From the Department of Medicine III, Grosshadern Hospital and
Helmholtz Zentrum München, Clinical Cooperative Group "Leukemia"
Ludwig-Maximilians-University, Munich
Chair: **Prof Dr. Wolfgang Hiddemann**

AML1-ETO COLLABORATES WITH THE HOMEOBOX GENE MEIS1 IN INDUCING ACUTE LEUKEMIA IN THE MOUSE BONE MARROW TRANSPLANTATION MODEL

Thesis Submitted for a Doctoral Degree in Human Biology
at the Faculty of Medicine Ludwig-Maximilians-University,
Munich,Germany

Submitted by

Mahalakshmi Naidu, Vegi

From

Anakapalli, India

2009

Aus der Medizinischen Klinik und Poliklinik III am Klinikum Großhadern
und Helmholtz Zentrum München, Klinische Kooperations-Gruppe "Leukämie"
der Ludwig-Maximilians-Universität München,
Vorstand: **Prof Dr. Wolfgang Hiddemann**

DIE GEMEINSAME EXPRESSION DES FUSIONSGENS AML1-ETO UND DES HOMEOBOX-KOFAKTORS MEIS1 INDUZIERT AKUTE LEUKAEMIEN IM MAUSTRANSPLANTATIONSMODELL

Dissertation zum Erwerb des Doktorgrades der Humanbiologie
an der Medizinischen Fakultät der Ludwig-Maximilians-
Universität zu München, Deutschland

Vorgelegt von
Mahalakshmi Naidu, Vegi

Aus
Anakapalli, India
2009

With permission from the Faculty of Medicine, University of Munich, Germany

Supervisor/Examiner: **PD Dr. med. Christian Buske**
Second Examiner:

Co-Examiners: Prof. Dr Ernst Schmid
                     Priv.Doz.Dr.Tim M. Stromm

Co-Supervisor:

Dean: **Prof. Dr. med. H.C.M. Reiser**
             FACR, FRCR

Date of Submission: 09.07.09

Date of Oral Examination: 13.02.09

Dedicated to my

Dear Parents……………………….

Contents

INTRODUCTION

Normal and Malignant Hematopoiesis

The human body produces a large number of various kinds of cells during the course of a lifetime. For instance, blood is composed of red blood cells, white blood cells and platelets. Blood cell production is a continuous process, keeping the body metabolism constant during stress or illness or trauma. This process of blood cell production and homeostasis is called *hematopoiesis* (Figure 1). In humans, this process begins in the yolk sac in the first weeks of embryonic development. Between 3rd and 7th month of gestation, stem cells migrate to the fetal liver and then to the spleen where these two organs play a major role in hematopoiesis. Later on, the bone marrow (BM) becomes the major hematopoietic organ and hematopoiesis ceases in the liver and spleen. Malfunctioning of normal hematopoietic development can lead to malignancies like myelo-proliferative disorders, leukemia, aplastic anaemia, lymphoma, myelodysplasia, and inborn errors of metabolism (Weissman et al., 2001). Leukemia results from the deregulation of the normal hematopoietic system due to the acquisition of mutations in hematopoietic progenitors and is characterized by the accumulation of immature blasts that fail to differentiate (Figure 2). Based on the natural course, leukemia can be subdivided into acute and chronic leukemia. Evidence shows that many pathways that are deregulated in cancer may also regulate normal stem cell development (Domen et al., 1998). Other signalling pathways associated with oncogenesis, such as the Notch, Sonic hedgehog (Shh) and Wnt signalling pathways are also involved in the regulation of normal hematopoietic cell self renewal(Taipale and Beachy, 2001).

Acute leukemia is a heterogeneous disease that occurs due to genetic alterations like translocations involving oncogenes and transcription factors, activation of signal transduction pathways and alterations of growth factor receptors. However, many *in vivo* models have postulated that the development of cancer is a stepwise process where somatic mutations give rise to a transformed clonal population. One of the most characterized leukemia types is acute myeloid leukemia (AML) which accounts for about 30% of all adult leukemias (Parkin, 2001; Parkin et al., 2001a; Parkin et al., 2001b).
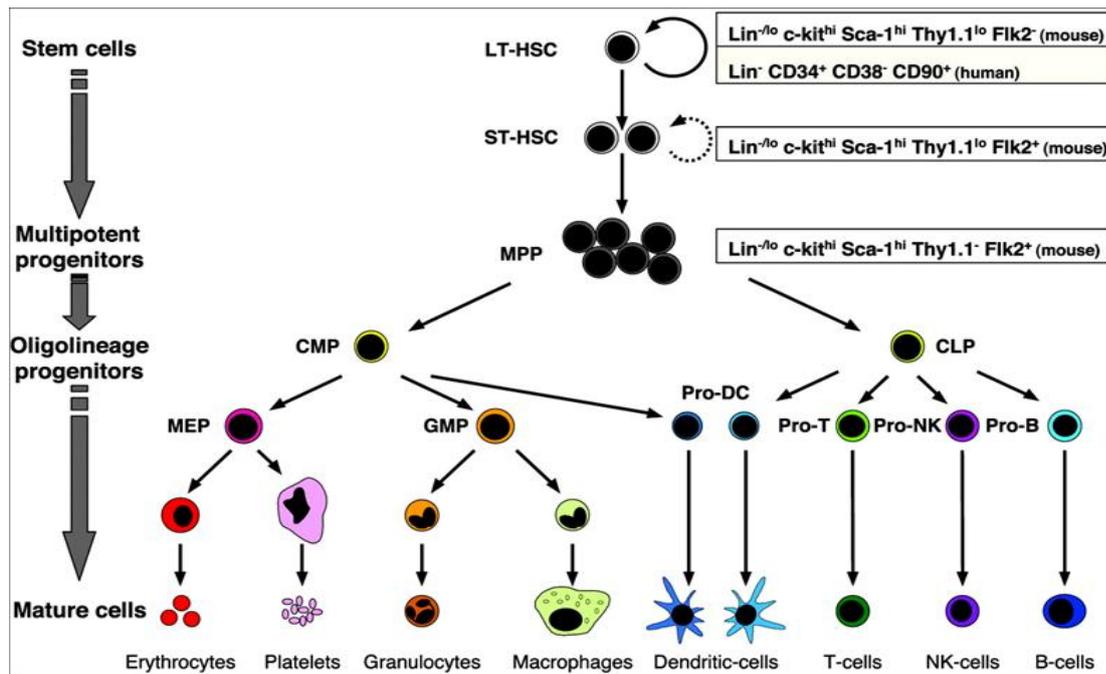
Figure 1: Hematopoietic and progenitor cell lineages:

The hematopoietic hierarchy consists of the hematopoietic stem cells (HSC), the multipotent progenitors (MPPs) and the more downstream progenitors, the common myeloid and the common lymphoid progenitor (CMP and CLP) respectively. Collectively, these give rise to all the mature cells of the hematopoietic lineage (Passegue et al., 2003).

AML

AML is characterized by the accumulation of large numbers of myeloid blasts arrested at varying stages of differentiation. AML cell populations are quite heterogeneous and the heterogeneity is due to the fact that malignant cells have divergent differentiation capacity and that most probably different target cells are transformed in different patients. In addition, for the leukemic clone to eventually become dominant, changes that confer a proliferative or self renewal advantage must occur. A common feature to all AML cases is an arrest in differentiation leading to an accumulation of more than 20% blast cells in the bone marrow (Gilliland and Tallman, 2002).

The leukemia stem cell concept

Every functional specialized mature blood cell is derived from a common blood cell termed the hematopoietic stem cell (HSC). In 1961, Till and McCulloch reported the existence of HSCs for the first time as a population of clonogenic bone marrow cells capable of generating myelo-erythroid colonies in spleen of lethally irradiated mice which could also be re-

transplanted into secondary recipients (Spangrude et al., 1988). Since then, a lot of progress has been made in the identification and functional, biochemical and immunophenotypic characterization of the HSC population. One of the defining characteristics of HSCs is the ability to self-renew as well as to differentiate into the multiple cell types that constitute the blood system, a process termed pluripotency. The self-renewal of HSCs can be symmetrical, thereby producing two daughter HSCs, or asymmetrical, resulting in one daughter HSC and one differentiated cell. As HSCs mature from the long-term self-renewing pool to MPPs, they progressively lose their potential to self-renew but become more mitotically active. Enriched human stem/progenitor cell populations show telomere shortening with age as do mouse LT-HSCs that undergo many divisions during serial transplantation (Allsopp et al., 2001). Normal hematopoietic development is critically dependent on a tightly regulated balance between their self renewal and differentiation properties. Perturbations in this balance can result in leukemia or other hematological malignancies (Warner et al., 2004) (Figure 2).
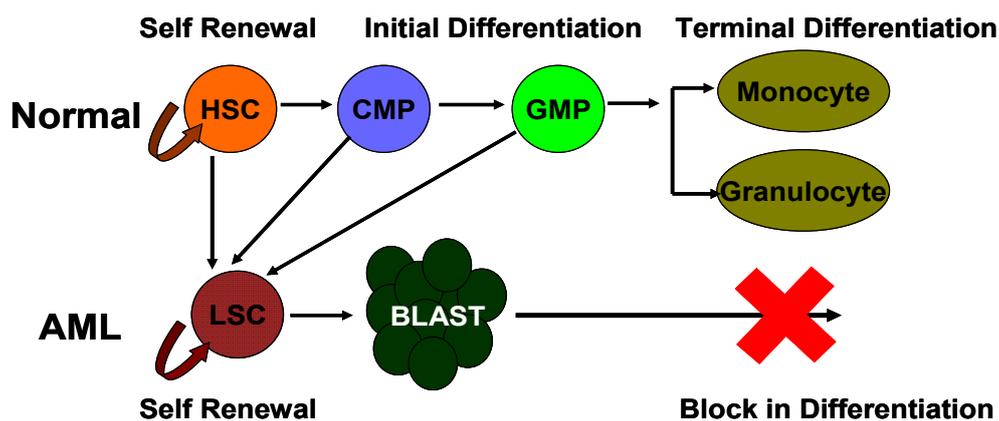
Figure 2: Schematic diagram illustrating the stages of hematopoiesis during normal and leukemic phases. In normal hematopoiesis the HSCs which have self renewal capacity initially differentiate into common myeloid progenitors and granulocyte-monocyte progenitors (CMP/GMP) finally producing monocytes and granulocytes whereas during hematopoietic malignancies like AML, leukemic stem cells (LSC) which have the same properties of HSCs like self renewal have a block in their differentiation leading to blast cell population. (Adapted from Rosenbauer et al., 2005).

There are two models of tumor propagation. One model, termed the stochastic model, assumes that all malignant cells give rise to daughter cells with identical tumorigenic properties. The other model, termed the cancer stem cell (CSC) model, proposes that not all cells within the tumor are malignant but only a defined subset of these neoplastic cells can give rise to the bulk tumor {Wang, 2005 #9165}. Recent studies have demonstrated that the CSC hypothesis holds true in several human tumors (Al-Hajj et al., 2003; Passegue et al.,

2003; Singh et al., 2003). Studies on leukemia have led the way in the characterization of the CSC. Since leukemic stem cells (LSC) share stem cell characteristics with normal HSCs, it is believed that both these cells are critically controlled by a set of genes like Wnt (Jamieson et al., 2004a; Jamieson et al., 2004b), JunB (Passegue et al., 2004), Bmi1 (Lessard and Sauvageau, 2003), Hox family (Antonchuk et al., 2002; Sauvageau et al., 1995), Notch family (Karanu et al., 2000; Varnum-Finney et al., 2000). The derivation of LSCs from HSCs depends on the transforming events due to which the stemness of these cells is preserved during transformation (Cozzio et al., 2003; Huntly and Gilliland, 2004). Since LSCs have several properties of HSCs like self renewal, proliferation and quiescence, these are the cells which are prone to transforming mutations and generation of leukemia. In contrast, although the common progenitors lose the property of self renewal, they could still be the targets for mutation through various other oncogenic pathways and re-gain self renewal property (Cozzio et al., 2003; Huntly and Gilliland, 2004). Mutations affecting the genetic stability of the stem pool could then induce a secondary mutation which also leads to disease progression. In this case HSCs would act as cells of origin of the tumor or pre-LSCs and then the LSCs would be the highly leukemogenic transformed progenitors. The identification of the cell of origin then becomes the most important aspect in several malignancies and designing treatment strategies that aim at eliminating these cells would be of primary importance.

Identification of LSCs

It has become possible to identify leukemic stem cells due to emerging techniques like long term *in vitro* culture assays, bone marrow transplantation models, and immunophenotyping assays. In addition to the regular karyotyping procedures, these assays are useful tools in assigning the leukemic properties. Several studies based on the above assays could show that AML LSCs have self renewal properties, a biological property of normal HSC, and also are characterized by multipotency, quiescence, and undergoing proliferation (Bonnet and Dick, 1997; Guan et al., 2003). They differ from normal cells by their deregulated proliferation. A rare population of $CD34^+CD38^-$ could induce leukemia in a SCID and NOD/SCID mouse model, indicating that these cells could be leukemia initiating cells, whereas the $CD38^+$ and Lin^+ fraction of cells, containing committed progenitors, did not generate leukemia. (Bhatia et al., 1997; Kondo et al., 2003; Terstappen et al., 1991) Further refining of immunophenotype showed that LSCs are $CD34^+CD38^-CD90^-IL-3R^+CD71^-HLA^-DR^-CD117^-$ (Blair et al., 1997; Jordan et al., 2000). Recent work using human CML samples revealed that LSCs can

originate from either normal HSCs or progenitor cells, depending on whether patients in chronic phase or blast crisis were analyzed (Blair et al., 1997; Jamieson et al., 2004a; Jamieson et al., 2004b). This was also demonstrated by different groups by targeting different AML specific oncogenes in highly purified progenitor cells, showing that downstream progenitors could also offer targets for oncogenic transformation in AML (Cozzio et al., 2003; Huntly and Gilliland, 2004, 2005) One way in which this could be possible is the initiation of a stem cell program in these transformed cells (Krivtsov and Armstrong, 2007). The generation of murine leukemia models allows the characterization of the mechanisms leading to transformation. In addition, it offers a valuable tool to identify and characterize the tumor propagating cells providing insights for the design of therapeutic strategies.

Mechanisms of transformation in AML

An AML is defined by the presence of more than 20 % leukemic blasts in the bone marrow (Gilliland and Tallman, 2002). According to the classification from the World Health Organization (WHO), molecular lesions are taken into account in classifying AML (Harris et al., 1999). Based on this, certain genetic and epigenetic alterations are associated with specific AML sub-types. Of the genetic alterations, chromosomal translocations, deletions, amplifications and inversions of chromosomal segments are common. In addition, mutations in several protein coding genes also have been reported in a significant proportion of AML such as the FLT3 and KIT tyrosine kinases and the nucleophosmin gene NPM1. Interestingly, more than 80% of myeloid leukemias have been attributed to, or associated with, one or more specific molecular lesions (Pandolfi, 2001). One of the conventional methods for identifying leukemia is karyotyping which enables the detection of some of the chromosomal deletions, inversions or translocations. Based on karyotyping, acute leukemias are classified as aberrant or non-aberrant karyotype. Aberrant karyotype constitutes 52% of all aberrations and comprises balanced (25%) and unbalanced (27%) chromosomal aberrations as illustrated in Table 1. t(8;21), inv(16) and t(15;17) are balanced karyotypes. Unbalanced aberrations include 5q-, 17q-, -5 and AML with complex karyotypes (Hiddemann et al., 2003). Normal karyotype includes NPM mutations (62%), FLT3 length mutations (35%), MLL-tandem duplications and CEBPα mutations (Table 1).

Translocations/inversions	Genes	Morphological association	Incidence
t(8;21)(q22;q22)	*AML1;ETO*	M2 with Auer rods	6%
inv(16)(p13q22)	*CBFB;MYH11*	M4Eo	7%
t(15;17)(q22;q11-21)	*PML;RARA*	M3/M3v	7%
t(9;11)(p22;q23)	*MLL;AF9*	M5	2%
t(6;11)(q27;q23)	*MLL;AF6*	M4 & M5	~1%
inv(3)(q21q26)	*EVI1;RPN1*	M1, M4, M6, M7?	~1%
t(6;9)(p23;q34)	*DEK;NUP214*	M2, M4	~1%
Chromosomal imbalances			
+8	…	M2, M4 & M5	9%
-7/7q-	…	No FAB preference	7%
-5/5q-	…	No FAB preference	7%
-17/17p-	*TP53*	No FAB preference	5%
-20/20q-	…	No FAB preference	3%
9q-	…	No FAB preference	3%
+22	…	M4, M4Eo	3%
+21	…	No FAB preference	2%
+13	…	M0, M1	2%
+11	*MLL1*	M1, M2	2%
Complex Karyotype			10%
Normal Karyotype			44%

Table 1: Common chromosomal aberrations in AML (Adapted from Estey and Dohner, 2006).

Chromosomal translocations

A detailed study on translocations has thrown light on the transcriptions factors involved in leukemia. Many of these rearrangements involve genes encoding transcription factors that have been shown to play an important role in hematopoietic lineage development. Thus, alteration of the transcriptional machinery appears to be a common mechanism leading to arrested differentiation (Pandolfi, 2001; Tenen, 2003) (Figure 3). One of the most common chromosomal translocations is the t(8;21) translocation which is present in 12% of AML subtype M2 (Speck and Gilliland, 2002). A detailed analysis of the t(8;21) translocation led to identification of the AML1 gene, a hematopoietic transcription factor on chromosome 21 (Miyoshi et al., 1991)and belonging to the core binding factor (CBF) family.
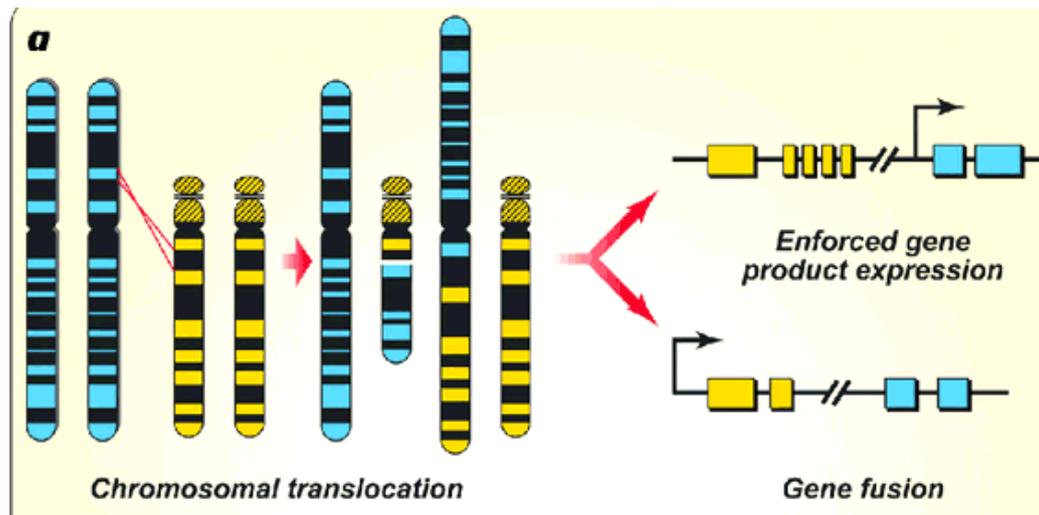
Figure 3: Schematic representation of chromosomal translocations.

Chromosomal translocations result in the exchange of chromosomal arms of same or different chromosomes (Rabbitts and Stocks, 2003).

The CBF family of transcription factors and AML1

The core binding factors (CBFs) are heterodimeric transcription factors which activate and repress transcription of key regulators of growth, survival and differentiation pathways. These are frequent targets of mutations and re-arrangements in human AMLs and ALLs. The CBF family consists of three distinct DNA binding CBFα units: RUNX1, RUNX2, RUNX3 and a common non DNA binding CBFβ subunit that is encoded by CBFB. RUNX1 or AML1 was the first mammalian CBF gene to be cloned. All RUNX proteins contain a runt homology DNA binding domain at the N-terminus which is highly homologous to the drosophila Runt protein which is involved in segmentation and sex determination (Romana et al., 1995). Runx1 (and by extension Cbfβ) is required for the differentiation of definitive hematopoietic progenitors and HSCs from a hemogenic endothelium in the mouse embryo (Miyoshi et al., 1991; Mukouyama et al., 2000). Besides the RUNT domain AML1 also contains a transactivation domain (Meyers et al., 1995) and a nuclear matrix attachment signal (NMTS) (Zeng et al., 1998). Mutations in the AML1 gene were shown to be associated with a number of malignant and premalignant conditions including acute myelogenous leukemia, childhood acute lymphocytic leukemia, familial platelet disorder, and myelodysplastic syndromes (Speck and Gilliland, 2002). AML1 is involved in many different chromosomal translocations, the most common ones being t(8;21)(q22;q22) (Downing et al., 1993; Erickson et al., 1992) and inv(16)(p13;q22) (Nucifora et al., 1993) which account for approximately 25% of adult AML. The t(12;21)(p13;q22) translocation is observed in 20–25% of pediatric

ALL (Liu et al., 1993). The AML1 gene generates three different spliced isoforms, AML1a, AML1b, and AML1c, where AML1a differs from AML1b and AML1c by the lack of C-terminus (Miyoshi et al., 1995).

ETO

ETO (also called MTG8 or CBFA2T1) is best known as the fusion partner of AML1 in leukemia carrying the t(8;21) translocation (Miyoshi et al., 1993). The ETO gene is located on chromosome 8q22. Earlier studies have revealed that ETO interacts with nuclear co-repressor proteins and have shown that these interactions enable it to play a critical role as transcriptional repressor by interacting with co-repressors like NCOR, SMRT, Sin3 and various other HDACs. It also acts as a negative regulator of AML1 transcriptional regulation (Gelmetti et al., 1998; Lutterbach et al., 1998; Wang et al., 1998)

AML1-ETO

AML1-ETO was first reported by Janet D. Rowley in a leukemic patient. It is associated with nearly 40% of cases of FAB-M2 AML. It is also observed in approximately 6% of AML M1 and, seldom present in AML M0, M4, M5, and other myeloproliferative syndromes (1990; Lai et al., 2005) (Figure 4). Cloning of the breakpoint regions of various chromosomal translocations has resulted in extensive studies on some transcription factors as fusion partners like AML1-ETO. The resulting fusion yields 177 amino acids (a.a) of AML1 with its N-terminal region containing the Runt domain (RHD) and 575 amino acids of the entire reading frame of ETO (Figure 5). Due to its similarities with drosophila nervy proteins, ETO has four domains named nervy homology domains (NHR1-4). It has 50% to 70% sequence homology with the drosophila homologue. The NHR1 domain is also known as TAF domain and resembles the TATA binding associated factors in humans as well as drosophila (TAF110) (Erickson, 1994) which indicates its role as a transcription factor. NHR2 is known as 'Hydrophobic Heptad Repeats' (HHR) essential for hetero- and homodimerizations (Gelmetti, 1998). NHR3 contains predicted coiled-coil structure (Minucci et al., 2000) and NHR4 myeloid-Nervy-DEAF1 homology domain (MYND) with two predicted zinc-finger motifs which are involved in protein–protein interaction ((Erickson et al., 1994; Gross and McGinnis, 1996). Moreover, structure–function analyses have revealed the NHR2 and NHR4 domains of ETO to be crucial for the activity of AML1–ETO in cellular assays, suggesting a

possible role for N-CoR or SMRT in repression by the oncoprotein (Gelmetti et al., 1998; Lutterbach et al., 1998).
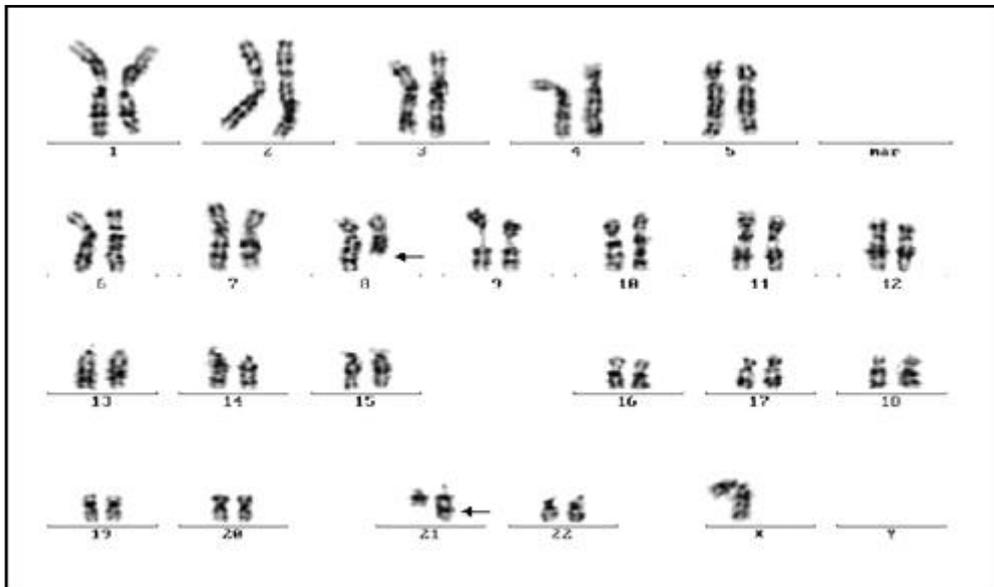
Figure 4: Karyogram of a t(8;21) translocation.

(Source: http://knm1.ibe.med.uni-muenchen.de/tumorzytogenetik/index.html)
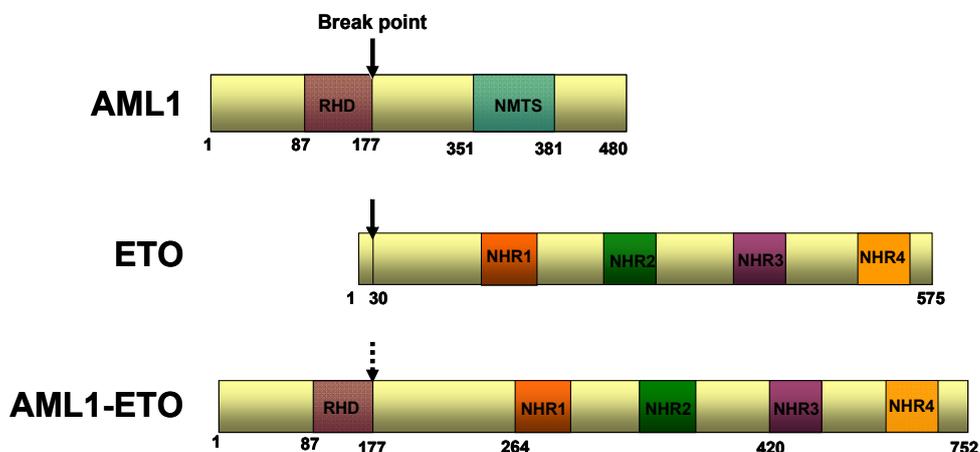
Figure 5: Domain structure of AML1, ETO and Fusion protein AML1-ETO.

Schematic diagram of the structure of AML1, ETO and the fusion gene AML1-ETO with functional domains involved in transcriptional activity. The AML1 gene has a distal VWRPY a.a site which is deleted when it fuses with ETO. The fusion gene forms a 754 a.a protein which has Runt domain at its N-terminal region on AML1 and ZNF domain at the C-terminal region

Two-hit model of leukemogenesis

A lot of the commonly occurring leukemia-associated fusion genes have been shown to be insufficient for transformation. In human leukemia, there are numerous cases in which a chromosomal translocation, co-expressed with an activating mutation or with an aberrant expression of proto-oncogenes, is detected. These observations favour a pathogenic model of the AML, in which the interaction of at least two different groups of genetic alterations are necessary for disease development (Gilliland, 2002) (Figure 6). This concept is supported by experimental data, which show that the fusion gene AML1-ETO alone is not sufficient, but can cooperate with unknown additional genetic alterations in order to induce leukemia. In a conditionally expressing AML1-ETO mouse model, only mice which had been treated additionally with the mutagen ENU developed AML, while the non treated group showed only minimal hematopoietic abnormalities (Higuchi et al., 2002). A very similar observation was reported with an hMRP8-AML1-ETO transgenic mouse model and a murine retroviral AML1-ETO model (de Guzman et al., 2002; Yuan et al., 2001). AML1-ETO co-expressed with tyrosine kinase FLT3-LM (Schessl, 2005) or Wilms tumour (WT1), a proto-oncogene could induce full blown leukemia (Nishida et al., 2006) in murine bone marrow transplantation models. Similarly, the TEL/PDGFRβ fusion gene cooperates with AML1/ETO in inducing AML in mice (Grisolano et al., 2003). These data clearly show that additional cooperating mutations are crucial for the pathogenesis of one the most frequent sub-types of AML.The translocation t(15;17) PML-RARA, commonly found in acute pro-myelocytic leukemias, is known to co-operate with BCL2 (Wuchter et al., 1999) or with activating FLT3 mutations (Kelly et al., 2002; Reilly, 2002) in inducing leukemia. Similarly, deregulation of Hox genes also results in leukemogenesis and Hox co-factor Meis1 is believed to be involved in accelerating leukemogenesis like in HoxA9 mediated leukemia (Alcalay et al., 2005). Meis1 also upregulates FLT3 in AML models of HOXA9 and NUP98-HOX (Palmqvist et al., 2006; Wang et al., 2005). The characterization of cooperating mutations is not only important for the understanding of the pathogenesis of these frequently occurring subtypes of leukemia, but also important for the development of purposeful therapies, since eradication of the leukemic clone can be attained only by a combination of innovative therapies and by antagonizing the oncogenic power of multiple genetic alterations. In this project we sought to analyze whether AML1-ETO can also collaborate with a HOX gene co-factor to induce leukemia in a murine model. The latter part of the project deals with the identification of leukemic stem cells in our murine leukemia model.
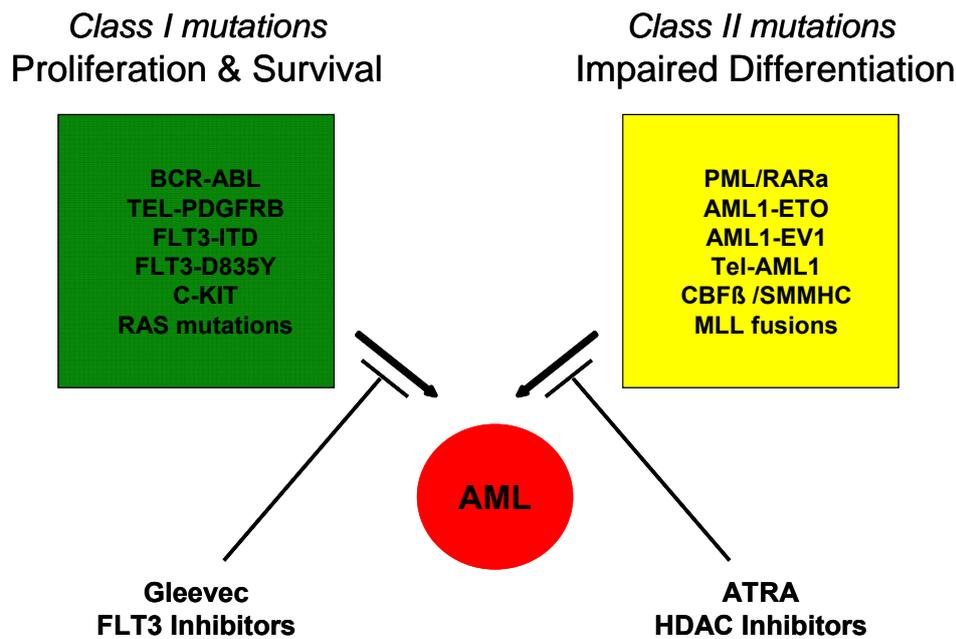
Figure 6: Diagram showing class I mutations and Class II mutations.

The Class I mutations which are involved in proliferation and Class II mutations which result in impaired differentiation cooperate with each other in inducing leukemia (Speck and Gilliland, 2002).

Leukemogenicity of AML1-ETO

Several lines of evidence suggest that translocation of AML1-ETO is alone not sufficient to induce leukemia. Murine models in AML1-ETO, FLT3-LM, WT1, and ICSBP deficient mice showed that mice transplanted with AML1-ETO did not succumb to leukemia (de Guzman et al., 2002; Nishida et al., 2006; Schessl et al., 2005; Schwieger et al., 2002). Recently it has been shown in a mouse model that a truncated form caused by a frame shift mutation that leads to an additional exon resulting in deletion of C-terminal region of AML1-ETO protein, is able to induce AML on its own (Yan et al., 2006). In another case, a 50-bp frame shift deletion in exon 2 of ETO was reported in a patient with M2 leukemia carrying t(8;21)(q22;q22). This deletion leads to the disruption of the open reading frame and the formation of a 31-residue truncated protein, having only 7 a.a residues in common with ETO of a full length fusion transcript (Lasa et al., 2002). The C-terminal region of ETO contains the MYND domain, comprising highly conserved zinc-finger-like protein motifs, and interacting with co-repressor proteins. It has been found that, instead of the MYND domain, an alternative last exon of MTG8 encoding 27 amino acids in-frame is expressed naturally in human adult testis and in several leukemia cell lines (Kozu et al., 2005). This type of

alternative splicing also occurred in the AML1-MTG8 fusion gene at high levels in leukemia cell lines with t(8;21) as well as in blast cells of leukemia patients with t(8;21).

Hox genes and Hox co-factors

The clustered *Hox* families of homeobox genes are evolutionarily highly conserved genes that act as key regulators of positional identity along the anterior–posterior body axis of animal embryos (Krumlauf, 1994). These genes contain a 61 a.a. helix-turn helix DNA binding domain which has been identified to play an important role in segmental patterning during embryogenesis as well as cell proliferation and differentiation during normal hematopoiesis (Thorsteinsdottir et al., 1997) and their deregulation leads to severe hematological malignancies. Hox genes belong to a family of transcription factors that contains 39 members clustered on four chromosomes (Krumlauf, 1994; McGinnis and Krumlauf, 1992). These Hox genes play a crucial role in pattern formation and tissue identity during embryogenesis. Homeobox (HB) genes are classified into 2 classes: class I includes clustered HB (HOX) genes, recognized for their role in anterio-posterior patterning during embryogenesis, while the class II divergent HB (non-HOX) genes are dispersed throughout the genome. These genes are believed to have arisen from gene duplication and have 13 paragroups which are further organized into 4 clusters (A-D) (Owens and Hawley, 2002). In normal hematopoiesis HOX genes of groups A and B are expressed in $CD34^+$ cells and are down regulated when the $CD34^+$ cells progress towards committed erythroid and myeloid progenitors (Sauvageau et al., 1994). HoxB4 is one of the important Hox genes which play a defined role in hematopoiesis. It is expressed in both primitive human and murine hematopoietic cells and it has been shown that overexpression of HoxB4 promotes HSC expansion both *in vivo* and *in vitro* without inducing leukemia (Antonchuk et al., 2002; Sauvageau et al., 1995). A recent report suggests that overexpression of HOXB4 induced leukemia in canines and macaques whereas those transplanted with empty vector did not show any signs of disease (Zhang et al., 2008). Several experiments performed by overexpressing Hox genes like Hoxa9, Hoxa10, Hoxb3, Hoxb6 or Hoxb8 showed growth advantage of retrovirally targeted cells and induction of long latency leukemia in transplanted mice (Fischbach et al., 2005; Kroon et al., 1998; Sauvageau et al., 1997; Thorsteinsdottir et al., 1997). These results strongly indicate that deregulated Hox genes can induce leukemic transformation.

HOX co-factors are itself homeobox genes belonging to the so called TALE (Three Amino Acid Extension Loop) family. These co-factors differ from HOX proteins in possessing an additional three amino acids between α motifs 1 and 2. HOX genes generally bind to Hox co-factors like PBX, Meis1 and PREP1/KNOX1 proteins in order to increase the specificity of the target sequence and for additional stability. In many experimental models co-expression of these TALE homeobox genes with HOX genes rapidly induces leukemia, whereas overexpression of the TALE members alone does not lead to any disease. In the following section Meis1 will be described in detail.

Myeloid Ectopic Integration Site 1 (Meis1)

Hox co-factors of the three amino acid loop class have been shown to interact directly with Hox proteins due to DNA binding affinity (Moens and Selleri, 2006) and are related to the PBX family (Burglin, 1997). In general, Hox proteins from paralogous groups 1–10 interact physically with PBX1, while those from paralogous groups 9–13 interact with MEIS1 (Shen et al., 1997)). At the N-terminal region, Meis1 has two α-helicase motifs which act as PBX biding sites and homeodomain at the c-terminal region (figure 12). Meis1 and PBX1 interact through these Meinox domains M1 and M2 (Mann and Affolter, 1998). Meis1 proteins co-operatively bind DNA with ABD-B class HOX proteins of groups 9 and 10 (Shanmugam et al., 1999).

It has been shown that Meis1 is required for establishing definitive HSCs in the embryo and is expressed in the aorta-gonad-mesonephros (AGM) mesenchyme in the hemogenic embryonic arterial endothelium and it is relevant to the Runx1 expressing populations (Azcoitia et al., 2005). As in the case of Hox genes, the expression level of *Meis1* in hematopoietic system corresponds to the differentiation stage. *Meis1* is highly expressed in HSCs and the expression levels decreases in progenitors and committed lineages (Pineault et al., 2002).

Meis1 was first observed as a transcript located at a site of common retroviral integrations in BXH2 mice (Moskow et al., 1995) and has been reported to be non-leukemogenic when expressed alone (Kroon et al., 1998) but to show transforming activity when fused to the transactivating domain of VP16 (Mamo et al., 2006). *Meis1* along with HOXA9 and HOXA7 is expressed in a vast majority of AML (Golub et al., 1999). In murine bone marrow transplantation models, Meis1 synergizes with different Hox genes and accelerates the onset

of AML. For example Meis1 collaborates with NUP98-HOXA9 (Kroon et al., 2001), NUP98-HOXD13 (Pineault et al., 2003) or HOXB6 (Fischbach et al., 2005) in accelerating AML.

In human leukemogenesis, MEIS1 has been identified as a putative collaborative gene in acute lymphocytic leukemias with the t(4;11) abnormality, where both *MEIS1* and *HoxA9* are upregulated (Imamura et al., 2002; Rozovskaia et al., 2001). Meis1 forms the most frequent cooperating protein with HoxA9 in inducing leukemia with a poor prognosis (Golub et al., 1999).
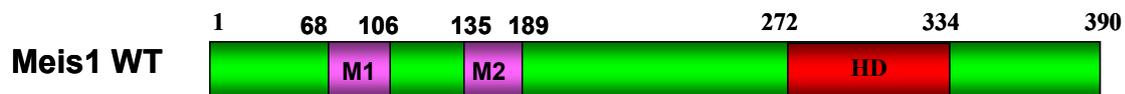
Figure 7: Domain structure of Meis1.
Meis1 is 390 amino acid protein with alpha helicase motifs at the N-terminal end and homeodomain region at the C-terminal region.

MATERIALS

Chemicals and Reagents

Agarose	(Sigma)
Acetic acid	(Sigma)
Calcium chloride	(Sigma)
Chloroform	(Sigma)
Dextran sulphate	(Sigma)
DNAzol	(GIBCO/BRL)
EDTA	(Sigma)
Ethidium bromide	(Sigma)
Formamide	(Sigma)
Fetal bovine serum	(PAN)
Glycine	(Sigma)
Giemsa	(Merck)
Hydrochloric acid	(Merck)
Isopropanol	(Merck)
Methocult M 3434	(Stemcell Technologies)
May-Grunwald's Eosin	(Merck)
Protamine sulfate	(Sigma-Aldrich)
Quikchange XL site-directed Mutagenesis Kit	(Stratagene)
Sodium chloride	(Sigma)
Sodium hydroxide	(Sigma)
Sodium dodecyl sulphate	(Sigma)
Skimmed milk	(Sigma)
Trizma base	(Sigma)
Trizol	(GIBCO/BRL)
5-Fluorouracil	(Medac)
ThermoScript RT-PCR Kit	(Invitrogen)
Megaprime DNA labeling system	(Amersham)
dNTP mix: 10 mM each of dATP, dTTP, dCTP and dGTP	(Invitrogen)

GFX PCR DNA gel band elution   (Amersham)

Microspin S-300 HR columns   (Amersham)

ECL Western blotting analysis system   (Amersham)

Antibodies

Name	Company	Label	Dilutions used
Gr-1	BD Pharmingen, Heidelberg	PE/APC	1:500
CD11b (Mac1)	BD Pharmingen, Heidelberg	PE/APC	1:800
Sca-1	BD Pharmingen, Heidelberg	PE	1:150
Ter119	BD Pharmingen, Heidelberg	PE	1:150
B220	BD Pharmingen, Heidelberg	PE/APC	1:200
CD4	BD Pharmingen, Heidelberg	PE	1:150
CD19	BD Pharmingen, Heidelberg	PE	1:200
CD117 (c-kit)	BD Pharmingen, Heidelberg	APC	1:500
CD8	BD Pharmingen, Heidelberg	APC	1:150
GFP	Molecular Probes Inc., OR		1:5000
Anti-His	Invitrogen, Carlsbad, CA	HRP	1:3000
Goat Anti-mouse	Invitrogen, Carlsbad, CA	HRP	1:2000

Cytokines

mIL3   (Tebu-bio)

mIL6   (Tebu-bio)

mSCF   (Tebu-bio)

Enzymes

EcoRI, XhoI, HpaI   (New England Biolabs)

T4 DNA Ligase   (New England Biolabs)

DNA Polymerase: Platinum Taq DNA polymerase kit   (Invitrogen)

DNaseI: DNase I DNA inactivating enzyme kit   Invitrogen)

Stock solutions and buffers

5-Fluorouracil: 50 mg/ml stock solution Medac, Hamburg, Germany. Working solution was 6 ml of the above solution mixed with 4 ml of phosphate buffered saline.

Avertin solution: Stock solution was prepared by adding 15.5 ml tert-amyl alcohol to 25 grams Avertin (2-2-2 Tribromoethanol), both procured from (Sigma-Aldrich, St. Louis, MO) and dissolved overnight. For working solution, 0.5 ml stock solution was added to 39.5 ml of cell culture grade phosphate buffered saline (PBS) and dissolved with a magnetic stirrer.

Calcium chloride solution for transfection: 2.5 M $CaCl_2$ (Sigma-Aldrich, St. Louis, MO) solution in water

Denaturation solution: 1.5M NaCl, 0.5N NaOH

Erythrocyte lysis buffer: 0.8% NH_4Cl with 0.1 mM EDTA (Stem Cell Technologies, Vancouver, Canada).

Formalin: 4% solution of formaldehyde (Sigma-Aldrich, St. Louis, MO) in water.

Giemsa: Giemsa's azure eosin methylene blue solution modified. (Merck KGaA, Darmstadt, Germany)

HEPES-Buffered Saline: (HBS) (Invitrogen, Carlsbad, CA).

May-Gruenwald: May-Gruenwald's Eosin-Methylene blue solution for microscopy (Merck KGaA, Darmstadt, Germany)

Pre-hybridization solution: Dissolve 0.2g skimmed milk and 2.0g dextran sulphate in 17ml water and add 6ml 20x SSC, 2ml formamide, 1ml 20% SDS and 80µl 500mM EDTA

20x SSC: Dissolve 3M sodium chloride and 0.3M sodium citrate in 800 ml de-ionized water and adjust pH to 7.0 and make up the volume to one liter

Telleyesnickzky's solution: 450 ml absolute ethanol + 25 ml glacial acetic acid + 25 ml formaldehyde.

Consumables

Cell strainer: BD Falcon 40 μm Nylon strainer for macerating the spleen and filtering the tissue (BD Biosciences, Palo Alto, CA)

Cell Scrapers: 25 cm sterile cell scrapers (Sarstedt, Newton, NC)

Cell culture pipettes: (2, 5, 10 and 25 ml): Sterile disposable pipettes (Corning Inc., Corning, NY)

Cell culture plates and dishes: Sterile 96 well, 24 well, 6 well plates (Sarstedt, Numbrecht, Germany) 100 mm x 20 mm dishes for adherent cells (Corning Inc., Corning, NY), and Petri dishes for suspension cells (Becton Dickinson Labware, Franklin Lakes, NJ) 150 mm x 20 mm dishes for adherent cells (Greiner Bione, Frickenhausen, Germany).

Cytospin slides: Marienfield pre-cleaned twin frosted slides for fixing single cell suspensions and blood smears (Marienfield, Lauda-Königshofen, Germany).

Cytospin filter cards: ThermoShandon thick white 5991022 filter cards for cytospins (Histocom AG, Zug, Switzerland).

Sterile Syringes: BD Plastipak 1 ml syringe (BD Biosciences, Palo Alto, CA) for injection of cells in mice and Kendall Monoject 3 ml syringes (Tyco Healthcare, UK) for bone marrow flushing and plating of CFCs. The stubs of 3 ml syringes were used to macerate the spleens of mice.

Sterile needles: 0.5 x 25 mm for injection of cell in mice i.v. and 0.55 x 25 mm (BD Microlance, Drogheda, Ireland) for bone marrow aspiration from living mice and flushing of bone marrow from extracted bones. 16-gauge bunt-end needles were used for dispensing and plating Methocult (CFC) media (Stem Cell Technologies, Vancouver, Canada).

Heparinized capillaries: (Microvette CB 300) plastic capillaries containing Lithium heparin for collection of blood (Sarstedt, Numbrecht, Germany).

