values were plotted against the position of the SNPs in the genome. In addition, copy number changes were defined by visual inspection of the SNP hybridization signal intensities. All copy number aberrations found were compared to those observed in a group of 231 healthy unrelated individuals, as well as to known regions of normal variation (Redon, 2006; http://projects.tcag.ca/variation/) to exclude those regions that show normal copy number variation (CNV). Copy number neutral loss of heterozygosity (LOH) was defined as areas consisting of stretches of >70 sequentially homozygous SNPs. The regions thus defined were ordered based on size and location (intrachromosomal or telomeric).

FISH

Fluorescent in situ hybridization (FISH) was performed to confirm the mono-allelic loss of the 4q24 locus. DNA from BAC-clone XXX spanning the *TET2* gene was isolated, labeled and hybridized according to standard procedures.

Sequence analysis

Sequence analysis was performed on PCR-amplified genomic DNA fragments spanning the entire coding region of all three isoforms (for primer sequences, see supplemental Table 1). All intron-exon boundaries were included to identify possible splice site aberrations. To avoid PCR and sequencing errors, all products were sequenced in two directions and whenever sequence variations were observed, these were confirmed by an independent PCR and sequencing procedure starting from the original patient DNA. To analyze whether the observed mutations were acquired, DNA from T cells was isolated of patients showing sequence variations in *TET2*.

Quantitative PCR

To assess the mRNA expression of *TET2* in different hematopoietic cell fractions and tissues, specific PCR- primers and probes were designed that allow to distinguish between the three different isoforms of *TET2* (Supplemental Table 5). Quantitative PCR was performed by real-time PCR using an Applied Biosystems Taqman 7900HT machine. Quantities were normalized using GAPDH expression as a reference and calculated using the $2^{\Delta CT}$ method. The hematopoietic cell fractions were isolated from bone marrow and blood from healthy controls after informed consent Granulocytes (n = 4) were isolated after Ficoll density gradient centrifugation. The other cell subtypes were isolated by FACS sorting using monoclonal antibodies directed at CD4 and CD8 (T-cells, n = 3), CD14 (monocytes, n = 3), CD19 (B-cells, n = 3), CD71 (erythroid cells, n = 3), CD34 (progenitor cells, n = 3), CD56 (NK-cells, n = 3). NB4 cells were cultured and exposed to ATRA as described previously (n = 3) (ref). Pooled RNA from all other tissues was commercially obtained (Clontech).

Splice-variants

To confirm the *TET2* splice-variants, primers pairs were designed covering all exon-exon bounderies (for primer sequences, see supplemental Table 2). PCR was performed on cDNA followed by sequencing. All products were sequenced in two directions.

Statistical analysis

occurrence of *TET2* mutations in different patient subgroups was determined. P-values were calculated using two-tailed Fisher exact test. P-value <0.05 was considered statistically significant.

Accession codes: TET2 Isoform 1: NM_001127208 / UniRef100Q6N021. Isoform 2: NM_017628 / UniRef100Q6N021-2. Isoform 3: UniRef100_Q6N021-3

Table 1: Characteristics of *TET2* mutations.

In 27/102 MDS patients, aberrations of *TET2* were observed in hematopoietic cells. Uniparental disomy (UPD), amplifications and deletions at 4q24 were analyzed using SNP-arrays. Genomic sequencing of the protein coding region and splice donor and acceptor sites revealed nonsense, missense and insertion/deletion mutations leading to an amino acid deletion and substitution or to a frameshift resulting in a stopcodon after 2-55 aminoacids (indicated as fsX followed by the number of aminoacids[20]) . Sequence aberrations were classified as heterozygous, homozygous or, in case 4q24 deletions were present on one allele, hemizygous by analyzing the intensity of the base pair signal in the sequence reaction. Whenever possible, purified T cells were analyzed to assess whether the observed mutations were acquired. UPN indicates "unique patient number". N/A indicates "not analyzed" due to lack of sufficiently available T cells.

Each of the mutations as identified in table 1 is identified by reference to isoform 1 of the TET2 protein.

ID	Classification				SNP array		Mutation analysis by genomic sequencing				
UPN	LAB	WHO	IPSS	Karyotype	UPD at 4q24	del at 4q24	Nonsense	Missense	Indel	Homo/ Hetero/ hemizygous	Acquired/ inherited
7	RAEB-t	RAEB-2	int-2	complex	no	4q23-q35.2		P1962L		heterozygous*	N/A
9	RA	RCMD	low	normal	no	no			C1271CfsX28 S1424fsX23	heterozygous heterozygous	N/A
11	RARS	RCMD-RS	int-1	+8	no	no		H1881Q		heterozygous	N/A
36	RARS	RCMD-RS	low	normal	no	no			995fsX13	heterozygous	N/A
44	RAEB	RAEB-2	int-2	normal	no	4q24.23	Q383X			hemizygous	acquired
46	RA	RCMD	low	normal	no	no	E961X	 R1896S	 F1285del	heterozygous heterozygous heterozygous	acquired
50	RA	RCMD	low	normal	no	no	Q417X			heterozygous	acquired
52	RAEB	RAEB-1	int-1	del(9)(q22q32)	4q	no	S842X			homozygous	acquired
59	RAEB	RAEB-1	int-2	-7	no	no	R506X			heterozygous	acquired
72	RAEB	RAEB-1	int-1	normal	no	no	R544X	 C1875R		heterozygous heterozygous	acquired
73	RAEB	RAEB-1	int-1	+8	4q12-q35.2	no		W1291R		homozygous	acquired
74	RAEB-t	AML MLD	high	normal	4q23-q35.2	no			Q1170RfsX55	homozygous	N/A
75	RAEB	RAEB-1	int-1	normal	no	no	E368X S1848X			heterozygous heterozygous	acquired
77	RAEB	RAEB-1	int-1	normal	no	no		L1398R	 L1240LfsX2	heterozygous heterozygous	acquired
79	RARS	RCMD-RS	int-1	normal	no	no		G1913D		heterozygous	acquired
81	RA	RCMD	int-1	normal	no	no			Q574fsX5	heterozygous	N/A
83	RA	MDS-U	low	normal	no	no	C1271X		 T1220PfsX6	heterozygous heterozygous	N/A
86	RA	RA	low	normal	no	no			N607fsX30 K1493fsX66	heterozygous heterozygous	N/A
87	RA	RA	low	normal	no	no			847fsX24	heterozygous	N/A
96	RA	RCMD	low	normal	no	no	R1216X	 R1214W		heterozygous heterozygous	N/A
97	RA	RCMD	int-1	normal	no	no		C1396W		heterozygous	N/A