Filtration units: Millex syringe driven filter units 0.22 μm and 0.45 μm filters (Millipore, Billerica, MA).

Oligonucleotides

All nucleotides were synthesized by Metabion AG, Martinsried, Germany.

Oligonucleotide	Sequence 5´ to 3´
ApMSCV For	CAG CCC TCA CTC CTT CTC TA
ApMSCV Rev	CTT GAC GAG CAT TCC TAG
AML1-ETO 2.2 For	ATG CGT ATC CCC GTA GAT GC
AML1-ETO 2.2 Rev	CTA GCG AGG GGT TGT CTC TAT G
AML1-ETO FL For	GCC CCA ACT TCC TCT GCT C
AML1-ETO FL Rev	GGG TTC CGG GGT GG TTG A
AML1-ETO BR For	ATG ACC TCA GGT TTG TCGGTC G
AML1-ETO BR Rev	TGA ACT GGT TCT TGG AGC TCC T
Meis1 FL For	ATG GAG TAG GCA TCC CCT CCA CG
Meis1 FL Rev	CAT GCC CAT ATT CAT GCC CAT TCC
β-2microglobin mouse	TGC TAT CCA GAA AAC CCC TC
β-2 microglobin	CGG CCA TAG TGT CAT GCT TA
LM-PCR-GFP-A	ACTTCAAGATCCGCCACAAC
Nested Linker Primer	TACGAGAATCGCTGTCCTCTCCTT
Vectorette_primer	CGAATCGTAACCGTTCGTACGAGAATCGCT
LM-PCR-GFP-C	ACATGGTCCTGCTGGAGTTC

Primers for V-DJ$_H$ and DJ$_H$ recombination

Oligonucleotide	Sequence 5´ to 3´
V$_H$ 7183	CGGTACCAAGAASAMCCTGTWCCTGCAAATGASC
V$_H$ 558	CGAGCTCTCCARCACAGCCTWCATGCARCTCARC
V$_H$ Q52	CGGTACCAGACTGARCATCASCAAGGACAAYTCC
J$_H$ 3	GTCTAGATTCTCACAAGAGTCCGATAGACCCTGG
C-mu-5'	TGGCCATGGGCTGCCTAGCCCGGGACTT
C-mu-3'	GCCTGACTGAGCTCACACAAGGAGGA
B rec chk fw1	ACGTCGACTTTTGTSAAGGGATCTACTACTGT
B rec chk fw2	ACGTCGACGCGGASSACCACAGTGCAACTG
B rec chk rev	GGGTCTAGACTCTCAGCCGGCTCCCTCAGGG

Plasmids

MIG/MIY: Murine stem cell virus derived vector. A bi-cistronic vector with an internal ribosomal entry site (IRES) and an EGFP or EYFP cassette 3' of the IRES) (kindly provided by Prof. Dr. R.K. Humphries, Vancouver, Canada).

pCDNA6/V5-His: Mammalian expression vector used for tagging genes (Invitrogen, Carlsbad, CA).

pEGFP-C1: Mammalian expression vector used for tagging genes as a C-terminal fusion to EGFP protein (Clontech, Ca).

Ecopac: A packaging vector coding for the gag, pol, and env viral proteins.

Molecular weight markers

1 kb ladder, 1 kb plus ladder and 100 bp ladder (Invitrogen, Carlsbad, CA)

Cell and tissue culture

Mammalian cell lines

293T	(human embryonic kidney cell line)
GP+E86	(mouse fibroblast cell line)
NIH 3T3	(mouse fibroblast cell line)

Media and supplements

Methylcellulose media: Methocult 3434 for the culture of myeloid CFC assays and Methocult 3630 for the pre-B CFC assays (Stem Cell Technologies, Vancouver, Canada).

Media: Dulbecco's Modified Eagle's Medium (DMEM) 4,5 g/l glucose, l-glutamine, sodium pyruvate and 3,7 g/l $NaHCO_3$ (PAN biotech GmbH, Aidenbach, Germany)

Fetal Bovine Serum (FBS): 0,2 μm-filtered mycoplasma screened (PAN biotech GmbH, Aidenbach, Germany)

Dulbecco's phosphate buffered saline (DPBS): without magnesium and calcium, sterile filtered (PAN biotech GmbH, Aidenbach, Germany)

Trypsin–EDTA: 1x in HBS without calcium and magnesium with EDTA (Invitrogen, Carlsbad, CA)

Penicillin/Streptomycin: Antibiotic solution with 10,000 u/ml Pen G sodium and 10,000 μg/ml Streptomycin sulfate in 0,85% saline. Used 5 ml per 500 ml medium bottle (Invitrogen, Carlsbad, CA)

Murine cytokines: mIL3, mIL6, mSCF, (lyophilized) (Tebu-bio, Offenbach, Germany)

Ciprofloxacin: Ciprofloxacin 400 solution, (Bayer AG, Leverkusen, Germany)

Instruments

DNA Cross linker	(GS Gene linker BIO-RAD)
Fluorescence Activated Cell Sorting	(BD FACSVantage SE System)
Flox Cytometry	(BD FACS Calibur System)
Cytospin	(Cytospin 2 Shandon)
Microscope	(Leitz Diavert Inverted Microscope Ernst Leitz Wetzlar GmbH)
quantitative PCR Cycler	TaqMan 7900HT, instrument serial No: 201381

Patient samples

Mononuclear cells prepared from diagnostic bone marrow or peripheral blood (PB) samples from 48 adult AML patients were analyzed. The AML cases were classified according to the French-American-British criteria and the World Health Organization classification (Varela et al., 1985). The study was approved by the ethics committees of all participating institutions, and informed consent was obtained from all patients before they entered the study in accordance with the Declaration of Helsinki (**http://www.wma.net/e/policy/b3.htm**). As a control, bone marrow mononuclear cells (BM MNCs; CellSystem, St Katharinen, Germany) from healthy individuals were analyzed. Cytomorphology, cytochemistry, cytogenetics, and molecular genetics were applied in all cases.

Mice

Parental strain mice were bred and maintained at the GSF animal facility. Donors of primary BM cells were > 12-week-old (C57Bl/6Ly-Peb3b x C3H/HeJ) F1 (PebC3) mice and recipients were > 8–12 week old (C57Bl/6J x C3H/HeJ) F1 (B6C3) mice.

METHODS

Patient samples

cDNA was prepared from BM or PB samples of 48 AML1-ETO patients. AML1-ETO patients with either secondary mutations (n=22) and those negative for any mutation (n=26), were selected for real-time PCR.

Cloning of constructs

For retroviral gene transfer into mouse bone marrow (BM) cells, different gene products were subcloned into the multiple cloning site of the modified MSCV 2.1 vector (Pineault et al., 2003), upstream of the internal ribosomal entry site (IRES) and the enhanced green or yellow fluorescent protein (GFP/YFP) gene. As a control the MIG vector was used.
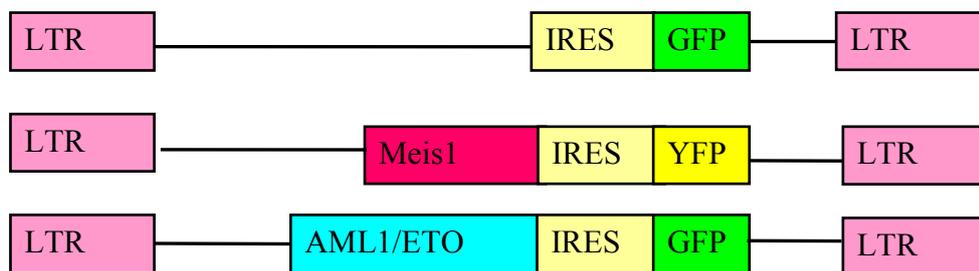
Figure 8a: Schematic diagram of constructs.
AML1-ETO, Meis1, EGFP and AML1-ETO mutants were cloned into the multiple cloning sites of the modified MSCV 2.1 vector upstream of the internal ribosomal entry site (IRES) and the enhanced green or yellow fluorescent protein (GFP/YFP) gene.

The AML1-ETO (AE) cDNA was kindly provided by S.W.Hiebert (Vanderbilt University School of Medicine, Nashville, Tennessee, USA).This 2.2Kb gene has been subcloned by blunt end ligation into the *Hpa I* site in the multiple cloning site (MCS) of the MIG vector. The AE mutants ΔTAF, ΔHHR, Δ540 were also provided by S.W.Hiebert (Vanderbilt University School of Medicine, Nashville, Tennessee, USA) and subcloned into MSCV vector. The dead mutant AE L148D construct was generated by using the QuickChange Site-Directed Mutagenesis Kit (Stratagene) according to the manufacturer's instructions.

The Meis1 (*M*yeloid *E*ctopic *I*ntegration *S*ite 1) construct was kindly provided by Dr. Keith Humphries (Terry Fox Laboratory, British Columbia Cancer Agency, Vancouver, British Columbia, Canada.). The 1.2kb *Meis1* gene was sub-cloned into the multiple cloning sites (MCS) of the MIG vector using the enzymes *EcoRI* and *Xho1*.
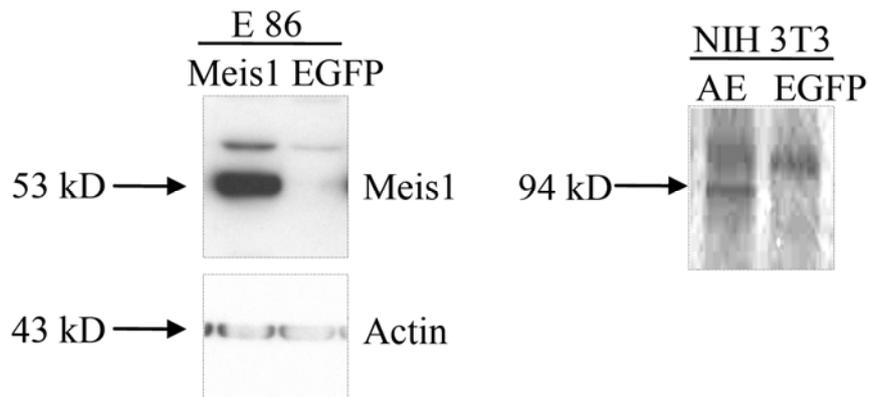
Fig 8b. **Protein expression of AML1-ETO and Meis1**: Protein expression of AML1-ETO and Meis1 were observed by immunoblotting of the whole protein lysate. Protein was made from the AML1-ETO transfected on NIH 3T3 cell line and untransfected NIH 3T3 cell line as negative control. Similarly protein expression for Meis1 was performed from the Meis1 GP+E86 cell line used for transduction of murine bone marrow. As a negative control, a GP+E86 cell line transduced with the empty vector EGFP was used.

Preparation of high titre stable virus producing cell lines

A total of 1.2x 10^6 293T cells were plated in a 15 cm dish and on the following day used for transient transfection. Medium was changed 4 hours prior to the transfection. 30 μg plasmid DNA of the gene of interest, and the retroviral packaging construct Ecopac were added to sterile water (to make up a total volume of 1 ml) and 100 μl 2.5M $CaCl_2$ was added drop wise to the water-DNA mixture. This was added slowly to a tube containing 1 ml sterile HBS (pH 7.2). After gentle mixing and incubating at room temperature for 3-4 minutes, this mixture was added drop wise to the 15 cm dish. The dish was carefully placed in a 37°C incubator. The medium was changed after about 12 hours. The virus conditioned medium (VCM) was collected from the cells every 12 hours and filtered with a 0.45 μm Millipore filter and stored as VCM at −80°C for later use or used directly to transduce GP+E86 fibroblasts or murine bone marrow.

5×10^4 GP+E86 fibroblasts were plated into 6 well plates one day prior to retroviral transduction. For transduction, 500 µl or 1 ml of fresh or frozen VCM from transfected 293T cells was layered on top GP+E86 cells with the addition of 10 µg/ml protamine sulphate. Fresh medium was added to the cells after 4 hours and the transduction procedure was repeated every 12 hours for four times. The cells were expanded for two days after final transduction to allow the GFP expression. GFP expressing cells were sorted using the fluorescence activated cell sorter (FACS), propagated and used as stable virus producing cell lines to transduce murine bone marrow. The following cell lines were prepared using this method: E86-GFP, E86-AML1-ETO, E86-Meis1, E86-AML1-ETOΔTAF, E86-AML1-ETOΔ540, E86-AML1-ETO L148D.

In some cases where viral titres of bulk cell lines were low, single cells were sorted into 96 well plates and after expansion their viral titres are determined on NIH-3T3 cells. Clones producing highest titres were expanded, frozen and used for experiments.

Titration was performed by plating 2×10^5 NIH3T3 cells per well in 6 well plates and layering them with 500 µl VCM the next day with the addition of 10 µg/ml protamine sulphate. Fresh medium was added after 3-4 hours to stop transduction. This was performed every 12 hours (thrice totally). Two days following transduction with VCM, cells were analyzed for GFP expression at the FACS Calibur.
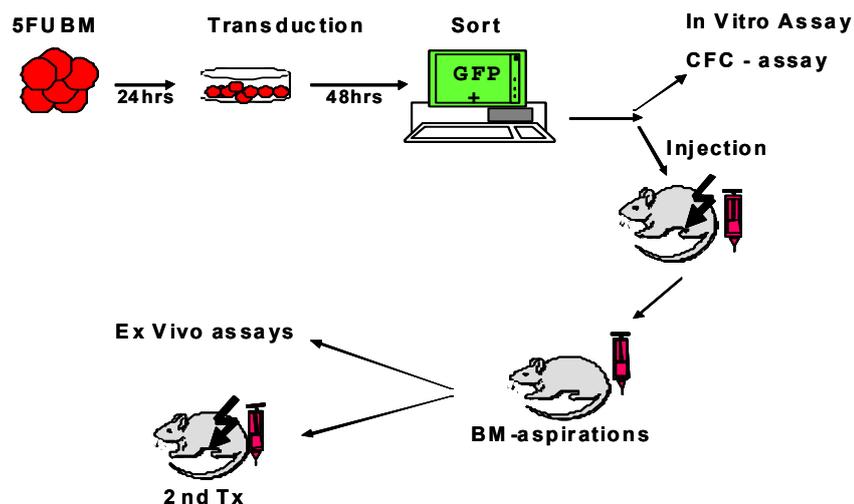
Figure 9: Schematic diagram of Bone marrow transplantation model.
5-FU treated BM was transduced with the gene of interest and sorted after 48hrs for GFP positive cells. Sorted cells were used for *in vitro* or in vivo experiments. GFP+ cells were injected into mice and the engraftment

levels of these mice were checked every four weeks. Furthermore, BM from leukemic mouse was transplanted into secondary recipients.

Retroviral transduction of primary bone marrow

Parental strain mice were bred and maintained at the GSF animal facility. Donors of primary BM cells were > 8-wk-old (C57BL/6Ly-Pep3b x C3H/HeJ) F_1 (PepC3) mice. The mice were fed with autoclaved chow and supplied with drinking water containing ciprofloxacin and acetic acid and housed in individually vented cage systems. 150 milligrams of 5-Fluorouracil (5-FU) was injected per kg of mouse body weight to eliminate cycling cells and to enrich for hematopoietic progenitor cells (Figure 9). On the fifth day following 5-FU injection, these mice were sacrificed and their femurs and tibia flushed with serum-supplemented medium to obtain bone marrow cells. This bone marrow cells were pre-stimulated by culturing for 2 days in a cytokine cocktail (10 ng/ml mIL-6, 6 ng/ml mIL-3 and 100 ng/ml murine stem cell factor) in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 15% FBS. On day 3, transduction was performed by layering the bone marrow cells on top of adhered GP^+E86 cell lines (co-culture). The cell lines were irradiated for 40cGy and plated on adherent 150 mm x 20 mm dishes one day prior to the transduction. 10 µg/ml protamine sulphate was always added to the medium during viral transduction. On day 5, following transduction for two days, bone marrow was removed gently but completely without disturbing the adhered monolayer of the GP+E86 cell line. Bone marrow was cultured in DMEM 15% FBS and 2 more days were allowed for GFP expression. On day 7, GFP positive cells were sorted by FACSVantage and used for bone marrow transplantation or for in vitro culture (Figure 9). Bone marrow was always cultured in DMEM 15% FBS medium supplemented with 10 ng/ml mIL-6, 6 ng/ml mIL-3 and 100 ng/ml murine stem cell factor.
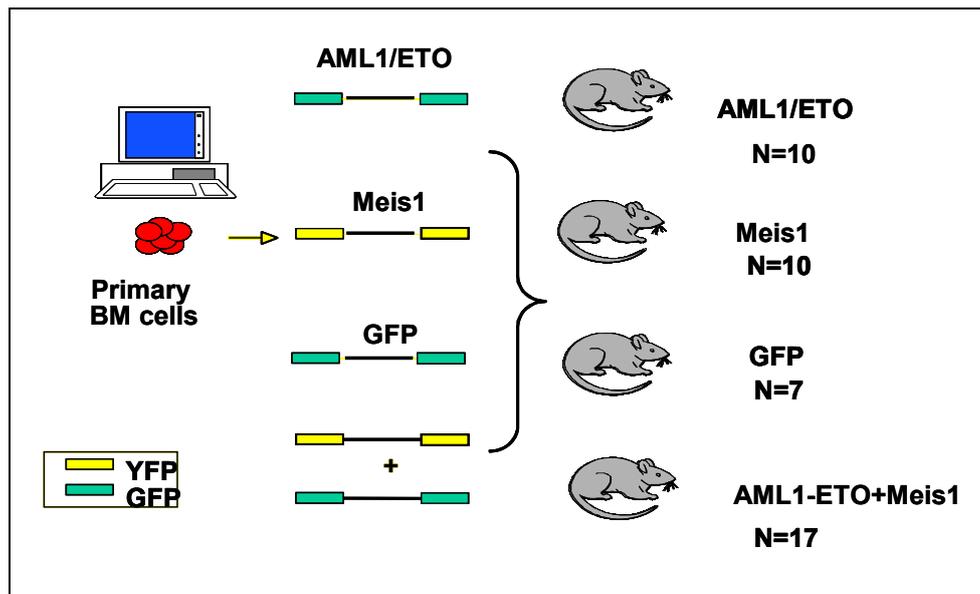
Figure 10: Experimental design of bone marrow transplantation.

This is a representation of the experimental set of mice from the AML1-ETO, Meis1 and the EGFP arms. Following cohorts of mice were injected: AML1-ETO+ Meis1 (17), AML1-ETO (n=10), Meis1 (n=10).

Bone marrow transplantation and assessment of mice

Recipients were >8 to 12 wk old (C57BL/6J x C3H/HeJ) F_1 (B6C3) mice. These mice were lethally irradiated (850 cGy) a few hours prior to receiving BM transplants. Transduced BM or BM from leukemic mice was injected with or without addition of mock transduced or non-transduced bone marrow cells intravenously into the tail vein of mice using a sterile 0.5 x 25 mm needle. Mice were assessed periodically for signs of leukemic symptoms by blood withdrawal from the tail vein using sterile scalpels and bone marrow aspiration from the tibia of anaesthetized animals or by the observance of symptoms that included crouching, frizzled body hair, paleness in the feet, heavy breathing and disturbed gait. Mice were considered moribund when one of these symptoms was starkly visible. In this study BM cells transduced with AML1-ETO, Meis1 and the EGFP arms were injected into mice as shown in figure 10.

For secondary, $1x10^6$ bone marrow cells from a leukemic primary were injected with the addition, to each mouse, of $1x10^6$ bone marrow cells from a syngenic wild type mouse (non-transduced mock cells).

Moribund mice were sacrificed by CO_2 asphyxiation and BM was aspirated as described before. Spleens were dissected and macerated to produce single cell suspensions and peripheral blood was drawn with a sterile 0.5 x 25 mm needle by puncturing the heart immediately after sacrificing the mice. Red blood cell (RBC) lysis for peripheral blood, bone marrow and spleen cells was performed by incubating the cells in ammonium chloride buffer for 10 minutes at room temperature.

Exp. No.	Gene	No cells Tx	Mock*
3944b	AE+Meis1	2.00×10^5	3.00×10^5
3944a	AE+Meis1	2.00×10^5	3.00×10^5
3912	AE+Meis1	1.25×10^5	3.00×10^5
3912#2	AE+Meis1	1.25×10^5	3.00×10^5
3933	AE+Meis1	unsorted	-
3933 #4	AE+Meis1	unsorted	-
3933 #2	AE+Meis1	unsorted	-
3918#2	AE+Meis1	1.30×10^5	5.00×10^5
3974#2	AE+Meis1	unsorted	-
3974#3	AE+Meis1	unsorted	-
4406B	AE+Meis1	1.00×10^5	3.00×10^5
4406B#2	AE+Meis1	1.00×10^5	3.00×10^5
4441A	AE+Meis1	unsorted	-
4525	AE+Meis1	5.00×10^4	3.00×10^5
4441B#2	AE+Meis1	unsorted	-
4441A#3	AE+Meis1	unsorted	-
4323	AE+Meis1	unsorted	-
4696	AE+Meis1	unsorted	-
4415#1	AEΔTAF+Meis1	1.2×10^6	-
4415#2	AEΔTAF+Meis1	1.2×10^6	-
4415#3	AEΔTAF+Meis1	1.2×10^6	-
4571A#1	AEΔTAF+Meis1	1.2×10^6	-
4571A#1	AEΔTAF+Meis1	1.2×10^6	-
4585#1	AE148D+Meis1	1.0×10^6	-
4585#2	AE148D+Meis1	1.0×10^6	-
4585#3	AE148D+Meis1	1.0×10^6	-
4757#1	AE148D+Meis1	3.0×10^6	-
4757#1	AE148D+Meis1	3.0×10^6	-
4757#1	AE148D+Meis1	3.0×10^6	-
4664#1	AE148D	1.0×10^6	-
4664#2	AE148D	1.0×10^6	-

4664#3	AE148D	1.0×10^6	-
4664#4	AE148D	1.0×10^6	-
4768#1	AEΔTAF	2.0×10^6	-
4768#2	AEΔTAF	2.0×10^6	-
4768#3	AEΔTAF	2.0×10^6	-
4768#4	AEΔTAF	2.0×10^6	-
4768#5	AEΔTAF	2.0×10^6	-

Table 2: Injection of transduced and mock bone marrow in AML1-ETO+ Meis1, AEΔTAF+Meis1, AE148D+Meis1, AE148D, AEΔTAF. Mice transplanted with unsorted cells were injected with a total of 1.0×10^6 cells. *Mock indicates mock-transduced GFP negative BM cells. Tx=transplanted.

Exp. No.	Primary/secondary	Gene	No cells Tx
4519	1°	Meis1	4.40×10^5
4638	1°	Meis1	1.00×10^6
4638#2	1°	Meis1	1.00×10^6
4638#3	1°	Meis1	1.00×10^6
4638#4	1°	Meis1	1.00×10^6
4638#5	1°	Meis1	1.00×10^6
4752A	1°	MIG	5.00×10^5
4752A#2	1°	MIG	5.00×10^5
4760	1°	MIG	5.00×10^5
4760#2	1°	MIG	6.00×10^5
4760#3	1°	MIG	6.00×10^5
4760#4	1°	MIG	6.00×10^5
4760#5	1°	MIG	6.00×10^5
4747	1°	AE	5.00×10^5
4747#2	1°	AE	5.00×10^5
4747#3	1°	AE	5.00×10^5
4747#4	1°	AE	5.00×10^5
4747#5	1°	AE	5.00×10^5

Table 3: Injection of transduced and mock AML1-ETO, Meis1 and empty vector cells into lethally irradiated mice. Tx=transplanted.

Exp. No.	Primary/secondary	gene	no cells Tx	NT
4155#1	2° (3918#2)	AE+Meis1	1.00×10^6	1.00E+06
4155#2	2°(3918#2)	AE+Meis1	1.00×10^6	1.00E+06
4155#3	2°(3918#2)	AE+Meis1	1.00×10^6	1.00E+06
4155#4	2°(3918#2)	AE+Meis1	1.00×10^6	1.00E+06
4700	2° (4525)	AE+Meis1	1.00×10^6	1.00×10^6
4700#2	2° (4525)	AE+Meis1	1.00×10^6	1.00×10^6
4828	2° (4323#3 SP)	AE+Meis1	1.20×10^6	1.00×10^6
4828#2	2° (4323#3 SP)	AE+Meis1	1.20×10^6	1.00×10^6
4828#3	2° (4323#3 SP)	AE+Meis1	8.00×10^5	1.00×10^5
4828#4	2° (4323#3 SP)	AE+Meis1	8.00×10^5	1.00×10^5

Table 4: Tabular representation of the number of cells injected into each mouse. NT indicates non transduced GFP-negative cells from a syngenic healthy mouse. Tx=transplanted.

FACS analysis of murine cells

Single cell suspensions of cells were stained with various fluorescence-conjugated antibodies. Staining was performed in PBS with the fluorescence-conjugated antibodies using a 1: 200 dilution for each antibody. Samples were incubated at 4°C for 20 minutes and subsequently washed with PBS to remove excess antibody. Cells were centrifuged and after decanting the supernatant, resuspended in FACS buffer (2% fetal bovine serum and 2 µg/ml propidium iodide in phosphate-buffered-saline). Antibodies used for FACS were labelled with phycoerythrin for Gr-1, CD11b (Mac1), Sca-1, Ter119, CD4, CD19, and allophycocyanin conjugated CD11b (Mac-1), CD117 (c-kit), B220, and CD8. Fluorescence was detected using a FACS Calibur flow cytometer and analyzed using the CellQuest software. Dead cells were gated out by high PI staining and forward light scatter.

Ex vivo proliferation and CFC assays

Cell proliferation was assessed in DMEM supplemented with 15% FBS 10 ng/ml mIL-6, 6 ng/ml mIL-3 and 100 ng/ml murine stem cell factor (standard medium). Quantification of clonogenic progenitors was performed by plating cells in methylcellulose supplemented with cytokines (Methocult M3434). Replating was performed every week in appropriate dilutions.

Cytospin preparations and Wright-Giemsa staining

Cytospins of single cell suspensions were performed by resuspending cells in PBS at a concentration of 2-6x 10^5 cells per 200 µl and this volume was introduced into the cytospin apparatus. The cells were permanently fixed on glass slides by centrifugation at 500 rpm for 10 minutes and subsequently air-dried. Modified Wright Giemsa staining was performed by immersing the slides in an undiluted solution of May-Grunwald stain for 5 minutes followed by immersion of slides in 1:50 diluted Giemsa stain for 1 hour. Slides were dipped in water to remove excess stain between the two staining steps and after the staining procedure and air-dried for observance under the inverted light microscope.

For histological analysis, the peritoneum of sacrificed leukemic mice was dissected so as to expose all organs (as shown in figure 11) and most of the blood drained by cutting the peritoneal artery and absorbing the blood with a tissue paper. The mice were fixed in an aqueous solution of formaldehyde (10% v/v) and sections of selected organs were prepared and hematoxylin-eosin stained using standard protocols.

Figure 11: Photographic representation of fixing of diseased mouse for histopathology.

RNA and genomic DNA isolation and cDNA preparation

The TRIzol method for RNA isolation described by the manufacturer was used to extract RNA with the addition of 1 ml of TRIzol solution per million cells. Equal amounts of RNA as quantified by a spectrophotometer were added to each reaction (in a set) used for cDNA preparation for the semi-quantitative PCRs. Each sample was treated with DNaseI for prevention of genomic DNA contamination in cDNA samples. This was performed for each sample prior to cDNA preparation according to the manufacturer's instructions.

Genomic DNA was isolated from a minimum of 1×10^7 cells for Southern blotting from various murine organ cells using the DNAZOL reagent and the protocol for the same according to the manufacturer. Genomic DNA for the V-DJ and D-J PCRs was isolated using the DNeasy mini kit using supplied methods. Genomic DNA was resuspended in sterile water and quantified using a spectrophotometer after proper dissolution.

cDNA was prepared from DNaseI treated RNA. First-strand cDNA synthesis was done with ThermoScript kit. In a 20 µl reaction volume, 1 µg RNA and 1 µg of oligo (dT) were mixed to a final volume of 11 µl and incubated 10 minutes at 70°C. Then, 4 µl of 5 X first-strand buffer, 2 µl of DTT 0.1 mol/L, 1 µl of 10 mmol/L deoxynucleoside triphosphate mix, and 2 µl of ThermoScript reverse transcriptase were added. The sample was incubated 1 hour at 42°C and used for PCRs.

Southern blot analysis

Southern blot analysis to assess proviral integration was performed by isolating DNA from bone marrow, spleen and peripheral blood of leukemic mice using DNAzol reagent as recommended by manufacturer. Southern blot was performed using standard protocols. DNA was digested with *EcoRI*, which cuts the proviral DNA once, to release a fragment specific to the proviral integration site. To check the full-length integration, DNA was digested with Nhe I, which cuts in the long terminal repeats (LTRs) to release the proviral genome. After digestion DNA was loaded on a 0.7% agarose gel with 0.5 µg/ml ethidium bromide. After electrophoresis, the DNA was depurinated by soaking the gel in 0.2 N HCL for 8 minutes, and subsequently for 45 minutes in denaturation buffer. After denaturation, the DNA was

transferred on zeta-Probe GT membrane by capillary action in a 10X transfer buffer. Cross-linking of the DNA with the membrane was done by incubating the membrane at 150 mjoule in a UV gene linker. The probe used was a 700 bp GFP fragment, which was digested out from the pEGFP-C1 plasmid and labeled with α-^{32}P dCTP using Megaprime DNA labeling system. Probe was purified using Microspin S-300 HR columns. Hybridization was done with α-^{32}P GFP labeled overnight at 62^{0}C. After two rounds of washing the membrane was dried, covered with a plastic film and put in a cassette for exposure of the film. The film was put on the membrane in a dark room and the exposure was done at variable exposing times between 48 hours and one week, depending on the visualization signal observed.

Western blotting (Immunoblotting)

Sample preparation and cell lysis (total cell extract)

Proof of protein expression was performed using E86 AML1-ETO and Meis1 cell lines. The cells were lysed using 150 µl RIPA buffer with fresh added protease inhibitors and detached using a cell culture scraper. The cells with RIPA buffer were transferred to an Eppendorf microcentrifuge tube and mixed by inversion for 30 minutes at 4°C. After the homogenization, the sample was centrifuged at 14000 rpm for 30 minutes. After centrifugation, the supernatant was transferred to a new Eppendorf tube and either frozen at -80°C, or kept on ice for determination of protein concentration. As a control, 293T cells from an 80% confluent 15 mm cell culture dish (between 5 and 10x 10^7 cells) were transiently transfected with 10 µg of AML1-ETO and Meis1. Lysates were prepared using the method described above.

Determination of protein concentration

The method used for measuring the protein concentration was the Bradford method. The assay is based on the observation that the absorbance maximum for an acidic solution of Coomassie Brilliant Blue G-250 shifts from 465 nm to 595 nm when binding to protein occurs. Both hydrophobic and ionic interactions stabilize the anionic form of the dye, causing a visible color change. The assay is useful since the extinction coefficient of a dye-albumin complex solution is constant over a 10-fold concentration range. The protein concentration of the sample was determined by comparison to values obtained for the measure of the known

range of protein standards. The protein standard used was Bovine Serum Albumin (BSA). Six different albumin concentrations (2.5 µg, 5 µg, 10 µg, 15µg, 20µg and 25µg) were diluted in distilled water to a final volume of 800 µl. One microliter of cell lysate was diluted in distilled water for the measure. 200 µl of Protein Assay solution was added to the tubes. The tubes were incubated at RT for 15 minutes and the content was further transferred to polystyrol cuvettes. A determination of the standard curve of the spectrophotometer with distilled water and the protein standards was done using the specific program for protein in the spectrophotometer. The sample was measured following the standard curve determination.

SDS PAGE of cell extracts

Total cell extract (TCE) proteins were separated on a denaturing gel consisting of 8% Tris-glycine gel was used and a 5% stacking gel. The concentration of the separation gel was chosen considering the sizes of wt AML1-ETO protein (94kDa) and Meis1 (about 53 kDa) as indicated in molecular protocols ((Sambrook and Gething, 1989). The sample was homogenized and diluted 1:1 with 2x loading buffer and heated in a boiling water bath for 10 minutes. 80 µg protein was loaded on each gel lane. The electrophoresis was performed under100 v for 1hour and 30 minutes in the cold room at 4°C.

Protein blotting

After the electrophoresis, the gel was taken from the cassette and washed once with TBS. For the blotting, the wet system was used. To permit a better transfer of large molecular weight proteins as AML1-ETO, Meis1, with a predicted size of about 94kDa, and 53kDa, a PVDF membrane was chosen. The membrane was wetted in methanol for 30 seconds, rinsed in distilled water and equilibrated in transfer buffer for 10 minutes. The system was assembled putting a sponge on the bottom of the sandwich (in contact to the negative pole), a 0.8 mm filter paper in contact to the sponge, and the gel over the paper. A 10 ml pipette was used to eliminate the eventually formed air bubbles. On the membrane, another filter paper was put, a second sponge and the chamber was closed. The PVDF membrane was oriented to the positive pole to permit the protein (negatively charged) to migrate from the gel to the membrane (on the positive pole). The transfer system was submitted to constant amperage of 250 milliamp for 4 hours at 4°C with agitation. The observation of the high molecular weight proteins of the pre-stained protein standard on the membrane was an indicator of successful transfer.

Protein detection on the blotting membrane with HRP-marked antibodies

The antibody-detection of protein was performed following the instructions of the antibody's supplier. After the transfer, the membrane was blocked to prevent non-specific binding of antibodies to the membrane by incubating with BlottA buffer for one hour at room temperature or overnight at 4°C in constant shaking. The membrane was further washed once with TBS for five minutes and incubated with the primary antibody at 1:1000 dilution in BlottoA overnight. The concentration used for the antibodies was adjusted according to the intensity and background. After incubation with the primary antibody, the membrane was washed three times with TBS with 0.05% Tween-20 (TBST). The secondary antibody conjugated with Horse Radish Peroxidase (HRP) was diluted 1:2000 in BlottoA and put on the membrane for 45 to 90 minutes incubation at room temperature. The membrane was rinsed with distilled water, washed again tree times with TBST and once with TBS for 5 minutes under agitation. To detect the antibodies on the membrane a commercial chemiluminescence kit was used according to the manufacturer's instructions. After washing, the ECL detection solution was put on the membrane for 90 seconds. The membrane was dried, covered with a plastic film and put in a cassette for exposure of the film. The film was put on the membrane in a dark room and the exposure was done at variable exposing times between 15 seconds and 10 minutes, depending on the visualization signal observed.

Polymerase chain reactions (PCRs)

Quantitative PCR

Meis1 expression was assayed by the TaqMan real-time quantitative polymerase chain reaction (RQ-PCR) method in total human bone marrow (BM), and AML1-ETO patient samples. Meis1 and FLT3 probes were used from Applied Biosystems (Foster City, CA; assay IDs: Meis1, Hs00180020_m1; FLT3, Hs00975657_m1). Quantification of *Meis1* expression was performed by RQ-PCR with Applied Biosystems primers. For normalization, the TATA binding protein (TBP) gene was used. Reactions were with 1.0 μL of cDNA in a total reaction volume of 20 μL by using an ABI PRISM 7900 Sequence Detection System (Applied Biosystems).

PCR to check gene expression in mice

PCR was performed to check the expression of *AMLl-ETO* and *Meis1* in leukemic mouse BM, spleen and PB. All the cDNA and genomic DNA samples were initially amplified for house keeping genes like β-2 microglobin and then with gene specific primers for AML1-ETO, Meis1 etc. Additionally, AML1-ETO expresssion was detected by breakpoint PCR and *Meis1* expression by primers flanking the MIG vector MCS.

PCR for V-D-J recombination status

D-J_H rearrangements in the Ig locus were detected by a PCR strategy employing two upstream degenerate primers binding 50 of the DFL/DSP element or the DQ52 element. The reverse primer was complementary to a binding site downstream of the JH4 segment. All three primers were used in a single PCR reaction in a multiplex PCR and the following reaction used in germ line configuration, the DQ52 and JH4A primers will amplify the 2.15-kb germ line fragment. D-J_H1, D-J_H2, D-J_H3, and D-J_H4 rearrangements involving D_{FL}, D_{SP}, or D_{Q52} elements will be detected by the emergence of bands of ~1.46, ~1.15, ~0.73, and ~0.20 kb, respectively. The amplification protocol was an initial denaturation at 94°C for 1 minute followed by 35 cycles of 1 minute at 94°C, 1 minute at 60°C, and 1 minute 45 seconds at 72°C. Final extension was carried out at 72°C for 10 minutes. The concentration of genomic DNA taken was always between 20-300 ng/μl.

The PCR assay for V to D-J rearrangement employs three degenerate primers oligo nucleotides (each in separate reactions) homologous to the conserved framework region 3 (FR3) sequences of the three V_H gene families (V_H 7183, V_H558 and V_H Q52) and the J_H reverse primer. This results in amplified VDJ rearrangements of ~1,058, ~741, or ~333 nucleotides. Wild-type murine spleen genomic DNA was used as a positive control and genomic DNA from the myeloid cell line 32D as a rearrangement negative control. PCR was carried out after an initial denaturation step of 94°C for 4 min for 35 cycles with 94°C for 1 minute, 60°C for 1 minute and 72°C for 1 minute 45 seconds. Final extension was carried out for 7 minutes at 72°C. All PCR products were evaluated on a 1.5% agarose gel by gel electrophoresis.

Identification of retroviral integration sites

The identification of the genomic integration sites of the retroviral constructs by Bubble PCR [1]. For the determination of the retroviral integration sites, a linker mediated (LM)-PCR was performed (Schessl et al., 2005). Aliquots of the cell lysates from leukemic mice were digested with PstI (New England Biolabs), and the fragments were then ligated overnight at room temperature to a double stranded bubble linker prior to performing a first PCR (PCR-A) on 10 µl (one-tenth) of the ligation product using a linker-specific Vectorette primer and an LTR-specific primer. The bubble linker contains a 30-nucleotide non-homologous sequence in the middle region that prevents binding of the linker primer in the absence of minus strand generated by the LTR-specific primer. A 1 µl aliquot of the PCR-A reaction (one-fifteenth) is then used as a template for a second nested PCR (PCR-B) using an internal LTR-specific primer and the same linker-specific vectorette primer as it is used in PCR-A. 10 µl (one-half) of the final PCR-B product is then separated by electrophoresis using 2% agarose TAE gel. Individual bands are excised, purified and then cloned into PCR2.1 before sequencing the integration site of the retrovirus.

Statistical analysis

Data were evaluated by using the t-test for dependent or independent samples (Microsoft EXCEL). Differences with P values < 0.05 were considered statistically significant. For calculations of frequency of leukemia propagating cells, the L-Calc software was used. Cell numbers were entered as doses, the number of mice per cohort as test and the number of mice dead as the response for the frequency calculation.

RESULTS

Expression of *Meis1* in AML1-ETO positive patient samples

We analyzed 39 AML1-ETO positive AML samples harbouring various other mutations like C-KITD816, FLT3-LM, FLT3D835, NRAS codon 12/13/61 etc. These samples were analysed for *MEIS1* expression using qRT-PCR probes. The expression levels were compared with normal human bone marrow. Samples were normalized with the house keeping gene TBP. Most of the 39 samples along with Kasumi, a t(8;21) positive cell line, showed a low expression of *MEIS1*. However, three samples expressed comparably high levels of *MEIS1*. After normalization with the house keeping gene TBP, t(8;21) cases and Kasumi showed average ΔC_T values of 8.47 and 10.33 respectively, whereas normal bone marrow cells showed a ΔC_T value of 1.11. The data are shown in Fig. 12. The higher expression of *MEIS1* in t(8;12) positive samples did not correlate with the presence of other cooperating mutations such as FLT-LM or CKITD816. Furthermore, there was no correlation between high *MEIS1* expression and poorer treatment outcome in these samples.
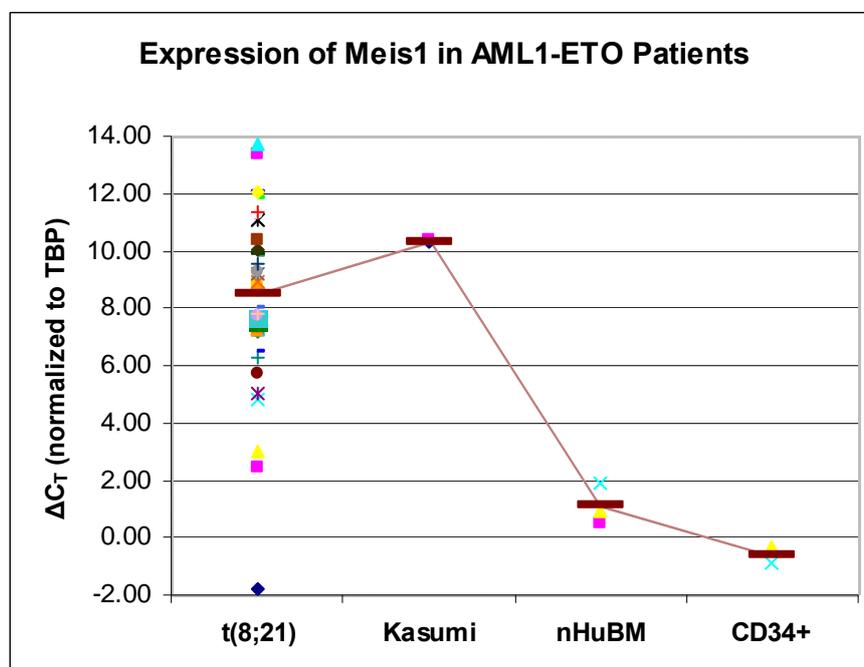
Figure 12: Quantification of *MEIS1* expression in AML1-ETO patients by qRT-PCR.

The expression levels of *MEIS1* in patients with AML1-ETO translocation. The expression level of the gene is inversely correlated with the ΔC_T value. (nHuBM= Normal Human BM, CD34+= Normal BM CD34+)

Transplantation of mice with BM transduced with AML1-ETO (AE) and Meis1

Mice were injected with bone marrow cells expressing AML1-ETO, Meis1, EGFP or AML1-ETO together with Meis1. Transplanted mice were monitored for symptoms like pale legs, lethargic, breathing problems and frizzled body hair. Moribund mice were then sacrificed for further analysis. Peripheral blood, spleen and bones were taken from these mice and were subjected to cell count like RBCs, WBCs from blood, spleen and bone marrow. Before crushing the spleen parameters like weight, length and width were calculated. After taking the values of all these hematological parameters the cells from blood, spleen and bone marrow were checked for engraftment levels by analyzing for GFP positive cells. Various organs of the mice were fixed with formalin and sent for histopathology. Sacrificed mice transplanted with EGFP, Meis1 and AE showed median BM engraftment levels of 35% ± 27.8, 67% ± 13.6 and 56% ± 11.1 respectively. The AML1-ETO+Meis1 mice showed significantly higher levels of engraftment (median = 93% ±3) sacrificed between 27 to 384 days (n=14) as compared to AE (p<0.03) sacrificed between 90 to 262 days (n=3), Meis1 (p<0.02) sacrificed between 312 to 465 days (n=3) and EGP (p<0.001) sacrificed between 90 to 272 days (n=3). A representative picture of engraftment in BM, spleen and PB of mice transplanted with Meis1+AE transduced cells is shown in Figure 13.
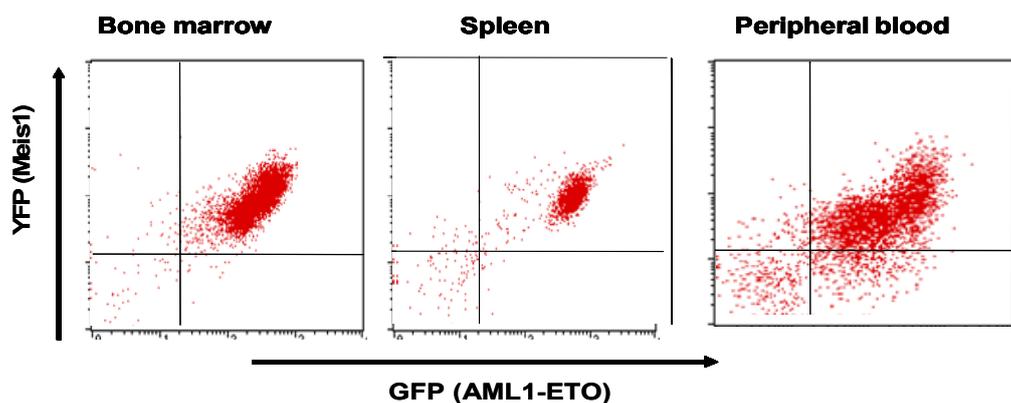
Figure 13: AML1-ETO+Meis1 mice were highly engrafted in BM, SP and PB. Diseased mice were sacrificed and analyzed with flow cytometry to check the engraftment levels of transplanted cells in BM, spleen and PB. This figure shows a representative plot of AE + Meis1 mouse that was analyzed for both YFP (Meis1) and GFP (AML1-ETO) fluorescence.

Survival curves of transplanted mice

The survival of transplanted mice was monitored for 400 days. Mice injected with bone marrow transduced with EGFP (n= 7), Meis1 (n=10) or AML1-ETO (n=10) remained free from disease 365 days after transplantation. However, mice injected with both AML1-ETO and Meis1 (n=14) transduced BM cells died within a range of 27 to 384 days with a median latency of 102 days (Figure 14). Pathological examination revealed death due to leukemia or myeloproliferative syndrome (MPS). Secondary mice transplanted with BM derived from leukemic AE+Meis1 mice died after a median latency of 57 days.
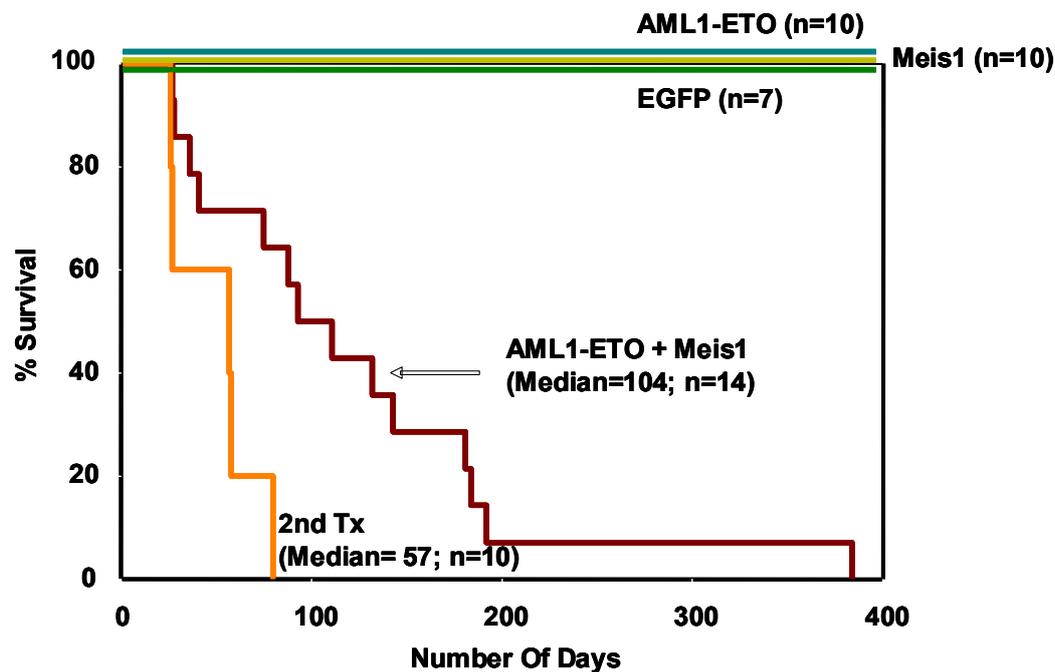
Figure 14: Kaplan–Meier survival curves of mice transplanted with BM cells expressing AML1-ETO (n=10), Meis1 (n=10), control EGFP (n=7) and Meis1+ AML1-ETO (n=14).

The curve also includes the survival of secondary recipient mice (n=10), transplanted with BM from leukemic primary Meis1+ AML1-ETO mice.

Characterization of diseased mice

WBC and RBC counts

Examination of the peripheral blood of mice injected with BM cells transduced with AML1-ETO + Meis1 showed hyperleukocytosis with a very high number of WBCs and low RBC counts when compared to control mice. The RBCs showed a decrease in number and the WBC count was significantly increased with Meis1+AML1-ETO transplanted mice (mean

8.59×10^{8}/ml ± 4.86 and WBC's with 9.00×10^{7}/ml± 1.51) compared to control mice (mean 5.72×10^{6}/ml± 2.91 and $5.21+0^{9}$/ml± 1.93) as shown in Table 5.

Exp. No.	Gene	WBC	RBC
3944b	AE+Meis1	4.00×10^{6}	1.00×10^{9}
3944a	AE+Meis1	4.00×10^{6}	7.50×10^{8}
3912#2	AE+Meis1	2.60×10^{7}	1.00×10^{9}
3933	AE+Meis1	5.00×0^{7}	2.00×10^{9}
3933 #2	AE+Meis1	3.00×07	2.50×10^{8}
3918#2	AE+Meis1	2.10×08	3.50×10^{8}
4525	AE+Meis1	1.70×10^{7}	1.25×10^{9}
4441A#3	AE+Meis1	5.00×10^{7}	1.00×10^{9}
4441A	AE+Meis1	2.25×10^{7}	6.00×10^{8}
4323#3	AE+Meis1	6.50×10^{7}	5.00×10^{8}
4696	AE+Meis1	5.12×08	7.50×10^{8}
4752B#1	AE	2.35×10^{7}	1.2×10^{9}
4752B#5	AE	1.60×10^{7}	7.00×10^{9}
4747#2	AE	1.30×10^{7}	8.00×10^{9}
4638#5	Meis1	8.50×10^{6}	7.20×10^{9}
4638#1	Meis1	1.10×10^{7}	6.50×10^{9}
4519	Meis1	7.00×10^{6}	5.35×10^{9}
4760#2	EGFP	7.00×10^{6}	5.50×10^{9}
4752A	EGFP	1.10×10^{7}	4.50×10^{9}
4760#1	EGFP	5.00×10^{6}	5.25×10^{9}

Table 5: Table showing the hematological parameters of various mice.

Pathology of transplanted mice

Moribund mice were sacrificed and examined for pathological signs by autopsy. Macroscopic examination revealed marked splenomegaly in the mice transplanted with AE+Meis1 BM cells with numerous colonies on the spleen. A list of spleen weight and size is depicted in Table 6.

Exp. No.	Gene	Weight (in mg)	Size (in mm)
3944b	AE+Meis1	296	20x5
3944a	AE+Meis1	300	20x6
3912#2	AE+Meis1	232	20x5
3933 (2610)	AE+Meis1	410	21x6
3933 (1609)#4	AE+Meis1	250	20x5
3933 (1411)#2	AE+Meis1	117	16x4
3918#2	AE+Meis1	400	26x4
4406B#2	AE+Meis1	270	20x3
4155	AE+Mesi1	373	22x4
4155#2	AE+Mesi1	321	20x5
4323#3	AE+Mesi1	348	22x4
4696	AE+Mesi1	603	29x6
4441A#3	AE+Mesi1	300	25x4
4525	AE+Mesi1	290	22x3
4760#1	EGFP	167	15x4
4760#2	EGFP	135	14x3
4752A	EGFP	52.2	13x4
4519	Meis1	126	10x1
4638#5	Meis1	131	13x4
4638#1	Meis1	60	13x2
4747#2	AE	223	18x3
4752B#1	AE	400	27x5
4752B#5	AE	260	23x6

Table 6: Table showing spleen weight and size of the transplanted mice.

For histopathological examination, sacrificed mice were fixed by immersing in 4% buffered formalin solution. Histopathological analyses of various organs revealed severe blast infiltration. The hematopoietic organs like bone marrow, spleen, liver, lymphnodes of these mice demonstrated perivascular infiltration with leukemic cells. Positivity for myeloperoxidase (MPO) and chloracetate esterase (CAE) confirmed the myeloid nature of the

cells. Additional immuno-histological staining to detect myeloid or lymphoid nature of blasts, revealed the presence of lymphoid blasts in a subset of mice.
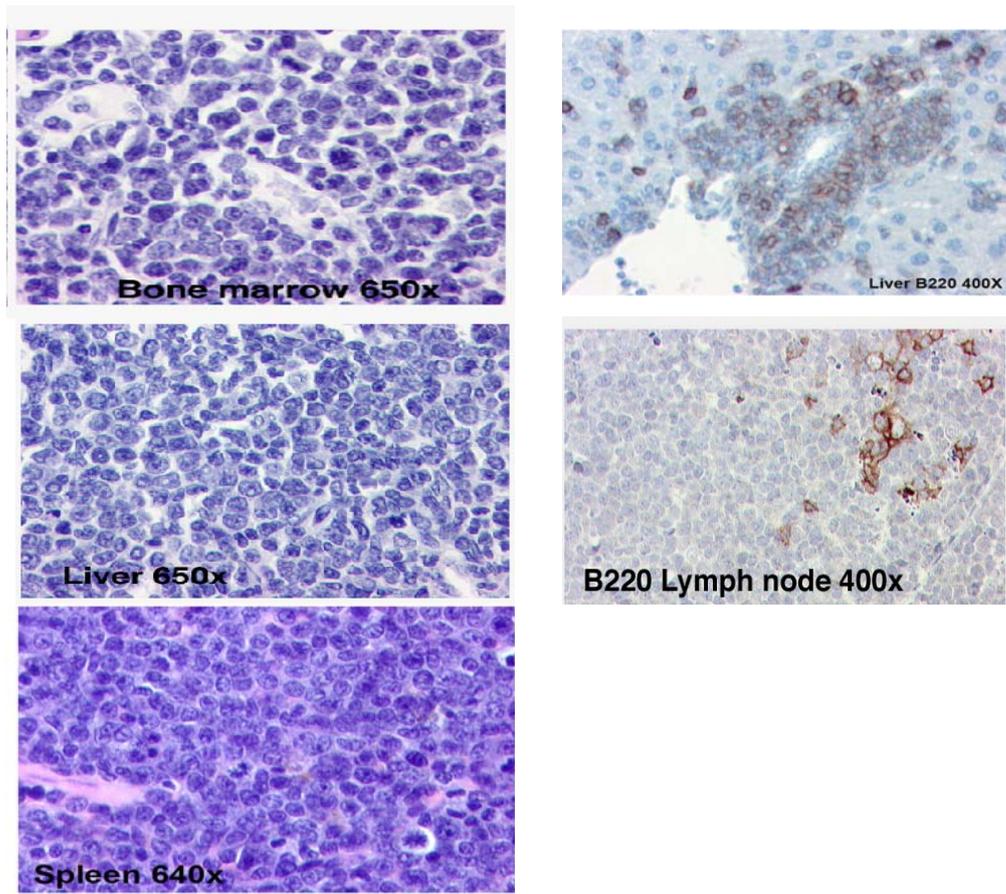
Figure 15: Histological analysis of different organs of a representative leukemic AML1-ETO + Meis1 mouse.

Bone marrow, lymph nodes, kidneys and lungs (Giemsa staining,) of the analyzed animals show infiltration with blast cells.

Morphological examination of infiltrating cells in the hematopoietic organs bone marrow, spleen and peripheral blood was performed by cytospin preparation and May-Grunwald-Giemsa staining. Representative cytospin preparations shown in figure 16 depict immature blast cells in BM and spleen of mice transplanted with Meis1+AML1-ETO transduced BM cells.

ALL   AML
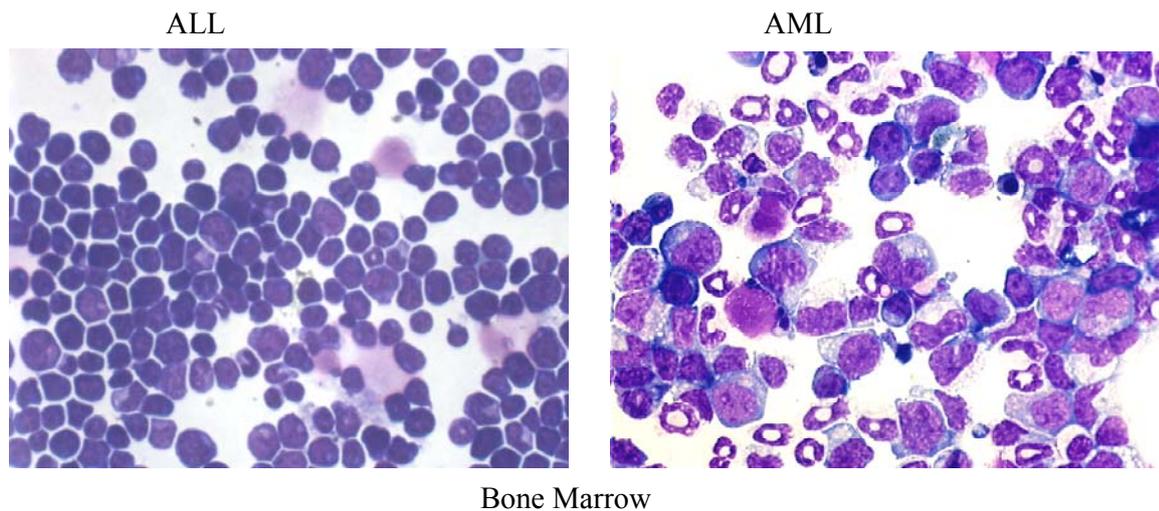
Bone Marrow

Figure 16: Determination of disease by the morphological analysis.
May-Grunwald-Giemsa–stained cytospin preparations of cells isolated from the bone marrow of AML1-ETO+Meis1 mice shows a number of large immature blast cells in both AML and ALL phenotypes.

The morphology, immunohistochemistry and the blast count in the bone marrow of diseased mice revealed three distinct categories of disease in these mice: out of 14 mice analysed a) 3 mice had developed a lethal myeloproliferative syndrome (MPS) with low blast counts (<20%). The myeloproliferative disease was not transplantable into secondary recipients. b) 7 mice suffered from AML with more than 20% blasts in the bone marrow (median= 24 ± 3.5 blasts/100 cells in BM). The BM of the leukemic mice with AML was hypercellular and occupied by immature myeloblasts with finely reticulated chromatin, 2 or more nucleoli, gray-blue cytoplasm, and delicate and azurophilic cytoplasmatic granules. c) 3 mice developed ALL with very high blast counts in the BM (median= 69±7.9 blasts/100 cells in BM).

Mouse Nr.	Exp.No.	Retroviral construct	% PB Blasts	% BM Blasts	% Spleen Blasts	Diagnosis
#1	4752A	EGFP	0	1	0	No disease
#2	4760	EGFP	0	3	1	No disease
#3	4519	Meis1	0	0	0	No disease
#4	4638	Meis1	0	1	0	No disease
#5	4747#2	AESCS	0	1	0	No disease
#6	4752B#5	AESCS	0	6	3	No disease

#7	3944B	AE+Meis1	22	20	21	AML
#8	3944a	AE+Meis1	15	15	19	MPS
#9	3933#1	AE+Meis1	19	21	NA	MPS
#10	3933#2	AE+Meis1	NA	NA	NA	MPS
#11	3933#4	AE+Meis1	NA	NA	NA	ALL
#12	4323#3	AE+Meis1	75	84	72	ALL
#13	3918#2	AE+Meis1	75	63	71	ALL
#14	4525	AE+Meis1	66	43	75	AML
#15	4696	AE+Meis1	ND	ND	ND	AML
#16	3912#2	AE+Meis1	22	31	29	AML
#17	4406B	AE+Meis1	32	78	81	AML
#18	4441A#1	AE+Meis1	ND	ND	ND	AML
#19	4441B#2	AE+Meis1	81	75	83	AML

Table 7: Blast percentage of mice transplanted with AML1-ETO+Meis1, EGFP, Meis1 and AML1-ETO.
Number of blast cells in bone marrow, spleen and peripheral blood of diseased mice and control mice engrafted with Meis1, AML1-ETO and EGFP cells.

Immunophenotyping of transplanted mice

Flow cytometry of surface antibody expression of BM, spleen and PB of AML1-ETO+Meis1 transplanted mice showed that a) seven out of 13 mice characterized as MPS and AML showed 70% of their cells co-expressing Gr-1$^+$ and Mac1$^+$ (mean=84.7%±9.39 in BM); three mice belonging to the AML group expressed the c-Kit marker in a high proportion of cells (mean = 80.78±11.8% in BM) with very few cells being positive for myeloid markers (mean=11.2±9.9 in BM). b) Three mice developed ALL leukemia with high percentage of the cells expressing both the Mac1$^+$ surface marker and the lymphoid associated B220 antigen (mean = 40.16±19.9%).
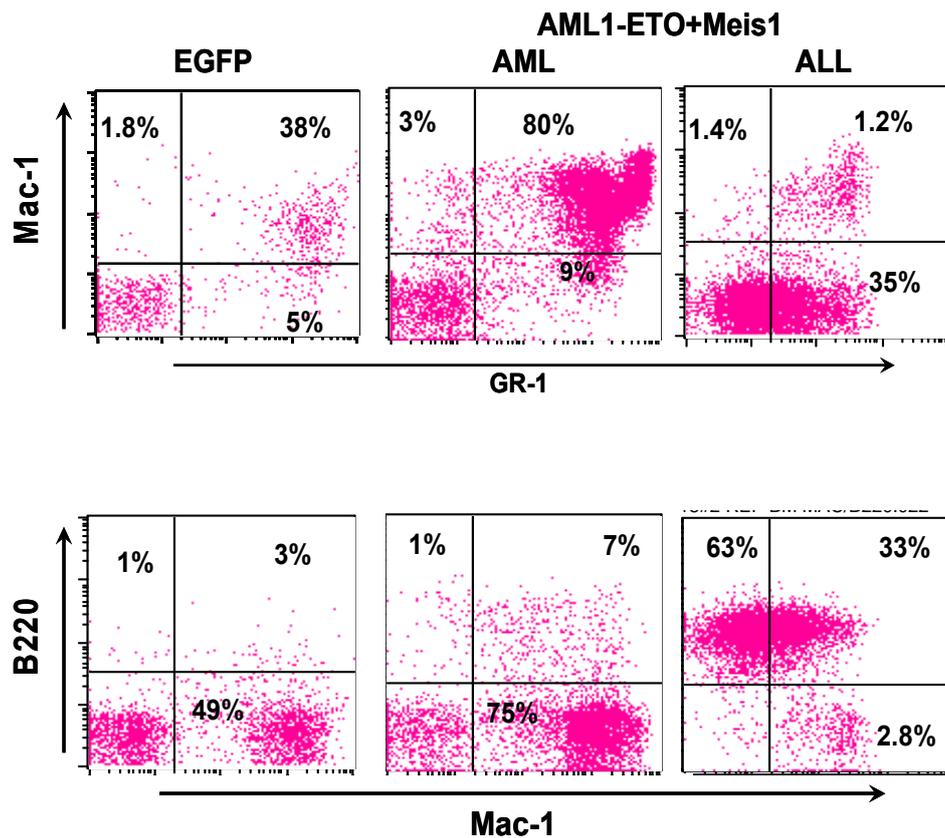
Figure 17: Immunophenotype of cells isolated from the bone marrow of mice that received transplants.
Expression of the myeloid markers Gr-1 and Mac-1 and the lymphoid marker B220 on cells isolated from the bone marrow of different representative animals that received transplants of BM cells transduced with AML1-ETO+Meis1. The proportion of positive cells within the EGFP$^+$ or EYFP$^+$ compartment is indicated. The plots show representative FACS profiles from BM cells of mice with different types of leukemia in comparison with cells from GFP control animals, with indication of the proportion of positive cells within the GFP$^+$/YFP$^+$ compartment.

Transplantation of secondary recipient mice

The self-renewal properties of leukemic cells derived from AML1-ETO+Meis1 transduced cells were tested by transplantation into lethally irradiated secondary recipients. All recipient mice transplanted with 1x10^6 of cells from leukemic mice developed leukemia and died with latencies between 27 and 80 days (median = 57 days; Table 7 & Figure 14). The secondary recipients showed the same disease as the primary mice.

Exp. No	Gene	Cells transplanted	Mock	Days of survival	Disease
4155#1	AE+Meis1	1.00×10^6	1.00×106	27	ALL
4155#2	AE+Meis1	1.00×10^6	1.00×10^6	27	ALL
4155#3	AE+Meis1	1.00×10^6	1.00×10^6	26	ALL
4155#4	AE+Meis1	1.00×10^6	1.00×10^6	26	ALL
4700#1	AE+Meis1	1.00×10^6	1.00×10^6	80	AML
4700#2	AE+Meis1	1.00×10^6	1.00×10^6	80	AML
4828#1	AE+Meis1	1.20×10^6	1.00×10^6	58	ALL
4828#2	AE+Meis1	1.20×10^6	1.00×10^6	58	ALL
4828#3	AE+Meis1	8.00×10^5	1.00×10^6	57	ALL
4828#4	AE+Meis1	8.00×10^5	1.00×10^6	57	ALL

Table 8 List of secondary mice.

Mice transplanted with leukemic cells from primary mice into secondary recipients developed aggressive leukemia with a median latency of 57 days. Mice with exp.nos 4155, 4700, 4828 were transplanted with primary leukemic BM cells from exp.nos 3918, 4525, 4323.

Southern blot analysis of the proviral integration pattern in leukemic cells from AML1-ETO+ Meis1 mice

In order to check the clonality of the disease in AE+Meis1 transplanted mice, we performed Southern blot analysis. For this, DNA from BM cells derived from leukemic mice was digested with *EcoRI*, which cuts once within the proviral sequence and once within the integrated locus, and was subjected to electrophoresis and hybridization with EGFP probes. We observed several bands of different sizes and intensities of proviral integration signals in the different hematopoietic organs such BM, spleen and PB, indicating oligoclonal nature of the disease (Figure.18).
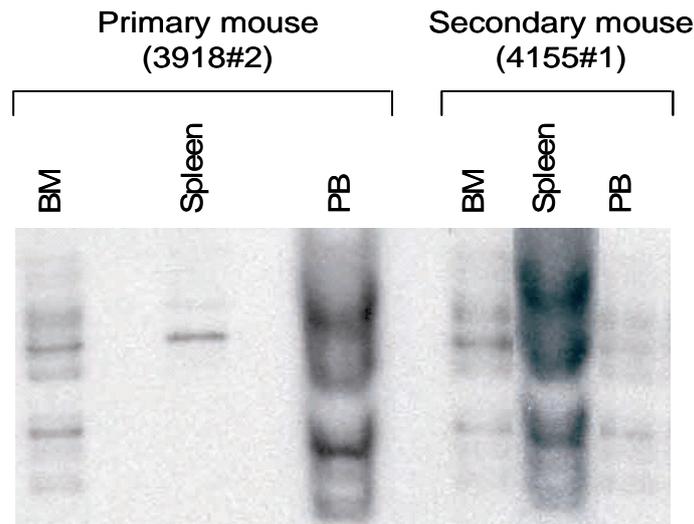
Figure 18. Southern blotting shows multiple integrations of the provirus in diseased mice.
Southern blot analyses of genomic DNA of bone marrow, spleen, peripheral blood from a primary mice and a secondary recipient to detect clonal proviral integrations. DNA was digested with *EcoRI*, which cuts once in the proviral sequence, and blots were hybridized to a GFP/YFP probe.

Frequency of clonogenic cells in leukemic mice - Colony Forming Units (CFU)

Cells isolated from BM, spleen and PB of AE+Meis1 leukemic mice and control mice were plated in methylcellulose assays. Compared to the control, cells derived from AE+Meis1 leukemic mice showed increased proliferative and replating potential (Figure 19). The mean frequency of clonogenic cells in the BM of AE+Meis1 (with AML phenotype) mice was 55.58 ± 40.43 CFU/1×10^4 (N=3) input cells and 220.3 ± 16.63 in BM of AE+Meis1 (with ALL phenotype) (N=2) for the BM as compared to 13.3 ± 8.5 CFU/1×10^4 input cells for the EGFP control (N=3). Replating the primary CFC into secondary CFC yielded 1831 ± 809 CFU/1×10^4 input cells in the AE+Meis1(with AML) and 4475 ± 2925 CFU/1×10^4 input cells in the AE+Meis1(with ALL) arm as compared to 6.6 ± 3.0 CFU/1×10^4 in the EGFP control (N=3). Tertiary CFC assays derived from replating of secondary CFC showed 3991 ± 862 CFU/1×10^4 in the AE+Meis1(with AML) and 5000 ± 0 CFU/1×10^4 in the AE+Meis1(with ALL) arm as compared to 1CFU/1×10^4 in the EGFP control.
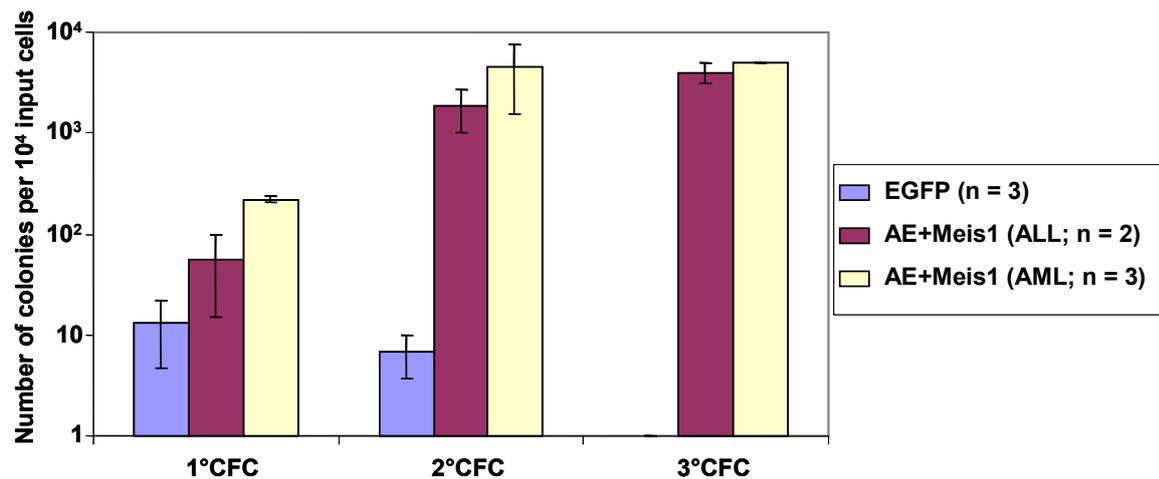
Figure 19 The bar graphs of ex-vivo assay of bone marrow cells isolated from mice transplanted with Meis1+AML1-ETO and EGFP.

The colonies in the methylcellulose dishes were counted every 7 days and were serially replated from the primary CFCs cells.

IgH D-J rearrangements can be detected in myeloid populations of cells with AML1-ETO/Meis1 positive ALL

Leukemic mice were partly characterized by expression of both myeloid and lymphoid antigen (Figure 20A). This finding prompted investigations of the IgH rearrangement status in the different types of leukemias: whereas all AML cases as well as one analyzed MPD case showed germline configuration, all the ALL cases were positive for IgH D-JH rearrangements as shown in Table 9 (Figure 20B).

	Exp.No.	Gr1+ (%)	Mac1+ (%)	Gr1+Mac1 (%)	B220+ (%)	Gr1+B220+ (%)	Mac1+B220+ (%)	IgH DJ rearrangement
AML	12#1	83.83	94.16	79.72	34.95	27.29	39.20	ND
	12#2	85.36	94.80	80.66	32.66	25.17	35.36	negative
	44b#1	90.10	89.78	87.07	6.02	ND	ND	negative
MPS	33#1	76.96	75.00	68.99	7.83	3.15	7.04	negative
	44a#1	98.90	98.40	98.64	15.24	6.01	10.19	negative
ALL	33#2	93.24	74.69	71.30	67.24	64.72	62.72	positve
	18#2	1.93	42.10	1.47	87.55	0.50	33.10	positive
	23#3	10.99	20.83	8.23	94.39	5.96	24.67	positive

Table 9 IgH Rearrangements in AML1-ETO/Meis1 positive leukemias.
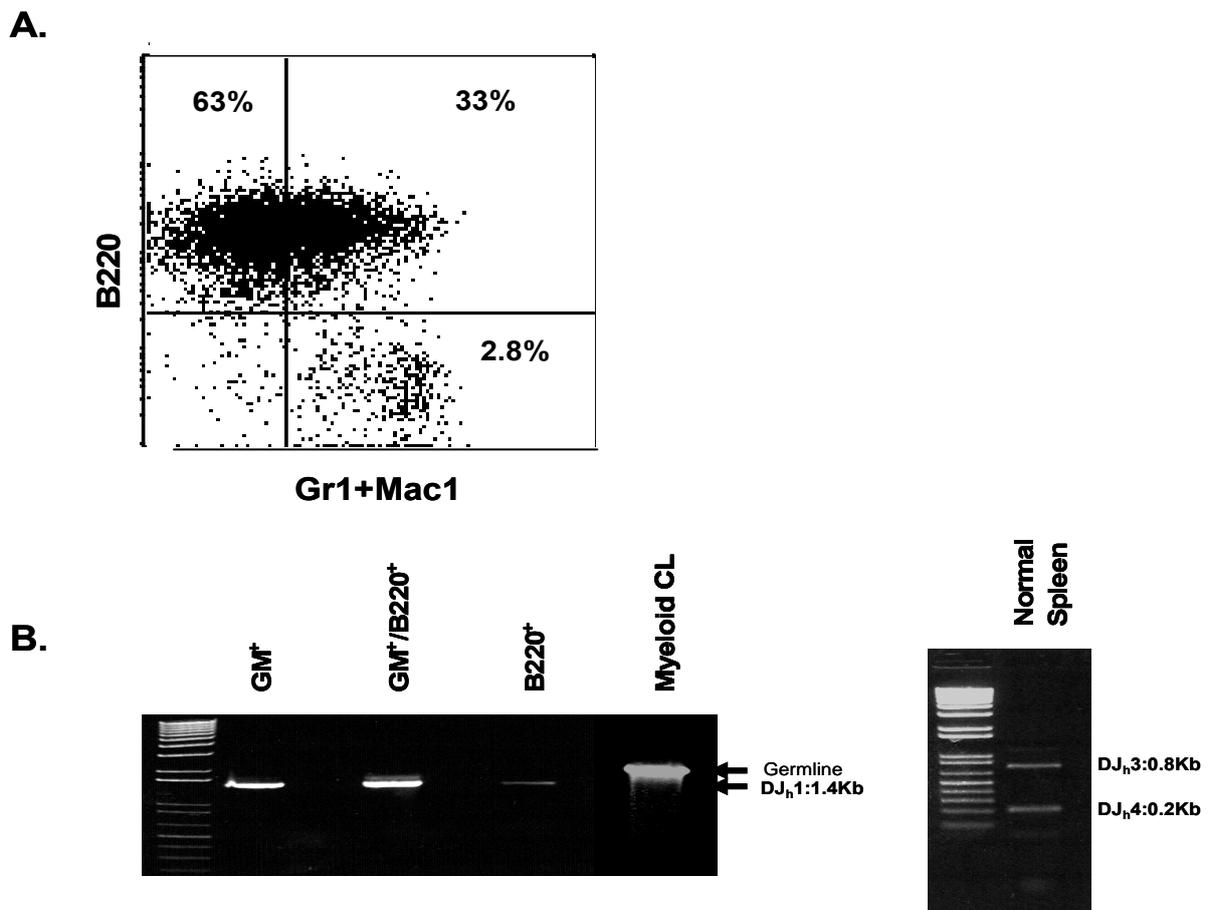
Figure 20 Analysis of IgH DJ rearrangement in different subpopulations of an ALL mouse.

A) Spleen cells of AML1-ETO/Meis1 positive ALL co-expressed both Gr-1+Mac-1+(MM)/B220+ markers. B) The subpopulations B220+/MM-, B220+/MM+, B220-/MM+ were sorted and analysed for DJ rearrangement. The myeloid cell line 32D showed germline configuration and served as a control. Normal spleen CD19+/B220+ cells served as a control for DJ rearrangements.

The B220$^+$ population has the highest frequency of leukemia propagating cells

Since the B (B220$^+$/Gr-1$^-$/Mac-1$^-$), B/M (B220$^+$/Gr-1$^+$Mac-1$^+$) and myeloid marker MM (B220$^-$/Gr-1$^+$Mac-1$^+$) subpopulations were present in all ALL cases, we determined the frequency of the leukemia propagating cell (LPC) in each of the highly purified sub-populations by limiting-dilution secondary transplantation assays. These assays determined the frequency of LPC as 1 in 211 in the B population and 1 in 2,765 cells in the B/M population, whereas no LPC were detectable in the MM population (Table 10). This

demonstrates that the B220+/MM- population in the leukemic mice contains the highest number of cells that are able to propagate leukemia in transplanted mice, hereby supporting the argument that these are the leukemic stem cell candidates in AML1-ETO+Meis1 induced ALL.

Cell Number	B	No. of leukemic mice	Median Latency	B/M	No. of leukemic mice	Median Latency	M/M	No. of leukemic mice	Median Latency
	tested			tested			tested		
50,000	3	3	37	1	1	45	0	0	0
10,000	2	2	39	4	4	45	4	0	0
1,000	4	4	43.5	4	1	61	4	0	0
100	3	1	68	4	0	0	4	0	0

Table 10: Frequency of leukemia propagating cells.

The median latency of mice injected with subpopulations sorted for MM and B220 markers with various dilutions. Mice injected with higher dilutions of B+ population- and B+/MM+ succumbed to leukemia whereas those with B-/MM+ did not show any leukemia. B:(B220+/Gr-1-/Mac-1-), B/M: (B220+/Gr-1+Mac-1+) and myeloid marker MM:(B220-/Gr-1+Mac-1+) subpopulations

Identification of retroviral integration sites in diseased mice

To analyze whether proviral integration might play a role in the development of leukemia in transplanted mice 22 retroviral integration sites were subcloned and sequenced from 3 leukemic mice; all 10 sites were unique, and thus there was no indication of a common integration site associated with the leukemic transformation. Moreover, most of these sites were intergenic or not linked to known genes as shown in Table 11. The remaining sites were within introns in a 5′ to 3′ orientation and most likely led to gene knock down rather than activation. One of the mice showed integration at the genomic region between Hoxb5 and Hoxb4, as well as other multiple integrations targeting hypothetical proteins.

Exp. No	Disease	No.of clones seq.		Chromosome	Gene
3933#1	AML	5	4b1	intron, 9	Transducin like enhancer protien3 and ribosomal protien p1
			4b2	9	hypothetical
			4b3	4	polyhomeotic like-2 (binds with bmi1)
			4b4	4	hypothetical
			5b3	9	hypothetical
3933 #4	AML	4	1a2	intron, 5	hypothetical
			2a1	intron, 5	hypothetical
			2a3	intron,4	hypothetical
			2a4	intron,17	hypothetical
3918#2	ALL	13	7d2	Chromosome. 11	Hoxb4,Hoxb5
			8d2	Chromosome, 8	ringfinger protien
			8d3	Chromosome, 17	luc7 holog like
			8d4	Chromosome,13	semaphorin
			9d2	Chromosome, 15	brain abundant membrane attached signal protien
			9d4	Chromosome, 15	
			10d1	Chromosome, 15	
			10d2		
			10d3		
			10d4		
			11d1		
			12d4		
			13d4		

Table 11: List of integration sites of leukemic mice:

LM-PCR analysis revealed the integration sites of various AML1-ETO+Meis1 mice that developed AML.

AML-ETO L148D mutant does not induce leukemia when co-expressed with Meis1

After these initial results demonstrating a co-operation between Meis1 and AML1-ETO in inducing leukemia *in vivo,* we aimed at analyzing the functional relevance of various domains for A-E induced leukemia. We subcloned AML1-ETO mutants such as AML1-ETO L148D (inactivating the DNA binding domain) in which the leucine (L) residue at the DNA binding region of AML1 was substituted with aspartic acid (D) (Figure 22) (kind gift from S.W.Hiebert). When bone marrow cells expressing *AML1-ETO* L148D and Meis1 were injected into mice there was no incidence of leukemia development upto an observation period of 365 days (N=5) in these mice, similar to mice transplanted with BM expressing the

L148 construct alone (n=5) (Figure 23). Thus, an intact AML1 DNA binding domain is critical for AML1-ETO to collaborate with Meis1.
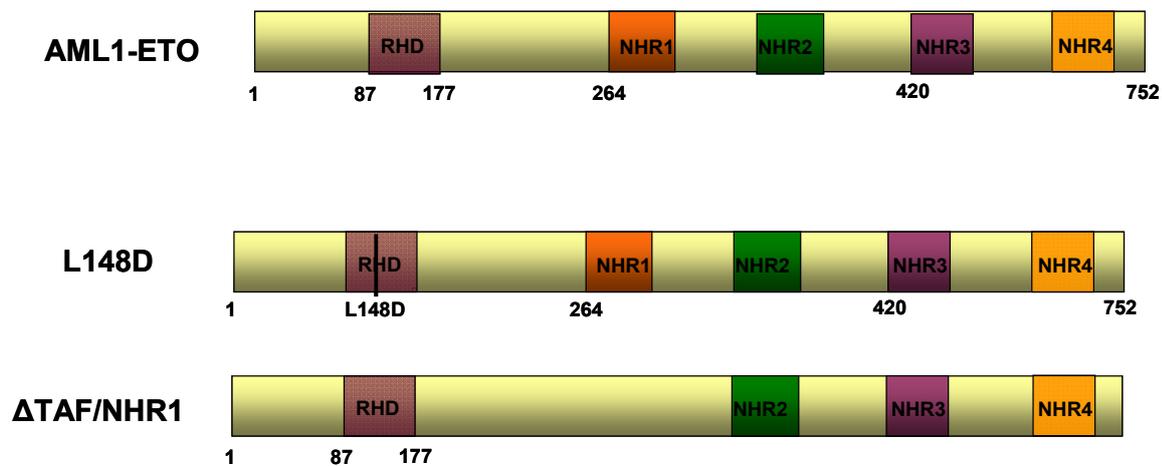
Figure 21: Structure function analysis of the hematopoietic activity of AML1-ETO.
AML1-ETO mutants were generated and functionally tested by co-expressing them with Meis1BM expressing the AML1-ETO L148D mutant and Meis1 (n=5) as well as BM expressing AML1-ETOL148D alone (n=5) were transplanted into mice. In addition mice were transplanted with AML1-ETO ΔTAF plus Meis1 (n=4) or AML1-ETO ΔTAF alone (n=3).

AML1-ETO ΔTAF/NHR1 is critical in inducing leukemia

Another mutant of AML1-ETO made by deleting the TAF domain of the ETO region (277-344 amino acids of AML1-ETO [AML1-ETO ΔTAF]; Figure 22)was subcloned into the MIG vector (Lutterbach et al., 1998). We injected BM cells transduced with AML1-ETO ΔTAF and Meis1 together into lethally irradiated mice. Interestingly, the mutant strongly collaborated with Meis1 in inducing AML. These mice showed high BM leukemic engraftment with a median latency of leukemia of 112 days ranging from 112 to 153 days (Figure 23). This shows that the TAF domain has functional relevance in the normal hematopoiesis.
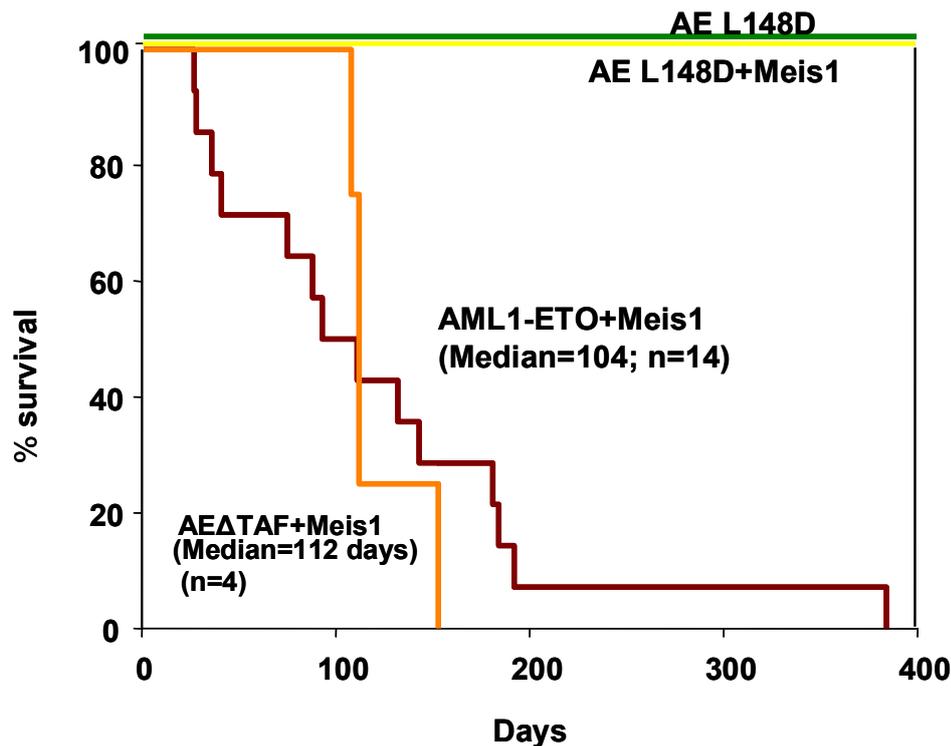
Figure 22: Survival curve of mice transplanted with AML1-ETO L148D and AML1-ETO ΔTAF.
Mice transplanted with AML1-ETO L148D+ Meis1 (N=5) and AML1-ETO L148D (N=5) alone did not disease whereas mice transplanted with AML1-ETO ΔTAF along with Meis1 (N=4) developed AML with a median latency of 112 days. Mice injected with AML1-ETO, Meis1 or EGFP alone also did not show any leukemia.

Detection of spliced variants of AML1-ETO in diseased mice

In order to check the expression of *AML1-ETO* and *Meis1* transcripts in transplanted mice, we performed RT-PCR analysis of RNA derived from sacrificed leukemic mice using different sets of primers (Figure 24). Whereas *Meis1* expression was detectable in the BM cells by RT-PCR, we failed to detect any PCR product with primers designed to detect full length *AML1-ETO*, but got positive results using primers spanning the breakpoint region of AML1-ETO. To investigate this further, we performed genomic PCR on BM samples of diseased mice. The genomic DNA PCR showed alternatively spliced fusion transcripts of *AML1-ETO* in all mice tested (Figure 25). Sequencing of the PCR products confirmed spliced fusion transcripts. These mice predominantly showed truncated *AML1-ETO* at the region of the TAF domain caused by an in-frame deletion of sequences in the TAF domain. This spliced variant could collaborate with *Meis1* to cause AML in transplanted mice as described before.
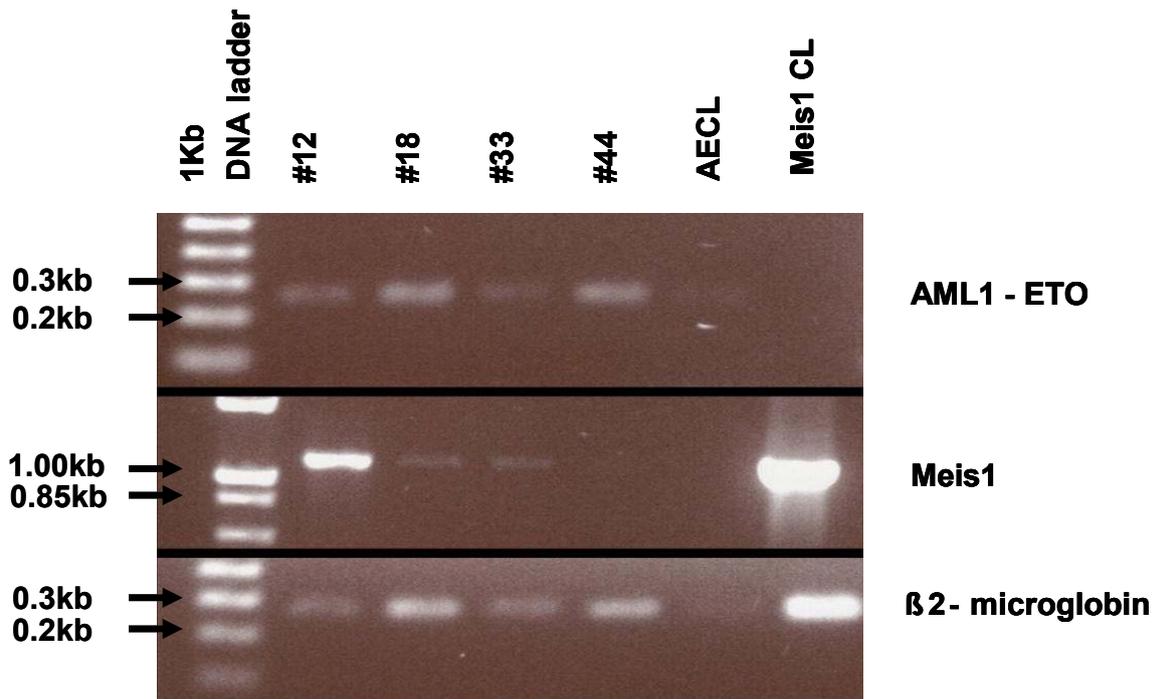
Figure 23: RT-PCR showing expression of *AML1-ETO* and *Meis1* in mice co-expressing both.

RT-PCR to determine the expression of *AML1-ETO* and *Meis1*: PCR was performed using the AML1-ETO specific primers spanning the break point region and for *Meis1* expression with *Meis1* specific primers. Expression of both genes was confirmed by sequencing the PCR product in all experiments shown. Normalization was performed with the housekeeping gene ß2-microglobin. E86 AML1-ETO cell line (AECL), Meis1 cell line (Meis1CL)
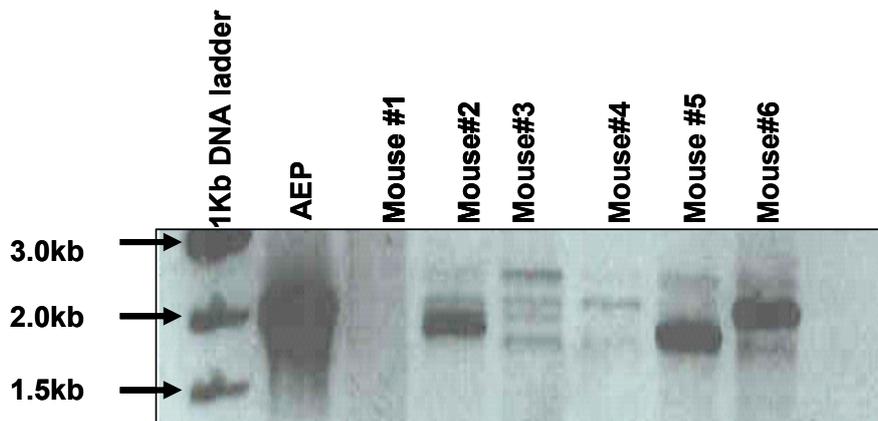
Figure 24: PCR showing different spliced variants of *AML1-ETO* in transplanted mice.