							I1873T		heterozygous	
98	RA	RA	low	normal	no	no		P555fsX33	heterozygous	N/A
100	RA	RCMD	low	normal	4q22.3-q35.2	no		H682EfsX10	homozygous	N/A
101	RA	RA	low	normal	no	no		K1911_L1916	heterozygous	N/A
102	RA	RCMD	low	normal	no	no	R1261L		heterozygous	N/A
103	RAEB	RAEB-1	int-1	normal	no	no		L1819fsX T164fsX5	heterozygous heterozygous	N/A
104	RAEB-t	AML	high	Complex**	No	no		S826fsX14	heterozygous	N/A

* Sequencing showed both the missense as well as the wildtype sequence (albeit with lower signal intensity. As the SNP array showed loss of one allele, the wild type sequence might have arisen from remaining non-clonal bone marrow cells.

**Including deletion of chromosome 4 in 2/39 cells analyzed by conventional karyotyping.

Table 2 : PCR primers used to identify splice variants

Primer	Forward/Reverse	Location
tgtaaaacgacggccagtACAGAAGGTGGGCCGGGGCGG	Forward	1A
tgtaaaacgacggccagtAGAACTCGGTCAATTTCCCAGTT	Forward	1A
tgtaaaacgacggccagtGGGAGAAACAGAACTCGGTC	Forward	1A
tgtaaaacgacggccagtGTCGGGTCTTTAAAAATACAGG	Forward	1A
TAAAGGGAGATAGAGACGCG	Forward	1A
TATTGATGCGGAGGCTAGGC	Forward	1B
CAAACATTCAGCAGCACACC	Forward	2
CCTTACTGCTCTTCCTGGATC	Reverse	2
CCTGGATCATGTCCTATTGGCT	Reverse	2
GACACCAACCAAAAGAACAGC	Forward	3A
TGACTAGACAAACCACTGCTGC	Forward	3A
GTGCGTTTTATTCCTCCATTTT	Reverse	3A
caggaaacagctatgaccGGAGCTTTGTAGCCAGAGGT	Reverse	3A
caggaaacagctatgaccTGTGCGTTTTATTCCTCCATTTT	Reverse	3A
CATGATTTCTCTAATAGCTGCCACA	Reverse	3C
CAATAGGACATCCCTGAGAACTT	Reverse	4
GCAGAAGCAGCAGTGAAGAGAA	Forward	5
CGCACCAAACACAGTAGCTT	Reverse	5

GGTCATCCCCAAGCAGCTTA	Reverse	6
CTGGCAAACTTACATCCATTGTAGTAC	Reverse	6
GAAACTGGAGTCTCATTTGCAAAAC	Forward	7
CAAGTTTCTTATATGTTGGTGCCATAA	Reverse	7
CCGTCTGGGTCTGAAGGAAG	Forward	8
TCTAGTTTCCTTTGTCGGCAAGTC	Reverse	9
GACTCTGTCTGAGGGTGATGTGG	Reverse	10

Table 3: sequencing primers

Primer	Forward/ reverse	Location
tgtaaaacgacggccagtCAGTTTGCTATGTCTAGGTATTCCGA	Forward	Intron 2/3A
caggaaacagctatgaccAGAAGGTTCACTAACTGTGCGTTTT	Reverse	Exon 3A
tgtaaaacgacggccagtTTCAACTAGAGGGCAGCCTTG	Forward	Exon 3A
caggaaacagctatgaccTGTGCGTTTTATTCCTCCATTTT	Reverse	Exon 3A
tgtaaaacgacggccagtCAGAATAGTCGTGTGAGTCCTGAC	Forward	Exon 3A
caggaaacagctatgaccGCAATGGAAACACAATCTGGA	Reverse	Exon 3A
tgtaaaacgacggccagtGAACACACACATGGTGAACTCC	Forward	Exon 3A
caggaaacagctatgaccAATTGTGATGGTGGTGGTGG	Reverse	Exon 3A
tgtaaaacgacggccagtTCCAGGGAACCACAAAGCTAG	Forward	Exon 3A
caggaaacagctatgaccGCTTGAGGTGTTCTGACATTGG	Reverse	Exon 3A
tgtaaaacgacggccagtACATGTATGCAGCCCTTCTCC	Forward	Exon 3A
caggaaacagctatgaccGGGAATCTGCTCTTTGTTGAAA	Reverse	Exon 3A
tgtaaaacgacggccagtACCAACATCTCCAGTTCCAA	Forward	Exon 3A
caggaaacagctatgaccATGCACTTGATTTCATGGTCT	Reverse	Exon 3A
tgtaaaacgacggccagtCAAATGGGACTGGAGGAAGT	Forward	Exon 3A
caggaaacagctatgaccGTTTGCTGCTGTTCTTGCTT	Reverse	Exon 3A
tgtaaaacgacggccagtCAGAAGGACACTCAAAAGCATG	Forward	Exon 3A
caggaaacagctatgaccTTGCTGCTCTAAAGCTGGG	Reverse	Exon 3A
tgtaaaacgacggccagtGAGAATCCACCTGCAAGCTG	Forward	Exon 3A
caggaaacagctatgaccTTTCACAAGACACAAGCATCG	Reverse	Exon 3A
tgtaaaacgacggccagtGGCCACATTTTCCTAATAGATCAGTCCA	Forward	Intron 3B/3C
caggaaacagctatgaccCTGCTTTGTGTGTGAAGGCTG	Reverse	Intron 3C/4
tgtaaaacgacggccagtTTCATTTCTCAGGATGTGGTCATAG	Forward	Intron 3C/4
caggaaacagctatgaccCCAATTCTCAGGGTCAGATTTACA	Reverse	Intron 4/5
tgtaaaacgacggccagtGTTGCCCTAATTGTGATCTAAACATG	Forward	Intron 4/5
caggaaacagctatgaccAGATTGGGCTTTCCTATCAGTGG	Reverse	Intron 5/6
tgtaaaacgacggccagtGGGTTCTACTTAACTGGGTATTTTCCA	Forward	Intron 5/6
caggaaacagctatgaccAGCTTACCAAGATATGTCATATTGTTCAC	Reverse	Intron 6/7
tgtaaaacgacggccagtGGATTCAAAATGTAAGGGAATAATC	Forward	Intron 6/7
caggaaacagctatgaccGCAAGTGGTTTCAACAATTAAGAGGA	Reverse	Intron 7/8
tgtaaaacgacggccagtCCATGTCAAGATATTTGCTCTATTTTGT	Forward	Intron 7/8
caggaaacagctatgaccCAGCCATGTGGAACTGTGAGTC	Reverse	Intron 8/9
tgtaaaacgacggccagtCCACCAACACAAATCTGAATACTGA	Forward	Intron 8/9
caggaaacagctatgaccAAGTTGATGGGGGCAAAACC	Reverse	Intron 9/10
tgtaaaacgacggccagtTCTTTGCTTAATGGGTGTCGTATATC	Forward	Intron 9/10
caggaaacagctatgaccTTGGGTATTCTGATTCAAAAGCC	Reverse	Exon 10
tgtaaaacgacggccagtCCAGCCCTATGAACTTCTATTCCA	Forward	Exon 10