PCR from genomic DNA to determine the spliced variants using AML1-ETO specific primers, the AML1-ETO plasmid served as control (AEP).

We then performed PCR to detect alternatively spliced variants in human AML. We screened 14 patients with t(8;21) by RT-PCR and sequencing. Novel spliced variants were detected in

most of the patient samples analysed (Figure 26). One variant which occurred due to a frame shift mutation resulting in an extra exon 9 (AML1-ETO9a) which leads to a C-terminal deletion, the other variant being an out of frame deletion of 243 bp in the ETO region immediately after few amino acids of its N-terminal region leading to the disruption of c-terminal region of ETO. The variants are depicted schematically in Figure 27. These results underline the importance of studying spliced variants in leukemogenesis in t(8;21) cases.
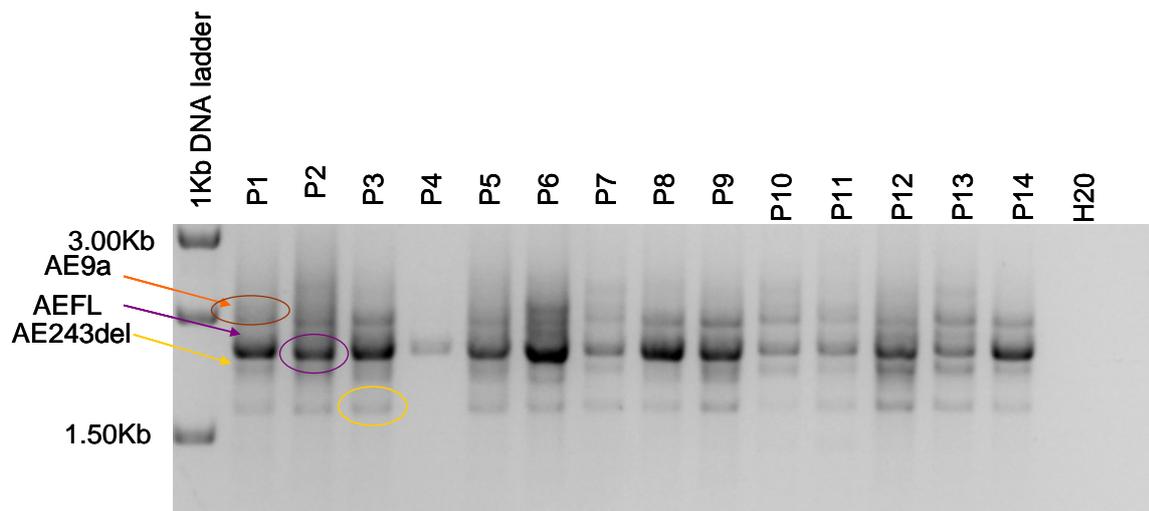
Figure 25: RT-PCR determining splicing in *AML1-ETO* positive patients.

An RT-PCR was performed to determine the splicing variants in *AML1-ETO* patients. PCR products were run on 1% agarose gel, the bands were purified and sequenced yielding 3 different variants.
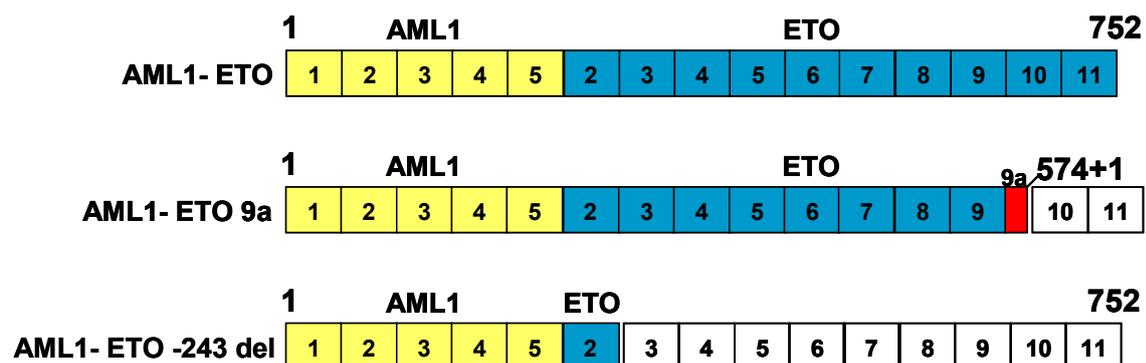
Figure 26: Schematic representation of spliced variants of *AML1-ETO* observed in t(8;21) positive patients.

DISCUSSION

AML results from the malignant transformation of self renewing pluripotent HSC or their progenitors which acquire the self renewing program of stem cells. However, AML is a heterogeneous disorder, in which non-random, somatically acquired chromosomal abnormalities are the most commonly observed genetic irregularities (Buske and Humphries, 2000; Rowley, 1999). The genes which are most commonly affected by chromosomal translocations in AML patients are transcription factors which normally play a crucial role in hematopoietic differentiation. The t(8;21) translocation is one of the common translocations identified in leukemia and fuses chromosome 8q22 to chromosome 21q22 leading to the formation of the *AML1-ETO* fusion gene. Although this translocation is found in around 12% AML cases(Downing, 1999; Nucifora and Rowley, 1995), and more than 40% of AML M2 subtype (Bitter et al., 1987), it was demonstrated in various experimental models that AML1-ETO is not able to induce leukemia on its own (Rhoades et al., 2000; Yuan et al., 2001). The fact that AML1-ETO alone is not leukemogenic is further supported by findings that non-leukemic *AML1-ETO* expressing progenitor cells can be isolated from healthy individuals as well as AML patients in remission, which suggests that additional mutations in these AML1-ETO–positive progenitors are necessary for the transformation into leukemia-initiating cells (Miyamoto et al., 2000; Nucifora et al., 1993). Recently published data from our lab and others demonstrate that AML1-ETO collaborates with secondary mutations such as FLT3-LM (Schessl et al., 2005), TEL/PDGFβR (Grisolano et al., 2003) and WT1 (Nishida et al., 2006), to rapidly induce leukemia. However, as FLT3-LM is present in only 8.1% of patients with AML1-ETO rearrangement (Schessl et al., 2005) it could be that there are several other secondary oncogenic events involved in the induction of leukemia in AML1-ETO positive patients. Several lines of evidences exist that MEIS1, a Hox co-factor which plays an important role in hematopoiesis (Pineault et al., 2002), is highly upregulated in MLL-related leukemias (Wong et al., 2007). Additionally, the importance of MEIS1 in human leukemogenesis is underscored by the finding that it is frequently up-regulated in primary AML and ALL samples (Imamura et al., 2002; Rozovskaia et al., 2001). In this thesis we therefore sought to assess the role of Meis1 as a collaborative gene of AML1-ETO. Meanwhile it has been shown that Meis1 plays an important role in accelerating leukemia in Hox gene mediated leukemogenesis when co-expressed with Hoxa9, Hoxa7, and NUP98-Hox fusion genes (Kroon et al., 1998; Lawrence et al., 1999; Pineault et al., 2002). We found that *MEIS1* was expressed in a subset of AML1-ETO positive AML samples with a level of

expression similar to that of normal human bone marrow. Thus, in order to study the functional significance of t(8;21) in leukemogenesis and its possible collaboration with Meis1 overexpression as a secondary event, we co-expressed the AML1-ETO fusion gene with Meis1. Expression of both these genes together but not alone, led to the induction of acute leukemias with various phenotypes. The reason that these mice showed different disease types could be due to the fact that different cell types have been targeted, as already shown in the AML1-ETO+FLT3-LM model where the fusion gene and FLT3 collaborate in inducing myeloid, B-cell and T-cell leukemia (Schessl et al., 2005).

Mice injected with Meis1 transduced bone marrow cells did not show any sign of hematopoietic perturbation or development of disease, demonstrating that Meis1 on its own is non-leukemogenic, confirming previous observations (Kroon et al., 1998). In addition mice transplanted with BM expressing AML1-ETO alone also did not develop leukemia but showed increased spleen size. It has been already reported that expression of AML1-ETO in HSC induces non-lethal long latency myeloproliferation (Fenske et al., 2004). It is also known that AML1-ETO inhibits the differentiation of myeloid progenitors, resulting in an increase in their replating ability and expansion of clonogenic myeloid progenitors *in vitro* ((Higuchi et al., 2002; Okuda et al., 1998) and induced myelodysplasia *in vivo*, in which immature myeloid cells were significantly increased in the BM and spleen (de Guzman et al., 2002; Fenske et al., 2004; Grisolano et al., 2003). These results indicated that AML1-ETO–expressing hematopoietic progenitors remained responsive to normal *in vivo* homeostatic controls, were at pre-leukemic stages, and required additional genetic mutations or events to gain full leukemic transformation. In our mouse model co-expression of Meis1 with AML1-ETO readily transformed hematopoietic progenitor cells resulting in acute leukemia, indicating that Meis1 over expression could be one of the additional oncogenic events required.

Perturbation of hematopoiesis by expression of AML1-ETO results in an increase in the replating capacity of murine clonogenic progenitors (Higuchi et al., 2002). Similarly, in our model the *ex-vivo* analysis of BM, spleen and PB from diseased mice using methylcellulose CFU assays could show colonies formed of tightly aggregated cell clusters that were entirely composed of immature myeloblasts. Successful serial replating of these cultures proved that blast cells expressing both AML1-ETO and Meis1 are highly clonogenic and showed replating capacity. In comparison to EGFP control mice, the clonogenic potential of cells

derived from leukemic mice were increased manifold. Southern blot analyses confirmed the oligoclonal nature of the disease in mice, indicating that the combination of AML1-ETO and Meis1 is highly leukemogenic. Insertional mutagenesis mediated by integration of the retroviral construct could potentially constitute an oncogenic event (Kustikova et al., 2005). To rule out the role of insertional mutagenesis in our model, we analyzed the genomic DNA from BM samples of leukemic mice. Analyses of the diseased mice showed integration into intragenic regions or introns of genes, more likely resulting in their knock down than in their activation except one mouse in which we found intergenic integrations between Hoxb4 and Hoxb5.

In order to study the role of AML1-ETO domains, we did functional studies of AML1-ETO mutants with regard to their collaborative potential with Meis1. To test the relevance of the AML1 DNA binding domain in this respect, mice were transplanted with AEL148D (with a point mutation in the DNA binding region of AML1) together with Meis1: none of the mice developed disease. These data are in line with published data showing that this mutant lacks the ability to collaborate with FLT3-LM (Schessl et al., 2005) or TEL/PDGFRβ (Grisolano et al., 2003) to induce leukemia. The PCR analysis of *AML1-ETO* in mice transplanted with AML1-ETO + Meis1 transduced BM cells showed truncated *AML1-ETO* variants with an in-frame deletion of the TAF domain (also called NERVY homology region 1) of ETO. The TAF domain has been shown to be critical for the E protein silencing by ETO (Plevin et al., 2006). It has also been reported that this domain is responsible for protein-protein interaction with N-CoR (nuclear receptor co-repressor) (Lausen et al., 2004) When we co-expressed ΔTAF mutant with Meis1, it resulted in AML suggesting that this domain is dispensable and not important in AML1-ETO induced pathogenesis.

Another important observation in our study was the occurrence of a significant number of cells in ALL mice displaying both myeloid as well as lymphoid markers (B220$^+$ and Mac-1$^+$ Gr-1$^+$), and a considerable cell population that displayed only lymphoid markers (B220$^+$/Mac-1$^-$/Gr-1$^-$). Since the cell populations observed here showed expression of a B-lymphoid associated marker we investigated the presence of IgH genomic rearrangements by PCR. Indeed we found that all three cell fractions were positive for the genomic DJ rearrangements. In order to determine the presence of a hierarchy in the leukemic cell population, we performed leukemic CRU assay in one of these ALL mice to determine the frequency of leukemia initiating cell in various cell populations. The frequency of leukemia

initiating cell as assessed by the transplantability of leukemia was significantly higher in the $B220^+/Mac1^-/Gr1^-$ fraction as compared to the other two fractions ($B220^+/myeloid\ marker^+$ and the $B220^-/\ myeloid\ marker^+$).

The fact that myeloid leukemias and leukemias with high c-kit marker expression were frequently observed in our models recapitulates previous observations from other murine models of AML1-ETO positive leukemia (Nishida et al., 2006; Schessl et al., 2005; Yan et al., 2006). In other mice, we saw that lymphoid markers were expressed in addition to myeloid markers as mentioned above. There could be multiple reasons for the presence of these lymphoid marker expressing cells in our AML1-ETO/Meis1 mice. One possible explanation is that AML in some of our mice is propagated by a leukemia-initiating cell with lymphoid characteristics similar to the already published data with the leukemogenic CALM-AF10 fusion gene (Deshpande et al., 2006). An alternative possibility could be that aberrant lymphoid marker expression is initiated in the leukemic cells of the AML1-ETO/Meis1 mice. It is worth noting here that a high percentage of AML1-ETO positive AMLs show lymphoid marker expression. It would be interesting to identify the LSC in these human leukemias to identify whether these AMLs derive, as in some of our mice, from a lymphoid marker expressing LSC. This would have important therapeutic implications for the design of LSC specific drugs in these leukemias.

This thesis also describes the presence of certain spliced variants of *AML1-ETO*. It is known from published data that *AML1-ETO* splicing occurs resulting in variants which play role in inducing leukemia. Characterization of some t(8;21) patients showed deletions at the 5'end region of ETO encompassing from 260kb to 2Mb (van de Locht et al., 1994). Similarly, in a t(8;21) patient with M2 subtype, it was observed that the reading frame of the transcript was disrupted due to a 50bp frame shift mutation (Godon et al., 2002; Lasa et al., 2002; van de Locht et al., 1994). These results show that *AML1-ETO* undergoes splicing events resulting in isoforms which might lead to pathogenesis. In our model, some leukemic mice showed spliced variants along with full length *AML1-ETO*. The most frequent variant we observed was the in-frame deletion of TAF domain resulting in truncation of *AML1-ETO*. When we screened patients positive for t(8;21) translocation for spliced variants we observed two different variants in almost all the samples except those with complete remission. Both variants observed in patients were expressed along with full length *AML1-ETO* and it has been already shown that one of the variant (AML1-ETO9a) induces leukemia on its own in

mouse models (Yan et al., 2006). These results suggest that spliced variants of AML1-ETO may induce leukemia and that these variants itself could also act as one of the cooperating oncogenic partners for full-length AML1-ETO in these patients.

Taken together, the data of this theses provide the first functional evidence of a leukemogenic collaboration of the translocation AML1-ETO with the Hox gene co-factor Meis1, extending our knowledge about the biology of one of the most frequent subtypes of human AML

SUMMARY

One of the most common features of leukemia is the formation of reciprocal chromosomal rearrangements often leading to the formation of aberrant fusion genes. Even though several of these fusion genes have been shown to directly contribute to malignancy, there is ample evidence to indicate the requirement of additional oncogenic events in leukemias involving these fusion genes. AML1-ETO, a product of the t(8;21) translocation is one of the common and recurrent fusion genes observed in AML. In some of the AML1-ETO positive AML, other co-operating mutations like mutations in various receptor tyrosine kinases such as FLT-3, c-KIT and NRAS have been reported. However most of the AML1-ETO patients do not harbor any mutations in these genes suggesting that there are several other unknown partner genes involved in t(8;21) mediated leukemias. The identification of these additional co-operating mutations is necessary for the development of therapeutic strategies for t(8;21) associated leukemias. Meis1 is a Hox co-factor which plays an important role in accelerating leukemia mediated by several leukemogenic Hox genes and HOX gene fusions such as HOXA9 and the fusion gene NUP98-HOXD13. Considering its importance in leukemogenesis, we sought to determine if this homeobox gene can also collaborate with leukemia associated non-Hox fusion proteins to induce leukemogenesis. Therefore, the aim of this thesis was to determine whether the homeobox gene Meis1 can also collaborate with AML1-ETO in inducing leukemia.

In summary we could show that AML1-ETO and Meis1 could induce leukemia when expressed in tandem, but not alone, in bone marrow cells. Acute leukemias with various phenotypes such as AML, ALL and myeloproliferative disease could be observed in different mice which was confirmed by characteristic features of perturbed hematological parameters, splenomegaly, and a high percentage of cells expressing myeloid markers in AML and lymphoid marker B220 in ALL. These leukemias were also transplantable as assessed by secondary transplantation. Moreover, immuno-histochemical staining and morphology of the cells demonstrated the presence of blast cells in hematopoietic organs of the diseased mice unlike the control mice. Moreover, *ex vivo* analyses of cells from AML1-ETO+Meis1 transplanted leukemic mice showed that these cells were serially replatable, thereby indicating high clonogenic potential. In AML1-ETO+/Meis1+ ALL developing in transplanted mice, the leukaemia propagating cell displayed a $B220^{+}/MM^{-}$ phenotype.

In conclusion, we demonstrated for the first time that the commonly occurring chromosomal translocation AML1-ETO induces acute leukemia in collaboration with Meis1, a Hox gene co-factor, in the murine bone marrow transplantation model.

ZUSAMMENFASSUNG

Ein wesentliches Merkmal von Leukämien besteht im Auftreten von reziproken Translokationen, die oftmals zur Bildung von Fusionsproteinen führen. Obwohl bereits eine direkte Beteiligung einiger dieser Fusionsproteine an der Entstehung von Leukämien gezeigt werden konnte, bestehen vielseitige Hinweise auf die Mitwirkung weiterer onkogener Ereignisse an der Entstehung der Leukämie. AML1-ETO, resultierend aus der Translokation t(8;21), ist eines der am häufigsten beobachteten Fusionsproteine bei der AML. In einigen Fällen von AML1-ETO positiver AML wurden kooperierende Mutationen, unter anderem in verschiedenen Rezeptortyrosinkinasen wie FLT3, c-KIT und RAS, gefunden. Jedoch besitzt die überwiegende Mehrheit an AML1-ETO positiven Patienten keine derartige Mutationen, was für eine Beteiligung weiterer, noch unbekannter Gene an der t(8;21) assoziierten Leukämie spricht. Die Identifizierung dieser kooperierenden Mutationen ist grundlegend für die Entwicklung neuer therapeutischer Strategien zur Behandlung der t(8;21) assoziierten Leukämie.

In Leukämien, welche durch leukämogene Hox-Gene sowie Hox-Fusionsgene wie HOXA9-Fusionen oder NUP98-HOXD13 ausgelöst werden, spielt der Hox-Kofaktor Meis1 eine bedeutende Rolle in der Beschleunigung des Krankheitsverlaufes. Auf Grund der Bedeutsamkeit dieses Homeobox-Gens in der Leukämogenese stellte sich uns die Frage, ob Meis1 gleichermaßen mit leukämogenen, Nicht-Hox-Fusionsprotein, kooperiert. Ziel dieser Arbeit war es daher zu untersuchen, in wie weit das Homeobox-Gen Meis1, zusammen mit AML1-ETO, an der Induktion von Leukämie beteiligt ist.

Während die Expression eines dieser Gene allein keinen Effekt zeigt, konnten wir nachweisen dass eine gemeinsame Expression beider Gene in Knochenmarkzellen Leukämie induziert. Erkrankte Mäuse zeigten phänotypische Merkmale von AML, ALL und myeloproliferativer Erkrankung, die durch charakteristische Kennzeichen wie veränderte hämatologische Parameter, Splenomegalie und der Expression von myeloiden Markern in AML bzw. von lymphoiden Markern in ALL bestätigt wurden. Die Transplantabilität der Leukämien wurde durch sekundäre Transplantationen belegt. Desweiteren konnte mittels immunohistochemicher Färbung und Morphologie der Zellen ein Vorhandensein von Blasten in den hämatopoetischen Organen der erkrankten Mäusen nachgewiesen werden. *Ex vivo* Anlaysen von Zellen AML1-ETO + Meis1 transplantierter Mäuse zeigten das Vorhandensein

eines hohen klonogen Potentials. Die leukämie-propagierende Zelle in der AML1-ETO+
/Meis1+ ALL zeigte einen B220+/MM- Phänotyp.

Zusammenfassend konnten wir erstmalig zeigen, dass die häufig auftretende chromosomale
Translokation AML1-ETO in Kollaboration mit dem Hox-kofaktor Meis1 in der Lage ist im
Maus-Modell Leukämie zu induzieren.

REFERENCES

(1990). Acute myelogenous leukemia with an 8;21 translocation. A report on 148 cases from the Groupe Francais de Cytogenetique Hematologique. Cancer Genet Cytogenet *44*, 169-179.

Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J., and Clarke, M.F. (2003). Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A *100*, 3983-3988.

Alcalay, M., Tiacci, E., Bergomas, R., Bigerna, B., Venturini, E., Minardi, S.P., Meani, N., Diverio, D., Bernard, L., Tizzoni, L*., et al.* (2005). Acute myeloid leukemia bearing cytoplasmic nucleophosmin (NPMc+ AML) shows a distinct gene expression profile characterized by up-regulation of genes involved in stem-cell maintenance. Blood *106*, 899-902.

Allsopp, R.C., Cheshier, S., and Weissman, I.L. (2001). Telomere shortening accompanies increased cell cycle activity during serial transplantation of hematopoietic stem cells. J Exp Med *193*, 917-924.

Antonchuk, J., Sauvageau, G., and Humphries, R.K. (2002). HOXB4-induced expansion of adult hematopoietic stem cells ex vivo. Cell *109*, 39-45.

Azcoitia, V., Aracil, M., Martinez, A.C., and Torres, M. (2005). The homeodomain protein Meis1 is essential for definitive hematopoiesis and vascular patterning in the mouse embryo. Dev Biol *280*, 307-320.

Bhatia, M., Wang, J.C., Kapp, U., Bonnet, D., and Dick, J.E. (1997). Purification of primitive human hematopoietic cells capable of repopulating immune-deficient mice. Proc Natl Acad Sci U S A *94*, 5320-5325.

Bitter, M.A., Le Beau, M.M., Rowley, J.D., Larson, R.A., Golomb, H.M., and Vardiman, J.W. (1987). Associations between morphology, karyotype, and clinical features in myeloid leukemias. Hum Pathol *18*, 211-225.

Blair, A., Hogge, D.E., Ailles, L.E., Lansdorp, P.M., and Sutherland, H.J. (1997). Lack of expression of Thy-1 (CD90) on acute myeloid leukemia cells with long-term proliferative ability in vitro and in vivo. Blood *89*, 3104-3112.

Bonnet, D., and Dick, J.E. (1997). Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat Med *3*, 730-737.

Burglin, T.R. (1997). Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals. Nucleic Acids Res *25*, 4173-4180.

Buske, C., and Humphries, R.K. (2000). Homeobox genes in leukemogenesis. Int J Hematol *71*, 301-308.

Cozzio, A., Passegue, E., Ayton, P.M., Karsunky, H., Cleary, M.L., and Weissman, I.L. (2003). Similar MLL-associated leukemias arising from self-renewing stem cells and short-lived myeloid progenitors. Genes Dev *17*, 3029-3035.

de Guzman, C.G., Warren, A.J., Zhang, Z., Gartland, L., Erickson, P., Drabkin, H., Hiebert, S.W., and Klug, C.A. (2002). Hematopoietic stem cell expansion and distinct myeloid developmental abnormalities in a murine model of the AML1-ETO translocation. Mol Cell Biol *22*, 5506-5517.

Deshpande, A.J., Cusan, M., Rawat, V.P., Reuter, H., Krause, A., Pott, C., Quintanilla-Martinez, L., Kakadia, P., Kuchenbauer, F., Ahmed, F.*, et al.* (2006). Acute myeloid leukemia is propagated by a leukemic stem cell with lymphoid characteristics in a mouse model of CALM/AF10-positive leukemia. Cancer Cell *10*, 363-374.

Domen, J., Gandy, K.L., and Weissman, I.L. (1998). Systemic overexpression of BCL-2 in the hematopoietic system protects transgenic mice from the consequences of lethal irradiation. Blood *91*, 2272-2282.

Downing, J.R. (1999). The AML1-ETO chimaeric transcription factor in acute myeloid leukaemia: biology and clinical significance. Br J Haematol *106*, 296-308.

Downing, J.R., Head, D.R., Curcio-Brint, A.M., Hulshof, M.G., Motroni, T.A., Raimondi, S.C., Carroll, A.J., Drabkin, H.A., Willman, C., Theil, K.S.*, et al.* (1993). An AML1/ETO fusion transcript is consistently detected by RNA-based polymerase chain reaction in acute myelogenous leukemia containing the (8;21)(q22;q22) translocation. Blood *81*, 2860-2865.

Erickson, P., Gao, J., Chang, K.S., Look, T., Whisenant, E., Raimondi, S., Lasher, R., Trujillo, J., Rowley, J., and Drabkin, H. (1992). Identification of breakpoints in t(8;21) acute myelogenous leukemia and isolation of a fusion transcript, AML1/ETO, with similarity to Drosophila segmentation gene, runt. Blood *80*, 1825-1831.

Erickson, P.F., Robinson, M., Owens, G., and Drabkin, H.A. (1994). The ETO portion of acute myeloid leukemia t(8;21) fusion transcript encodes a highly evolutionarily conserved, putative transcription factor. Cancer Res *54*, 1782-1786.

Estey, E., and Dohner, H. (2006). Acute myeloid leukaemia. Lancet *368*, 1894-1907.
Fenske, T.S., Pengue, G., Mathews, V., Hanson, P.T., Hamm, S.E., Riaz, N., and Graubert, T.A. (2004). Stem cell expression of the AML1/ETO fusion protein induces a myeloproliferative disorder in mice. Proc Natl Acad Sci U S A *101*, 15184-15189.

Fischbach, N.A., Rozenfeld, S., Shen, W., Fong, S., Chrobak, D., Ginzinger, D., Kogan, S.C., Radhakrishnan, A., Le Beau, M.M., Largman, C.*, et al.* (2005). HOXB6 overexpression in murine bone marrow immortalizes a myelomonocytic precursor in vitro and causes hematopoietic stem cell expansion and acute myeloid leukemia in vivo. Blood *105*, 1456-1466.

Gelmetti, V., Zhang, J., Fanelli, M., Minucci, S., Pelicci, P.G., and Lazar, M.A. (1998). Aberrant recruitment of the nuclear receptor corepressor-histone deacetylase complex by the acute myeloid leukemia fusion partner ETO. Mol Cell Biol *18*, 7185-7191.

Gilliland, D.G. (2002). Molecular genetics of human leukemias: new insights into therapy. Semin Hematol *39*, 6-11.

Gilliland, D.G., and Tallman, M.S. (2002). Focus on acute leukemias. Cancer Cell *1*, 417-420. Godon, C., Proffitt, J., Dastugue, N., Lafage-Pochitaloff, M., Mozziconacci, M.J., Talmant, P., Hackbarth, M., Bataille, R., and Avet-Loiseau, H. (2002). Large deletions 5' to the ETO breakpoint are recurrent events in patients with t(8;21) acute myeloid leukemia. Leukemia *16*, 1752-1754.

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A.*, et al.* (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science *286*, 531-537.

Grisolano, J.L., O'Neal, J., Cain, J., and Tomasson, M.H. (2003). An activated receptor tyrosine kinase, TEL/PDGFbetaR, cooperates with AML1/ETO to induce acute myeloid leukemia in mice. Proc Natl Acad Sci U S A *100*, 9506-9511.

Gross, C.T., and McGinnis, W. (1996). DEAF-1, a novel protein that binds an essential region in a Deformed response element. EMBO J *15*, 1961-1970.

Guan, Y., Gerhard, B., and Hogge, D.E. (2003). Detection, isolation, and stimulation of quiescent primitive leukemic progenitor cells from patients with acute myeloid leukemia (AML). Blood *101*, 3142-3149.

Harris, N.L., Jaffe, E.S., Diebold, J., Flandrin, G., Muller-Hermelink, H.K., Vardiman, J., Lister, T.A., and Bloomfield, C.D. (1999). World Health Organization classification of neoplastic diseases of the hematopoietic and lymphoid tissues: report of the Clinical Advisory Committee meeting-Airlie House, Virginia, November 1997. J Clin Oncol *17*, 3835-3849.

Hiddemann, W., Haferlach, T., Schoch, C., Behre, G., Kohlmann, A., Kern, W., Feuring-Buske, M., and Bohlander, S. (2003). New insights into the biology of acute myeloid leukemia and their impact on treatment. Verh Dtsch Ges Pathol *87*, 72-78.

Higuchi, M., O'Brien, D., Kumaravelu, P., Lenny, N., Yeoh, E.J., and Downing, J.R. (2002). Expression of a conditional AML1-ETO oncogene bypasses embryonic lethality and establishes a murine model of human t(8;21) acute myeloid leukemia. Cancer Cell *1*, 63-74.

Huntly, B.J., and Gilliland, D.G. (2004). Blasts from the past: new lessons in stem cell biology from chronic myelogenous leukemia. Cancer Cell *6*, 199-201.

Huntly, B.J., and Gilliland, D.G. (2005). Leukaemia stem cells and the evolution of cancer-stem-cell research. Nat Rev Cancer *5*, 311-321.
Imamura, T., Morimoto, A., Takanashi, M., Hibi, S., Sugimoto, T., Ishii, E., and Imashuku, S. (2002). Frequent co-expression of HoxA9 and Meis1 genes in infant acute lymphoblastic leukaemia with MLL rearrangement. Br J Haematol *119*, 119-121.

Jamieson, C.H., Ailles, L.E., Dylla, S.J., Muijtjens, M., Jones, C., Zehnder, J.L., Gotlib, J., Li, K., Manz, M.G., Keating, A.*, et al.* (2004a). Granulocyte-macrophage progenitors as candidate leukemic stem cells in blast-crisis CML. N Engl J Med *351*, 657-667.

Jamieson, C.H., Weissman, I.L., and Passegue, E. (2004b). Chronic versus acute myelogenous leukemia: a question of self-renewal. Cancer Cell *6*, 531-533.

Jordan, C.T., Upchurch, D., Szilvassy, S.J., Guzman, M.L., Howard, D.S., Pettigrew, A.L., Meyerrose, T., Rossi, R., Grimes, B., Rizzieri, D.A.*, et al.* (2000). The interleukin-3 receptor alpha chain is a unique marker for human acute myelogenous leukemia stem cells. Leukemia *14*, 1777-1784.

Karanu, F.N., Murdoch, B., Gallacher, L., Wu, D.M., Koremoto, M., Sakano, S., and Bhatia, M. (2000). The notch ligand jagged-1 represents a novel growth factor of human hematopoietic stem cells. J Exp Med *192*, 1365-1372.

Kelly, L.M., Liu, Q., Kutok, J.L., Williams, I.R., Boulton, C.L., and Gilliland, D.G. (2002). FLT3 internal tandem duplication mutations associated with human acute myeloid leukemias induce myeloproliferative disease in a murine bone marrow transplant model. Blood *99*, 310-318.

Kondo, M., Wagers, A.J., Manz, M.G., Prohaska, S.S., Scherer, D.C., Beilhack, G.F., Shizuru, J.A., and Weissman, I.L. (2003). Biology of hematopoietic stem cells and progenitors: implications for clinical application. Annu Rev Immunol *21*, 759-806.

Kozu, T., Fukuyama, T., Yamami, T., Akagi, K., and Kaneko, Y. (2005). MYND-less splice variants of AML1-MTG8 (RUNX1-CBFA2T1) are expressed in leukemia with t(8;21). Genes Chromosomes Cancer *43*, 45-53.

Krivtsov, A.V., and Armstrong, S.A. (2007). MLL translocations, histone modifications and leukaemia stem-cell development. Nat Rev Cancer *7*, 823-833.

Kroon, E., Krosl, J., Thorsteinsdottir, U., Baban, S., Buchberg, A.M., and Sauvageau, G. (1998). Hoxa9 transforms primary bone marrow cells through specific collaboration with Meis1a but not Pbx1b. EMBO J *17*, 3714-3725.

Kroon, E., Thorsteinsdottir, U., Mayotte, N., Nakamura, T., and Sauvageau, G. (2001). NUP98-HOXA9 expression in hemopoietic stem cells induces chronic and acute myeloid leukemias in mice. EMBO J *20*, 350-361.

Krumlauf, R. (1994). Hox genes in vertebrate development. Cell *78*, 191-201.
Kustikova, O., Fehse, B., Modlich, U., Yang, M., Dullmann, J., Kamino, K., von Neuhoff, N., Schlegelberger, B., Li, Z., and Baum, C. (2005). Clonal dominance of hematopoietic stem cells triggered by retroviral gene marking. Science *308*, 1171-1174.

Lai, Y.Y., Qiu, J.Y., Jiang, B., Lu, X.J., Huang, X.J., Liu, Y.R., Shi, Y., Dang, H., He, Q., and Lu, D.P. (2005). [Analysis of characteristics of 72 cases of t(8;21) acute myeloid leukemia]. Beijing Da Xue Xue Bao *37*, 245-248.

Lasa, A., Nomdedeu, J.F., Carnicer, M.J., Llorente, A., and Sierra, J. (2002). ETO sequence may be dispensable in some AML1-ETO leukemias. Blood *100*, 4243-4244.
Lausen, J., Cho, S., Liu, S., and Werner, M.H. (2004). The nuclear receptor co-repressor (N-CoR) utilizes repression domains I and III for interaction and co-repression with ETO. J Biol Chem *279*, 49281-49288.

Lawrence, H.J., Rozenfeld, S., Cruz, C., Matsukuma, K., Kwong, A., Komuves, L., Buchberg, A.M., and Largman, C. (1999). Frequent co-expression of the HOXA9 and MEIS1 homeobox genes in human myeloid leukemias. Leukemia *13*, 1993-1999.

Lessard, J., and Sauvageau, G. (2003). Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells. Nature *423*, 255-260.

Liu, P., Tarle, S.A., Hajra, A., Claxton, D.F., Marlton, P., Freedman, M., Siciliano, M.J., and Collins, F.S. (1993). Fusion between transcription factor CBF beta/PEBP2 beta and a myosin heavy chain in acute myeloid leukemia. Science *261*, 1041-1044.

Lutterbach, B., Westendorf, J.J., Linggi, B., Patten, A., Moniwa, M., Davie, J.R., Huynh, K.D., Bardwell, V.J., Lavinsky, R.M., Rosenfeld, M.G.*, et al.* (1998). ETO, a target of t(8;21) in acute leukemia, interacts with the N-CoR and mSin3 corepressors. Mol Cell Biol *18*, 7176-7184.

Mamo, A., Krosl, J., Kroon, E., Bijl, J., Thompson, A., Mayotte, N., Girard, S., Bisaillon, R., Beslu, N., Featherstone, M.*, et al.* (2006). Molecular dissection of Meis1 reveals 2 domains required for leukemia induction and a key role for Hoxa gene activation. Blood *108*, 622-629. Mann, R.S., and Affolter, M. (1998). Hox proteins meet more partners. Curr Opin Genet Dev *8*, 423-429.

McGinnis, W., and Krumlauf, R. (1992). Homeobox genes and axial patterning. Cell *68*, 283-302.

Meyers, S., Lenny, N., and Hiebert, S.W. (1995). The t(8;21) fusion protein interferes with AML-1B-dependent transcriptional activation. Mol Cell Biol *15*, 1974-1982.

Minucci, S., Maccarana, M., Cioce, M., De Luca, P., Gelmetti, V., Segalla, S., Di Croce, L., Giavara, S., Matteucci, C., Gobbi, A.*, et al.* (2000). Oligomerization of RAR and AML1 transcription factors as a novel mechanism of oncogenic activation. Mol Cell *5*, 811-820.

Miyamoto, T., Weissman, I.L., and Akashi, K. (2000). AML1/ETO-expressing nonleukemic stem cells in acute myelogenous leukemia with 8;21 chromosomal translocation. Proc Natl Acad Sci U S A *97*, 7521-7526.

Miyoshi, H., Kozu, T., Shimizu, K., Enomoto, K., Maseki, N., Kaneko, Y., Kamada, N., and Ohki, M. (1993). The t(8;21) translocation in acute myeloid leukemia results in production of an AML1-MTG8 fusion transcript. EMBO J *12*, 2715-2721.

Miyoshi, H., Ohira, M., Shimizu, K., Mitani, K., Hirai, H., Imai, T., Yokoyama, K., Soeda, E., and Ohki, M. (1995). Alternative splicing and genomic structure of the AML1 gene involved in acute myeloid leukemia. Nucleic Acids Res *23*, 2762-2769.
Miyoshi, H., Shimizu, K., Kozu, T., Maseki, N., Kaneko, Y., and Ohki, M. (1991). t(8;21) breakpoints on chromosome 21 in acute myeloid leukemia are clustered within a limited region of a single gene, AML1. Proc Natl Acad Sci U S A *88*, 10431-10434.

Moens, C.B., and Selleri, L. (2006). Hox cofactors in vertebrate development. Dev Biol *291*, 193-206.

Moskow, J.J., Bullrich, F., Huebner, K., Daar, I.O., and Buchberg, A.M. (1995). Meis1, a PBX1-related homeobox gene involved in myeloid leukemia in BXH-2 mice. Mol Cell Biol *15*, 5434-5443.

Mukouyama, Y., Chiba, N., Hara, T., Okada, H., Ito, Y., Kanamaru, R., Miyajima, A., Satake, M., and Watanabe, T. (2000). The AML1 transcription factor functions to develop and maintain hematogenic precursor cells in the embryonic aorta-gonad-mesonephros region. Dev Biol *220*, 27-36.

Nishida, S., Hosen, N., Shirakata, T., Kanato, K., Yanagihara, M., Nakatsuka, S., Hoshida, Y., Nakazawa, T., Harada, Y., Tatsumi, N.*, et al.* (2006). AML1-ETO rapidly induces acute myeloblastic leukemia in cooperation with the Wilms tumor gene, WT1. Blood *107*, 3303-3312.

Nucifora, G., Larson, R.A., and Rowley, J.D. (1993). Persistence of the 8;21 translocation in patients with acute myeloid leukemia type M2 in long-term remission. Blood *82*, 712-715.
Nucifora, G., and Rowley, J.D. (1995). AML1 and the 8;21 and 3;21 translocations in acute and chronic myeloid leukemia. Blood *86*, 1-14.

Okuda, T., Cai, Z., Yang, S., Lenny, N., Lyu, C.J., van Deursen, J.M., Harada, H., and Downing, J.R. (1998). Expression of a knocked-in AML1-ETO leukemia gene inhibits the establishment of normal definitive hematopoiesis and directly generates dysplastic hematopoietic progenitors. Blood *91*, 3134-3143.

Owens, B.M., and Hawley, R.G. (2002). HOX and non-HOX homeobox genes in leukemic hematopoiesis. Stem Cells *20*, 364-379.

Palmqvist, L., Argiropoulos, B., Pineault, N., Abramovich, C., Sly, L.M., Krystal, G., Wan, A., and Humphries, R.K. (2006). The Flt3 receptor tyrosine kinase collaborates with NUP98-HOX fusions in acute myeloid leukemia. Blood *108*, 1030-1036.

Pandolfi, P.P. (2001). Transcription therapy for cancer. Oncogene *20*, 3116-3127.
Parkin, D.M. (2001). Global cancer statistics in the year 2000. Lancet Oncol *2*, 533-543.
Parkin, D.M., Bray, F., Ferlay, J., and Pisani, P. (2001a). Estimating the world cancer burden: Globocan 2000. Int J Cancer *94*, 153-156.

Parkin, D.M., Bray, F.I., and Devesa, S.S. (2001b). Cancer burden in the year 2000. The global picture. Eur J Cancer *37 Suppl 8*, S4-66.

Passegue, E., Jamieson, C.H., Ailles, L.E., and Weissman, I.L. (2003). Normal and leukemic hematopoiesis: are leukemias a stem cell disorder or a reacquisition of stem cell characteristics? Proc Natl Acad Sci U S A *100 Suppl 1*, 11842-11849.
Passegue, E., Wagner, E.F., and Weissman, I.L. (2004). JunB deficiency leads to a myeloproliferative disorder arising from hematopoietic stem cells. Cell *119*, 431-443.

Pineault, N., Buske, C., Feuring-Buske, M., Abramovich, C., Rosten, P., Hogge, D.E., Aplan, P.D., and Humphries, R.K. (2003). Induction of acute myeloid leukemia in mice by the human leukemia-specific fusion gene NUP98-HOXD13 in concert with Meis1. Blood *101*, 4529-4538.

Pineault, N., Helgason, C.D., Lawrence, H.J., and Humphries, R.K. (2002). Differential expression of Hox, Meis1, and Pbx1 genes in primitive cells throughout murine hematopoietic ontogeny. Exp Hematol *30*, 49-57.

Plevin, M.J., Zhang, J., Guo, C., Roeder, R.G., and Ikura, M. (2006). The acute myeloid leukemia fusion protein AML1-ETO targets E proteins via a paired amphipathic helix-like TBP-associated factor homology domain. Proc Natl Acad Sci U S A *103*, 10242-10247.

Rabbitts, T.H., and Stocks, M.R. (2003). Chromosomal translocation products engender new intracellular therapeutic technologies. Nat Med *9*, 383-386.

Reilly, J.T. (2002). Class III receptor tyrosine kinases: role in leukaemogenesis. Br J Haematol *116*, 744-757.

Rhoades, K.L., Hetherington, C.J., Harakawa, N., Yergeau, D.A., Zhou, L., Liu, L.Q., Little, M.T., Tenen, D.G., and Zhang, D.E. (2000). Analysis of the role of AML1-ETO in leukemogenesis, using an inducible transgenic mouse model. Blood *96*, 2108-2115.

Romana, S.P., Mauchauffe, M., Le Coniat, M., Chumakov, I., Le Paslier, D., Berger, R., and Bernard, O.A. (1995). The t(12;21) of acute lymphoblastic leukemia results in a tel-AML1 gene fusion. Blood *85*, 3662-3670.

Rosenbauer, F., Koschmieder, S., Steidl, U., and Tenen, D.G. (2005). Effect of transcription-factor concentrations on leukemic stem cells. Blood *106*, 1519-1524.

Rowley, J.D. (1999). The role of chromosome translocations in leukemogenesis. Semin Hematol *36*, 59-72.

Rozovskaia, T., Feinstein, E., Mor, O., Foa, R., Blechman, J., Nakamura, T., Croce, C.M., Cimino, G., and Canaani, E. (2001). Upregulation of Meis1 and HoxA9 in acute lymphocytic leukemias with the t(4 : 11) abnormality. Oncogene *20*, 874-878.

Sambrook, J., and Gething, M.J. (1989). Protein structure. Chaperones, paperones. Nature *342*, 224-225.

Sauvageau, G., Lansdorp, P.M., Eaves, C.J., Hogge, D.E., Dragowska, W.H., Reid, D.S., Largman, C., Lawrence, H.J., and Humphries, R.K. (1994). Differential expression of homeobox genes in functionally distinct CD34+ subpopulations of human bone marrow cells. Proc Natl Acad Sci U S A *91*, 12223-12227.

Sauvageau, G., Thorsteinsdottir, U., Eaves, C.J., Lawrence, H.J., Largman, C., Lansdorp, P.M., and Humphries, R.K. (1995). Overexpression of HOXB4 in hematopoietic cells causes the selective expansion of more primitive populations in vitro and in vivo. Genes Dev *9*, 1753-1765.

Sauvageau, G., Thorsteinsdottir, U., Hough, M.R., Hugo, P., Lawrence, H.J., Largman, C., and Humphries, R.K. (1997). Overexpression of HOXB3 in hematopoietic cells causes defective lymphoid development and progressive myeloproliferation. Immunity *6*, 13-22.

Schessl, C., Rawat, V.P., Cusan, M., Deshpande, A., Kohl, T.M., Rosten, P.M., Spiekermann, K., Humphries, R.K., Schnittger, S., Kern, W., *et al.* (2005). The AML1-ETO fusion gene and the FLT3 length mutation collaborate in inducing acute leukemia in mice. J Clin Invest *115*, 2159-2168.

Schwieger, M., Lohler, J., Friel, J., Scheller, M., Horak, I., and Stocking, C. (2002). AML1-ETO inhibits maturation of multiple lymphohematopoietic lineages and induces myeloblast transformation in synergy with ICSBP deficiency. J Exp Med *196*, 1227-1240.

Shanmugam, K., Green, N.C., Rambaldi, I., Saragovi, H.U., and Featherstone, M.S. (1999). PBX and MEIS as non-DNA-binding partners in trimeric complexes with HOX proteins. Mol Cell Biol *19*, 7577-7588.

Shen, W.F., Montgomery, J.C., Rozenfeld, S., Moskow, J.J., Lawrence, H.J., Buchberg, A.M., and Largman, C. (1997). AbdB-like Hox proteins stabilize DNA binding by the Meis1 homeodomain proteins. Mol Cell Biol *17*, 6448-6458.

Singh, S.K., Clarke, I.D., Terasaki, M., Bonn, V.E., Hawkins, C., Squire, J., and Dirks, P.B. (2003). Identification of a cancer stem cell in human brain tumors. Cancer Res *63*, 5821-5828. Spangrude, G.J., Heimfeld, S., and Weissman, I.L. (1988). Purification and characterization of mouse hematopoietic stem cells. Science *241*, 58-62.

Speck, N.A., and Gilliland, D.G. (2002). Core-binding factors in haematopoiesis and leukaemia. Nat Rev Cancer *2*, 502-513.

Taipale, J., and Beachy, P.A. (2001). The Hedgehog and Wnt signalling pathways in cancer. Nature *411*, 349-354.

Tenen, D.G. (2003). Disruption of differentiation in human cancer: AML shows the way. Nat Rev Cancer *3*, 89-101.

Terstappen, L.W., Huang, S., Safford, M., Lansdorp, P.M., and Loken, M.R. (1991). Sequential generations of hematopoietic colonies derived from single nonlineage-committed CD34+CD38- progenitor cells. Blood *77*, 1218-1227.

Thorsteinsdottir, U., Sauvageau, G., and Humphries, R.K. (1997). Hox homeobox genes as regulators of normal and leukemic hematopoiesis. Hematol Oncol Clin North Am *11*, 1221-1237.
van de Locht, L.T., Smetsers, T.F., Wittebol, S., Raymakers, R.A., and Mensink, E.J. (1994). Molecular diversity in AML1/ETO fusion transcripts in patients with t(8;21) positive acute myeloid leukaemia. Leukemia *8*, 1780-1784.

Varela, B.L., Chuang, C., Woll, J.E., and Bennett, J.M. (1985). Modifications in the classification of primary myelodysplastic syndromes: the addition of a scoring system. Hematol Oncol *3*, 55-63.

Varnum-Finney, B., Xu, L., Brashem-Stein, C., Nourigat, C., Flowers, D., Bakkour, S., Pear, W.S., and Bernstein, I.D. (2000). Pluripotent, cytokine-dependent, hematopoietic stem cells are immortalized by constitutive Notch1 signaling. Nat Med *6*, 1278-1281.

Wang, G.G., Pasillas, M.P., and Kamps, M.P. (2005). Meis1 programs transcription of FLT3 and cancer stem cell character, using a mechanism that requires interaction with Pbx and a novel function of the Meis1 C-terminus. Blood *106*, 254-264.

Wang, J., Hoshino, T., Redner, R.L., Kajigaya, S., and Liu, J.M. (1998). ETO, fusion partner in t(8;21) acute myeloid leukemia, represses transcription by interaction with the human N-CoR/mSin3/HDAC1 complex. Proc Natl Acad Sci U S A *95*, 10860-10865.

Warner, J.K., Wang, J.C., Hope, K.J., Jin, L., and Dick, J.E. (2004). Concepts of human leukemic development. Oncogene *23*, 7164-7177.

Weissman, I.L., Anderson, D.J., and Gage, F. (2001). Stem and progenitor cells: origins, phenotypes, lineage commitments, and transdifferentiations. Annu Rev Cell Dev Biol *17*, 387-403.

Wong, P., Iwasaki, M., Somervaille, T.C., So, C.W., and Cleary, M.L. (2007). Meis1 is an essential and rate-limiting regulator of MLL leukemia stem cell potential. Genes Dev *21*, 2762-2774.

Wuchter, C., Karawajew, L., Ruppert, V., Buchner, T., Schoch, C., Haferlach, T., Ratei, R., Dorken, B., and Ludwig, W.D. (1999). Clinical significance of CD95, Bcl-2 and Bax expression and CD95 function in adult de novo acute myeloid leukemia in context of P-glycoprotein function, maturation stage, and cytogenetics. Leukemia *13*, 1943-1953.

Yan, M., Kanbe, E., Peterson, L.F., Boyapati, A., Miao, Y., Wang, Y., Chen, I.M., Chen, Z., Rowley, J.D., Willman, C.L., *et al.* (2006). A previously unidentified alternatively spliced isoform of t(8;21) transcript promotes leukemogenesis. Nat Med *12*, 945-949.

Yuan, Y., Zhou, L., Miyamoto, T., Iwasaki, H., Harakawa, N., Hetherington, C.J., Burel, S.A., Lagasse, E., Weissman, I.L., Akashi, K., *et al.* (2001). AML1-ETO expression is directly involved in the development of acute myeloid leukemia in the presence of additional mutations. Proc Natl Acad Sci U S A *98*, 10398-10403.

Zeng, C., McNeil, S., Pockwinse, S., Nickerson, J., Shopland, L., Lawrence, J.B., Penman, S., Hiebert, S., Lian, J.B., van Wijnen, A.J., *et al.* (1998). Intranuclear targeting of AML/CBFalpha regulatory factors to nuclear matrix-associated transcriptional domains. Proc Natl Acad Sci U S A *95*, 1585-1589.

Zhang, X.B., Beard, B.C., Trobridge, G.D., Wood, B.L., Sale, G.E., Sud, R., Humphries, R.K., and Kiem, H.P. (2008). High incidence of leukemia in large animals after stem cell gene therapy with a HOXB4-expressing retroviral vector. J Clin Invest *118*, 1502-1510.

ACKNOWLEDGEMENTS

I take this opportunity to thank all those people who have made this dissertation possible and because of whom my graduate experience has been one that I will cherish forever.

My heartiest gratitude is to my supervisor, Dr. Christian Buske. I have been fortunate to have an advisor who gave me the utmost freedom to work in his esteemed lab. He helped me a lot with my work with his valuable suggestions and guided me when ever I expected his help during critical times. His patience and support helped me to overcome many crisis situations and finish this dissertation. I am very much thank full to him for giving me an opportunity to work with him, for holding me to a high standard research and come out learning a lot from him. It's indeed a great pleasure to work with Dr. Michaela Feuring-Buske with whom I was working during my initial days of work. It was Michaela from whom I learnt hardworking and perfectness with regards to my work techniques.

I would like extend my sincere thanks to Prof. Dr W.Hiddemann for his support. I would like to thank Prof. Dr Stefan Bholander for his valuable suggestions and for being always available for scientific discussions that helped me to sort out the technical details of my work.

It is indispensable to mention about my colleagues who helped me through out my PhD tenure First people to be mentioned here are, initially my seniors and then friends Aniruddha, Farid and Vijay without whom my thesis would have been incomplete. I still can not forget my starting days in Germany, who helped me a lot in getting acquainted to the environment, work culture and made me feel at home. Their insightful comments and scientific discussions at different stages of my research were thought-provoking and they helped me to refine my ideas. They not only helped me during my work but were also of great help personally with respect to mental support. I am very much grateful to Vijay for his valuable suggestions and experimental discussion which enabled me to over come many hurdles with my experiments. My heartiest thanks to Aniruddha for helping every time and working with him has been always fun and I admire him for his knowledge in every aspect and mostly for his sense of humor due to which no wonder we are very good friends. It's my great pleasure to have a friend and colleague like Farid who has been of immense help. I am very much thankful to him for his timely help in correcting my thesis and giving valuable suggestions. Natalia, always kept the whole lab environment lively and was very close to us through out her working time in our lab. Tina in helping me with my mouse experiments and most

importantly always helped me in translating German letters to English. It is with her, that I have learnt lot of new techniques with regards to my project and no doubt that I have learnt organizing thing from her. Monica is always energetic in the lab and was very helpful in teaching the techniques and helping out. I am very much thankful to Kosti for his kind help when ever I needed help with either with mouse experiments or writing me letters in German. He has been a big source of re-gaining confidence when ever my experiment failed. And a big thanks to Bianka who did a great job sorting cells. …To Christiane and Pawan for help with my experiments and discussions. Nicole for helping with ordering chemicals and sometimes getting things faster when needed very urgently. My heartfelt thanks to Silvia for her helping hand, especially with my experiments and for her crucial help in shaping up my thesis during the last days. I am thankful for her spontaneous help with German language in handling administration work. Many thanks to Carolla, Deepak dada, Dinesh, Alex, Luciana and Medhani. I am also thankful to the animal house facility caretakers.

Many friends have helped me stay sane through these difficult years. I am very much thankful to all my friends in Pfingstrosenstrasse who had been with me and my family during good and bad phases. I. would like to mention special thanks to all of them namely Pratibha, Pankaj, Madhu, Anagha, Purvi, and special mention to Ritu babhiji who helped me during my initial days and then my family and my heart full wishes to her cute little kids Shivam and Aditya.

It's my pleasure to have friends like Avinash, Joshi, and Sridhar who helped me a lot while working in CCMB and in Germany. I greatly value their friendship and I deeply appreciate their belief in me. Avinash had been of great help and without him I would have had tough time in recent times. He is now more of a family member than a friend. I would like to mention my good friends from CCMB Prasad kola, Sabri and Satya. I would like to mention about my professor Dr. P. V Raju in Andhra University and my guides in CCMB Dr. B.M Reddy and Dr. K.Thangaraj who gave me opportunity to work in such an esteemed lab CCMB, Hyderabad.

Most importantly, none of this would have been possible without the love and patience of my family. Their support and care helped me overcome setbacks and allowed me to stay focused on my graduate study. I would like to express my heart-felt gratitude to my family most importantly to my dear parents, my beloved wife Sandhya and my little angel Bhargavi. I am grateful to my all family members in India, VK.Naidu, Saroja, Sekhar, Srinu, thank full to my

Vegi, Mahalakshmi Naidu

CCG-Leukemia, Helmholtz Zentrum Muenchen,

GSF-Research Center for Environment and Health

Marchioninistrasse, 25. 81377 Munich, Germany

e-mail: **naidu.vegi@helmholtz-muenchen.de**

Phone: +49-89-7099406   Mobile: +49-179-6813090

Educational qualifications:

2004- Present	**PhD student** Ludwig-Maximilians-Universität (LMU), Dept of Medicine. III. Munich, Germany
1998- 2000	Master of Science in Human Genetics from Andhra University, Visakhapatnam, India ,passed with First Division
1995-1998	B.Sc (Bachelors) Nagarjuna University, Guntur with First Division
1993-95	Board of Intermediate Education (10+2), Andhra Pradesh with First Division

Employment:

2001-2004   Project research fellow in Centre for Cellular and Molecular Biology, Hyderabad, India

Research Experience

Presently working as a PhD student in Dr. Christian Buske's lab in project titled 'AML1-ETO collaborates with Meis1, a Hox gene co-factor in inducing acute leukemia in murine bone marrow transplantation model'

Project JRF from March 2001 to January 2004 at Centre for Cellular and Molecular Biology (CCMB), Hyderabad, India,

May 2000 to November 2000**,** research experience for the partial fulfilment of M.Sc. dissertation at the Department of Human Genetics, Andhra University, Visakhapatnam, India.

Conferences and presentations:

2008   Poster presentation at the International Society for Stem Cell Research (ISSCR-2008) held in Philadelphia, USA,

2005   Best poster award in the *Wissenschaftlisches Symposium* in Kloster Irsee.

2003   Best poster presentation award on ***"DNA fingerprinting technologies"*** on the CSIR open day, 2003 held at CCMB, Hyderabad.

Publications:

Sridhar Vempati, Carola Reindl, Ulla Wolf,Ruth Kern, Konstantin Petropoulos, **Vegi M. Naidu**, Christian Buske, Wolfgang Hiddemann, Tobias M. Kohl, and Karsten Spiekermann Transformation by oncogenic mutants and ligand dependent activation of FLT3-WT requires the tyrosine residues 589 and 591. **(in press)**

Rawat VP, Thoene S, **Naidu VM**, Arseni N, Heilmeier B, Metzeler K, Petropoulos K, Deshpande A, Quintanilla-Martinez L, Bohlander SK, Spiekermann K, Hiddemann W, Feuring-Buske M, Buske C. Over expression of CDX2 perturbs HOX gene expression in murine progenitors depending on its N-terminal domain and is closely correlated with deregulated HOX gene expression in human acute myeloid leukemia. Blood**. 2008** Jan 1;111(1):309-19

B.Mohan Reddy, A.N. Srikar Reddy, T. Nagaraju, L.V. K. S. Bhaskar**,** H.P. Singh, **V. M. Naidu**, K. Thangaraj, A.Govardhan Reddy, A. Nirmala, Lalji Singh. Anthropological Perspective of the Single Nucleotide Polymorphisms in the NPY and DRD2 Genes among the Socio-Economically Stratified Populations of Andhra Pradesh, India. International Journal of Human Genetics, 07(4): 277-284; **2007**

Kumar V, Langstieh BT, Madhavi KV, **Naidu VM**, Singh HP, Biswas S, Thangaraj K, Singh L, Reddy BM. Global patterns in human mitochondrial DNA and Y-chromosome variation caused by spatial instability of the local cultural processes. PLoS Genet. **2006** Apr;2(4).

Reddy BM, **Naidu VM**, Madhavi VK, Thangaraj LK, Kumar V, Langstieh BT, Venkatramana P, Reddy AG, Singh L. Microsatellite diversity in Andhra Pradesh, India: genetic stratification versus social stratification.Hum Biol. **2005** Dec;77(6):803-23.

Reddy BM, **Naidu VM**, Madhavi VK, Thangaraj K, Langstieh BT, Venkataramana P, Kumar V, Singh L. STR data for the Amp FlSTR Profiler Plus loci among 27 populations of different social

hierarchy from southern part of Andhra Pradesh, India.Forensic Sci Int. **2005** Apr 20;149(1):81-97

Naidu V M, Ramesh M, Rao T V, Veerraju P (**2002**); Study of some genetic markers in Sishta Karanam population, Andhra Pradesh, India. *Ind. J. Hum.Gen*,8(1): 26-28.

References:

PD. Dr. med. Christian Buske
Dept. of Medicine III, CCG-Leukemia,
(LMU)University Hospital- Grosshadern,
Helmholtz Zentrum Muenchen,
GSF-Research Center for Environment and Health,
Marchioninistrasse, 25. 81377 Munich, Germany
e-mail: buske@helmholtz-muenchen.de

Prof. Dr. med. Stefan Bohlander
CCG-Leukemia, Dept. of Medicine III,
Helmholtz Zentrum Muenchen,
GSF-Research Center for Environment and Health,
Marchioninistrasse, 25. 81377 Munich, Germany
e-mail: sbohlan@gwdg.de

PD Dr. med. Michaela Feuring-Buske
Dept. of Medicine III, CCG Leukemia
(LMU)University Hospital- Grosshadern,
Helmholtz Zentrum Muenchen,
Marchioninistrasse 15, 81377 Munich, Germany
e-mail: feuring@helmholtz-muenchen.de

Personal Details:

Name: Vegi, Mahalakshmi Naidu

Date of birth: 25.11.1977

Nationality: Indian

blood

Overexpression of CDX2 perturbs HOX gene expression in murine progenitors depending on its N-terminal domain and is closely correlated with deregulated HOX gene expression in human acute myeloid leukemia

Vijay P. S. Rawat, Silvia Thoene, Vegi M. Naidu, Natalia Arseni, Bernhard Heilmeier, Klaus Metzeler, Konstantin Petropoulos, Aniruddha Deshpande, Leticia Quintanilla-Martinez, Stefan K. Bohlander, Karsten Spiekermann, Wolfgang Hiddemann, Michaela Feuring-Buske and Christian Buske

---

Updated information and services can be found at:
http://bloodjournal.hematologylibrary.org/cgi/content/full/111/1/309

Articles on similar topics may be found in the following *Blood* collections:
Gene Expression (1085 articles)
Hematopoiesis (2373 articles)
Neoplasia (4028 articles)

Information about reproducing this article in parts or in its entirety may be found online at:
http://bloodjournal.hematologylibrary.org/misc/rights.dtl#repub_requests

Information about ordering reprints may be found online at:
http://bloodjournal.hematologylibrary.org/misc/rights.dtl#reprints

Information about subscriptions and ASH membership may be found online at:
http://bloodjournal.hematologylibrary.org/subscriptions/index.dtl

Overexpression of CDX2 perturbs *HOX* gene expression in murine progenitors depending on its N-terminal domain and is closely correlated with deregulated *HOX* gene expression in human acute myeloid leukemia

Vijay P. S. Rawat,[1,2] Silvia Thoene,[1,2] Vegi M. Naidu,[1,2] Natalia Arseni,[1,2] Bernhard Heilmeier,[1] Klaus Metzeler,[1] Konstantin Petropoulos,[1,2] Aniruddha Deshpande,[1,2] Leticia Quintanilla-Martinez,[3] Stefan K. Bohlander,[1,2] Karsten Spiekermann,[1,2] Wolfgang Hiddemann,[1,2] Michaela Feuring-Buske,[1,2] and Christian Buske[1,2]

[1]Department of Medicine III, Klinikum Grosshadern and [2]Clinical Cooperative Group (CCG) Leukemia, National Research Center for Environment and Health (GSF), Munich; and [3]Institute of Pathology, GSF, Neuherberg, Germany

The mechanisms underlying deregulation of *HOX* gene expression in AML are poorly understood. The ParaHox gene *CDX2* was shown to act as positive upstream regulator of several *HOX* genes. In this study, constitutive expression of *Cdx2* caused perturbation of leukemogenic *Hox* genes such as *Hoxa10* and *Hoxb8* in murine hematopoietic progenitors. Deletion of the N-terminal domain of Cdx2 abrogated its ability to perturb *Hox* gene expression and to cause acute myeloid leukemia (AML) in mice. In contrast inactivation of the putative Pbx interacting site of Cdx2 did not change the leukemogenic potential of the gene. In an analysis of 115 patients with AML, expression levels of *CDX2* were closely correlated with deregulated *HOX* gene expression. Patients with normal karyotype showed a 14-fold higher expression of *CDX2* and deregulated *HOX* gene expression compared with patients with chromosomal translocations such as t(8:21) or t(15;17). All patients with AML with normal karyotype tested were negative for *CDX1* and *CDX4* expression. These data link the leukemogenic potential of Cdx2 to its ability to dysregulate *Hox* genes. They furthermore correlate the level of *CDX2* expression with *HOX* gene expression in human AML and support a potential role of CDX2 in the development of human AML with aberrant *Hox* gene expression. (Blood. 2008; 111:309-319)

© 2008 by The American Society of Hematology

Introduction

In recent years, substantial progress has been made in understanding the biology of acute myeloid leukemia (AML). One of the pathogenetic hallmarks of AML are chromosomal translocations generating leukemogenic fusion genes that often act as aberrant transcription factors.[1] The second key genetic characteristics in AML are mutations, particularly those found in patients with normal karyotype and affecting the receptor tyrosine kinase FLT3 or the nucleophosmin protein (NPM1). Beside these structural genetic changes, large-scale gene expression analyses of cDNA samples from patients with AML have demonstrated that deregulated expression of nonaltered genes characterizes many AML cases. The most prominent example for this is the deregulated expression of homeobox genes in AML.[2-4] Homeobox genes form a highly conserved family of transcription factors known to be key regulators of normal hematopoietic stem cell and progenitor development.[5] Several studies have demonstrated that aberrant *HOX* gene expression profoundly perturbs normal murine and human hematopoietic development and causes leukemia in mice.[5-9] The aberrant expression of homeobox genes such as *HOXA9* and *HOXA10* is strongly associated with certain AML subtypes characterized by *MLL* fusion genes, *NPM1* mutations (NPMc⁺), and by more rare translocations such as the translocation t(10;11)(p13q14) generating the *CALM-AF10* fusion gene.[4,10-13] All together, deregulated homeobox gene expression characterizes more than every

third case of AML. So far, it is largely unknown how the aberrant expression of homeobox genes is initiated in the malignant clone. In cases with 11q23 chromosomal translocations, it is thought that aberrant function of the *MLL* gene, a known positive upstream regulator of *HOX* gene expression, is responsible for the perturbed expression of these key regulatory genes of early hematopoietic development.[14] In contrast, the aberrant *HOX* gene expression in patients with AML with normal karyotype and *NPM1* mutation is not well understood.[15] In particular, the patients with NPMc⁺ AML demonstrate that aberrant *HOX* gene expression cannot be just explained by the stage of differentiation at which the leukemic clone is arrested: NPMc⁺ patients are CD34⁻ in more than 95% of patients, and represent therefore a cell stage in which *HOX* genes are normally silenced.[8,16]

Another gene family critically involved in *Hox* gene regulation is the family of the so-called ParaHox genes, comprising the different "caudal-related homeobox genes" such as *CDX1*, *CDX2*, and *CDX4*, and the *GSH2* homeobox gene.[17] Several experimental systems have demonstrated that loss of *Cdx2* causes homeotic alterations and posterior shifts in *Hox* expression domains,[18] and that consensus-binding sites for the 3 Cdx homologs are present in the promoters of multiple *Hox* genes.[19-22] Expression of Cdx2 is tightly restricted to intestinal development in the adult.[23] Aberrant expression of *CDX2* is associated with intestinal metaplasia,[24,25]

Barrett epithelium,[26] and gastric carcinoma.[27] It was previously demonstrated in a single patient with AML carrying the translocation t(12;13)(p13;q12) that this translocation can induce ectopic expression of *CDX2* in adult hematopoietic cells beside the expression of the *ETV6-CDX2* fusion gene generated by the translocation.[28] We have previously shown in a murine leukemia model that the ectopic expression of the *Cdx2* and not the *ETV6-CDX2* fusion gene is the key transforming event in this type of leukemia.[29]

We now demonstrate that the ParaHox gene *Cdx2* dysregulates expression of leukemogenic *Hox* genes such as *Hoxa10 or Hoxb8* in murine hematopoietic progenitors. Furthermore, we show that loss of the ability of Cdx2 to perturb *Hox* gene expression by deletion of its N-terminal transactivation domain is paralleled by the inability of the gene to induce AML in vivo. We extended our analyses to human patients with AML and demonstrate that high expression levels of *CDX*2 were closely associated with *HOX* gene dysregulation in human AML.

Methods

Patient samples

Mononuclear cells prepared from diagnostic bone marrow or peripheral blood (PB) samples from 115 adult patients with AML were analyzed. The AML cases were classified according to the French-American-British criteria and the World Health Organization classification.[30] The study was approved by the ethics committees of all participating institutions, and informed consent was obtained from all patients before they entered the study in accordance with the Declaration of Helsinki (http://www.wma.net/e/policy/b3.htm). As a control, bone marrow mononuclear cells (BMMCs; CellSystem, St Katharinen, Germany) from healthy individuals were analyzed. Cytomorphology, cytochemistry, cytogenetics, and molecular genetics were applied in all cases as described.

Cytogenetics, FISH analysis, and molecular analysis

Cytogenetic analyses were performed using standard techniques. For fluorescence in situ hybridization (FISH), commercially available *AML1-ETO, PML-RARa, MLL-AF4, MLL-AF9, MLL-AF10, MLL-AF6, CBFβ-MYH11*, or *BCR-ABL* probes were used according to the manufacturer's instructions (Vysis, Bergisch-Gladbach, Germany).[31,32]

Microarray

Affymetrix HGU-133 A and B microarrays (Santa Clara, CA) were used to compare the expression of *HOX* genes in clinical specimens from patients with various subtypes of AML and in normal human bone marrow samples. The expression of 22 genes from the *HOX* gene cluster, represented by 29 different probesets on the microarrays, were analyzed. RNA extraction, cDNA preparation, in vitro transcription, hybridization, and microarray scanning were performed according to standard protocols as recommended by Affymetrix and as published previously.[33] Data analysis was performed using the R 2.4.0 software package (www.R-project.org) and routines from the biostatistics software repository "Bioconductor."[34] Raw microarray data were normalized using the variance stabilizing normalization (vsn) procedure,[35] and probe-set expression values were calculated by the median polish method. For the comparison between normal bone marrow and AML specimens with normal karyotype and unmutated nucleophosmin gene, empirical P values and the local false discovery rate for each gene were calculated using the successive exclusion procedure implemented in the Twilight software package.[36] To visualize the differences in *HOX* gene expression, a heatmap showing the expression of the 22 selected genes (29 probesets) in 75 clinical samples was constructed. Unsupervised hierarchic clustering using Euclidean distances was performed to group patient samples according to the similarity of their *HOX* gene expression profiles.

Quantitative PCR and LM-PCR

Expression of *CDX1*, *CDX2*, and *CDX4* was assayed by the TaqMan real-time quantitative polymerase chain reaction (RQ-PCR) method in total human bone marrow (BM), cord blood cells and mouse BM subpopulations. *CDX1, CDX2*, and *CDX4* primer and probes were used from Applied Biosystems (Foster City, CA; assay IDs: *CDX1*, Hs00156451 m1; *CDX2*, Hs01078080 m1; and *CDX4*, Hs01085517 m1). Quantification of *CDX2* expression was performed by RQ-PCR with Applied Biosystems primers. For normalization, the TATA binding protein (*TBP*) gene was used. Reactions were run in triplicates with 2.5 μL of cDNA in a total reaction volume of 20 μL by using an ABI PRISM 7900 Sequence Detection System (Applied Biosystems).

For quantification of *Hox* gene expression in vitro, 5-FU murine BM progenitors were transduced with MSCV-*Cdx2*-IRES-*EYFP*, MSCV-ΔN-*Cdx2*-IRES-*EGFP*, or MSCV-IRES-*EGFP*. After transduction, EGFP+ or EYFP+ cells were cultured in Dulbecco modified Eagle medium (DMEM) supplemented with 15% fetal bovine serum (FBS), IL-3, IL-6, and stem cell factor (SCF) for 24 hours. Expression levels of *HOX* genes were determined by the RQ-PCR method as described (Applied Biosystems), and fold expression was determined by the $r\rho C_T$ method. For the linker-mediated PCR (LM-PCR), integrated long-terminal repeats (LTRs) and flanking genomic sequences were amplified and then isolated using a modification of the bubble LM-PCR strategy as previously described.[13,37]

cDNA constructs and retroviral vectors

The cDNA of *Cdx2* was kindly provided by D. G. Gilliland (Division of Hematology/Oncology, Harvard Medical School, Boston, MA) and N. Cross (Department of Hematology, Hammersmith Hospital, London, United Kingdom). Cloning strategies and retroviral vectors were used as described before.[29]

Flow cytometry and histology

Immunophenotypic analysis of murine single-cell suspensions was performed as previously described.[13] Antibodies used for fluorescence-activated cell sorting (FACS) were labeled with phycoerythrin for Gr1, CD11b (Mac1), Sca-1, Ter119, CD4, CD19, and allophycocyanin-conjugated CD11b (Mac-1), CD117 (c-kit), B220, and CD8 (BD Pharmingen, Heidelberg, Germany). For histologic analyses, sections of selected organs were prepared and stained at the Academic Pathology Laboratory (GSF, Munich, Germany) using standard protocols as previously described.[29] Photographs from cytospins and colony formation were taken with an Axiovert 135 microscope (Zeiss, Goettingen, Germany), Plan-Neofluar 5×/0.15 NA, equipped with a CoolSNAP camera (Photometrics, Tucson, AZ). Openlab software (Improvision, Coventry, United Kingdom) was used for image processing. Histologic section images were acquired using a Hitachi camera HW/C20 (Hitachi, Tokyo, Japan) installed in a Zeiss Axioplan microscope (Zeiss, Jena, Germany) using Intellicam software (Matrox Electronic Systems, Middlesex, United Kingdom). Plan-Neofluar 10×/0.30 NA, 20×/0.50 NA, and 40×/0.75 NA objectives as well as a Plan Apochromat 63×/1.40 oil objective were used. Images were processed using Adobe Photoshop (Adobe Systems, San Jose, CA).

In vitro assays

GP+ E86 cells, NIH 3T3 cells, and 293T cells were grown in DMEM medium with 10% FBS and 1% penicillin/streptomycin (pen/strep) in a humidified incubator at 37°C and 5% CO_2 (complete medium). Primary murine BM cells were plated in complete medium consisting of DMEM supplemented with 15% FBS, 1% pen/strep, 6 ng/mL IL-3, 10 ng/mL IL-6, and 100 ng/mL SCF (Tebu-bio, Offenbach, Germany). IL-3–dependent cell populations from leukemic *Cdx2* mice were cultured in vitro in RPMI 20% FBS supplemented with IL-3 (10 ng/mL). Differentiation of clonogenic progenitors was analyzed by plating cells in methylcellulose supplemented with cytokines (primary colony-forming cell [CFC] assays, 500 input cells per dish; Methocult M3434; StemCell Technologies, Vancouver, BC). Serial CFC assays were performed by replating appropriate aliquots of cells obtained by harvesting all of the cells present in the previous CFC assay.

Mice and retroviral infection of primary BM cells

Primary mouse BM cells were transduced as previously described.[29] For transduction of Cdx2, W167A-Cdx2 and ΔN-Cdx2 cells were cocultured in complete medium with irradiated (40 Gy from a ^{137}Cs γ-radiation) Cdx2, W167A-Cdx2, and ΔN-Cdx2 producer cells. All transductions were performed with the addition of 5 μg/mL protamine sulfate.

BM transplantation and assessment of mice

Following transduction, cells were cultured in complete medium for 48 hours. After this, lethally irradiated (0.80 Gy) mice were given transplants of highly purified EYFP$^+$/EYFP$^+$ cells alone (2.5-3 × 10^5 cells per mouse transduced with *Cdx2* and W167A-*Cdx2,* 4-5 × 10^5 cells per mouse transduced with ΔN-*Cdx2*) without helper cells (FACSVantage; Becton Dickinson, San Jose, CA). Lethally irradiated secondary recipients (0.80 Gy) were injected with 10^6 BM cells from a primary diseased mouse with an equal number of nontransduced BM cells from a syngenic normal animal.

Statistical analysis

Data were evaluated using the *t* test for dependent or independent samples (SPSS, Chicago, IL). Differences with *P* values less than .05 were considered statistically significant.

Results

Cdx2-induced up-regulation of *Hox* gene expression depends on its N-terminal transactivation domain

As Cdx2 was shown to function as a positive *HOX* gene upstream regulator, it was analyzed whether ectopic expression of *Cdx2* induced up-regulation of *Hox* genes in murine hematopoietic progenitor cells. For this, *Hox* gene expression in murine progenitor cells isolated from 5-fluorouracil (5-FU)–treated bone marrow was determined by quantitative PCR (qPCR) 24 hours after successful retroviral transduction of the cells with the MSCV-*Cdx2*-IRES-*EYFP* or the MSCV-IRES-*EGFP* (control) constructs. Before the qPCR analysis, cells were highly purified by expression of EYFP or EGFP. Ectopic *Cdx2* expression induced significant up-regulation of *Hox* genes with leukemogenic potential such as *Hoxb3*, *Hoxb6*, and *Hoxb8* or *Hoxb9*, *Hoxa10*, *Hoxb5*, and *Hoxa7*. In contrast, *Hoxb4* or *Hoxd13* did not show any major changes in their expression levels (Figure 1A). Thus, these results indicated that ectopic CDX2 is able to up-regulate expression of leukemogenic *Hox* genes in adult hematopoietic progenitors after a short time interval. Furthermore, we also analyzed whether ectopic *Cdx2* expression would alter *Hox* gene expression in the progeny of clonogenic progenitors performing CFC assays in vitro: in comparison with the EGFP control, primary colonies derived from Cdx2-transduced progenitors showed substantial up-regulation of leukemogenic *Hox* genes such as *Hoxa7*, *Hoxa9*, *Hoxa10*, *Hoxb6*, and *Hoxb8* (Figure 1B).

It was then tested whether deletion of this domain would affect the ability of Cdx2 to deregulate *Hox* gene expression. Retroviral expression of the ΔN-Cdx2 mutant lacking the N-terminal transactivation domain (1 amino acid [aa]–179 aa) did not induce up-regulation of *Hoxb8*, *Hoxa9*, and *Hoxa10* in contrast to nonmutated Cdx2. Furthermore, deletion of the N-terminal domain clearly limited the ability of Cdx2 to up-regulate expression of *Hox* genes such as *Hoxa7*, *Hoxa9*, and *Hoxb3*, whereas the expression of the nonleukemogenic *Hox* genes *Hoxb4* and *Hoxd13* were not changed by both Cdx2 constructs compared with the control (Figure 1A,B).
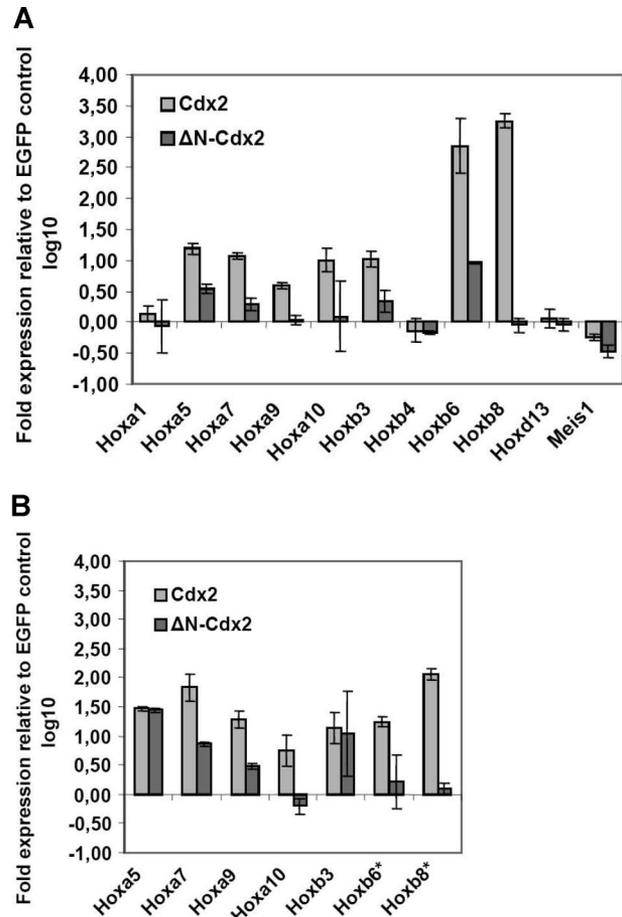
Figure 1. Aberrant expression of *Cdx2* up-regulates *Hox* gene expression in murine bone marrow progenitors. (A) Fold expression levels of *Hox* genes in murine BM progenitors induced by ectopic expression of *Cdx2* or the ΔN-*Cdx2* mutant (ΔN) referred to the expression level in BM progenitors transduced with the EGFP control vector 48 hours after the end of transduction. (B) Fold expression levels of *Hox* genes in primary colonies expressing *Cdx2*, ΔN-*Cdx2*, or the empty control vector (EGFP). The fold expression was calculated by the r$_P$C$_T$ method based on the expression level of the *Hox* genes in cells transduced with the EGFP control. *Expression of *Hoxb6* and *Hoxb8* were not detectable in EGFP control cells. Error bars indicate the standard deviations (SD).

The transforming potential of Cdx2 is associated with its ability to perturb *Hox* gene expression

To analyze whether the transforming potential of Cdx2 would correlate with its ability to perturb *Hox* gene expression in normal hematopoietic progenitor cells, serial replating and in vivo transplantation assays were performed with murine BM progenitor cells expressing *Cdx2*, ΔN-*Cdx2*, or *EGFP* alone. Cells were highly purified based on EGFP or EYFP expression before plating into methylcellulose. In the first CFC assay, expression of *Cdx2* resulted in a significantly higher number of colonies compared with EGFP (138 ± 18 vs 65 ± 11 per 500 initially plated cells, respectively; n = 8; *P* < .02) and a significantly higher yield of cells generated per 500 initially plated cells (2.2 × 10^6 ± 4.4 × 10^5 vs 1.6 × 10^5 ± 5.2 × 10^4; n = 5; *P* < .01; Figure 2A,B). Furthermore and in contrast to the control, colonies expressing Cdx2 were serially replatable (at second CFC assay 14 000 ± 2403 vs 374 ± 14 EGFP; n = 8; *P* < .002) with a significant higher yield of cells compared with the control (4.5 × 10^8 ± 1.8 × 10^8 vs 7.4 × 10^6 ± 4.0 × 10^6 EGFP; n = 8; *P* < .001; Figure 2A,B). After 3 rounds of replating, on average 60% (± 10%) of the Cdx2-expressing colonies were classified as CFU blasts (Figure
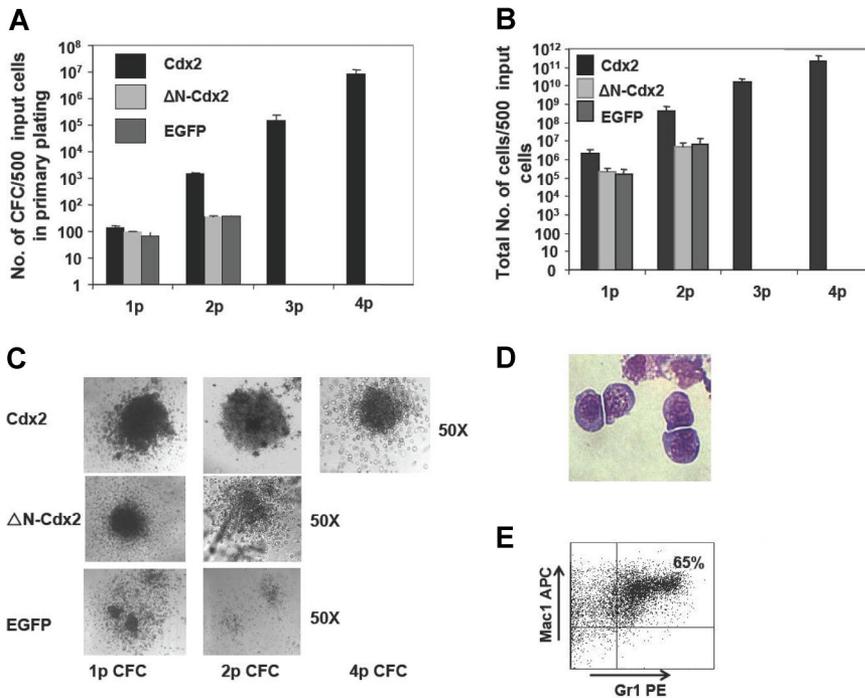
Figure 2. Cdx2 confers self-renewal properties to murine 5-FU BM progenitors. (A) Serial replating capacity of BM progenitors constitutively expressing *Cdx2*, ΔN-*Cdx2*, or the empty control vector (EGFP). (B) Yield of cells generated in the serial replating assays of BM progenitors constitutively expressing *Cdx2*, ΔN-*Cdx2*, or the empty control vector (EGFP). (C) Morphology of colonies obtained from Cdx2-, ΔN-*Cdx2*–, and EGFP-expressing BM cells in replating assays. (D) Blast morphology of cells after the fourth replating (May-Grunwald-Giemsa–stained cytospin preparations). (E) Coexpression of the myeloid markers Gr1 and Mac1 on Cdx2+ cells obtained from the fourth round of replating. p indicates plating. Error bars indicate SD.

2C). This was confirmed by cytospin preparations of colonies demonstrating a primitive myelomonoblastic morphology of the clonogenic cells (Figure 2D). Immunophenotypic characterization of cells at that time point documented coexpression of the myeloid markers Gr1 and Mac1 in 70% (± 10%) of the cells (n = 3) and positivity for Sca-1 in 50% (± 7%) of the cells, respectively, confirming the primitive phenotype of the colonies (Figure 2E). Cells isolated from tertiary CFC assays showed unlimited growth in liquid culture supplemented with IL-3. In order to test the leukemic potential of cells isolated from tertiary CFC assays, lethally irradiated mice received transplants of 1×10^6 (n = 4): all the mice developed AML after a median latency time of 8 weeks (data not shown). Thus, ectopic *Cdx2* expression conferred leukemogenic properties to hematopoietic progenitors after serial replating in vitro.

Although in the primary CFC assay expression of ΔN-*Cdx2* in hematopoietic progenitor cells increased the colony number and the yield of cells compared with the EGFP control (97 ± 4 vs 65 ± 11 EGFP per 500 initially plated cells, respectively; n = 5; $P < .03$), N-terminal deletion resulted in a significant loss of hematopoietic activity compared with Cdx2 (97 ± 4 vs 138 ± 18; n = 8; $P < .01$; Figure 2A). In addition, hematopoietic progenitors expressing the N-terminal deleted mutant did not achieve any serial replating capacity after the first round of replating or leukemogenic potential after propagation in methylcellulose as observed for ectopic Cdx2 expression (Figure 2B). Taken together, these data indicated that loss of the transforming activity of Cdx2 in vitro was paralleled by its loss to up-regulate expression of leukemogenic *Hox* genes.

To confirm the crucial role of the N-terminal transactivation domain for AML development in vivo, murine hematopoietic progenitors constitutively expressing Cdx2 or ΔN-Cdx2 were highly purified by EYFP+ or EGFP+ expression, respectively, and injected into lethally irradiated recipient mice directly after sorting (2.5-3 × 10^5 cells per mouse transduced with *Cdx2* and W167A-*Cdx2*, 4-5 × 10^5 cells per mouse transduced with ΔN-*Cdx2*). Mice that received transplants of BM cells expressing *Cdx2* became

moribund after a median of 116 days (n = 25) after transplantation (Figure 3). Furthermore, inactivation of the putative Pbx-interacting site of *Cdx2* (W167A-*Cdx2* mutant) did not change the phenotype or the time until disease development significantly (median latency time, 172 days; n = 14; Figure 3). All the mice showed elevated peripheral white blood cell count, suffered from splenomegaly, and were anemic. More detailed hematologic analyses demonstrated that the animals had developed AML with a high percentage of blasts in BM, spleen and PB (Table 1). Histologic sections demonstrated infiltration of myeloid blasts in multiple nonhematopoietic organs, including the testis. Immunohistochemistry showed positivity of the blasts for myeloperoxidase and chloracetate esterase (Figure 4) and negativity for B220 and CD3 (data not shown), indicating the myeloid nature of the blast population. Immunophenotypic characterization of PB, BM, and spleen in diseased mice confirmed the predominance of myeloid Mac1+ and Gr-1+ cells and the reduction of lymphoid cells compared with the EGFP control mice (Figure 5; Table 1). Analysis of the clonality of the disease by Southern blotting demonstrated
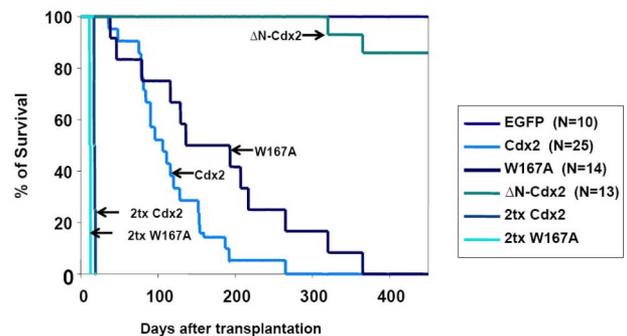
Figure 3. Survival of mice that received transplants. Survival curves of mice that received transplants of BM cells expressing *Cdx2* (n = 25), the *W167A-CDX2* mutant (n = 14), or the ΔN-*Cdx2* mutant (n = 13). The control group received transplants of BM infected with the EGFP control retrovirus (n = 10). The survival time of secondary recipient mice that received transplants of BM from diseased primary animals from the *Cdx2* or *W167A-CDX2* cohort is indicated. Tx indicates transplantation.

Table 1. Hematologic parameters of experimental mice

Retroviral construct	Mice analyzed, no.	Median day of killing (range)	Mean RBC, $\times 10^9$/mL ($\pm$ SD)	Mean WBC, $\times 10^6$/mL ($\pm$ SD)	Mean spleen weight, mg ($\pm$ SD)	BM, % blasts ($\pm$ SD)	Spleen, % blasts ($\pm$ SD)	PB, % blasts ($\pm$ SD)	Lymphoid/ myeloid ratio in PB
EGFP	6	123.5 (85-127)*	5.67 ($\pm$ 0.93)	3.37 ($\pm$ 2.61)	156 ($\pm$ 59)†	0	0	0	5:1
Cdx2	15	111 (37-229)	1.22 ($\pm$ 0.43)	22.87 ($\pm$ 14.68)	546 ($\pm$ 195)	45 ($\pm$ 15)	38 ($\pm$ 15)	18 ($\pm$ 10)	2:1
W167A	6	128.5 (46-381)	0.93 ($\pm$ 0.35)	33.5 ($\pm$ 32.25)	516 ($\pm$ 233)	50 ($\pm$ 12)	34 ($\pm$ 9)	27 ($\pm$ 2)	2:1
ΔN-Cdx2	3	330 (309-350)	4.4 ($\pm$ 0.75)	5.3 ($\pm$ 1.42)	210 ($\pm$ 90)	6	0	0	4:1

*A total of 4 of 6 healthy EGFP mice were killed for analysis. A total of 2 of the 6 control mice were analyzed by bone marrow biopsy and bleeding.
†Average weight from 4 EGFP mice.

different intensities and patterns of proviral signals in the different hematopoietic organs consistent with an oligoclonal nature of the disease (Figure S1, available on the *Blood* website; see the Supplemental Materials link at the top of the online article). Sequencing of retroviral integration sites (n = 9; 3 independent transplantation experiments) in the diseased mice that received transplants of BM cells expressing *Cdx2* or W167A-*Cdx2* did not show any recurrent integration sites, arguing against insertional mutagenesis as a key factor in this disease model (Table 2).