caggaaacagctatgaccCGGAGCTGCACTGTAGTTATGG	Reverse	Exon 10
tgtaaaacgacggccagtCAAACATGGACTATAAAAATGGTGAAC	Forward	Exon 10
caggaaacagctatgaccAAGACGAGGGAGATCCTGGTG	Reverse	Exon 10
tgtaaaacgacggccagtGCCGTGGCTCCAACTCAT	Forward	Exon 10
caggaaacagctatgaccGTGACCTTTCCCCACTGCC	Reverse	Exon 10
tgtaaaacgacggccagt	Forward	Universal sequencing primer
caggaaacagctatgacc	Reverse	Universal sequencing primer

Table 4: SNPs in healthy controls vs MDS

SNP (protein)	Type of substitution	All MDS patients (n =102) n (%)	Healthy controls (n =104) n (%)	P-value 1	MDS patients with TET2 mutation (n =27) n (%)	P-value 2
p.I1762V	non-synonymous	67 (66)	59 (57)	0.20	18 (67)	1.0
p.L1721W	non-synonymous	18 (18)	24 (23)	0.39	5 (19)	1.0
p.M1701I	non-synonymous	3 (2.9)	1 (0.96)	0.37	2 (7.4)	0.17
p.H1778R	non-synonymous	4 (3.9)	5 (4.8)	1.0	1 (3.7)	1.0
p.K1171K	synonymous	1 (0.98)	2 (1.9)	1.0	0 (0)	1.0
p.L34F	non-synonymous	4 (3.9)	1 (0.96)	0.21	1 (3.7)	1.0
p.G355D	non-synonymous	12 (12)	12 (12)	1.0	3 (11)	1.0
p.P363L	non-synonymous	11 (11)	7(6.7)	0.33	2 (7.4)	0.72
p.P29R	non-synonymous	3 (2.9)	4 (3.8)	1.0	1 (3.7)	1.0
p.V218M	non-synonymous	4 (3.9)	4 (3.8)	1.0	1 (3.7)	1.0
p.Y867H	non-synonymous	1 (0.98)	3 (2.9)	0.62	0 (0)	1.0
p.P1723S	non-synonymous	1 (0.98)	3 (2.9)	0.62	0 (0)	1.0
p.H924R	non-synonymous	0 (0)	1 (0.96)	1.0	0 (0)	-
p.S1039S	synonymous	0 (0)	2 (1.9)	0.50	0 (0)	-
p.R123H	non-synonymous	0 (0)	1 (0.96)	1.0	0 (0)	-
p.G429R	non-synonymous	0 (0)	1 (0.96)	1.0	0 (0)	-
p.H949R	non-synonymous	0 (0)	1 (0.96)	1.0	0 (0)	-
p.Q1084P	non-	0 (0)	1 (0.96)	1.0	0 (0)	-

	synonymous					
p.H1380H	synonymous	0 (0)	2 (0.96)	0.50	0 (0)	-
p.S1791S	synonymous	0 (0)	1 (0.96)	1.0	0 (0)	-

P-value 1: Healthy controls compared to MDS patients

P-value 2: MDS patients with TET2 mutation compared to MDS patients with wtTET2

Each of the mutations as identified in table 4 is identified by reference to isoform 1 of the TET2 protein.

Table 5: patient characteristics

Characteristic	TET2 mutated n (%)	TET2 wild type n (%)
All patients	27 (26)	75 (74)
Sex ratio (Male/female) (p = 0.64)	2.4	1.7
Median age at diagnosis (p = 0.45)	67 yrs	62 yrs
FAB Classification		
- RA	13 (37)	22 (63)
- RARS	3 (18)	14 (82)
- RAEB	8 (27)	22 (73)
- RAEB-t	3 (15)	17 (85)
WHO classification		
- RA	2 (20)	8 (80)
- RCMD	10 (45)	12 (55)
- RARS	0 (0)	6 (100)
- RCMD-RS	3 (25)	9 (75)
- RAEB1	7 (41)	10 (59)
- RAEB2	2 (8.7)	21 (91)
- Isolated 5q-	0 (0)	1 (100)
- AML	2 (20)	8 (80)
- MDS-U	1 (100)	0 (0)
IPSS category		
- Low	12 (41)	17 (59)
- Int-1	10 (29)	25 (71)
- Int-2	3 (13)	21 (87)
- High	2 (14)	12 (86)
Cytogenetics		
- Normal	21 (32)	44 (68)
- 1-2 aberrations	4 (16)	21 (84)
- Complex	2 (17)	10 (83)
Cytogenetics by IPSS category		
- Good risk	21 (31)	46 (69)
- Intermediate	3 (17)	15 (83)
- Poor risk		

	3 (18)	14 (78)
Blast count in bone marrow		
- <5%	17 (31)	37 (69)
- 5-10%	7 (35)	13 (65)
- 11-19%	1 (5)	19 (95)
- 20-30%	2 (25)	6 (75)
Cytopenias		
- 0/1	17 (39)	27 (61)
- 2/3	10 (17)	48 (83)

Table 6: Q-PCR primers and probe

Primer	Forward/ reverse	Location
AATTTATTGGATACACCTGTCAAGACTC	Forward	Exon 3A
GATAAACGCCATGTGTCTCAGTACA	Reverse	Exon 3B
ACCTGCTCCTAGATGGGTATAAAAAG	Reverse	Exon 3C
TTTACCCTTCTGTCCAAACCTACAC	Reverse	Exon 3A/4
TATGATTTCCCATCTTGC	Probe	Exon 3A

Reference List

1. Nimer,S.D. Myelodysplastic syndromes. *Blood* **111**, 4841-4851 (2008).

2. Aul,C., Bowen,D.T., & Yoshida,Y. Pathogenesis, etiology and epidemiology of myelodysplastic syndromes. *Haematologica* **83**, 71-86 (1998).