In contrast, mice transplanted with hematopoietic progenitor cells expressing the ΔN-*Cdx2* (n = 13) did not develop any disease in 11 of 13 patients (Figures 3-5; Table 1). A total of 2 mice developed AML after a very long latency time of 365 and 400 days after transplantation (Figure 3). To exclude that ΔN-Cdx2 perturbed hematopoiesis without obvious clinical symptoms, 3 mice without disease manifestations were killed 120 days after transplantation. Engraftment with ΔN-*Cdx2*–positive cells was 90% ($\pm$ 10%) at that time point. There were no signs of splenomegaly, and the histopathology of the organs showed a normal tissue architecture (Figure 4). Moreover, cytospin preparations from BM and spleen showed differentiated myeloid and lymphoid cells, and immunophenotyping did not demonstrate a myeloid infiltration in the spleen as seen in mice that received transplants of Cdx2 or the W167A mutant (Figures 4,5). In summary, these experiments indicated that deletion of the N-terminal transactivation domain of Cdx2 eliminates its ability of the protein to dysregulate *Hox* gene expression and to transform hematopoietic progenitors in vitro and in vivo.

CDX2 is highly expressed in patients with AML with normal karyotype

As the experimental data in the murine system suggested that ectopic expression of CDX2 is linked to dysregulated *HOX* gene expression in AML, we focused on patients with normal karyotype and *NPM1* mutation (NPMc$^+$ AML), previously shown to aberrantly express *HOX* genes[12] (Table S1). We confirmed these data using oligonucleotide microarray analysis and extended these findings to the patient group with normal karyotype without the *NPM1* mutation (NPMc$^-$). A total of 24 patients with normal karyotype, 12 patients with NPMc$^+$ AML and 12 patients with NPMc$^-$ AML, was analyzed. Patients with normal karyotype and *NPM1* mutation were characterized by aberrant expression of multiple *HOXA* cluster genes such as *HOXA10*, *HOXA9*, and *HOXA7*, and *HOXB* cluster genes such as *HOXB5* and *HOXB6* (Figure 6A-B). However, patients with NPMc$^-$ AML with normal karyotype also showed dysregulated *HOX* gene expression compared with patients with AML expressing the *PML-RARA* (n = 20) or *AML1-ETO* (n = 20) fusion gene or compared with normal healthy donors (n = 11).
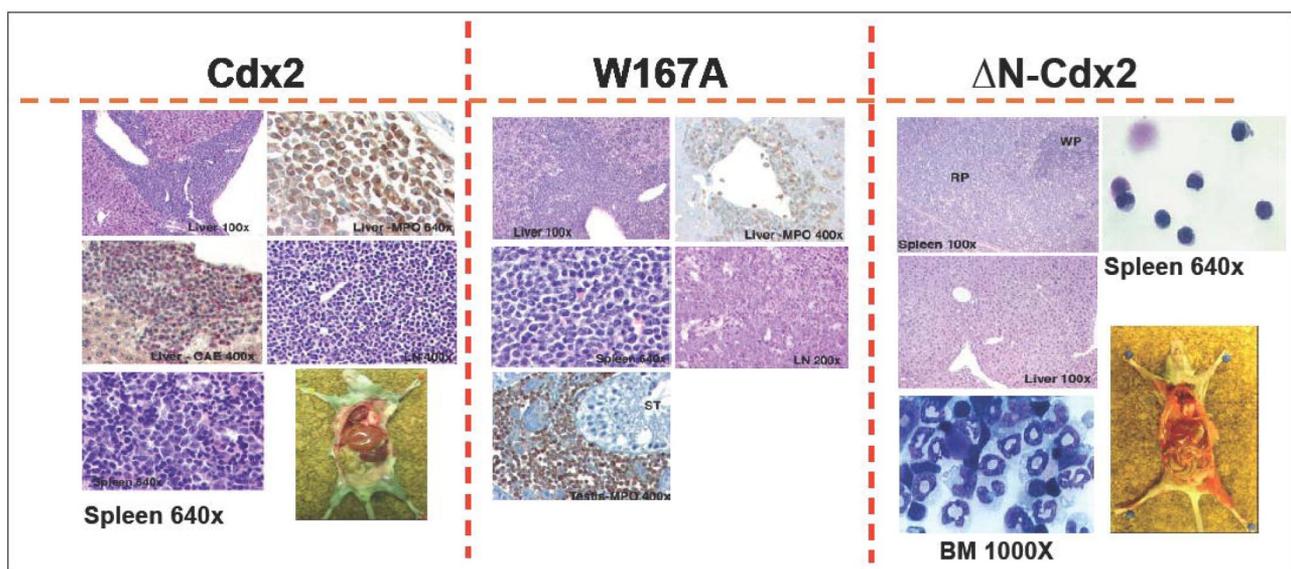
Figure 4. Histology of mice that received transplants. Immunohistologic analyses of different organs of a representative leukemic *Cdx2* or *W167A-Cdx2* mouse compared with a healthy animal from the Δ*N-Cdx2* cohort. The spleen (Giemsa staining, ×640) of the analyzed Cdx2 and W167A-Cdx2 animals shows an infiltration with blast cells in contrast to the ΔN-Cdx2 animal that received a transplant. The liver of the Cdx2 and W167A-Cdx2 mice demonstrates perivascular infiltration with leukemic cells. Positivity for myeloperoxidase (MPO) and chloracetate esterase (CAE) confirmed the myeloid nature of the cells (×100 and ×400). May-Grunwald-Giemsa–stained cytospin preparations of cells isolated from the spleen (×640) or BM (×1000) of Δ*N-Cdx2* mice that received transplants show mature lymphoid and myeloid cells, respectively. LN indicates lymph node; RP, red pulp; WP, white pulp; and CAE, chloracetate esterase.
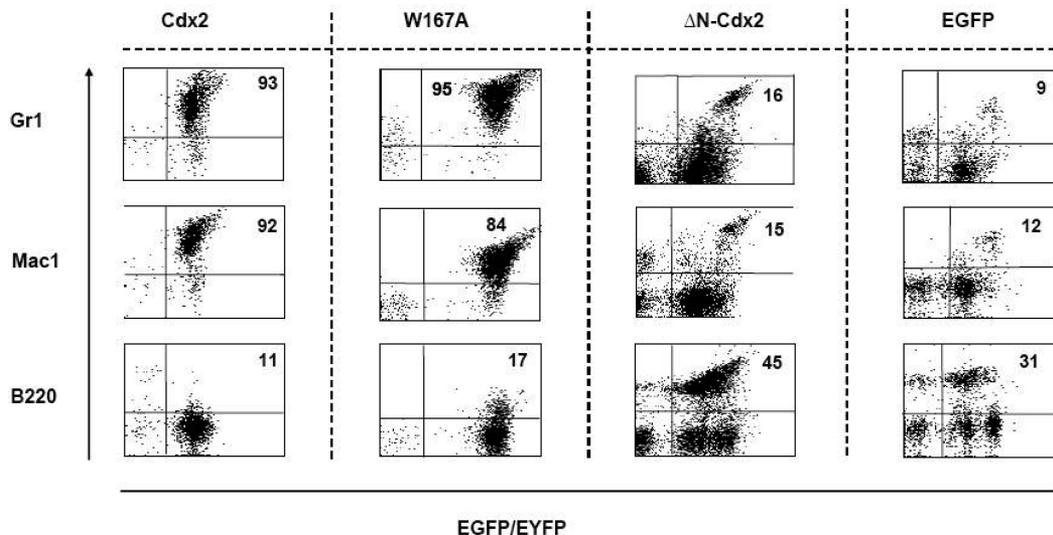
Figure 5. Immunophenotype of cells isolated from the spleens of mice that received transplants. Expression of the myeloid markers Gr1 and Mac1 and the lymphoid marker B220 on cells isolated from the spleen of representative animals that received transplants of BM cells transduced with different constructs as indicated. The proportion of positive cells within the EGFP+ or EYFP+ compartment is indicated.

We performed unsupervised hierarchic clustering using the expression levels of 21 genes from the *HOX* gene cluster in the 64 patient samples and 11 healthy controls: in this analysis, all 12 NPMc+ AML samples with normal karyotype and 9 of the 12 NPMc− samples with normal karyotype formed 1 main cluster. The other main cluster consisted of all the samples with t(8;21) or t(15;17) translocations, the 11 normal bone marrow samples, and the 3 NPMc− samples with normal karyotype, which had shown no perturbation of *HOX* gene expression (Table S1). Of note, only 1 of the 21 patients with AML with normal karyotype showed an up-regulation of the *MLL* gene, which is known to be an upstream regulator of *HOX* genes (Figure 6A).

As AML with normal karyotype separated from the other cytogenetic AML subgroups by their pattern of *HOX* gene expression, we analyzed expression of *CDX2* and other members of the *CDX* gene family in normal human and murine hematopoietic cells and different AML subgroups using quantitative PCR. *CDX1*, *CDX2*, and *CDX4* were not detectable in normal human BM (n = 3), CD34+ human BM (n = 3), and human cord blood cells (n = 2). We also could not detect the *CDX2* transcript in human CD34+/CD38−, CD34+/CD38+, and CD34−/CD38+ BM cells from healthy donors in up to 45 cycles of qRT-PCR (n = 3). *Cdx1* and *Cdx2* were also not detectable in murine samples, whereas *Cdx4* was expressed in murine BM and splenic cells as previously

reported[39] (Table S2; n = 3). A total of 71 patients with normal karyotype (AML NPMc+ = 45 patients; NPMc− = 26 patients) was analyzed for *CDX2* expression. Of the patients with NPMc+ AML, 89% showed aberrant expression of *CDX2* as did 88% of the patients without the NPM1 mutation (Figure 7A; Tables 3, S1). Sequencing of the complete coding region of *CDX2* in 5 patients with NPMc+ AML did not show any mutations or deletions. To test whether *FLT3* mutation would affect expression levels of *CDX2* in the patients with normal karyotype, we analyzed the 4 subgroups characterized by *NPM1* mutation with our without *FLT3* mutation (AML NPMc+ ± *FLT3* mutation) and the NPMc− patients with or without *FLT3* mutation (AML NPMc− ± *FLT3* mutation): as illustrated in Table 4 there was no major difference in the expression level between the different patient cohorts.

We extended this analysis to 44 patients with abnormal karyotype and detected aberrant *CDX2* expression in 64% (28 of 44) of the patients: 12 of 24 patients with the translocation t(8;21)(q22;q22), 10 of 10 patients with the translocation t(15;17)(q22;q11), 3 of 4 patients with inv16, and 3 of 6 patients with *MLL*-associated translocations showed expression of *CDX2* (Figure 7B; Tables 3, S1).

Importantly, when the expression level of the *CDX2* was compared between patients with AML with normal and abnormal karyotypes, there was a more than 14-fold higher expression level

Table 2. Identity of retroviral integration sites in diseased mice

No.	Gene	Description	Genomic location	Experimental group (mouse no.)
1	Intron of A930004K21Rik between exons 2 and 3	NA	2E5	3438 B1
2	Intron of D930015E06Rik between exons 26 and 27	NA	3F1	3432 B1
3	Intergenic region	NA	11B1.3	3998 B3
4	Intron of PhC2 between exons 8 and 9	"Polycomb" group (PcG) genes	4D2.2	3998 B3
5	Intergenic region	NA	2B	3998 B4
6	Intron of Armc2 between exons 1 and 2	NA	10B2	3478 B2-1
7	Intergenic region	Armadillo repeat containing 2	6A3.3*	3478 B2-2
8	Intron of Pag1 between exons 1 and 2	Phosphoprotein associated with glycosphingolipid-enriched microdomains 1; pag1	3A1*	3478 B2-3
9	Exon 1 of Pigb	Phosphatidylinositol glycan, class b	9D	4057A

NA indicates not applicable.

*Identified in or near regions (~50 kb) described as common integration sites (CISs) in the Retrovirus Tagged Cancer Gene database.[38]
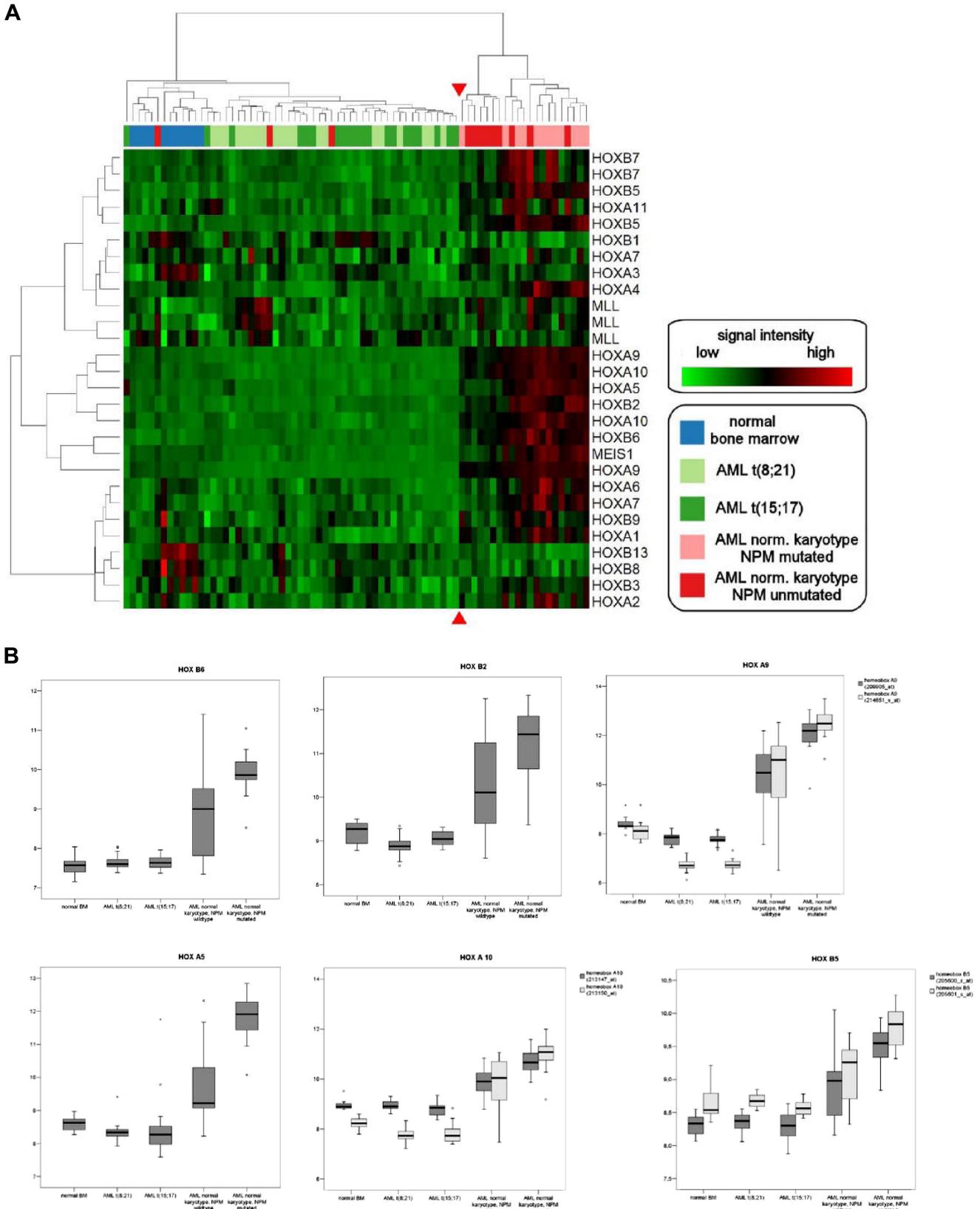
Figure 6. Hierarchic clustering of *CDX2*-positive AML samples according to *HOX* gene expression. (A) Unsupervised hierarchic clustering according to *HOX* gene expression demonstrates *HOX* gene deregulation in patients with normal karyotype with or without *NPM1* mutation compared with samples with abnormal karyotype or normal bone marrow samples. The red arrows highlight the boundary between the 2 main clusters. The genes and samples were permutated. (B) Expression of individual *HOX* genes in *CDX2*-positive AML samples with normal and abnormal karyotype. Log expression levels of 6 different *HOX* genes in 75 clinical samples are shown in box-and-whisker plots. Expression was determined by Affymetrix HGU-133 A and B microarrays. The plots show the normalized expression values in normal BM samples (n = 11), AML with t(8;21) (n = 20), AML with t(15;17) (n = 20), and AML with normal karyotype with (n = 12) and without *NPM1* mutation (n = 12). The bar indicates the median expression levels and the box shows the 25th and 75th percentiles, while the whiskers show the maximum and minimum values. Outliers (values that are more than 1.5 interquartile ranges above the 75th or below the 25th percentile) are represented by open circles.
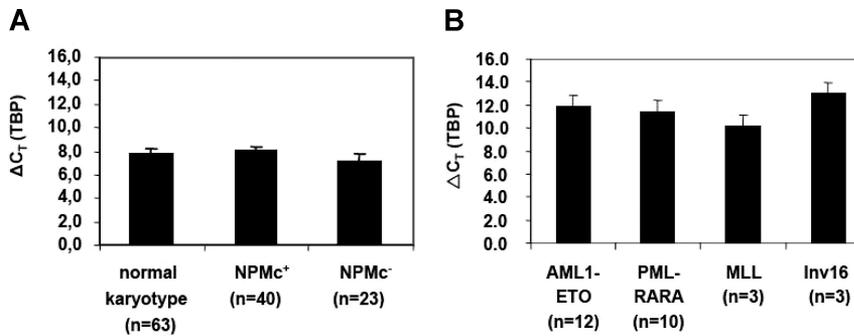
Figure 7. Quantification of *CDX2* expression in patients with AML by RQ-PCR. (A) Expression levels of *CDX2* in patients with AML with a normal karyotype with (NPMc+) or without *NPM1* mutation (NPMc−) or (B) in different AML subgroups with abnormal karyotype. A total of 115 samples were analyzed. The number of patients who were positive for *CDX2* expression and were therefore evaluated for expression levels are indicated. ΔC_T values were obtained by normalization for the housekeeping gene *TBP*; the mean values ($\pm$ SD) are shown. The expression level of the gene is inversely correlated with the ρC_T value.

in the patient group with normal karyotype (n = 42) compared with the group with aberrant karyotype (n = 28)($\emptyset\rho C_T$ 8.23 vs $\emptyset\rho C_T$ 11.66, respectively; $P < .001$). A total of 3 NPMc− patients with normal karyotype showed the same low level of expression of *CDX2* (range, ρC_T 10.55-11.55) as patients with AML with aberrant karyotype. Of note, these were the same 3 patients shown before to have no perturbation of *HOX* gene expression and thereby fall into the same cluster as patients with AML with t(8;21) or t(15;17) according to *HOX* gene expression.

When the *CDX2* expression level was analyzed in the different cytogenetic subgroups of patients with aberrant karyotype, patients characterized by expression of the *AML1-ETO*, *PML-RARA*, and *CBFβ-Myh11* showed uniformly low expression levels. Interestingly, the 3 patients with *MLL* fusion, a subgroup known to be characterized by *HOX* gene dysregulation, showed a higher expression of *CDX2* compared with the t(8;21) or t(15;17) cytogenetic subgroups. These data indicated that high expression levels of *CDX2* are associated with *HOX* gene dysregulation in AML. To analyze whether expression of other members of the *CDX* gene family might be associated with *HOX* gene dysregulation in patients with normal karyotype, we determined transcription of *CDX1* and *CDX4* in 23 patients of this AML subgroup: in contrast to *CDX2*, *CDX1* and *CDX4* were not detectable in any of the patients tested, pointing to a key role of CDX2 in this patient group.

Discussion

Leukemias are initiated by a minor fraction of leukemic stem cells (LSCs) that have maintained or reacquired the capacity for indefinite proliferation through accumulated genetic alterations mutations and/or epigenetic changes.[40] Aberrant expression of homeobox genes, detectable in more than every third case of AML, is thought to contribute to the infinite self-renewal properties of LSCs. The molecular mechanisms that mediate aberrant *HOX* gene expression in leukemias are known only for a minority of cases, namely those involving rearrangements of

specific *HOX* genes or rearrangements of the trithorax group gene *MLL*.[14,39,41-43] In this study, we now demonstrate that the *Cdx2* gene is able to up-regulate several HoxA and HoxB cluster genes such as *Hoxa5*, *Hoxa7*, *Hoxa9*, *Hoxa10*, or *Hoxb8*; *Hoxb6* and *Hoxb3* previously have been shown to induce AML in mice or perturb normal hematopoietic development.[7,44-48] Interestingly, Cdx2 did not change the expression of *Hox* genes with no reported leukemogenic potential such as *Hoxb4* or *Hoxb13*, suggesting that Cdx2 is able to perturb in particular expression of Hox genes with transforming potential. Our findings are consistent with data in nonhematopoietic cells that documented the ability of *Cdx* genes to alter *Hox* gene expression patterns.[49-51] The ability of Cdx2 to dysregulate *Hox* gene expression was clearly dependent on its N-terminal transactivation domain. This is in line with data that previously showed that the N-terminal transactivation domain, in contrast to the portion C-terminal of the homeodomain, is necessary for transcriptional activation of downstream target genes such as *Hox* genes and directly interacts with the transcriptional cofactor CBP.[52-54] Furthermore, it was shown that deletion of the N-terminal transactivation domain did not only abrogate the ability of Cdx2 to perturb *Hox* gene expression, but also eliminated the transforming activity of the gene in vitro and in vivo. Particularly in serial replating assays, considered to be a surrogate test for the self-renewal of hematopoietic progenitors, the N-terminally deleted mutant did not transfer infinite self-renewal properties to transduced progenitor cells as observed for Cdx2. Furthermore and in contrast to the ΔN mutant, Cdx2-expressing cells acquired leukemia-initiating potential in mice that received transplants after serial replating. This clear difference in the transforming potential between Cdx2 and its N-terminal deleted mutant was further confirmed by BM transplantation assays, showing rapid development of AML in mice that received transplants of Cdx2 in contrast to mice that received transplants of BM cells expressing the ΔN mutant. Inactivation of the putative Pbx-interacting site in Cdx2, however, did not change the transforming potential of the gene. These findings parallel data on the transforming activity of Cdx4: deletion of its N-terminal transactivation domain but not the inactivation of the

Table 3. Ectopic expression of *CDX2* in patients with AML

Patient group	Samples tested for *CDX2* transcript, no.	*CDX2*-positive samples, no.	Positive samples, %
NK, NPM1c+	45	40	89
NK, NPM1c−	26	23	88
AML1-ETO	24	12	50
PML-RARα	10	10	100
MLL fusions	6	3	50
inv16	4	3	75

Percentage of patients positive for *CDX2*.
NK indicates normal karyotype.

Table 4. Expression level of *CDX2* in patients with AML and normal karyotype

AML-NK subgroup analysis	Average ΔC_T ($\pm$ SD)
NPM+/FLT3+	8.26 ($\pm$ 2.84)
NPM+/FLT3−	7.74 ($\pm$ 2.27)
NPM−/FLT3+	7.27 ($\pm$ 2.91)
NPM−/FLT3−	7.18 ($\pm$ 2.48)

NK indicates normal karyotype; and FLT3, FLT3 mutation.

putative Pbx interacting site resulted in loss of its hematopoietic activity and ability to perturb *Hox* gene expression.[39] Like Cdx2, Cdx4 up-regulated leukemogenic *Hox* genes such as 5′-located HoxA cluster genes or HoxB cluster genes such as *Hoxb3* or *Hoxb8*.[39] Taken together, these data indicate that the key role of the N-terminal transactivation domain for the alteration of *Hox* gene expression is conserved between the different Cdx family members. However, in the murine BM transplantation model, Cdx4 had a comparably low leukemogenic potential compared with Cdx2, with only half of the animals developing AML after a long-latency time of 300 days after transplantation.[39] The reason for this is not yet understood, but it is interesting to note that in contrast to Cdx2, Cdx4 also up-regulated *Hox* genes such as the nonleukemogenic *Hoxb4*. Therefore, it might be that ectopic expression of Cdx2 induces a more pronounced shift toward the expression of 5′-located leukemogenic *Hox* genes compared with Cdx4.

Strikingly, in human AML, high *CDX2* expression levels were clearly correlated with perturbed *HOX* gene expression: whereas ectopic expression of *CDX2* was detected in AML subtypes with our without aberrant *HOX* gene expression, high transcript levels of the *HOX* upstream regulator were closely associated with *HOX* gene dysregulation. This was in particular demonstrated for the large group of patients with AML and normal karyotype independent of their *NPM1* mutational status, counting for 50% of all patients with human AML. These patients had a more than 14-fold higher expression of *CDX2* compared with t(8;21)- or t(15;17)-positive patients. The association of high *CDX2* expression and *HOX* gene perturbation was further underlined by the finding that the 3 NPMc− patients with low *CDX2* expression did not show *HOX* gene perturbation and fell into the same group as the t(8;21)- or t(15;17)-positive patients with AML when unsupervised clustering according to *HOX* gene expression was performed. Of note, in our study, expression of other members of the *CDX* gene family were not associated with dysregulated *HOX* gene expression in human AML with normal karyotype, as all 23 samples tested were negative for expression of *CDX1* and *CDX4*. In a recent report, Bansal et al detected expression of *CDX4* in 3 of 16 patients with AML and normal karyotype, indicating that expression of *CDX4* can occur in this AML subtype, but at low frequency.[39]

Taking the well-known role of Cdx2 as an upstream regulator of *Hox* genes and the close correlation between high expression levels of *CDX2* and *HOX* gene perturbation in human AML into account, it is intriguing to speculate that the initiation of high *CDX2* expression levels might be a key step in the development of AML with aberrant *HOX* gene expression. This concept would provide a model for the biology of the large group of patients suffering from AML with normal karyotype. Particularly in the NPMc+ patients, which are more than 95% CD34−, induction of *HOX* gene dysregulation by high expression levels of *CDX2* would be an intriguing explanation, because in this AML subtype aberrant *HOX* gene expression cannot be explained by the accumulation of CD34+ myeloid blasts, which also express high levels of multiple *Hox* genes during normal hematopoiesis. Our data are in line with a most recent report that also analyzed the role of ectopic CDX2 in human AML[55]: based on the observation that high-level amplification of the *CDX2* locus can occur in patients with AML with complex karyotype, patients with AML with aberrant and normal karyotypes were evaluated for ectopic *CDX2* expression. As in our data set, the vast majority of patients with normal karyotype or translocation t(15;17) showed ectopic expression of *CDX2*. In contrast to our data, in which 12 of the 24 patients with AML1-ETO were negative for *CDX2*, 8 of the 10 patients reported by Scholl et al were positive. Consistent with the presented data, this report also documented that expression levels substantially vary between different genetically defined AML subgroups, but did not correlate *CDX2* expression levels with *HOX* gene deregulation in AML.

Of note, in the report by Scholl et al, Cdx2 was able to up-regulate *HoxB8* transcript levels, although to a much lesser extent compared with the up-regulation of *HoxB8* observed in this study. Furthermore, the authors did not see any change in expression of leukemogenic HoxA cluster genes such as *HoxA9*. In addition, *Hox* genes associated with leukemogenesis such as *HoxA10* or *HoxB3* were even down-regulated compared with the control.[55] In contrast, this study could demonstrate that *Cdx2* expression is associated with up-regulation of leukemogenic *Hox* genes in the murine experimental system as well as in patients with AML. Interestingly, recent data described up-regulation of *HOXA9*, *HOXA2*, and *HOXA7* after stable transfection of the esophageal squamous epithelial cell line HET1A with *CDX2*, and increased *CDX2* and *HOX* gene expression in primary tissues of patients with esophageal cancer.[51,56] This suggests that CDX2-induced up-regulation of leukemogenic *HOX* genes might be a common mechanism in the development of cancer.

However, despite the intriguing correlation between the expression levels of *CDX2* and perturbed *HOX* gene expression in human AML, it is still uncertain to which extent the observed deregulation of this gene family is caused by CDX2: thus, although the transcript levels of *CDX2* were comparable between NPMc+ and NPMc− patients, there was generally a higher expression level of individual *HOX* genes in patients with the *NPM1* mutation. Therefore, other not-yet-defined mechanisms might be responsible for *HOX* gene perturbation in these patients. Another aspect is that CDX2 most probably is not exerting its transforming activity solely through induction of aberrant *HOX* gene expression. Another key question is how the ectopic expression of CDX2 is induced in human AML. Analyses of the promoter region of *CDX2* in patients with AML did not show any mutations or hypomethylation as an explanation for the aberrant expression of the gene.[55] Another possible explanation could be that constitutive activation of upstream regulators of *CDX2* would induce high expression levels of the gene in human AML. It was shown that *Cdx* genes are upstream regulated by the Wnt/β-catenin signaling pathway, the retinoic acid signaling pathway, and the FGF pathway.[57-59]

Taken together, our data underline that aberrant expression of *CDX2* is widespread in human AML. In addition, they show that high expression of this gene closely correlates with aberrant *HOX* gene expression in patients with AML, supporting a model in which CDX2 plays an important role in the development of AML with dysregulated *HOX* gene expression.

Acknowledgments

Authorship

References

1. Rowley JD. The critical role of chromosome translocations in human leukemias. Annu Rev Genet. 1998;32:495-519.

2. Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science. 1999;286:531-537.

3. Armstrong SA, Staunton JE, Silverman LB, et al. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. Nat Genet. 2002;30:41-47.

4. Debernardi S, Lillington DM, Chaplin T, et al. Genome-wide analysis of acute myeloid leukemia with normal karyotype reveals a unique pattern of homeobox gene expression distinct from those with translocation-mediated fusion events. Genes Chromosomes Cancer. 2003;37:149-158.

5. Buske C, Humphries RK. Homeobox genes in leukemogenesis. Int J Hematol. 2000;71:301-308.

6. Thorsteinsdottir U, Kroon E, Jerome L, Blasi F, Sauvageau G. Defining roles for HOX and MEIS1 genes in induction of acute myeloid leukemia. Mol Cell Biol. 2001;21:224-234.

7. Buske C, Feuring-Buske M, Antonchuk J, et al. Overexpression of HOXA10 perturbs human lymphomyelopoiesis in vitro and in vivo. Blood. 2001;97:2286-2292.

8. Sauvageau G, Lansdorp PM, Eaves CJ, et al. Differential expression of homeobox genes in functionally distinct CD34+ subpopulations of human bone marrow cells. Proc Natl Acad Sci U S A. 1994;91:12223-12227.

9. Abramovich C, Pineault N, Ohta H, Humphries RK. Hox genes: from leukemia to hematopoietic stem cell expansion. Ann N Y Acad Sci. 2005;1044:109-116.

10. Okada Y, Jiang Q, Lemieux M, Jeannotte L, Su L, Zhang Y. Leukaemic transformation by CALM-AF10 involves upregulation of Hoxa5 by hDOT1L. Nat Cell Biol. 2006;8:1017-1024.

11. Kohlmann A, Schoch C, Dugas M, et al. New insights into MLL gene rearranged acute leukemias using gene expression profiling: shared pathways, lineage commitment, and partner genes. Leukemia. 2005;19:953-964.

12. Alcalay M, Tiacci E, Bergomas R, et al. Acute myeloid leukemia bearing cytoplasmic nucleophosmin (NPMc+ AML) shows a distinct gene expression profile characterized by up-regulation of genes involved in stem-cell maintenance. Blood. 2005;106:899-902.

13. Deshpande AJ, Cusan M, Rawat VP, et al. Acute myeloid leukemia is propagated by a leukemic stem cell with lymphoid characteristics in a mouse model of CALM/AF10-positive leukemia. Cancer Cell. 2006;10:363-374.

14. Slany RK. When epigenetics kills: MLL fusion proteins in leukemia. Hematol Oncol. 2005;23:1-9.

15. Verhaak RG, Goudswaard CS, van Putten W, et al. Mutations in nucleophosmin (NPM1) in acute myeloid leukemia (AML): association with other gene abnormalities and previously established gene expression signatures and their favorable prognostic significance. Blood. 2005;106:3747-3754.

16. Falini B, Nicoletti I, Martelli MF, Mecucci C. Acute myeloid leukemia carrying cytoplasmic/mutated nucleophosmin (NPMc+ AML): biologic and clinical features. Blood. 2007;109:874-885.

17. Brooke NM, Garcia-Fernandez J, Holland PW. The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. Nature. 1998;392:920-922.

18. Chawengsaksophak K, de Graaff W, Rossant J, Deschamps J, Beck F. Cdx2 is essential for axial elongation in mouse development. Proc Natl Acad Sci U S A. 2004;101:7641-7645.

19. Mlodzik M, Gehring WJ. Expression of the caudal gene in the germ line of Drosophila: formation of an RNA and protein gradient during early embryogenesis. Cell. 1987;48:465-478.

20. Charite J, de Graaff W, Consten D, Reijnen MJ, Korving J, Deschamps J. Transducing positional information to the Hox genes: critical interaction of cdx gene products with position-sensitive regulatory elements. Development. 1998;125:4349-4358.

21. Gehring WJ. Homeo boxes in the study of development. Science. 1987;236:1245-1252.

22. Davidson AJ, Ernst P, Wang Y, et al. cdx4 mutants fail to specify blood progenitors and can be rescued by multiple hox genes. Nature. 2003;425:300-306.

23. Suh E, Traber PG. An intestine-specific homeobox gene regulates proliferation and differentiation. Mol Cell Biol. 1996;16:619-625.

24. Satoh K, Mutoh H, Eda A, et al. Aberrant expression of CDX2 in the gastric mucosa with and without intestinal metaplasia: effect of eradication of Helicobacter pylori. Helicobacter. 2002;7:192-198.

25. Ishikawa A, Sasaki M, Ohira S, et al. Aberrant expression of CDX2 is closely related to the intestinal metaplasia and MUC2 expression in intraductal papillary neoplasm of the liver in hepatolithiasis. Lab Invest. 2004;84:629-638.

26. Eda A, Osawa H, Satoh K, et al. Aberrant expression of CDX2 in Barrett's epithelium and inflammatory esophageal mucosa. J Gastroenterol. 2003;38:14-22.

27. Silberg DG, Sullivan J, Kang E, et al. Cdx2 ectopic expression induces gastric intestinal metaplasia in transgenic mice. Gastroenterology. 2002;122:689-696.

28. Chase A, Reiter A, Burci L, et al. Fusion of ETV6 to the caudal-related homeobox gene CDX2 in acute myeloid leukemia with the t(12;13)(p13;q12). Blood. 1999;93:1025-1031.

29. Rawat VP, Cusan M, Deshpande A, et al. Ectopic expression of the homeobox gene Cdx2 is the transforming event in a mouse model of t(12;13)(p13;q12) acute myeloid leukemia. Proc Natl Acad Sci U S A. 2004;101:817-822.

30. Bennett JM, Catovsky D, Daniel MT, et al. Proposed revised criteria for the classification of acute myeloid leukemia: a report of the French-American-British Cooperative Group. Ann Intern Med. 1985;103:620-625.

31. Schnittger S, Weisser M, Schoch C, Hiddemann W, Haferlach T, Kern W. New score predicting for prognosis in PML-RARA+, AML1-ETO+, or CBFBMYH11+ acute myeloid leukemia based on quantification of fusion transcripts. Blood. 2003;102:2746-2755.

32. Schnittger S, Schoch C, Dugas M, et al. Analysis of FLT3 length mutations in 1003 patients with acute myeloid leukemia: correlation to cytogenetics, FAB subtype, and prognosis in the AMLCG study and usefulness as a marker for the detection of minimal residual disease. Blood. 2002;100:59-66.

33. Haferlach T, Kohlmann A, Schnittger S, et al. Global approach to the diagnosis of leukemia using gene expression profiling. Blood. 2005;106:1189-1198.

34. Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004;5:R80.

35. Huber W, von Heydebreck A, Sultmann H, Poustka A, Vingron M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics. 2002;18:S96-S104.

36. Scheid S, Spang R. A stochastic downhill search algorithm for estimating the local false discovery rate. IEEE/ACM Trans Comput Biol Bioinform. 2004;1:98-108.

37. Imren S, Fabry ME, Westerman KA, et al. High-level beta-globin expression and preferred intragenic integration after lentiviral transduction of human cord blood stem cells. J Clin Invest. 2004;114:953-962.

38. National Cancer Institute–Frederick. Retrovirus Tagged Cancer Gene database. http://RTCGD.ncifcrf.gov. Accessed February 2007.

39. Bansal D, Scholl C, Frohling S, et al. Cdx4 dysregulates Hox gene expression and generates acute myeloid leukemia alone and in cooperation with Meis1a in a murine model. Proc Natl Acad Sci U S A. 2006;103:16924-16929.

40. Passegue E, Weisman IL. Leukemic stem cells: where do they come from? Stem Cell Rev. 2005;1:181-188.

41. Armstrong SA, Golub TR, Korsmeyer SJ. MLL-rearranged leukemias: insights from gene expression profiling. Semin Hematol. 2003;40:268-273.

42. Nakamura T, Largaespada DA, Lee MP, et al. Fusion of the nucleoporin gene NUP98 to HOXA9 by the chromosome translocation t(7;11)(p15;p15) in human myeloid leukaemia. Nat Genet. 1996;12:154-158.

43. Pineault N, Buske C, Feuring-Buske M, et al. Induction of acute myeloid leukemia in mice by the human leukemia-specific fusion gene NUP98-HOXD13 in concert with Meis1. Blood. 2003;101:4529-4538.

44. Knoepfler PS, Sykes DB, Pasillas M, Kamps MP. HoxB8 requires its Pbx-interaction motif to block differentiation of primary myeloid progenitors and of most cell line models of myeloid differentiation. Oncogene. 2001;20:5440-5448.

45. Crooks GM, Fuller J, Petersen D, et al. Constitutive HOXA5 expression inhibits erythropoiesis and increases myelopoiesis from human hematopoietic progenitors. Blood. 1999;94:519-528.

46. Sauvageau G, Thorsteinsdottir U, Hough MR, et al. Overexpression of HOXB3 in hematopoietic cells causes defective lymphoid development and progressive myeloproliferation. Immunity. 1997;6:13-22.

47. Thorsteinsdottir U, Sauvageau G, Hough MR, et al. Overexpression of HOXA10 in murine hematopoietic cells perturbs both myeloid and lymphoid differentiation and leads to acute myeloid leukemia. Mol Cell Biol. 1997;17:495-505.

48. Fischbach NA, Rozenfeld S, Shen W, et al. HOXB6 overexpression in murine bone marrow immortalizes a myelomonocytic precursor in vitro and causes hematopoietic stem cell expansion and acute myeloid leukemia in vivo. Blood. 2005;105:1456-1466.

49. Subramanian V, Meyer BI, Gruss P. Disruption of the murine homeobox gene Cdx1 affects axial skeletal identities by altering the mesodermal expression domains of Hox genes. Cell. 1995;83:641-653.

50. Suh E, Chen L, Taylor J, Traber PG. A homeodomain protein related to caudal regulates intestine-specific gene transcription. Mol Cell Biol. 1994;14:7340-7351.

51. Liu T, Zhang X, So CK, et al. Regulation of Cdx2 expression by promoter methylation, and effects of Cdx2 transfection on morphology and gene expression of human esophageal epithelial cells. Carcinogenesis. 2007;28:488-496.

52. Calon A, Gross I, Davidson I, et al. Functional interaction between the homeoprotein CDX1 and the transcriptional machinery containing the TATA-binding protein. Nucleic Acids Res. 2007;35:175-185.

53. Taylor JK, Levy T, Suh ER, Traber PG. Activation of enhancer elements by the homeobox gene Cdx2 is cell line specific. Nucleic Acids Res. 1997;25:2293-2300.

54. Lorentz O, Suh ER, Taylor JK, Boudreau F, Traber PG. CREB-binding protein interacts with the homeodomain protein Cdx2 and enhances transcriptional activity. J Biol Chem. 1999;274:7196-7199.

55. Scholl C, Bansal D, Dohner K, et al. The homeobox gene CDX2 is aberrantly expressed in most cases of acute myeloid leukemia and promotes leukemogenesis. J Clin Invest. 2007;117:1037-1048.

56. Takahashi O, Hamada J, Abe M, et al. Dysregulated expression of HOX and ParaHOX genes in human esophageal squamous cell carcinoma. Oncol Rep. 2007;17:753-760.

57. Simon M, Grandage VL, Linch DC, Khwaja A. Constitutive activation of the Wnt/beta-catenin signalling pathway in acute myeloid leukaemia. Oncogene. 2005;24:2410-2420.

58. Karajannis MA, Vincent L, Direnzo R, et al. Activation of FGFR1beta signaling pathway promotes survival, migration and resistance to chemotherapy in acute myeloid leukemia cells. Leukemia. 2006;20:979-986.

59. Piazza F, Gurrieri C, Pandolfi PP. The theory of APL. Oncogene. 2001;20:7216-7222.

Transformation by Oncogenic Mutants and Ligand-Dependent Activation of FLT3 Wild-type Requires the Tyrosine Residues 589 and 591

Sridhar Vempati,[1,2] Carola Reindl,[1,2] Ulla Wolf,[1,2] Ruth Kern,[1,2] Konstantin Petropoulos,[1,2] Vegi M. Naidu,[1,2] Christian Buske,[1,2] Wolfgang Hiddemann,[1,2] Tobias M. Kohl,[1,2] and Karsten Spiekermann[1,2,3]

Abstract

Purpose: Mutations in the receptor tyrosine kinase FLT3 are found in up to 30% of acute myelogenous leukemia patients and are associated with an inferior prognosis. In this study, we characterized critical tyrosine residues responsible for the transforming potential of active FLT3-receptor mutants and ligand-dependent activation of FLT3-WT.

Experimental Design: We performed a detailed structure-function analysis of putative autophosphorylation tyrosine residues in the FLT3-D835Y tyrosine kinase domain (TKD) mutant. All tyrosine residues in the juxtamembrane domain (Y566, Y572, Y589, Y591, Y597, and Y599), interkinase domain (Y726 and Y768), and COOH-terminal domain (Y955 and Y969) of the FLT3-D835Y construct were successively mutated to phenylalanine and the transforming activity of these mutants was analyzed in interleukin-3-dependent Ba/F3 cells. Tyrosine residues critical for the transforming potential of FLT3-D835Y were also analyzed in FLT3 internal tandem duplication mutants (FLT3-ITD) and the FLT3 wild-type (FLT3-WT) receptor.

Result: The substitution of the tyrosine residues by phenylalanine in the juxtamembrane, interkinase, and COOH-terminal domains resulted in a complete loss of the transforming potential of FLT3-D835Y-expressing cells which can be attributed to a significant reduction of signal tranducer and activator of transcription 5 (STAT5) phosphorylation at the molecular level. Reintroduction of single tyrosine residues revealed the critical role of Y589 and Y591 in reconstituting interleukin-3-independent growth of FLT3-TKD-expressing cells. Combined mutation of Y589 and Y591 to phenylalanine also abrogated ligand-dependent proliferation of FLT3-WT and the transforming potential of FLT3-ITD-with a subsequent abrogation of STAT5 phosphorylation.

Conclusion: We identified two tyrosine residues, Y589 and Y591, in the juxtamembrane domain that are critical for the ligand-dependent activation of FLT3-WT and the transforming potential of oncogenic FLT3 mutants.

FLT3 is a member of the class III protein receptor tyrosine kinase family (RTK) that is characterized by five extracellular immunoglobulin-like domains, a juxtamembrane domain (JM), and two protein tyrosine kinase domains (TKD) split by an interkinase domain (IK; ref. 1). The class III receptors also include KIT, FMS, platelet-derived growth factor receptor-α (PDGFRA), and platelet-derived growth factor receptor-β (PDGFRB). Binding of FLT3 ligand (FL) to its receptor induces dimerization, phosphorylation, and subsequent activation of downstream signaling pathways such as signal tranducer and activator of transcription 5 (STAT5), Ras/mitogen-activated protein kinase (MAPK), and phosphatidylinositol 3-kinase/AKT (2–6). FLT3 has been shown to play an important role in normal hematopoiesis and is highly expressed in CD34[+] hematopoietic progenitor cells (2, 7–9).

Activating mutations of FLT3 are found in 30% of patients with acute myelogenous leukemia (AML) and are associated with an inferior clinical outcome (10–12). FLT3 internal tandem duplications (FLT3-ITD) represent one of the most frequent genetic alterations and occur in ∼20% to 25% of patients. These mutations have a variable length resulting in an elongated FLT3 protein with constitutive kinase activity and are associated with higher leukocyte counts at diagnosis (13). A second class of FLT3 mutations primarily occurs at the highly conserved residue D835 in the TKD and is present in 7% to 8% of all patients with AML (14, 15).

Fig. 1. Successive Y→F substitutions in the FLT3-TKD mutant indicate the critical role of Y589 and Y591 for the transforming potential of mutated FLT3. *A,* overview of Y→F mutations created in the FLT3-TKD-D835Y construct. *B,* Ba/F3 cells stably transduced with FLT3-WT, FLT3-D835Y, D835Y-KD, D835Y-F1 to D835Y-F8 were seeded at a density of 4×10^4/mL in the absence or presence of IL-3. Viable cells were counted after 72 h by trypan blue exclusion. The proliferation of cells in the presence of IL-3 was defined as 100% (control). SE of three independent experiments are shown. *C,* long-term proliferation assay of Ba/F3 cells expressing FLT3-WT, FLT3-D835Y, D835Y-F1 to D835Y-F8, D835Y-4F. Cells (4×10^4/mL) were seeded and counted every 3 d for 13 d. To avoid overgrowth, cells were split every 3 d. The proliferation of FLT3-TKD-expressing cells was defined as 100% (control). SE of three independent experiments are shown. The FLT3-D835Y-kinase dead (KD) mutant was used as a negative control.

We and others have recently shown that FLT3-ITD mutants enhance the proliferative potential of hematopoietic progenitor cells and collaborate with fusion oncogenes, such as the AML1-ETO or PML-RARα oncoproteins, to induce AML *in vivo* (16, 17). FLT3 selective tyrosine kinase inhibitors are currently in clinical trials for combined treatment with conventional chemotherapy (18, 19). Nevertheless, the underlying mechanism of transformation that is exerted by constitutively activated FLT3 remains elusive.