3. Aul,C., Giagounidis,A., & Germing,U. Epidemiological features of myelodysplastic syndromes: results from regional cancer surveys and hospital-based statistics. *Int. J. Hematol.* **73**, 405-410 (2001).

4. Greenberg,P. *et al.* International scoring system for evaluating prognosis in myelodysplastic syndromes. *Blood* **89**, 2079-2088 (1997).

5. Greenberg,P.L., Sanz,G.F., & Sanz,M.A. Prognostic scoring systems for risk assessment in myelodysplastic syndromes. *Forum (Genova.)* **9**, 17-31 (1999).

6. Anderson,J.E. *et al.* Allogeneic bone marrow transplantation for 93 patients with myelodysplastic syndrome. *Blood* **82**, 677-681 (1993).

7. de Witte,T. *et al.* Allogeneic bone marrow transplantation for secondary leukaemia and myelodysplastic syndrome: a survey by the Leukaemia Working Party of the European Bone Marrow Transplantation Group (EBMTG). *Br. J. Haematol.* **74**, 151-155 (1990).

8. de Witte,T. *et al.* Intensive chemotherapy followed by allogeneic or autologous stem cell transplantation for patients with myelodysplastic syndromes (MDSs) and acute myeloid leukemia following MDS. *Blood* **98**, 2326-2331 (2001).

9. List,A. *et al.* Efficacy of lenalidomide in myelodysplastic syndromes. *N. Engl. J. Med.* **352**, 549-557 (2005).

10. List,A.F., Vardiman,J., Issa,J.P., & DeWitte,T.M. Myelodysplastic syndromes. *Hematology. Am. Soc. Hematol. Educ. Program.*297-317 (2004).

11. Pedersen-Bjergaard,J., Pedersen,M., Roulston,D., & Philip,P. Different genetic pathways in leukemogenesis for patients presenting with therapy-related myelodysplasia and therapy-related acute myeloid leukemia. *Blood* **86**, 3542-3552 (1995).

12. Haase,D. *et al.* New insights into the prognostic impact of the karyotype in MDS and correlation with subtypes: evidence from a core dataset of 2124 patients. *Blood* **110**, 4385-4395 (2007).

13. Ebert,B.L. *et al.* Identification of RPS14 as a 5q- syndrome gene by RNA interference screen. *Nature* **451**, 335-339 (2008).

14. Pellagatti,A. *et al.* Haploinsufficiency of RPS14 in 5q- syndrome is associated with deregulation of ribosomal- and translation-related genes. *Br. J. Haematol.* **142**, 57-64 (2008).

15. Gondek,L.P., Dunbar,A.J., Szpurka,H., McDevitt,M.A., & Maciejewski,J.P. SNP array karyotyping allows for the detection of uniparental disomy and cryptic chromosomal abnormalities in MDS/MPD-U and MPD. *PLoS. ONE.* **2**, e1225 (2007).

16. Gondek,L.P. *et al.* Chromosomal lesions and uniparental disomy detected by SNP arrays in MDS, MDS/MPD, and MDS-derived AML. *Blood* **111**, 1534-1542 (2008).

17. Mohamedali,A. *et al.* Prevalence and prognostic significance of allelic imbalance by single-nucleotide polymorphism analysis in low-risk myelodysplastic syndromes. *Blood* **110**, 3365-3373 (2007).

18. Wang,L. *et al.* Genome-wide analysis of copy number changes and loss of heterozygosity in myelodysplastic syndrome with del(5q) using high-density single nucleotide polymorphism arrays. *Haematologica* **93**, 994-1000 (2008).

19. Lorsbach,R.B. *et al.* TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). *Leukemia* **17**, 637-641 (2003).

20. den Dunnen,J.T. & Antonarakis,S.E. Nomenclature for the description of human sequence variations. *Hum. Genet.* **109**, 121-124 (2001).

Claims

1. A method for diagnosing a myelodysplastic syndrome (MDS) or an acute myeloid leukemia (AML) in a subject, the method comprising the step of identifying the presence of a mutation in an allele of the TET2 gene of said subject by comparing said TET2 gene to a control TET2 gene being represented by SEQ ID NO:1, SEQ ID NO:2 or SEQ ID NO:3 or an equivalent thereof.

2. A method according to claim 1, wherein the mutation in an allele of the TET2 gene is a deletion, an insertion, an amplification, a copy neutral loss of heterozygosity, a mono- or a bi-allelic nonsense, a missense and/or a frameshift.

3. A mutation according to claim 1 or 2, wherein the mutation is an acquired mutation.

4. A method according to any one of claims 1 to 3, wherein at least one mutation is found in each allele of the TET2 gene.

5. A method according to any one of claims 1 to 4, wherein the mutation is as described in table 1.

6. A method according to any one of claims 1 to 5, wherein the deletion or amplification is at 4q24.

7. A method according to any one of claims 1 to 6, wherein the mutation leads to a TET2 protein which is altered compared to the TET2 protein represented by SEQ ID NO:4, SEQ ID NO:5 or SEQ ID NO:6.

8. A method according to claim 7, wherein the altered TET2 protein has an amino acid deletion and/or substitution and/or insertion and/or is a truncated TET2 protein.

9. A method according to any one of claims 1 to 8, wherein the mutation is determined with respect to DNA, mRNA, and/or protein obtained from the subject and using direct and/or indirect methods.

10. A method according to any one of claims 1 to 9, wherein the mutation is determined *ex vivo* in a sample obtained from the subject.

11. A method for treating a MDS or an AML in a subject, the method comprising pharmacologically decreasing the activity or the steady-state level of a TET2 protein produced from a mutated TET2 gene and/or expressing a wild type or functional TET2 gene, preferably being a control TET2 gene or an equivalent thereof.

12. A method according to claim 11, wherein the pharmaceutical composition is administered to a hematopoietic cell of the subject to be treated.

13. A TET2 gene or a nucleic acid construct comprising said gene for use as a medicament for treating a MDS or an AML, preferably wherein the TET2 gene is represented by SEQ ID NO:1, SEQ ID NO:2 or SEQ ID NO:3 or an equivalent thereof.

14. Use of a TET2 gene or a nucleic acid construct comprising said gene as defined in claim 13 for the manufacture of a medicament for in a subject, preferably in a method as defined in claims 11 or 12.

Fig 1a

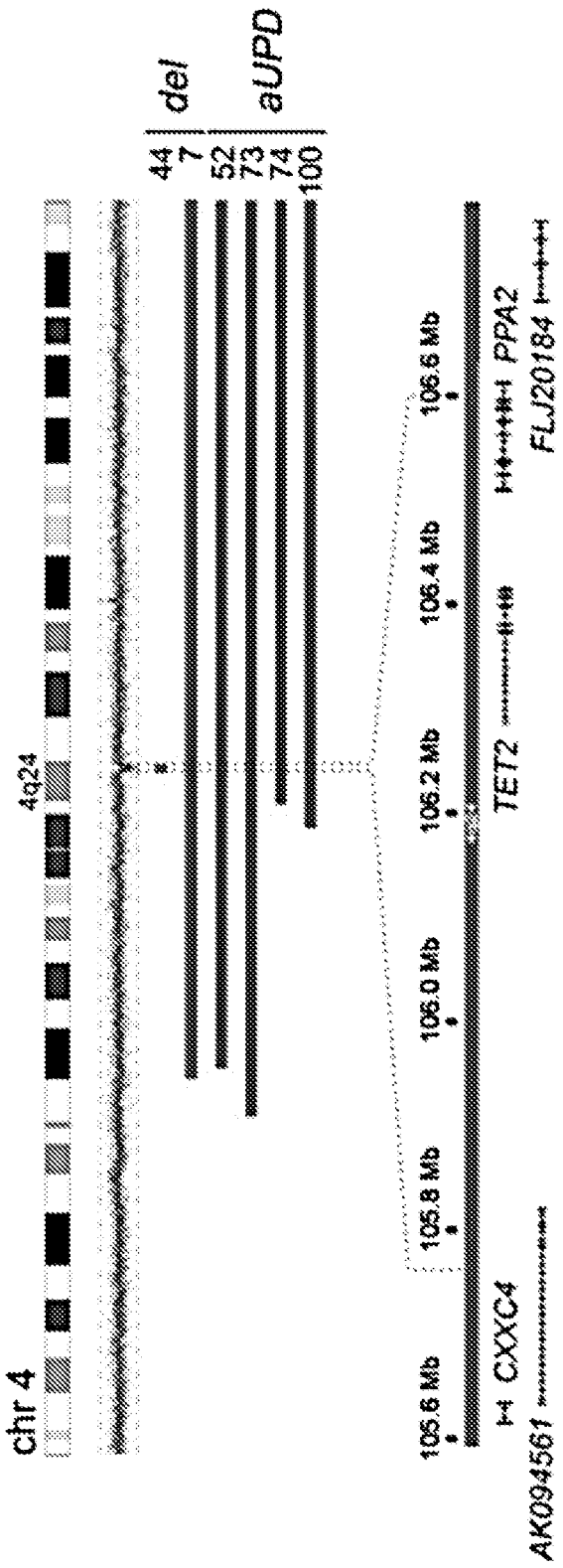

del
aUPD
44
7
52
73
74
100
chr 4
4q24
105.6 Mb
105.8 Mb
106.0 Mb
106.2 Mb
106.4 Mb
106.6 Mb
CXXC4
TET2
PPA2
FLJ20184
AK094561

Fig 1b

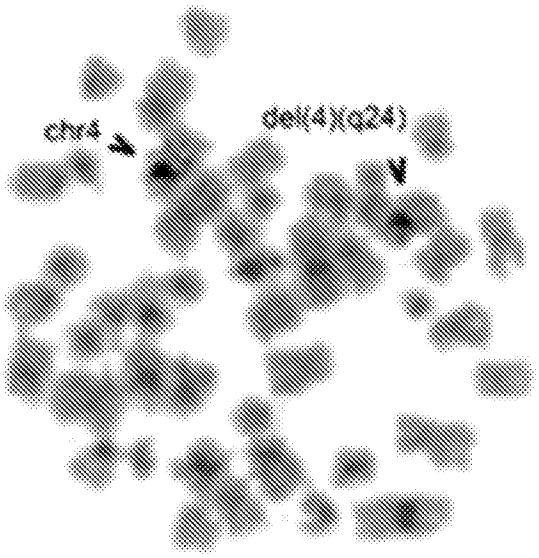

chr4
del(4)(q24)

Fig 2

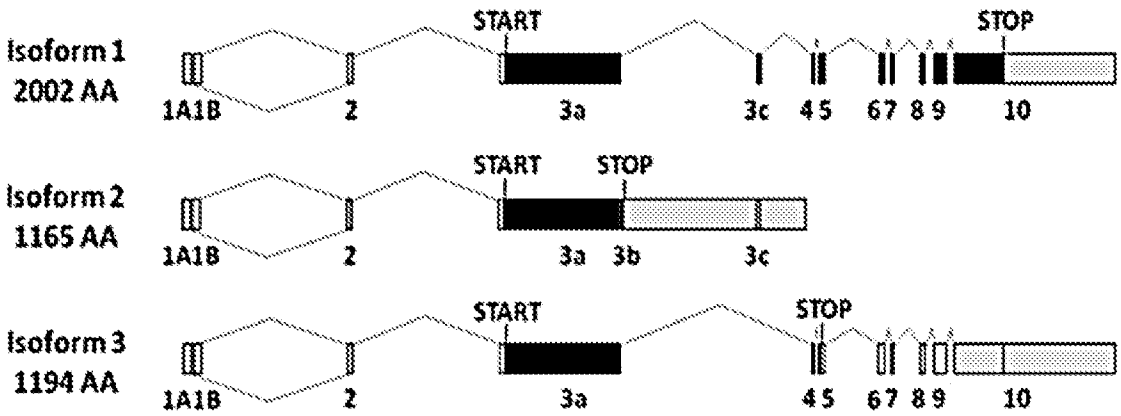

Isoform 1
2002 AA
START
STOP
1A1B
2
3a
3c
4 5
6 7
8 9
10
Isoform 2
1165 AA
START
STOP
1A1B
2
3a
3b
3c
Isoform 3
1194 AA
START
STOP
1A1B
2
3a
4 5
6 7
8 9
10