In the present study, we performed a detailed structure-function analysis of FLT3-TKD and FLT3-ITD receptor mutants to characterize the molecular mechanisms of FLT3-induced

transformation. We identified two tyrosine residues, 589 and 591, in the juxtamembrane region of FLT3 that are indispensable for the transforming potential of both FLT3-ITD and FLT3-TKD mutants and ligand-dependent activation of FLT3 wild-type (FLT3-WT).

Materials and Methods

Reagents and cell lines. Low-passage murine Ba/F3 cells were obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (Braunschweig, Germany) and maintained in RPMI 1640 with 10% fetal bovine serum and 10% WEHI conditioned

medium as a source of murine interleukin-3 (IL-3) when indicated. Recombinant human FL was purchased from Promokine and recombinant murine IL-3 from Biosource (Solingen, Germany).

DNA constructs and vectors. The FLT3-ITD-W51 construct contains a 7–amino acid duplicated sequence (REYEYDL) inserted between amino acids 601 and 602 of human FLT3-WT, and the FLT3-ITD-NPOS contains a 28–amino acid duplicated sequence (CSSDNEY-FYVDFREYEYDLKWEFPRENL) inserted between amino acids 611 and 612 of FLT3-WT(17). The FLT3-TKD carries substitution (point mutation) of aspartic acid to tyrosine at position 835 of FLT3-WT. All FLT3 constructs were subcloned in the MSCV-IRES-EYFP retroviral expression vector (kindly provided by R.K. Humphries, The Terry Fox Laboratory, British Columbia Cancer Agency, Vancouver, Canada).

In vitro *site directed mutagenesis, DNA sequencing, and nomenclature.* Point mutations were introduced into FLT3-WT cDNA, FLT3-ITD (FLT3-W51 and FLT3-NPOS), and FLT3-TKD by site-directed mutagenesis using the QuickChange kit from Stratagene as described previously (20). The correct sequence of all constructs was confirmed by complete nucleotide sequencing. Mutants with successive single tyrosine residues at the FLT3-D835Y background were named F1 to F8 (Fig. 1A), the mutant 4F contains four tyrosine-to-phenylalanine (Y→F) mutations in the JM domain (YF589/591/597/599). FLT3-D835Y-kinase dead (KD) indicates the mutant containing the K644R mutation leading to the complete loss of kinase activity as described previously (16). Mutations in the JM domain of FLT3-W51 and FLT3-NPOS were named according to the modified positions. All modifications of FLT3-ITD constructs were done in the wild-type, not the duplicated, DNA stretch (Fig. 3A).

Expression of CD135 by flow cytometry. Determination of FLT3 expression by FACS analysis was carried out as described previously (20).

Cell proliferation of Ba/F3 cells and assessment of apoptotic cell death by flow cytometry. IL-3-dependent Ba/F3 cells stably expressing the indicated constructs were seeded at a concentration of 4×10^4/mL in the presence or absence of IL-3 and FL as described previously (21). Assessment of apoptotic cells was carried out by Annexin V/7-aminoactinomycin D staining as recommended by the manufacturer (Annexin V-phycoerythrin apoptosis detection kit; Becton Dickinson) using a FACSCalibur flow cytometer (Becton Dickinson).

Antibodies. The following antibodies were used: anti-FLT3 antibody (Santa Cruz), anti-phosphorylated STAT5-Tyr[694] (New England Biolabs), anti-STAT5 (Santa Cruz), anti-pY (Santa Cruz), anti-phosphorylated p44/42 MAPK (New England Biolabs), and anti-p44/42 MAPK (New England Biolabs).

Transient transfection of 293 cells and stable transduction of Ba/F3 cells (22), immunoprecipitation and Western blot analysis (22), and GST pull-down (23) were done as described previously.

Results

Structure-function analysis of successive Y→F mutations in the FLT3-TKD background. In a screening approach to study the contribution of single tyrosine residues to the transforming potential of the FLT3 receptor, we successively mutated the tyrosine residues 566, 572, 589, 591, 597, 599, 725, 768, 955, and 969 of the intracellular domain of FLT3-TKD-D835Y to phenylalanine and named the constructs as D835Y-F1 to D835Y-F8 and D835Y-4F (Fig. 1A). The tyrosine residues in TKD1 (Y630, Y688, Y693, Y696, and Y702) and TKD2 (Y793, Y842, Y865, Y874, Y889, Y899, Y913, and Y919) are likely to be indispensable for the enzyme function as shown previously for the related KIT receptor (24); hence, these residues were not mutated. IL-3-dependent Ba/F3 cells were retrovirally transduced with the indicated constructs (F1-F8 and 4F) and stable expression was confirmed by CD135 staining as described in

Materials and Methods (data not shown). The transduced cell lines were characterized by short-term and long-term proliferation assays. In short-term proliferation assays, the cell lines carrying constructs D835Y-F1 (Y597F and Y599F) or D835Y-F2 (Y572F, Y597F, and Y599F) showed comparable or even increased proliferation rates compared with cell lines expressing the parental FLT3-TKD construct after IL-3 withdrawal for 72 h (Fig. 1B). In detail, the proliferation rates were 90% (D835Y-F1) and 117% (D835Y-F2) compared with cells expressing FLT3-TKD-D835Y (74%; Fig. 1B). The cell lines expressing the D835Y-F3 construct showed a significantly reduced proliferation after 72 h, suggesting that substitution of Y589 and Y591 impaired the transforming potential of FLT3-TKD (Fig. 1B). All the cell lines that harbor the substitutions of Y589 and Y591 (D835Y-F3 to D835Y-F8) showed a reduced transforming potential of 2.7% to 25.7% (Fig. 1B). The lowest transforming potential was observed in cell lines expressing D835Y-4F (data not shown) and D835Y-F8 (Fig. 1B). The effect was even more pronounced in long-term proliferation assays as cell lines expressing constructs D835Y-F3 to D835Y-F8 and D835Y-4F showed a very low transforming potential or completely died after 10 days in
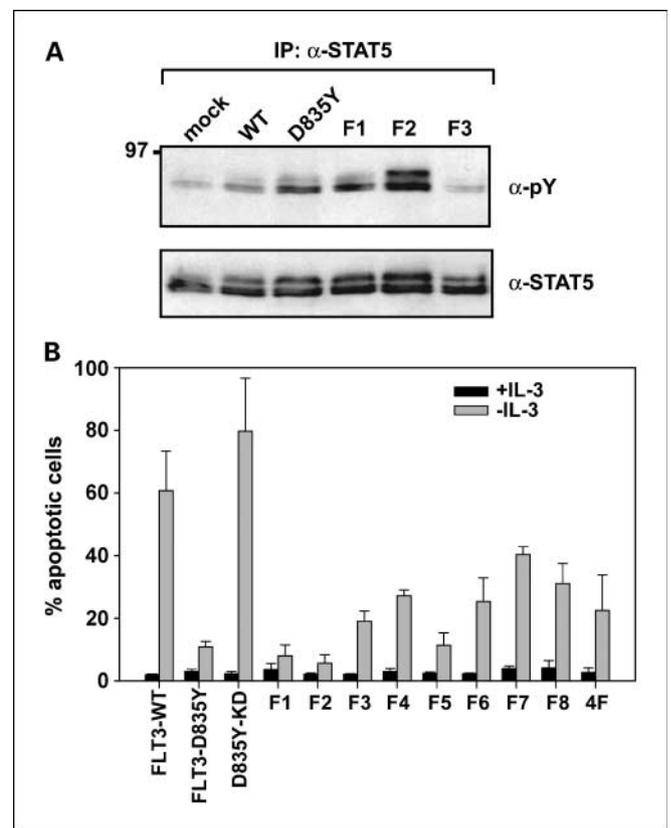
Fig. 2. Y589 and Y591 mediate the phosphorylation of STAT5 and resistance to apoptotic cell death in FLT3-D835Y-expressing cells. *A,* D835Y-F1 to D835Y-F3, FLT3-D835Y, FLT3-WT, or mock-transduced cell lines were starved for 24 h in the presence of 0.3% fetal bovine serum. Crude cell lysates were immunoprecipitated with STAT5 antibody, separated by SDS-PAGE, and blotted on a nitrocellulose membrane. Blots were then incubated with anti-phospho tyrosine antibody, stripped, and reblotted with anti-STAT5 antibody. *B,* Ba/F3 cell lines transduced with D835Y-F1 to D835Y-F8, D835Y-4F, unmanipulated FLT3-D835Y or FLT3-WT were cultured in the presence or absence of IL-3 for 48 h and stained with Annexin V and 7-aminoactinomycin D. The percentage of apoptotic cells was determined by FACS analysis.
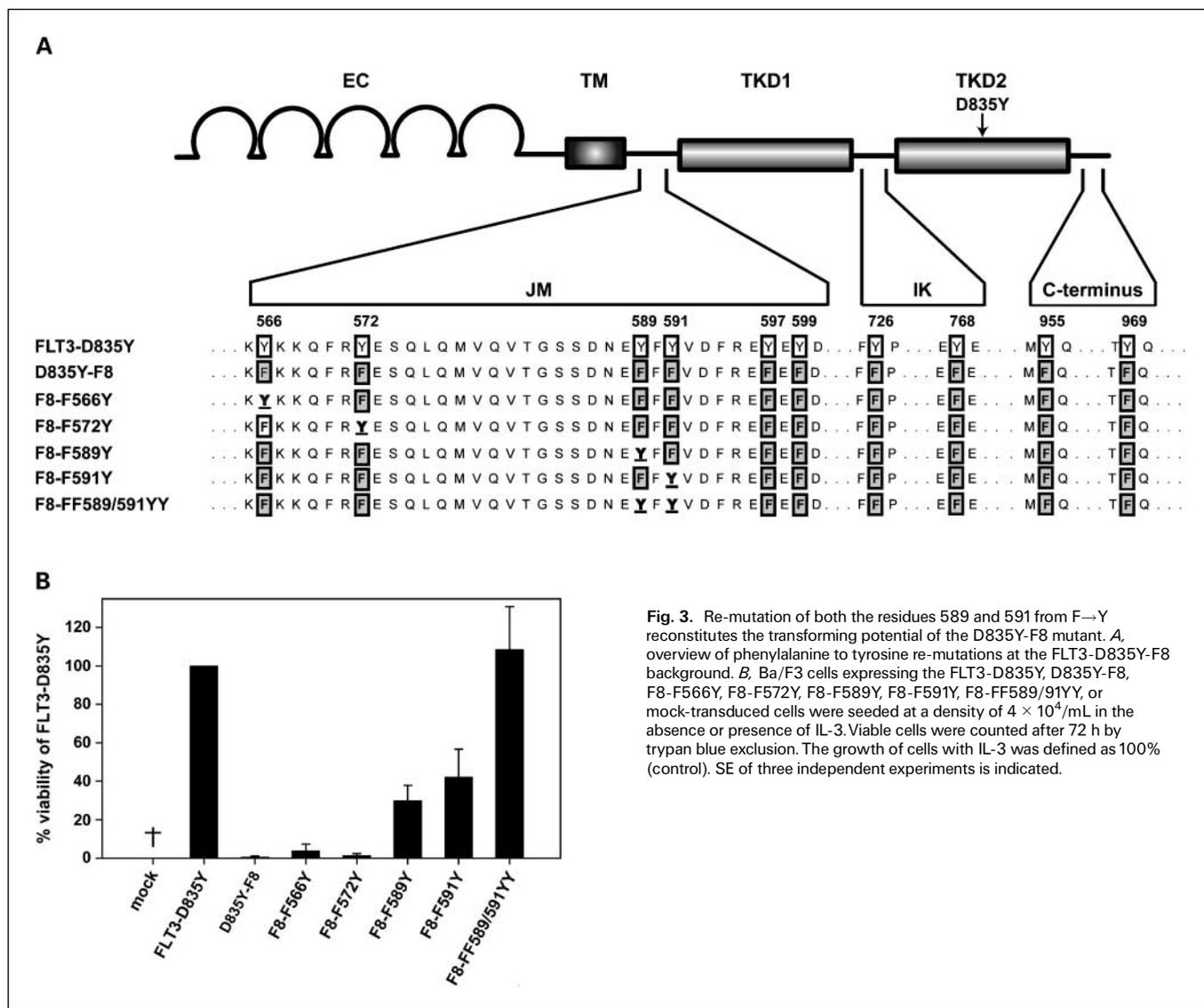
Fig. 3. Re-mutation of both the residues 589 and 591 from F→Y reconstitutes the transforming potential of the D835Y-F8 mutant. *A*, overview of phenylalanine to tyrosine re-mutations at the FLT3-D835Y-F8 background. *B*, Ba/F3 cells expressing the FLT3-D835Y, D835Y-F8, F8-F566Y, F8-F572Y, F8-F589Y, F8-F591Y, F8-FF589/91YY, or mock-transduced cells were seeded at a density of 4×10^4/mL in the absence or presence of IL-3. Viable cells were counted after 72 h by trypan blue exclusion. The growth of cells with IL-3 was defined as 100% (control). SE of three independent experiments is indicated.

culture, whereas the transforming potential of cell lines expressing D835Y-F2 did not significantly differ from cells expressing FLT3-TKD (Fig. 1C). Next, we analyzed the phosphorylation status of FLT3. Cells expressing FLT3-D835Y-F3 to FLT3-D835Y-F7 showed a successive reduction of the amount of phosphorylated FLT3 when compared with cells expressing nonmanipulated FLT3-TKD-D835Y (Supplementary Fig. S1), whereas no FLT3 phosphorylation was observed in D835Y-F8.

FLT3-TKD mutants carrying the Y→F substitutions of 589 and 591 show reduced STAT5 phosphorylation and increased apoptotic cell death. An important downstream signaling pathway of activated FLT3 is the STAT5 pathway. STAT5 has been shown to contribute essentially to the transforming potential of the activated FLT3 receptor *in vitro* and *in vivo* (25–27). To investigate the activation of the STAT5 signaling pathway, we prepared crude cell lysates of serum-starved Ba/F3 cells transduced with either vector control (mock), FLT3-WT, FLT3-D835Y, D835Y-F1, D835Y-F2, and D835Y-F3. Immunoprecipitated STAT5 was analyzed with specific antibody against phosphotyrosine (pY). We could clearly show that the

expression of D835Y-F3 did not show any phosphorylation of STAT5 compared with FLT3-TKD-D835Y, which is in accordance with the proliferation data (Fig. 2A). Conversely, the level of phosphorylated STAT5 was higher in the D835Y-F2 mutant that showed enhanced proliferation when compared with FLT3-D835Y (Fig. 2A).

Having shown that Y589 and Y591 play an important role in the transforming potential and STAT5 phosphorylation of FLT3-TKD mutants, we next analyzed the cell lines expressing D835Y-F1 to D835Y-F8, D835Y-4F, FLT3-TKD-D835Y, and D835Y-KD for resistance to apoptotic cell death after 48 h in the absence of IL-3. The apoptotic rate was lowest in the D835Y-F1 (7.9%) and D835Y-F2 (5.6%), whereas the highest levels were observed in D835Y-F7 (40%) and D835Y-F8 (32%) cells (Fig. 2B). FLT3-WT and D835Y-KD were used as positive controls and showed an apoptotic rate, of 60% and 80%, respectively. In all experiments, IL-3-stimulated cells were used as negative controls.

Re-mutation of residues F589 and F591 to tyrosine reconstitutes the transforming potential of the FLT3-D835Y-F8 mutant. To extend the finding that Y589 and Y591 play an important role in the transforming potential of FLT3-TKD, single or combined

phenylalanine residues were re-mutated to tyrosine in the D835Y-F8 mutant (Fig. 3A). Reintroduction of tyrosine at amino acid 566 (F8-F566Y) or amino acid 572 (F8-F572Y) induced only negligible IL-3-independent proliferation rates of 3.6% and 1.2%, respectively, when compared with unmanipulated FLT3-TKD-D835Y (Fig. 3B). In contrast, the reintroduction of tyrosine residues at amino acid 589 (F8-F589Y) or amino acid 591 (F8-F591Y) led to a partial reconstitution of the transforming potential of the mutated FLT3 receptor by

30% and 42%, respectively (Fig. 3B). Reintroduction of tyrosine at positions 589 and 591 (F8-FF589/591YY) completely reconstituted the transforming potential of the D835Y-F8 mutant (Fig. 3B).

Single Y→F mutations of Y589, Y591, or Y597 in the FLT3-ITD background do not reduce the transforming potential. To further characterize the roles of tyrosine residues 589 and 591 in the FLT3-ITD receptor, single Y→F mutations were introduced at amino acid 589 (FLT3-ITD-Y589F) or amino acid
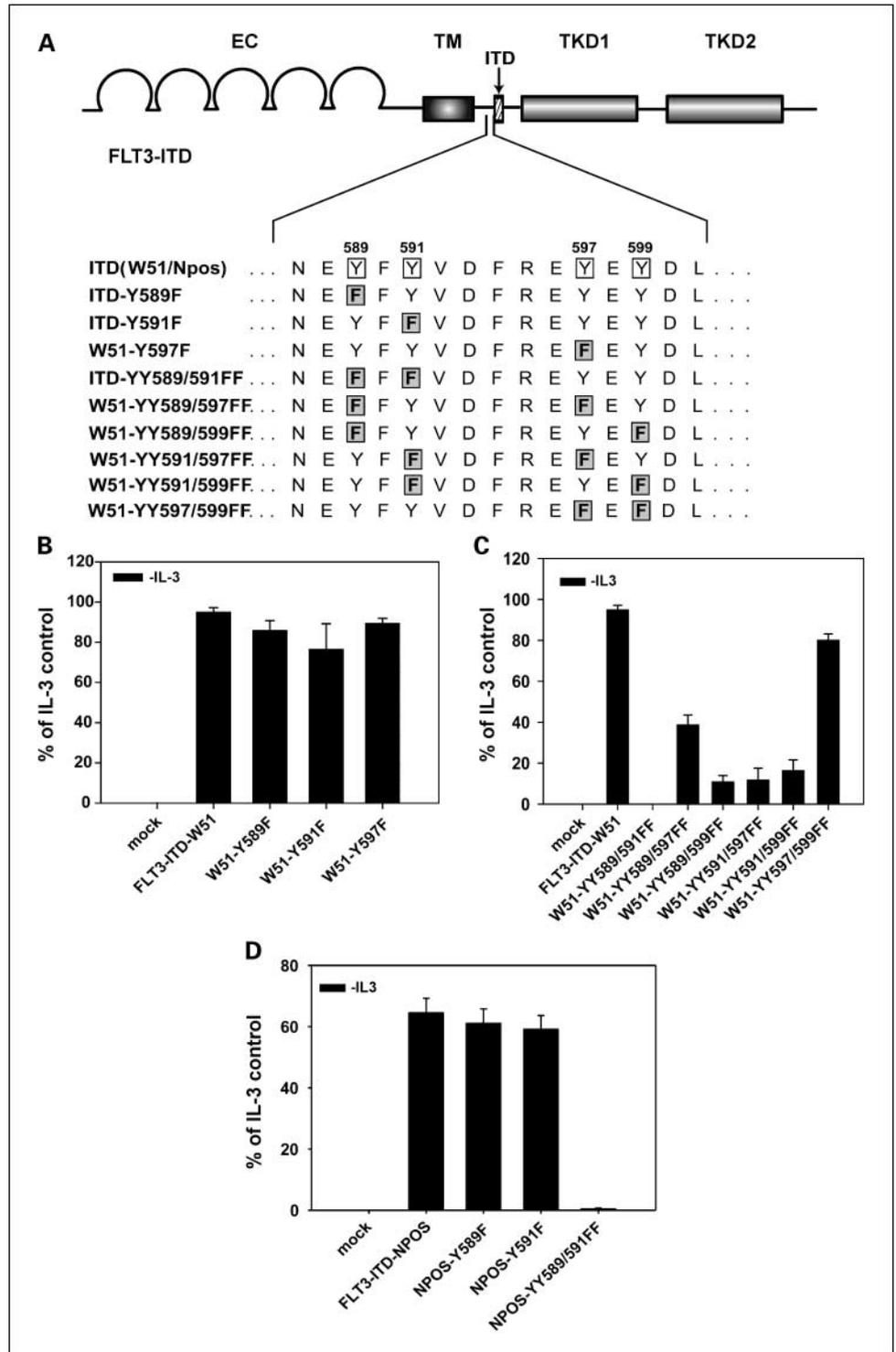
Fig. 4. Functional analysis of the role of tyrosine residues Y589, Y591, Y597, and Y599 in the FLT3-ITD background. *A,* overview of Y→F mutations generated in the FLT3-ITD background. *B,* Ba/F3 cells expressing the FLT3-W51, FLT3-W51-Y589F or FLT3-W51-Y591F, FLT3-W51-Y597F, or mock-transduced cells were seeded at a density of 4×10^4/mL in the absence or presence of IL-3. Viable cells were counted after 72 h by trypan blue exclusion. The proliferation of cells with IL-3 was defined as 100% (control). SE of three independent experiments is indicated. *C* and *D,* FLT3-ITD mutants were stably expressed in Ba/F3 cells and analyzed as described in *B.*
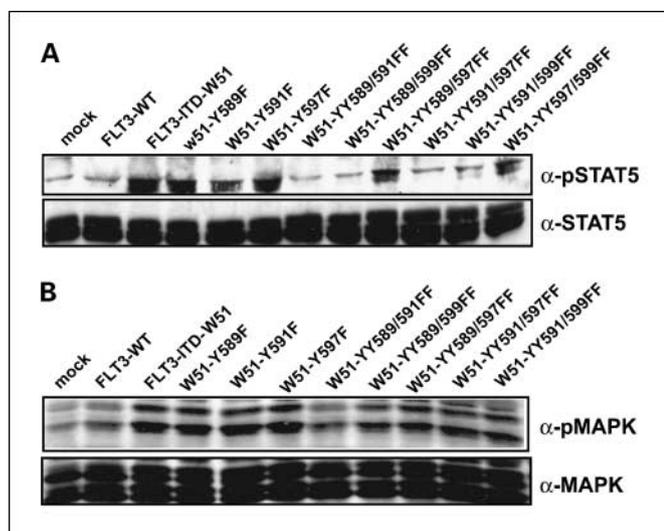
Fig. 5. Y589 and Y591 play an important role in the phosphorylation of STAT5 and MAPK pathways of FLT3-ITD. *A* and *B,* FLT3 constructs expressing cells were starved for 24 h in the presence of 0.3% fetal bovine serum. Crude cell lysates were subjected to Western blot analysis with an anti-phospho STAT5 or phospho-MAPK antibody, stripped, and reblotted with anti-STAT5 or MAPK antibody.

591 (FLT3-ITD-Y591F) in the FLT3-ITD-W51 and FLT3-ITD-NPOS constructs (Fig. 4A). All manipulations were done in the wild-type, not in the duplicated DNA stretch. As shown in Fig. 4B and D, cell lines expressing FLT3-ITD-Y589F or FLT3-ITD-Y591F showed no significant reduction of their transforming potential after IL-3 withdrawal when compared with unmanipulated FLT3-ITD cells. To extend this finding that single substitutions of tyrosine residues have no effect on transforming potential of FLT3-ITD-expressing cells, we generated another Y→F substitution mutant at amino acid 597 (FLT3-W51-Y597F; Fig. 4B). Cells expressing FLT3-ITD-Y597F show a comparable transforming potential to unmanipulated FLT3-W51 by phenylalanine.

Combined substitution of both Y589 and Y591 abrogates by phenylalanine the transforming potential of FLT3-ITD. As single Y→F mutants of amino acids 589, 591, and 597 had no effect on the transforming potential of FLT3-ITD-expressing cells, we generated a variety of double substitution mutants of JM domain tyrosine residues 589, 591, 597, and 599: FLT3-W51-YY589/591FF, FLT3-W51-YY589/597FF, and FLT3-W51-YY589/599FF in addition to FLT3-W51-YY591/597FF, FLT3-W51-591/99FF, and FLT3-W51-YY597/599FF (Fig. 4A). In proliferation assays, Ba/F3 cells expressing FLT3-W51-YY589/591FF were unable to proliferate in the absence of IL-3 (Fig. 4C). Except for FLT3-W51-YY597/599FF cell lines expressing W51-YY589/597FF, W51-YY589/599FF, W51-YY591/597FF, or W51-YY591/99FF showed a significant reduction in the transforming potential by 58% to 82% when compared with unmanipulated FLT3-ITD-W51 cells (Fig. 4C). Substitution of both Y589 and Y591 with phenylalanine (FLT3-NPOS-YY589/591FF) in a structurally different ITD (FLT3-ITD-NPOS) induced a phenotype identical to FLT3-W51-YY589/591FF-expressing cells (Fig. 4D).

Western blot analysis revealed a slight reduction of FLT3 phosphorylation in FLT3-W51-YY589/591FF cells when compared with FLT3-ITD-W51 cells (Supplementary Fig. S1).

Double Y→F substitution mutants of 589, 591, 597, and 599 show reduced activation of STAT5 and MAPK pathways. To investigate the activation of the STAT5 signaling pathway, we prepared whole-cell lysates of serum-starved Ba/F3 cells transduced with either vector control (MIY) or FLT3-WT, FLT3-ITD-W51, FLT3-W51-Y589F, FLT3-W51-Y591F, FLT3-W51-Y597F, FLT3-W51-YY589/591FF, FLT3-W51-YY589/597FF, FLT3-W51-YY589/599FF, FLT3-W51-YY591/597FF, FLT3-W51-591/99FF, or FLT3-W51-YY597/599FF. Single Y-F substitution and FLT3-ITD-W51-expressing cells showed comparable levels or slightly reduced levels (FLT3-W51-Y591F) of phosphorylated STAT5. The double substitution mutants FLT3-W51-YY589/591FF, FLT3-W51-YY589/599FF, FLT3-W51-YY591/597FF, and FLT3-W51-591/99FF showed reduced STAT5 phosphorylation. FLT3-W51-YY597/599FF and FLT3-W51-YY589/597FF showed a slightly reduced STAT5 phosphorylation, when compared with nonmanipulated FLT3-ITD-W51 or single Y→F substitution mutants (FLT3-W51-Y589F, FLT3-W51-Y591F, and FLT3-Y597F) cells (Fig. 5A).

Cells expressing FLT3-W51-YY589/599FF, FLT3-W51-YY591/597FF, and FLT3-W51-591/99FF showed a transforming potential of ~20% but did not show any STAT5 phosphorylation. To further analyze this phenomenon, we measured MAPK phosphorylation, another important downstream signaling pathway of the activated FLT3 receptor. Analysis of the lysates by immunoblotting with a specific antibody against phosphorylated MAPK showed that all the cells with double Y→F substitution mutants, except FLT3-W51-YY589/591FF, showed phosphorylation of MAPK but at a reduced rate compared with nonmanipulated FLT3-ITD-W51 or single Y→F substitution mutant cells (Fig. 5B).

These results show that the residues Y589 and Y591 play an important role in STAT5 and MAPK activation by the FLT3-ITD receptor.

Y589 and Y591 are indispensable for ligand-dependent signaling of the FLT3-WT receptor. Because combined Y→F mutations of Y589 and Y591 had severe effects on the transforming potential of FLT-TKD and FLT3-ITD expressing cells, we analyzed the role of Y589 and Y591 in the signaling properties of the FLT3-WT receptor. We generated the Y→F substitution mutant of both Y589 and Y591 (FLT3-WT-YY589/591FF) in FLT3-WT (Fig. 6A). Overexpression of FLT3-WT-YY589/591FF did not induce any IL-3-independent growth, but proliferation was totally abrogated in the presence of FL when compared with cells expressing FLT3-WT (Fig. 6B).

Discussion

In contrast to the studies published to date on structural motifs critical for the transformation mediated by FLT3 mutant receptors (28–30), our study presents data not only on FLT3-WT but also on FLT3-TKD and FLT3-ITD receptor mutants. Because the structure of the JM domain is altered by the insertion of an ITD (31), the study was focused on a FLT3 receptor with an activating mutation in the TKD domain that allows the analysis of all critical tyrosine residues. During the course of our experiments, we identified amino acids Y589 and Y591 as critical residues for the transformation exerted by FLT3-TKD. The essential role of tyrosine residues YY589/591 in mediating transformation was then confirmed in ITD mutants and the FLT3-WT receptor.

Our results indicate that a combined Y→F substitution of amino acids Y597, Y599, and Y572 in FLT3-TKD (F1 and F2) induce a higher proliferation rate in IL3-deprived Ba/F3 cells compared with unmanipulated FLT3-TKD (Fig. 1B and C). The FLT3 crystal structures shows that amino acids Y572 and Y599 are the only two tyrosine residues that form a tight pocket in FLT3 and might be responsible for maintenance of the inactive state of FLT3-WT (31). This finding is also in line with previous reports on the homologous Y533 in KIT showing the negative regulatory function of this residue (32). Mutation of Y533 (Y572 in FLT3) to alanine in wild-type KIT led to spontaneous phosphorylation of the receptor. The lack of phenolic hydroxyl oxygen in Y→F substituted 572 and 599 probably disrupts the interactions formed by these amino acids and therefore could account for the increased transforming potential of D835Y-F1 and D835Y-F2 compared with FLT3-TKD-D835Y.

Our results clearly show that BalF3 cell lines carrying FLT3-TKD constructs with Y→F substitution of Y589 and concomitantly Y591 (D835Y-F3-F8 and D835Y-4F) lead to a reduced transforming potential in short-term cultures and total loss of transforming potential in long-term cultures (Fig. 1B and C). To determine whether loss of FLT3 phosphorylation results in concomitant loss of transforming potential, we analyzed the FLT3 phosphorylation in FLT3-TKD-D835Y mutants. All the mutants D835Y-F3 to D835Y-F7 with little or no transforming potential in short-term and long-term assays exhibited FLT3 phosphorylation but at a reduced level when compared with nonmanipulated FLT3-TKD-D835Y (Supplementary Fig. S1). There was a total lack of phosphorylation in D835Y-F8 mutant in which all tyrosines in the JM, Ki and C-terminal region were mutated. The loss of the transforming potential in mutants D835Y-F3 and D835Y-4F, even in the presence of FLT3 phosphorylation, suggests that Y589 and Y591 act as docking sites for downstream signaling molecules.

Y589 and Y591 have been shown to be conserved in the related tyrosine kinases KIT (Y568, Y570; ref. 33), PDGFRA, (Y572, Y576; refs. 34, 35), and PDGFRB (Y579;Y581; ref. 36), and substitution of homologous residues with phenylalanine in these RTK's reduced their kinase activity (37–41). Previous studies on PDGFRB have shown that Y579 and Y581 homologous to Y589 and Y591 in FLT3 bind STAT5 (36). STAT5 is an important downstream signaling pathway of FLT3 and phosphorylated STAT5 has been detected in blasts of 20% to 80% of patients with AML (42–45). Our data clearly show that Y589 and Y591 play an important role in STAT5 autophosphorylation and apoptosis. All cell lines expressing activated FLT3 with Y→F substitutions of both Y589 and Y591 showed reduced STAT5 phosphorylation and an increased rate of apoptotic cell death after cytokine withdrawal (Fig. 2A and B). Our data is in contradiction to a recent report by Rocnik et al., in which no difference in the STAT5 phosphorylation was observed between FLT3-TKD-D835Y and FLT3-D835Y-YY589/591FF (29). A possible explanation for this discrepancy might be the use of different cell culture systems, like the usage of 32D cells by Rocnik et al. instead of Ba/F3 cells and starving of the cells for only 4 h, compared to 24 h starvation period in this report. STAT5 has been shown to induce the expression of the antiapoptotic protein BCL-xL (46) thereby protecting the cells from apoptosis (47). D835Y-F1 and D835Y-F2 cells showed high levels of phosphorylated STAT5 corresponding to a low

rate of apoptosis after IL-3 withdrawal (5-8%). In contrast, the D835Y-F7 and D835Y-F8, which showed no STAT5 autophosphorylation (data not shown), displayed a higher apoptotic rate after IL-3 withdrawal (32-40%; Fig. 2B).

To further confirm the role of Y589 and Y591 for the transforming potential of FLT3-TKD, we re-mutated single phenylalanine residues to tyrosine in the D835Y-F8 background (Fig. 3A). Single re-mutation of 566 or 572 (F8-F566Y or F8-F572Y) did not induce any significant IL-3-independent proliferation in BalF3 cells. The reintroduction of tyrosine at amino acid 589 (F8-F589Y) or amino acid 591 (F8-F591F) partially reconstituted the transforming potential of the D835Y-F8 mutant to 30% to 42% (Fig. 3B). Interestingly, reintroduction of tyrosines at both 589 and 591 (F8-FF589/591YY) totally reconstituted the transforming potential of the D835Y-F8 mutant. These results suggest that Y589 and 591 are two critical amino acids for the transforming potential of FLT3-TKD. FLT3 phosphorylation studies in the F8-F589Y, F8-F591Y, and FLT3-FF589/F591YY showed weak FLT3 phosphorylation (data not shown). This result points to a role of Y589 and Y591 as docking sites for downstream signaling molecules independent of FLT3 phosphorylation.

Next, we analyzed the role of Y589 and Y591 for the transforming potential of FLT3-ITD mutants in Ba/F3 cells. Single Y→F mutation of 589 (ITD-Y589F) and 591 (ITD-Y591F) in two structurally different FLT3-ITDs (W51/NPOS) did not significantly affect the transforming potential of cell lines expressing these constructs (Fig. 4B and D). However, Y→F substitution of both Y589 and Y591 (ITD-YY589/591FF) totally
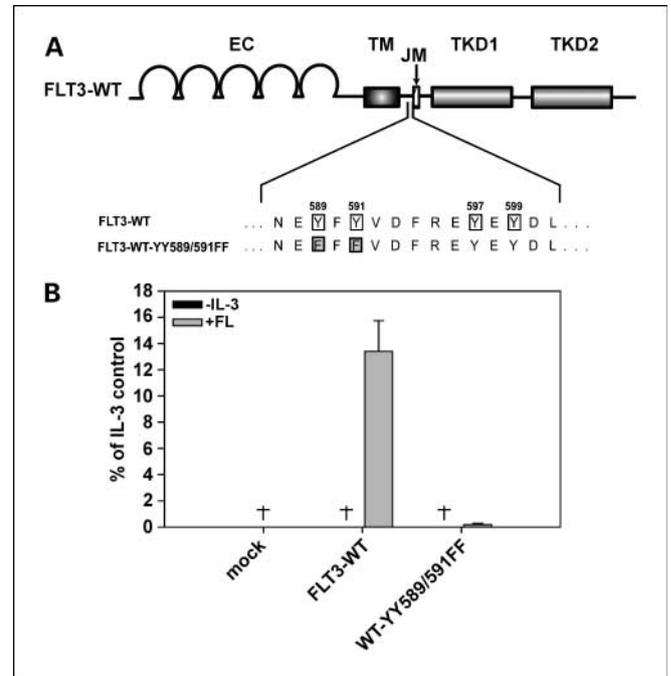
Fig. 6. Substitution of both Y589 and Y591 by phenylalanine abrogates the ligand-dependent activation of FLT3-WT in Ba/F3 cells. *A,* overview of Y→F substitutions of both 589 and 591 generated in FLT3WT. *B,* Ba/F3 cells expressing the FLT3-WT, FLT3-WT-YY589/591FF, or mock-transduced cells were seeded at a density of 4×10^4/mL in the absence or presence of IL-3 or FL (60 ng/mL). Viable cells were counted after 72 h by trypan blue exclusion. The proliferation of cells with FL was defined as 100% (control). SE of three independent experiments is indicated.

abolished the transforming potential of FLT3-ITD (Fig. 4C and D). Our data are discrepant to the results published by Kiyoi et al., where a FLT3-ITD carrying Y589, Y591 Y597, and Y599 mutated to phenylalanine did not show any difference in the transforming potential in 32D cells when compared with nonmanipulated FLT3-ITD (28). However, a similar study by Rocnik et al. in 32D cells expressing the mutant FLT3-ITD-YY589/591FF showed a reduced transforming potential when compared with nonmanipulated FLT3-ITD (29). To address these discrepancies, we expressed the mutants FLT3-ITD-W51/NPOS, FLT3-W51-YY589/591FF and NPOS-YY589/591FF in 32D cells. No factor-independent growth was observed in 32D cells expressing FLT3-W51/±NPOS-YY589/591FF confirming our data observed in Ba/F3 cells (Supplementary Fig. S2). Moreover, our results are in line with the study by Rocnik et al. reporting that transplanted mice with bone marrow cells carrying a FLT3-ITD-YY589/591FF construct showed no difference in survival compared with mice transplanted with FLT3-WT-expressing bone marrow. In contrast, transplantation with FLT3-ITD-transduced bone marrow cells led to a lethal myeloproliferative disease with short latency (29).

Next, we analyzed other combinations of tyrosine residues in the JM domain responsible for the transforming potential of FLT3-ITD. Hence, we analyzed the effect of Y→F substitution of different combinations of JM domain tyrosines (30) (Y589, Y591, Y597, and Y599). Cell lines expressing FLT3-W51-YY589/597FF, FLT3-W51-YY589/599FF, FLT3-W51-YY591/597FF, and FLT3-W51-591/99FF conferred a low transforming potential (Fig. 4C). FLT3-W51-YY597/599FF-expressing cells did not show an impaired IL-3-independent growth. These results were further supported by analyses of STAT5 phosphorylation (Fig. 5A). We observed a slightly reduced STAT5 phosphorylation in the cells expressing FLT3-W51-Y591F mutant when compared with FLT3-W51-Y589F, FLT3-W51-Y597F, and nonmanipulated FLT3-ITD-W51 (Fig. 5A). This observance could be explained by a recent study, in which *in vitro* mapping of autophosphorylated tyrosine sites in FLT3-ITD revealed Y591 as the sole autophosphorylated site in the JM domain. Also, Y591 along with Y589 were reported to be the putative binding sites for STAT5 in another study (29).

In accordance with the proliferation data, cell lines expressing FLT3-W51-YY589/591FF showed no STAT5 phosphorylation. Cell lines expressing FLT3-W51-YY589/597FF and FLT3-W51-YY597/599F showed a reduced STAT5 phosphorylation

when compared with unmanipulated FLT3-ITD (Fig. 5A). Surprisingly, cell lines expressing FLT3-W51-YY589/599FF, FLT3-W51-YY591/597FF, and FLT3-W51-591/99FF, which showed transforming potential of ~20%, did not show any STAT5 phosphorylation (Fig. 5A). Therefore, we further analyzed phosphorylation of MAPK, another important FLT3 downstream signaling molecule. All the cells expressing double Y→F substitution mutants of FLT3-ITD-W51, except FLT3-W51-YY589/591FF, showed MAPK phosphorylation but at lower rate compared with unmanipulated FLT3-ITD cells (Fig. 5B). These results suggest that combination of amino acid residues Y589 and Y591 in FLT3-ITD is important for STAT5 and MAPK phosphorylation, whereas residues Y597 and Y599 might be important for the structural maintenance of FLT3-ITD for STAT5 recruitment. Furthermore, these results suggest that multiple signaling pathways play a role for transformation of cells by FLT3 mutants. To analzse the mechanism of STAT5 phosphorylation, we performed coimmunoprecipitation assays and GST pull-down of FLT3 with SH2 domain of STAT5. We could not detect any direct interaction of STAT5 with FLT3 or FLT3 mutants (data not shown). In line with the proliferation data, FLT3-ITD-W51 showed an interaction with the SH2 domain of Src, whereas the FLT3-W51-YY589/591FF showed no interaction (Supplementary Fig. S3). In GST pull-down experiments, we confirmed direct interaction of STAT5 with Src-SH2 domain (Supplementary Fig. S3). Recent studies have shown the interaction of Src kinases with Y589 of FLT3-WT (30). Src kinases direct the phosphorylation of STAT5 (48). These data show that Y589 binds Src kinases, which can interact with an activated STAT5.

Y→F substitution of both 589 and 591 in the FLT3-WT background (FLT3-WT-YY589/591FF) totally abrogated the proliferation of Ba/F3 cells upon stimulation with FL (Fig. 6B) indicating that Y589 and Y591 in FLT3-WT are essential for FL-mediated proliferation.

In conclusion, we have identified Y589 and Y591 as the critical tyrosine residues required for STAT5 signaling and for the transforming phenotype of active FLT3-receptor mutants and FL-mediated proliferation of FLT3-WT.

Acknowledgments

References

1. Rosnet O, Birnbaum D. Hematopoietic receptors of class III receptor-type tyrosine kinases. Crit Rev Oncog 1993;4:595–613.
2. Lyman SD, Jacobsen SE. c-kit ligand and Flt3 ligand: stem/progenitor cell factors with overlapping yet distinct activities. Blood 1998;91:1101–34.
3. Srinivasa SP, Doshi PD. Extracellular signal-regulated kinase and p38 mitogen-activated protein kinase pathways cooperate in mediating cytokine-induced proliferation of a leukemic cell line. Leukemia 2002;16:244–53.
4. Zhang S, Broxmeyer HE. Flt3 ligand induces tyrosine phosphorylation of gab1 and gab2 and their association with shp-2, grb2, and PI3 kinase. Biochem Biophys Res Commun 2000;277:195–9.
5. Zhang S, Fukuda S, Lee Y, et al. Essential role of signal

transducer and activator of transcription (Stat)5a but not Stat5b for Flt3-dependent signaling. J Exp Med 2000;192:719–28.
6. Zhang S, Mantel C, Broxmeyer HE. Flt3 signaling involves tyrosyl-phosphorylation of SHP-2 and SHIP and their association with Grb2 and Shc in Baf3/Flt3 cells. J Leukoc Biol 1999;65:372–80.
7. Adolfsson J, Mansson R, Buza-Vidas N, et al. Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. Cell 2005;121:295–306.
8. Christensen JL, Weissman IL. Flk-2 is a marker in hematopoietic stem cell differentiation: a simple method to isolate long-term stem cells. Proc Natl Acad Sci U S A 2001;98:14541–6.
9. Sitnicka E, Buza-Vidas N, Larsson S, Nygren JM,

Liuba K, Jacobsen SE. Human CD34+ hematopoietic stem cells capable of multilineage engrafting NOD/SCID mice express flt3: distinct flt3 and c-kit expression and response patterns on mouse and candidate human hematopoietic stem cells. Blood 2003;102:881–6.
10. Schnittger S, Schoch C, Dugas M, et al. Analysis of FLT3 length mutations in 1003 patients with acute myeloid leukemia: correlation to cytogenetics, FAB subtype, and prognosis in the AMLCG study and usefulness as a marker for the detection of minimal residual disease. Blood 2002;100:59–66.
11. Thiede C, Steudel C, Mohr B, et al. Analysis of FLT3-activating mutations in 979 patients with acute myelogenous leukemia: association with FAB subtypes and identification of subgroups with poor prognosis. Blood 2002;99:4326–35.

12. Pollard JA, Alonzo TA, Gerbing RB, et al. FLT3 internal tandem duplication in CD34$^+$/CD33$^-$ precursors predicts poor outcome in acute myeloid leukemia. Blood 2006;108:2764–9.

13. Kottaridis PD, Gale RE, Frew ME, et al. The presence of a FLT3 internal tandem duplication in patients with acute myeloid leukemia (AML) adds important prognostic information to cytogenetic risk group and response to the first cycle of chemotherapy: analysis of 854 patients from the United Kingdom Medical Research Council AML 10 and 12 trials. Blood 2001;98:1752–9.

14. Abu-Duhier FM, Goodeve AC, Wilson GA, Care RS, Peake IR, Reilly JT. Identification of novel FLT-3 Asp^{835} mutations in adult acute myeloid leukaemia. Br J Haematol 2001;113:983–8.

15. Yamamoto Y, Kiyoi H, Nakano Y, et al. Activating mutation of D835 within the activation loop of FLT3 in human hematologic malignancies. Blood 2001;97:2434–9.

16. Schessl C, Rawat VP, Cusan M, et al. The AML1-ETO fusion gene and the FLT3 length mutation collaborate in inducing acute leukemia in mice. J Clin Invest 2005;115:2159–68.

17. Kelly LM, Kutok JL, Williams IR, et al. PML/RARα and FLT3-ITD induce an APL-like disease in a mouse model. Proc Natl Acad Sci U S A 2002;99:8283–8.

18. Stone RM, DeAngelo DJ, Klimek V, et al. Patients with acute myeloid leukemia and an activating mutation in FLT3 respond to a small-molecule FLT3 tyrosine kinase inhibitor, PKC412. Blood 2005;105:54–60.

19. Knapper S, Burnett AK, Littlewood T, et al. A phase 2 trial of the FLT3 inhibitor lestaurtinib (CEP701) as first line treatment for older patients with acute myeloid leukemia not considered fit for intensive chemotherapy. Blood 2006;108:3262–70.

20. Bagrintseva K, Schwab R, Kohl TM, et al. Mutations in the tyrosine kinase domain of FLT3 define a new molecular mechanism of acquired drug resistance to PTK inhibitors in FLT3-ITD-transformed hematopoietic cells. Blood 2004;103:2266–75.

21. Kohl TM, Schnittger S, Ellwart JW, Hiddemann W, Spiekermann K. KIT exon 8 mutations associated with core-binding factor (CBF)-acute myeloid leukemia (AML) cause hyperactivation of the receptor in response to stem cell factor. Blood 2005;105:3319–21.