Fig 3

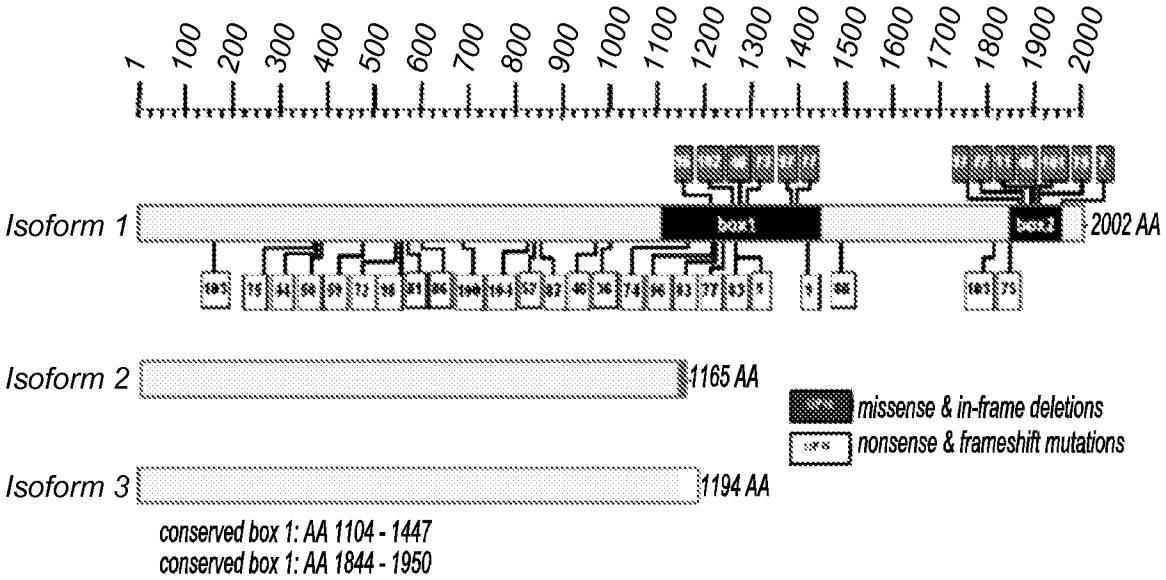

1
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
2000
Isoform 1
2002 AA
Isoform 2
1165 AA
Isoform 3
1194 AA
missense & in-frame deletions
nonsense & frameshift mutations
conserved box 1: AA 1104 - 1447
conserved box 1: AA 1844 - 1950

Fig 4a

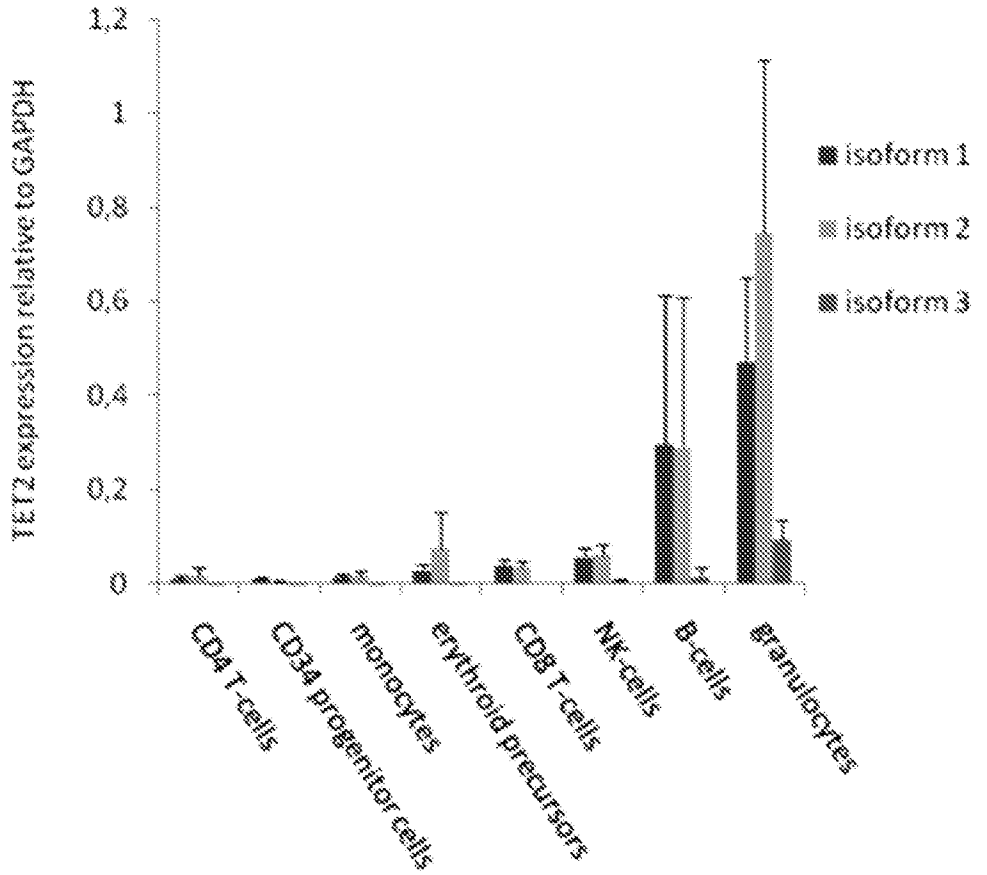

TET2 expression relative to GAPDH
1,2
1
0,8
0,6
0,4
0,2
0
isoform 1
isoform 2
isoform 3
CD4 T-cells
CD34 progenitor cells
monocytes
erythroid precursors
CD8 T-cells
NK-cells
B-cells
granulocytes

Fig 4b

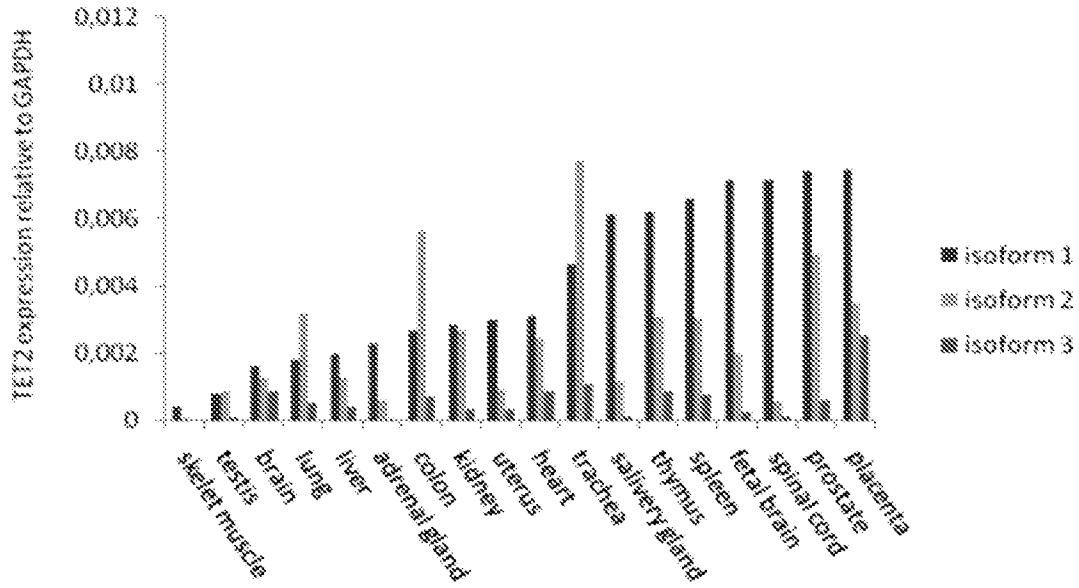

TET2 expression relative to GAPDH
0,012
0,01
0,008
0,006
0,004
0,002
0
skelet muscle
testis
brain
lung
liver
adrenal gland
colon
kidney
uterus
heart
trachea
salivery gland
thymus
spleen
fetal brain
spinal cord
prostate
placenta
isoform 1
isoform 2
isoform 3

Fig 4c

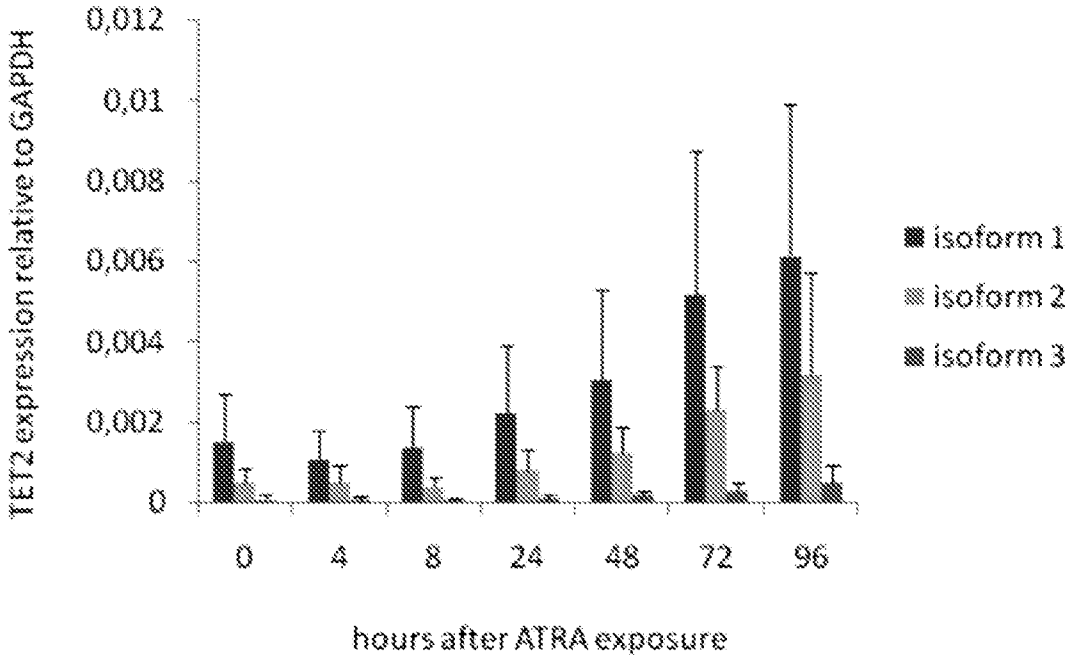
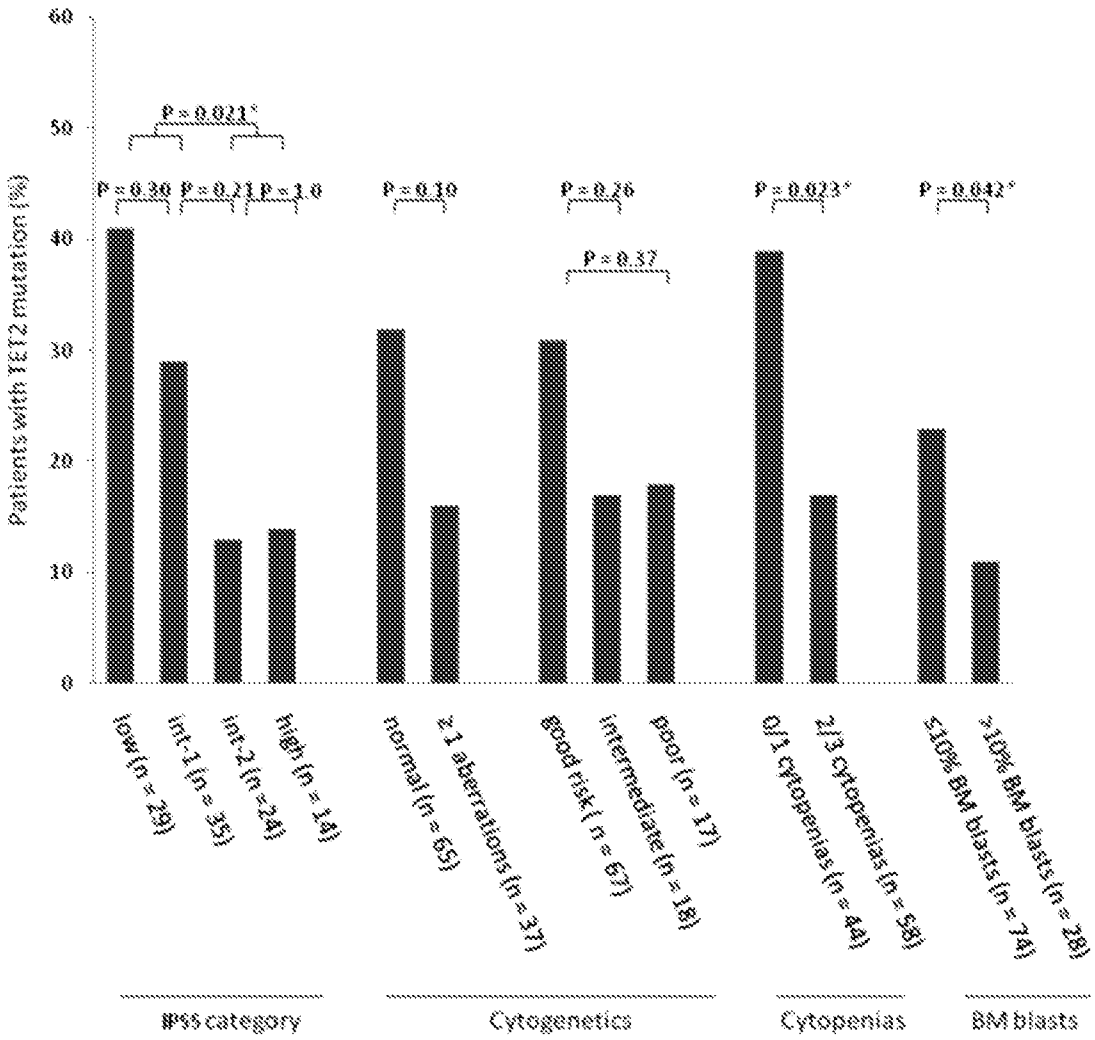

TET2 expression relative to GAPDH
0,012
0,01
0,008
0,006
0,004
0,002
0
0
4
8
24
48
72
96
hours after ATRA exposure
isoform 1
isoform 2
isoform 3

Fig 5


Patients with TET2 mutation (%)
60
50
40
30
20
10
0
P = 0.021*
P = 0.30
P = 0.71
P = 1.0
P = 0.10
P = 0.26
P = 0.37
P = 0.013*
P = 0.042*
low (n = 29)
int-1 (n = 35)
int-2 (n = 24)
high (n = 14)
normal (n = 65)
≥ 1 aberrations (n = 37)
good risk (n = 67)
intermediate (n = 18)
poor (n = 17)
0/1 cytopenias (n = 44)
2/3 cytopenias (n = 58)
≤10% BM blasts (n = 74)
>10% BM blasts (n = 28)
IPSS category
Cytogenetics
Cytopenias
BM blasts

INTERNATIONAL SEARCH REPORT

International application No
PCT/NL2010/050034

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, BIOSIS, WPI Data, MEDLINE, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DUPONT S. ET AL: "TET2 is a Novel Tumor Supressor Gene in Myeloproliferative Neoplasms: Identification of a Pre-JAK2 V617F Event" 50H ASH ANNUAL MEETING AND EXPOSITION, [Online] 6 December 2008 (2008-12-06), - 9 December 2008 (2008-12-09) XP002527300 San Francisco Retrieved from the Internet: URL:http://ash.confex.com/ash/2008/webprogram/Paper15853.html> [retrieved on 2008-12-09] the whole document ----- -/--	1-6,8-10

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
8 March 2010	12/03/2010

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Botz, Jürgen

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	TEFFERI, A. ET AL.: "Frequent TET2 mutations in systemic mastocytosis: clinical, KITD816V and FIP1L1-PDGFRA correlates" LEUKEMIA, vol. 23, no. 3, 5 March 2009 (2009-03-05), XP002527291 the whole document	1-14
X,P	TEFFERI, A. ET AL.: "TET2 mutations and their clinical correlates in polycythemia vera, essential thrombocythemia and myelofibrosis" LEUKEMIA, vol. 23, no. 3, 5 March 2009 (2009-03-05), pages 1-7, XP002527290 the whole document	1-14
A	VIGUIE F ET AL: "Common 4q24 deletion in four cases of hematopoietic malignancy: early stem cell involvement?" LEUKEMIA, vol. 19, no. 8, August 2005 (2005-08), pages 1411-1415, XP002527292 ISSN: 0887-6924 the whole document	1-14
A	LORSBACH ROBERT B ET AL: "TET1, a Member of a Novel Protein Family, Is Fused to MLL in Acute Myeloid Leukemia Containing the t(10;11)(q22;q23)." BLOOD, vol. 100, no. 11, 16 November 2002 (2002-11-16), page Abstract No. 2071, XP008105874 & 44TH ANNUAL MEETING OF THE AMERICAN SOCIETY OF HEMATOLOGY; PHILADELPHIA, PA, USA; DECEMBER 06-10, 2002 ISSN: 0006-4971 the whole document	1-14
A	DATABASE GENBANK [Online] 10 June 2008 (2008-06-10), STRAUSBERG, R.L.: "Homo sapiens tet oncogene family member 2, mRNA (cDNA clone IMAGE:40148086), complete cds" XP002527306 retrieved from NCBI Database accession no. BC150180 the whole document	1-14

-/--

INTERNATIONAL SEARCH REPORT

International application No
PCT/NL2010/050034

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
T	OLSEN ET AL.: "The implication of identifying JAK2V617F in myeloproliferative neoplasms and myelodysplastic syndromes with bone marrow fibrosis" JOURNAL OF HEMATOPATHOLOGY, vol. 1, no. 2, 2008, pages 111-117, XP002527331 the whole document	1-14

Form PCT/ISA/210 (continuation of second sheet) (April 2005)