22. Spiekermann K, Dirschinger RJ, Schwab R, et al. The protein tyrosine kinase inhibitor SU5614 inhibits FLT3 and induces growth arrest and apoptosis in AML-derived cell lines expressing a constitutively activated FLT3. Blood 2003;101:1494–504.

23. Greif PA, Tizazu B, Krause A, Kremmer E, Bohlander SK. The leukemogenic CALM/AF10 fusion protein alters the subcellular localization of the lymphoid regulator Ikaros. Oncogene 2007;27:2886–96.

24. Roskoski R, Jr. Structure and regulation of Kit protein-tyrosine kinase-the stem cell factor receptor. Biochem Biophys Res Commun 2005;338:1307–15.

25. Spiekermann K, Bagrintseva K, Schwab R, Schmieja K, Hiddemann W. Overexpression and constitutive activation of FLT3 induces STAT5 activation in primary acute myeloid leukemia blast cells. Clin Cancer Res 2003;9:2140–50.

26. Hayakawa F, Towatari M, Kiyoi H, et al. Tandem-duplicated Flt3 constitutively activates STAT5 and MAP kinase and introduces autonomous cell growth in IL-3-dependent cell lines. Oncogene 2000;19:624–31.

27. Mizuki M, Fenski R, Halfter H, et al. Flt3 mutations from patients with acute myeloid leukemia induce transformation of 32D cells mediated by the Ras and STAT5 pathways. Blood 2000;96:3907–14.

28. Kyoi H, Ohno R, Vedla R, Saito H, Naoe T. Mechanism of constitutive action of FLT3 with internal tandem duplication in the juxtamembrane domain. Oncogene 2002;21:2555–63.

29. Rocnik JL, Okabe R, Yu JC, et al. Roles of tyrosine 589 and 591 in STAT5 activation and transformation mediated by FLT3-ITD. Blood 2006;108:1339–45.

30. Heiss E, Masson K, Sundberg C, et al. Identification of Y589 and Y599 in the juxtamembrane domain of Flt3 as ligand-induced autophosphorylation sites involved in binding of Src family kinases and the protein tyrosine phosphatase SHP2. Blood 2006;108:1542–50.

31. Griffith J, Black J, Faerman C, et al. The structural basis for autoinhibition of FLT3 by the juxtamembrane domain. Mol Cell 2004;13:169–78.

32. Ma Y, Cunningham ME, Wang X, Ghosh I, Regan L, Longley BJ. Inhibition of spontaneous receptor phosphorylation by residues in a putative α-helix in the KIT intracellular juxtamembrane region. J Biol Chem 1999;274:13399–402.

33. Price DJ, Rivnay B, Avraham H. CHK down-regulates SCF/KL-activated Lyn kinase activity in Mo7e megakaryocytic cells. Biochem Biophys Res Commun 1999;259:611–6.

34. Gelderloos JA, Rosenkranz S, Bazenet C, Kazlauskas A. A role for Src in signal relay by the platelet-derived growth factor α receptor. J Biol Chem 1998;273:5908–15.

35. Hooshmand-Rad R, Yokote K, Heldin CH, Claesson-Welsh L. PDGFα-receptor mediated cellular responses are not dependent on Src family kinases in endothelial cells. J Cell Sci 1998;111:607–14.

36. Valgeirsdóttir S, Paukku K, Silvenoinnon O, Heldin CH, Claesson-Welch L. Activation of STAT5 by platelet-derived growth factor (PDGF) is dependent on phosphorylation cells in PDGF beta receptor juxtamembrane and kinase insert domains Oncogene 1998;16:605–15.

37. Baxter RM, Secrist JP, Vaillancourt RR, Kazlauskas A. Full activation of the platelet-derived growth factor β-receptor kinase involves multiple events. J Biol Chem 1998;273:17050–5.

38. Kimura Y, Jones N, Kluppel M, et al. Targeted mutations of the juxtamembrane tyrosines in the Kit receptor tyrosine kinase selectively affect multiple cell lineages. Proc Natl Acad Sci U S A 2004;101:6015–20.

39. Mori S, Ronnstrand L, Yokote K, et al. Identification of two juxtamembrane autophosphorylation sites in the PDGFβ-receptor; involvement in the interaction with Src family tyrosine kinases. EMBO J 1993;12:2257–64.

40. Rohde CM, Schrum J, Lee AW. A juxtamembrane tyrosine in the colony stimulating factor-1 receptor regulates ligand-induced Src association, receptor kinase function, and down-regulation. J Biol Chem 2004;279:43448–61.

41. Vaillancourt RR, Heasley LE, Zamarripa J, et al. Mitogen-activated protein kinase activation is insufficient for growth factor receptor-mediated PC12 cell differentiation. Mol Cell Biol 1995;15:3644–53.

42. Benekli M, Baer MR, Baumann H, Wetzler M. Signal transducer and activator of transcription proteins in leukemias. Blood 2003;101:2940–54.

43. Gouilleux-Gruart V, Gouilleux F, Desaint C, et al. STAT-related transcription factors are constitutively activated in peripheral blood cells from acute leukemia patients. Blood 1996;87:1692–7.

44. Hayakawa F, Towatari M, Iida H, et al. Differential constitutive activation between STAT-related proteins and MAP kinase in primary acute myelogenous leukemia. Br J Haematol 1998;101:521–8.

45. Xia Z, Baer MR, Block AW, Baumann H, Wetzler M. Expression of signal transducers and activators of transcription proteins in acute myeloid leukemia blasts. Cancer Res 1998;58:3173–80.

46. Calo V, Migliavacca M, Bazan V, et al. STAT proteins: from normal control of cellular events to tumorigenesis. J Cell Physiol 2003;197:157–68.

47. Vander Heiden MG, Chandel NS, Williamson EK, Schumacker PT, Thompson CB. Bcl-xL regulates the membrane potential and volume homeostasis of mitochondria. Cell 1997;91:627–37.

48. Okutani Y, Kitanaka A, Tanaka T, et al. Src directly tyrosine-phosphorylates STAT5 on its activation site and is involved in erythropoietin-induced signaling pathway. Oncogene 2001;20:6643–50.

Global Patterns in Human Mitochondrial DNA and Y-Chromosome Variation Caused by Spatial Instability of the Local Cultural Processes

Vikrant Kumar[1], Banrida T. Langstieh[1,2], Komal V. Madhavi[1], Vegi M. Naidu[1,3], Hardeep Pal Singh[1,4], Silpak Biswas[1], Kumarasamy Thangaraj[5], Lalji Singh[5], B. Mohan Reddy[1*]

1 Biological Anthropology Unit, Indian Statistical Institute, Hubsiguda, Hyderabad, India, 2 Department of Anthropology, North Eastern Hill University, Shillong, India, 3 GSF Hematologikum, Medizinische Klink III, Klinikum Grosshadern, München, Germany, 4 Kallam Anji Reddy Molecular Genetics Laboratory, L.V. Prasad Eye Institute, L.V. Prasad Marg, Banjara Hills, India, 5 Centre for Cellular and Molecular Biology, Hyderabad, India

Because of the widespread phenomenon of patrilocality, it is hypothesized that Y-chromosome variants tend to be more localized geographically than those of mitochondrial DNA (mtDNA). Empirical evidence confirmatory to this hypothesis was subsequently provided among certain patrilocal and matrilocal groups of Thailand, which conforms to the isolation by distance mode of gene diffusion. However, we expect intuitively that the patterns of genetic variability may not be consistent with the above hypothesis among populations with different social norms governing the institution of marriage, particularly among those that adhere to strict endogamy rules. We test the universality of this hypothesis by analyzing Y-chromosome and mtDNA data in three different sets of Indian populations that follow endogamy rules to varying degrees. Our analysis of the Indian patrilocal and the matrilocal groups is not confirmatory to the sex-specific variation observed among the tribes of Thailand. Our results indicate spatial instability of the impact of different cultural processes on the genetic variability, resulting in the lack of universality of the hypothesized pattern of greater Y-chromosome variation when compared to that of mtDNA among the patrilocal populations.

Introduction

The genetic patterns in human societies are often fashioned by their cultural practices. For example, it has been hypothesized that due to widespread phenomenon of patri-locality (a pattern of residence where the female spouse after marriage resides in the in-law's house) Y-chromosome variants tend to be more localized geographically than those of mitochondrial DNA (mtDNA) and the autosomes, and therefore high degree of inter-population genetic differences have been observed for the Y chromosome compared to the mtDNA [1–4]. Due to movement of females in patrilocal groups, the mtDNA diversity is assumed to be high within the populations and low between the populations, whereas the Y-chromosome diversity will be relatively low within the groups and high between the groups. This pattern is expected to be reversed in case of the matrilocal groups (a pattern of residence where the males after marriage reside in the in-law's house). Empirical evidence confirmatory to this hypoth-esis was subsequently provided by Oota et al. [5] among the three patrilocal and three matrilocal groups of Thailand. They found genetic diversity to be strikingly correlated with residence patterns suggesting the role of sex-specific patterns of migration in influencing the genetic patterns. In contrast, few other studies at the regional scale [6–8] show similar levels of differentiation for maternal and paternal lineages. There-fore, the patterns of genetic diversity at the local level may not reflect at the global scale, which is essentially an artifact of the sum total of differing local patterns. Concurrently, in a global survey, Wilder et al. [9] could not detect the signature of a

higher inter-population migration rate for females than for males. This is interpreted as due to lack of geographic stability of the behavioral customs of individual populations necessary to influence global genetic patterning. The norms governing the institution of marriage vary enormously among human populations of different regions or cultures [10–12], and different forms of social organization can impact patterns and levels of genetic diversity [13,14]. Therefore, the universality of the above hypothesis, i.e., the pattern of genetic variation vis-à-vis the residence pattern of spouses, is in question.

Implicit in the above hypothesis is the assumption that the population boundaries are permeable, permitting male/female spouses to move across their respective populations and become part of the gene pool of the new population to which the other spouse belongs. Only in such a scenario can the expectations of the above hypothesis hold, either in patrilocal or matrilocal societies. This situation, broadly

Synopsis

In most human societies, women traditionally move to their husband's home after marriage, and these societies are thus "patrilocal," but in a few "matrilocal" societies, men move to their wife's home. These social customs are expected to influence the patterns of genetic variation. They should lead to a localization of male-specific Y-chromosomal variants and wide dispersal of female-specific mitochondrial DNA variants in patrilocal societies and vice versa in matrilocal societies. These predicted patterns have indeed been observed in previous studies of populations from Thailand. Indian societies, however, are endogamous, so marriage should always take place within a population, and these different patterns of genetic variation should not build up. The authors have now analyzed ten patrilocal and five matrilocal Indian populations, and find that there is indeed little difference between the patrilocal and matrilocal societies. The authors therefore conclude that patterns of genetic variation in humans are not universal, but depend on local cultural practices.

speaking, approximates to isolation by distance mode of gene diffusion. On the other hand, for populations bound by rigid endogamy rules with their boundaries absolutely impermeable, neither patrilocality nor matrilocality can make any difference to their genetic variability, be it Y-chromosome or mtDNA, since the movement is restricted to within a population. The Indian subcontinent with its unique population structure and strictly defined endogamous castes, tribes, and religious groups is a case in point (Figure 1). The marriage interactions are restricted within an endogamous population consisting of the number of exogamous units/ clans between which marriages take place. We directly test the universality of the hypothesis delineated above and attempt to assess the spatial stability of the local cultural processes necessary to influence global patterning in two stages. In the first stage, we analyzed Y-chromosome short tandem repeat (Y-STR) and mtDNA hyper variable segment 1 (HVS1) sequence data from two groups of Indian tribes, comprised of five populations each, belonging to a broad linguistic family and with similar socio-economic status. The genetic data were obtained from the same set of populations and individuals making it appropriate for comparison. The populations included in this study are Maram, Khynriam, Pnar, Bhoi, and WarKhasi, the five matrilocal Khasi tribes of Meghalaya in the Northeastern part of India; and Asur, Bhumij, Kharia, Munda, and Santhal, the five patrilocal Mundari tribes of Eastern India, who along with the matrilocal Khasis, belong to the broad Austro-Asiatic linguistic family. At the second stage, to gauge the consistency in the genetic patterns within broad regional or cultural context, the same set of genetic data were generated on the five Dravidian language-speaking patrilocal caste populations from Andhra Pradesh (Akutota, Kapu, Panta, Pokanati, and Vanne) of Southern India and compared with the Austro-Asiatic matrilocal tribes.

The structure of populations considered in this study is characterized by numerous endogamous groups cohabiting as islands with no or negligible gene flow between them. Therefore, as the marital boundary of each population is impermeable, we intuitively expect that the pattern of genetic variability may not strictly follow the expectations of the aforesaid hypothesis, either in patrilocal or matrilocal

INDIAN POPULATION STRUCTURE

CASTES/TRIBES/RELIGIOUS GROUPS

SUBCASTES/SUBTRIBES/SUBGROUPS

BREEDING ISOLATES

EXOGAMOUS UNITS OF A POPULATION
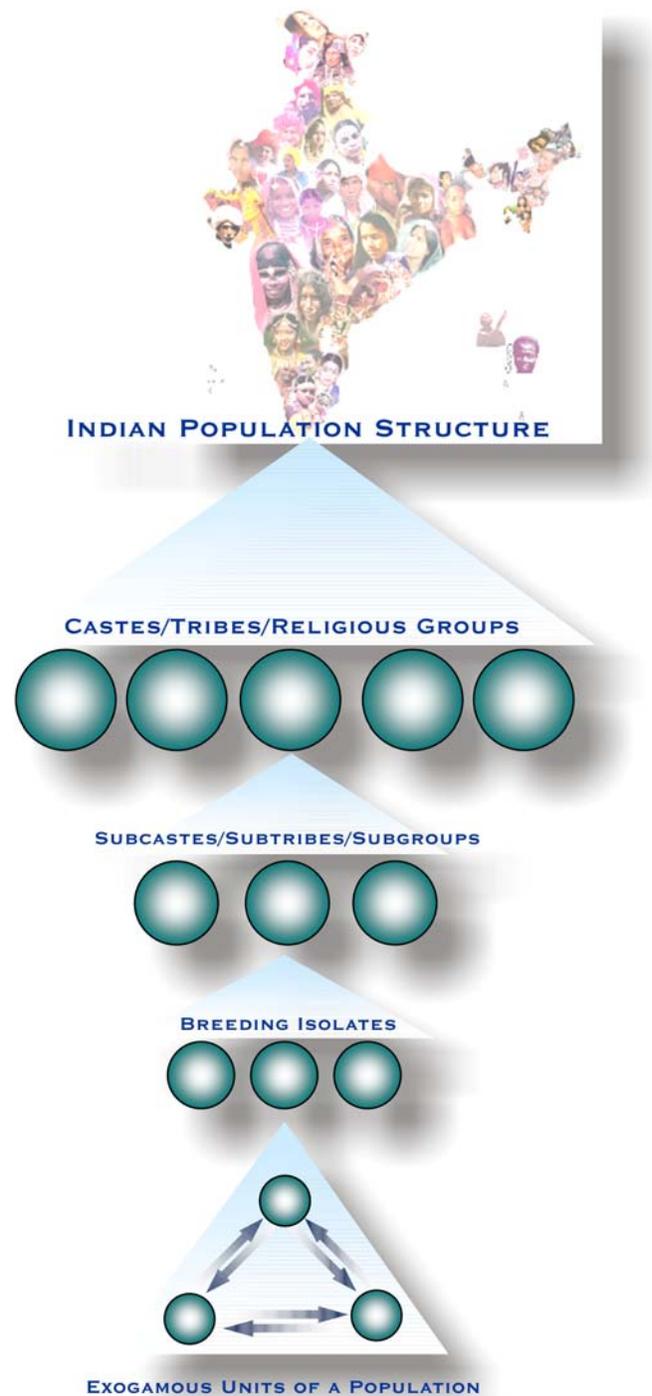
Figure 1. Schematic Representation of Indian Population Structure Characterized by Movement of Spouses Only within but Not among the Endogamous Groups

Each circle represents a population and its size represents the hierarchy. While the populations until the breeding isolates are all endogamous, the exogamous units refer to clans/lineages within a breeding isolate/ population.

DOI: 10.1371/journal.pgen.0020053.g001

groups. All three groups of populations have contiguous geographic distribution in their respective areas, which provide opportunity for exchange of mates, if the social norms permit, thus providing ideal study frame.
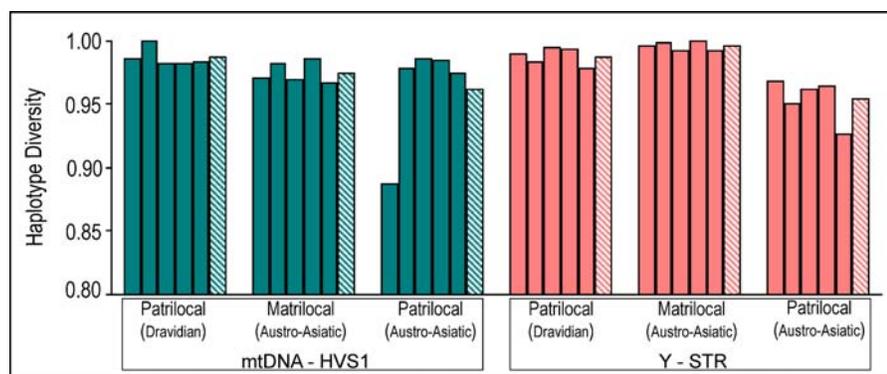
Figure 2. Haplotype Diversity in mtDNA (Green) and Y-STR (Pink) and Their Mean (Shaded Bar) in Five Dravidian and Five Austro-Asiatic Patrilocal and Five Austro-Asiatic Matrilocal Populations

From left to right, the Dravidian patrilocal groups (mtDNA sample size and Y-STR sample size) are Akhutota (32, 21), Kapu (22,16), Panta (37, 21), Pokanati (59, 25), and Vanne (32, 23); the Austro-Asiatic matrilocal groups are Maram (72, 58), Khynriam (95, 82), Pnar (69, 40), Bhoi (34, 30), and WarKhasi (31, 23); the Austro-Asiatic patrilocal groups are Asur (30, 28), Bhumij (40, 39), Kharia (21, 13), Munda (23, 23), and Santhal (39, 38).

DOI: 10.1371/journal.pgen.0020053.g002

Results

Within-group mtDNA diversity (Figure 2) is similar (Mann-Whitney U test, $p = 0.690$) for matrilocal Khasi tribes (0.975) versus patrilocal Mundari tribes (0.962), although the mean within-group Y-chromosome diversity of patrilocal Mundari tribal groups (0.954) is significantly lower (Mann-Whitney U test, $p = 0.008$) when compared with matrilocal Khasi (0.995). However, when we compare the patrilocal Dravidian caste groups with the matrilocal Khasi tribes we found similar and non-significant difference in the level of within-group diversity for both mtDNA ($p = 0.056$) and Y-chromosome ($p = .095$). The average values of genetic distance (Table 1) reflecting inter-group diversity (although smaller for mtDNA and larger for Y-chromosome among patrilocal Mundari groups than for matrilocal Khasi groups) are not statistically significantly different. Likewise, the average genetic distances in the Dravidian patrilocal groups are smaller for mtDNA and larger for Y-chromosome but not significantly so when compared with the matrilocal Khasi groups.

The index of probability of identity, which gives a quantitative measure of haplotype sharing between a pair of populations, further suggests, as against the hypothesis, that

the degree of Y-chromosome haplotype sharing (Table 2), although not significant, is substantially higher among the patrilocal Mundari groups when compared with the Matrilocal Khasi tribes, whereas the degree of mtDNA haplotype sharing is almost identical for both groups. On the other hand, we observe a very low level of mtDNA haplotype sharing among the patrilocal Dravidian groups compared with the matrilocal Khasi groups, while the level of Y-chromosome haplotype sharing is similar for both the groups. As per the hypothesis, a relatively lower degree of mtDNA haplotype sharing and greater degree of Y-chromosome haplotype sharing is expected among the matrilocal groups compared with the patrilocal groups. Overall, the results are not consistent with the universality of the hypothesis in question.

Discussion

The foregoing analysis of the results does not reflect higher migration rate of females and males, respectively, in the patrilocal and matrilocal populations, suggesting that the pattern of residence of the spouses has no bearing on the mtDNA and Y-chromosome variability in the populations, in which sex-specific migrations implicit in the hypothesis are

Table 1. Average Genetic Distance and Their Standard Error Based on mtDNA HVS1 and Y-STR among the Matrilocal and Patrilocal Groups

Genetic Distances	Patrilocal (Austro-Asiatic; Mundari) Average ± SE	Matrilocal (Austro-Asiatic; Khasi) Average ± SE	Patrilocal (Dravidian) Average ± SE	Mann-Whitney U Test (p: Two-Tailed[a])	
				Mundari versus Khasi	Dravidian versus Khasi
R_{st} (Y-STR)	0.100 ± 0.002	0.055 ± 0.001	0.114 ± 0.003	0.123	0.474
D_A (mtDNA)	0.128 ± 0.103	0.200 ± 0.095	0.142 ± 0.087	0.165	0.971

Genetic distances (d_A and R_{st}) and SE, based on 1,000 bootstrap replicates, were calculated using MEGA (http://www.megasoftware.net/mega3/mega.html) and RSTCALC (http://helios.bto.ed.ac.uk/evolgen/rst/rst.html).
[a]Calculated on the basis of genetic distance matrices.
SE, standard error.
DOI: 10.1371/journal.pgen.0020053.t001

Table 2. Index of Probability of Identity Based on mtDNA HVS1 and Y-STR among the Patrilocal and Matrilocal Groups

Genetic Markers	Patrilocal (Austro-Asiatic; Mundari) Average	Matrilocal (Austro-Asiatic; Khasi) Average	Patrilocal (Dravidian) Average	Mann-Whitney U Test (p: Two-Tailed)	
				Mundari versus Khasi	Dravidian versus Khasi
Y-STR	0.0116	0.0023	0.0011	0.260	0.029
mtDNA	0.0132	0.0117	0.0031	0.626	0.045

confined within the endogamous groups and do not usually transect the caste/tribal boundaries. However, a weak and non-significant trend of greater inter-group variation in Y-chromosome and lower variation in mtDNA in case of patrilocal groups, and greater mtDNA and lower Y-chromosome inter-group variation in matrilocal groups, which is consistent with the hypothesis, is observed. Nevertheless, the magnitude of differences, either intra- or inter-population observed in our study, are substantially smaller than what has been observed by Oota et al. [5] in Thailand, despite a relatively small number of samples and populations. The non-significant differences in the mean values of the genetic distances could have been due to two reasons: (1) either to small sample size; hence lacking sufficient power to correctly reject the null hypothesis, or (2) to small number of Y-STRs, which may not have adequate resolution. Therefore, we calculated power of the Mann-Whitney U test for the given sample sizes in the study and the results suggest that the test has > 99% power, even at alpha = 0.001, both for mtDNA and Y-chromosome. Additional analysis based on 15 Y-STRs suggests, contrary to the hypothesis, that the average genetic distance among the patrilocal groups was quite low (0.0469 ± 0.0009), albeit non-significantly ($p = 0.1$), as compared with the matrilocal groups (0.1024 ± 0.0024). Therefore, the hypothesized correlation of genetic diversity with the sex-specific migration patterns may not be applicable to the Indian situation, although it is observed elsewhere in certain populations whose marital boundaries are probably permeable.

One of the questions raised by Wilder et al. [9] is the extent to which local cultural practices influence genetic patterns at the regional and global scale. The groups we have considered in the present study have different cultural norms governing the rules of marriages compared to those studied by Oota et al. [5]; hence we find variation in the genetic patterns. Even within India, we find variation in the pattern depending on whether we compare the matrilocal Khasi tribes with the patrilocal Mundari tribal groups or with the patrilocal Dravidian caste groups. For example, the index of probability of identity shows very low values for both mtDNA and Y-chromosome haplotype sharing among the Dravidian castes when compared with the Austro-Asiatic tribes, either Mundari or Khasi (Table 2). This pattern is observed because the caste populations of India are considered to follow endogamy very strictly; hence their marital boundaries are highly rigid compared with the marital boundaries of the Indian tribes, particularly from Northeast India, suggesting the impact of varying cultural practices pertaining partic-

ularly to marriage, resulting in variable genetic patterns. Results of our study taken together with the previous studies, that have [6–8] or have not [1,3–5] detected sex-specific migration, suggest that the local cultural processes do not have spatial stability required to influence global patterning. Perhaps due to this, Wilder et al. [9] did not observe higher migration rates of females vis-à-vis males at the continental level, although most of the populations of the world follow patrilocality [15]. Therefore, the hypothesis of greater Y-chromosome vis-à-vis mtDNA variability due to patrilocality is not universal, as it can only be selectively applicable to populations with cultural norms that permit inter-group marriages; not to, for example, highly endogamous Indian populations. Pertinent to this are the two recent large-scale Indian studies [16,17] wherein the lack of spatial structure in the quantitative biological variables—anthropometry and dermatoglyphics— and traditional genetic markers was inferred to be consistent with population structure characterized by numerous endogamous groups cohabiting as islands with no or negligible gene flow between them; the monotonic decline in the spatial autocorrelation expected under the model of contiguous diffusion of genes is not evident in those data.

Materials and Methods

Blood samples from 636 individuals belonging to 15 populations were obtained for the above populations during 2000–2003 with informed written consent; DNA was extracted. The names of the populations along with their sample size are given in Figure 2. We analyzed 350 base pairs of the HVS1 of the mtDNA control region corresponding to positions 16050–16400 and six Y-STR loci (DYS19, DYS389I, DYS389b, DYS390, DYS391, and DYS393). Allele length for DYS389b was obtained by subtracting the allele length of DYS389I from DYS389II. The HVS1 sequences have been submitted to GenBank and are also available from the authors, as are the Y-STR data. To measure within-group variability we estimated haplotype diversity [18] for the HVS1 sequences and Y-STR haplotypes (Table S1), and calculated d_A distances [19] for the HVS1 sequences using the number of different sites model, and R_{ST} for the Y-STR haplotypes [20] as measures of between-group diversity. Further, we computed an index of probability of identity [21], which gives a quantitative measure of haplotype sharing between a pair of populations. To ascertain, for the given sample sizes, that the test has enough power at alpha = 0.05– 0.001, we computed power required for the Mann-Whitney U test. For this purpose, we decreased the sample sizes by 15% and used this sample size to compute power required for a t-test. This rule is based on the lower bound for the asymptotic relative efficiency (ARE) of the Mann-Whitney U test versus the t-distribution, which is 0.864. This says that no matter what the distribution is, the ARE of the Mann-Whitney U test can never be worse than 0.864 for a reasonable broad class of probability distributions. Inverting that gives an increase in the sample size by a factor of 1.157, and therefore the sample sizes were reduced by 15% [22]. To increase the

resolution, in addition to the six Y-STRs, we typed nine more Y-STRs (DYS388, DYS426, DYS437, DYS438, DYS439, DYS447, DYS448, DYS460, and H4; Table S2) in three populations each of Mundari patrilocal groups (Bhumij, Munda, and Santhal) and Khasi matrilocal groups (Khynriam, Maram, and Pnar) and recomputed genetic distances based on 15 Y-STR loci.

Supporting Information

Table S1. Y-Chromosome Haplotypes Based on Six Y-STRs for 15 Populations

Found at DOI: 10.1371/journal.pgen.0020053.st001 (946 KB DOC).

Table S2. Y-Chromosome Haplotypes Based on Nine Y-STRs for Six Populations

Found at DOI: 10.1371/journal.pgen.0020053.st002 (387 KB DOC).

Accession Numbers

The GenBank (http://www.ncbi.nlm.nih.gov) accession numbers for the sequence discussed in this paper are HVS1 (AY72095–AY721592).

References

1. Salem AH, Badr FM, Gaballah MF, Paabo S (1996) The genetics of traditional living. Y-chromosomal and mitochondrial lineages in the Sinai Peninsula. Am J Hum Genet 59: 741–743.
2. Seielstad MT, Minch E, Cavalli-Sforza LL (1998) Genetic evidence for a higher female migration rate in humans. Nat Genet 20: 278–280.
3. Perez-Lezaun A, Calafell F, Comas D, Mateu E, Bosch E, et al. (1999) Sex-specific migration pattern in Central Asian populations, revealed by analysis of Y-chromosome short tandem repeats and mtDNA. Am J Hum Genet 65: 208–219.
4. Oota H, Kitano T, Jin F, Yuasa I, Wang L, et al. (2002) Extreme mtDNA homogeneity in continental Asian populations. Am J Phys Anthrop 118: 146–153.
5. Oota H, Settheetham-Ishida W, Tiwawech D, Ishida T, Stoneking M (2001) Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. Nat Genet 29: 20–21.
6. Mesa NR, Mondragon MC, Soto ID, Parra MV, Duque C, et al. (2000) Autosomal, mtDNA, and Y-chromosome diversity in Amerinds: Pre- and post-Columbian patterns of gene flow in South America. Am J Hum Genet 67: 1277–1286.
7. Al-Zahery N, Semino O, Benuzzi G, Magri C, Passarino G, et al. (2003) Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations. Mol Phylogenet Evol 28: 458–472.
8. Fuselli S, Tarazona-Santos E, Dupanloup I, Soto A, Luiselli D, et al. (2003) Mitochondrial DNA diversity in South America and the genetic history of Andean highlands. Mol Biol Evol 20: 1682–1691.
9. Wilder JA, Kingan SB, Mobasher Z, Pilkington MM, Hammer MF (2004) Global patterns of human mitochondrial DNA and Y-chromosome structure are not influenced by higher migration rates of females versus males. Nat Genet 36: 1122–1125.
10. Levinson D (1996) Encyclopedia of world cultures. New York: MacMillan. 3855 p.
11. Korotayev AV (2003) Form of marriage, sexual division of labor, and postmarital residence in cross-cultural perspective: A reconsideration. J Anthropol Res 59: 69–89.
12. Marlowe FW (2004) Marital residence among foragers. Curr Anthropol 45: 277–284.
13. Austerlitz F, Heyer E (1998) Social transmission of reproductive behavior increases frequency of inherited disorders in a young-expanding population. Proc Natl Acad Sci U S A 95: 15140–15144.
14. Excoffier L, Schneider S (1999) Why hunter-gatherer populations do not show signs of Pleistocene demographic expansions. Proc Natl Acad Sci U S A 96: 10597–10602.
15. Burton ML, Moore CC, Whiting JWM, Romney AK (1996) Regions based on social structure. Curr Anthropol 37: 87–123.
16. Reddy BM, Demarchi DA, Malhotra KC (2001) Patterns of variation in a caste-cluster of Dhangars of Maharashtra, India. Coll Antropol 25: 425–442.
17. Reddy BM, Demarchi DA, Bharati S, Kumar V, Crawford MH (2004) Patterns of ethnic, linguistic, and geographic heterogeneity of palmar interdigital ridge counts in the Indian subcontinent. Hum Biol 76: 211–228.
18. Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. J Mol Evol 19: 153–170.
19. Nei M (1987) Molecular evolutionary genetics. New York: Columbia University Press. 333 p.
20. Slatkin M (1995) A measure of population subdivision based on micro-satellite allele frequencies. Genetics 139: 457–462.
21. Melton T, Peterson R, Redd AJ, Saha N, Sofro AS, et al. (1995) Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. Am J Hum Genet 57: 403–414.
22. Lehmann EL (1975) Nonparametrics: Statistical methods based on ranks. San Francisco: Holden-Day. 457 p.

Microsatellite Diversity in Andhra Pradesh, India: Genetic Stratification Versus Social Stratification

B. MOHAN REDDY,[1] V. M. NAIDU,[1] V. KOMAL MADHAVI,[1] K. THANGARAJ,[2] VIKRANT KUMAR,[1] B. T. LANGSTIEH,[3] P. VENKATRAMANA,[4] A. G. REDDY,[2] AND LALJI SINGH[2]

Abstract   DNA samples of 948 individuals belonging to 27 populations from southern Andhra Pradesh were analyzed for nine AmpF*l*STR Profiler Plus loci. The nature and extent of genomic diversity within and between these populations have been examined with reference to socioeconomic and geographic affiliations. The results suggest that the average heterozygosity is uniformly high in these populations (>0.80) and that the patterns of allele distributions are similar across the populations. The value of the coefficient of gene differentiation and the AMOVA and structure analysis results suggest that these populations are highly homogeneous. The neighbor-joining tree constructed using either D_A or F_{ST} distances suggests no intelligible pattern of population clusters based on ethnohistoric or geographic affiliations. All these observations suggest either a common recent origin of these populations or extensive gene flow across the populations that erased the original genetic differences. Given strict endogamy, the latter explanation can hold only if there has been unauthorized or unrecognized gene flow transecting the social boundaries. Nevertheless, the regression plot of average heterozygosity versus distance from the centroid (R_{ii}), based on Harpending and Ward's (1982) model, and the genetic distances computed between different hierarchical groups within Andhra Pradesh tend to support this conjecture. Overall, the results suggest lack of a significant degree of genetic stratification that is consistent with social stratification in Andhra Pradesh. Furthermore, the neighbor-joining tree based on comparative data from other Indian and continental populations brings out a single and compact cluster of all the Andhra populations that is clearly separated from the rest.

For historical reasons India presents enormous cultural, linguistic, and biological diversity (Karve 1961; Thapar 1966, 1995, 2003; Kosambi 1991; Gadgil et al.

1998; Krishnan and Reddy 1994; Majumder 1998; Reddy et al. 2004). Indian population structure is unique and characterized by the division of the population into strictly defined hierarchical castes, tribes, and religious groups, each of which are endogamous. By and large, these features of population structure apply to each linguistic region of the country, and languages form strong barriers to gene flow between such regions, even among members of apparently the same caste. It is therefore likely that populations that have the same caste identity but reside in extreme regions have had virtually no marital or cultural interaction between them and that populations from each such region might have evolved independently of populations from other regions.

The state of Andhra Pradesh, with its 70 million people organized into several endogamous castes, tribes, and religious groups, presents enormous variety in its populations, sociocultural patterns, and organization. Several anthropological investigations suggest that the populations of Andhra Pradesh that practice close consanguineous marriages prefer village endogamy and restrict marriage contacts to small distances and hence are highly inbred. This has probably led to a reduction in effective population size, creating breeding isolates within apparently single endogamous castes and subcastes (Reddy 2002).

These features of the Andhra Pradesh population structure are also unique and are restricted to each zone or region within the same linguistic state, such as Andhra Pradesh, which covers a large geographic area. The castes of different hierarchies from within each region have overlapping and/or contiguous distributions and interact closely in a symbiotic relationship with physical possibilities (historical or current) for exchange of genes; hence they provide an ideal study framework within which to test certain anthropological hypotheses, such as the alleged practice of hypergamy, and to study the patterns of gene flow among the hierarchical groups. These finer aspects of Indian population structure have not been hitherto adequately incorporated into the framework of earlier studies on the origin of Indian castes and the impact of Indian population structure on its genetic structure (Bamshad et al. 1996, 1998, 2001; Watkins et al. 1999; Clark et al. 2000; Kivisild et al. 1999a, 1999b, 2003; Cordaux et al. 2003, 2004; Basu et al. 2003).

We present here the results based on 9 autosomal short tandem repeat (STR) loci among the 27 populations, including 3 tribes inhabiting the southern region of Andhra Pradesh. In particular, we study the extent of genomic diversity and affinities between different hierarchical groups at the regional level in a linguistic area and examine whether the signatures of social stratification implicit among the populations are apparent in the observed genomic patterns. On the basis of published data, we also compare the affinities of the populations of Andhra Pradesh to those from other regions of India and elsewhere.

Materials and Methods

Intravenous blood samples (about 5 ml each) were collected, with the informed consent, from 948 individuals belonging to 27 endogamous groups distributed in the contiguous areas of the 6 southernmost districts of Andhra Pradesh
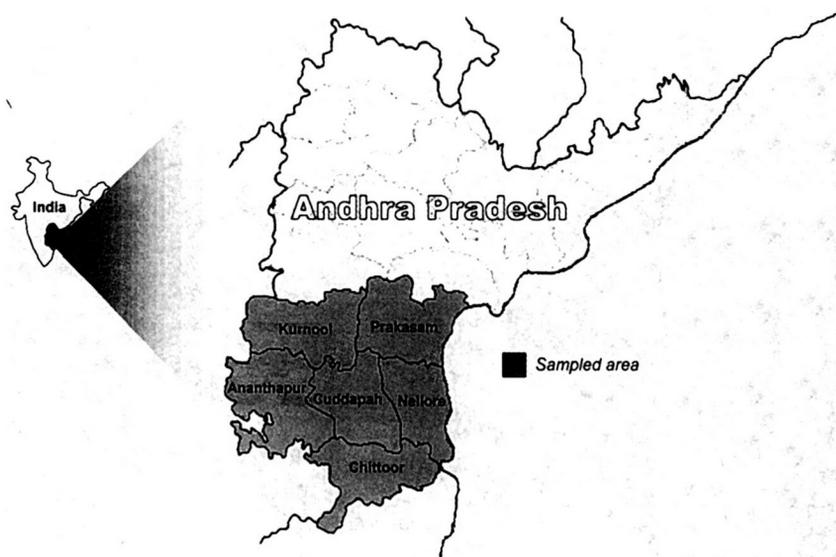
Figure 1. Map of Andhra Pradesh showing the study area and the six southern districts where the blood samples for the 27 populations were drawn.

(Chittoor, Cuddapah, Ananthapur, Kurnool, Prakasam, and Nellore) (Figure 1). The samples were drawn mostly from high school and college students, who represent a large number of surrounding villages and populations, and were supplemented with samples collected from the villages, particularly for tribes and lower castes. All the samples were anonymized and have no individual specific identity. The names of the populations studied, the number of samples drawn from each population, and their socioeconomic, geographic, and occupational backgrounds are furnished in Table 1. Although this study area can be considered culturally, linguistically, and geographically homogeneous, it is inhabited by a wide array of caste and tribal populations, representing almost the entire spectrum of socioeconomic variation in the state.

DNA was isolated from the samples following standard protocols (Sambrook et al. 1989). The AmpF/STR Profiler Plus kit (ABI, Foster City, California) was used to co-amplify nine STR loci (D3S1358, D8S1179, D5S818, VWA, D21S11, D13S317, FGA, D7S820, and D18S5) at a time, using an ABI Prism 377 automated DNA sequencer with the GeneScan and Genotyper software packages (Perkin Elmer, Wellesley, Massachusetts) to obtain the allele designations. All these loci are tetranucleotide repeats, and they are located on different chromosomes. To distinguish the alleles of different loci that have similar repeat lengths, we labeled each of the primers with one of the fluorescent dyes (5FAM, NED, and JOE), and the loci were amplified using a multiplex PCR.

Table 1. Studied Populations and Their Sample Size, Location, Occupation, and Socioeconomic Status

Population	Sample Size[a]	Sample Locations (Districts)	Traditional Occupation(s)	Socioeconomic Status (Caste or Tribe)[b]
Brahmin	56	Chittoor, Cuddapah, Kurnool	Service and religious cores	Upper
Kshatriya	84	Chittoor, Cuddapah,	Warriors	Upper
Vysya	40	Chittoor, Cuddapah, Nellore	Trade and business	Upper
Akuthota	58	Chittor (Chandragiri and Tirupati)	Land-owning agriculturists	Upper-middle
Kamma	98	Chittoor, Cuddapah	Land-owning agriculturists	Upper-middle
Kapu	40	Chittoor, Cuddapah, Kurnool	Land-owning agriculturists	Upper-middle
Pokanati	114	Chittoor, Cuddapah	Land-owning agriculturists	Upper-middle
Panta	82	Nellore	Land-owning agriculturists	Upper-middle
Vanne	64	Chittoor	Agriculturists	Upper-middle
Balija	76	Chittoor, Cuddapah, Kurnool	Traders and agriculturists	Lower-middle1
Ekila	62	Chittoor	Shepherds, agriculturists	Lower-middle1
Kurava	66	Chittoor, Cuddapah	Shepherds, agriculturists	Lower-middle1
Thogata	58	Chittoor, Cuddapah	Weavers and agriculturists	Lower-middle1
Yadava	52	Chittoor, Cuddapah, Kurnool, Nellore	Shepherds	Lower-middle1
Ediga	58	Chittoor, Cuddapah	Toddy tappers	Lower-middle2
Gandla	34	Chittoor, Cuddapah	Oil pressers	Lower-middle2
Jangam	48	Chittoor	Trade in pearls, beads, etc.	Lower-middle2
Chakali	42	Chittoor, Cuddapah	Washermen	Lower1
Mangali	38	Chittoor, Cuddapah, Kurnool	Barbers	Lower1
Vaddi	80	Chittoor, Cuddapah	Diggers and stone workers	Lower1
Madiga	94	Chittoor, Cuddapah, Kurnool, Anantapur	Scavengers and leather workers	Lower2/Pancham
Mala	190	Chittoor, Cuddapah, Kurnool, Ananthapur, Nellore	Scavengers, grave diggers, and menial workers	Lower2/Pancham
Erukala	62	Chittoor, Nellore, Ananthapur	Nomadic, basket weaving, hunter gatherers, pig herders	Proto-Australoid tribe
Sugali	116	Chittoor, Cuddapah, Ananthapur	Artisanry, cattle robbers	Europoid tribe
Yanadi	36	Chittoor, Nellore, Prakasam	Hunter-gatherers	Proto-"Australoid" tribe
Dudekula	84	Chittoor, Cuddapah	Cotton thread making	Muslims
Sheik	64	Chittoor, Cuddapah, Kurnool, Anantapur	Miscellaneous	Muslims

a. Number of chromosomes.
b. The socioeconomic categories are in decreasing order of hierarchy, except the two Muslim groups.

Statistical Methods.   The allele frequencies were estimated using a direct gene-counting procedure. Average heterozygosity and the coefficient of gene differentiation (G_{ST}) along with their standard errors were obtained following the method of Nei (1987). Genetic distances were computed using the modified Cavalli-Sforza distance (D_A) of Nei et al. (1983). The distances were also obtained using the stepwise weighted genetic distance measure (D_{SW}) of Shriver et al. (1995) and Reynolds's θ_{st}/F_{ST} (Reynolds et al. 1983). Although D_A is not linear with evolutionary time, it is the most efficient way to obtain the most correct phylogenetic relationships among closely related populations (Takezaki and Nei 1996). On the other hand, F_{ST} or θ_{st} (Reynolds et al. 1983) is a modified form of Cavalli-Sforza's chord distance, with the assumptions that there is no new mutation and that all gene frequency changes are due to genetic drift. Constant and equal population sizes were not assumed. Therefore this may be the most appropriate measure of genetic distance for the regional populations of Andhra Pradesh, which probably have a short evolutionary history.

The neighbor-joining algorithm (Saitou and Nei 1987) was used to construct the phylogenetic trees. Computations were performed using the NJBAFD program (supplied by N. Takezaki, National Institute of Genetics, Mishima, Japan), DISPAN, and PHYLIP, version 3.573.

The multidimensional scaling plots of the populations, which are based on the genetic distances, were created with the help of SPSS, version 7.5. Hardy-Weinberg equilibrium (Guo and Thompson 1992) and population genetic structure, as inferred from the analysis of molecular variance (AMOVA) (Excoffier et al. 1992), were performed using Arlequin, version 2.00 (Schneider et al. 1997). The AMOVA used in Arlequin is essentially similar to other approaches based on the analysis of variance of the gene frequencies but takes into account the number of mutations between molecular haplotypes. A hierarchical analysis of variance partitions the total variance into covariance components resulting from intra-individual differences, interindividual differences, interpopulation differences, and/or intergroup differences. The significance of these differences is tested using a nonparametric permutation approach (Excoffier et al. 1992) by permuting haplotypes, individuals, or populations among individuals, populations, or groups of populations, respectively. The population structure was hierarchically defined into seven groups considering the socioeconomic criteria and the popular perception in the area based on the traditional occupations and relative ranking of different caste groups (see Table 1). However, the two Muslim groups were kept in a separate and additional category.

We have also performed a Markov chain Monte Carlo analysis of population structure using the program Structure, version 2 (available at http://pritch .bsd.uchicago.edu), which implements a model-based clustering method for grouping individuals in populations and for identifying migrants and admixed individuals. This approach assumes a model with K populations, each of which is characterized by a set of allele frequencies at each locus; individuals are assigned to a population or jointly to two or more populations if their genotypes

indicate that they are admixed. Individuals are grouped into populations in such a way so as to achieve Hardy-Weinberg equilibrium and linkage equilibrium (Pritchard et al. 2000). Because the populations analyzed in this study were closely related in their linguistic affiliations and geographic distribution, we used the improved model of allele-frequency correlations elucidated by Falush et al. (2003) without a priori population information. The burn-in lengths (the duration of a simulation run, before collecting data, to minimize the effect of the starting configuration) for the analysis were chosen in such a way so that the summary statistics would converge, and the simulation run was quite long (8×10^6 runs) to get accurate estimates of parameters.

The regression model of Harpending and Ward (1982) was also used to study the possible effects of genetic drift and admixture on the substructured Telugu populations. In this model the average heterozygosity of the ith population (H_i) should be equal to the overall mean heterozygosity of the entire population (in this case, populations of Andhra Pradesh) H_t, multiplied by $(1 - R_{ii})$, where R_{ii} is the genetic distance of a particular population from the gene frequency centroid. If gene flow from outside the region varies in amount from population to population, then this linear relationship no longer holds. Isolated groups will be less heterozygous than the linear prediction and hence will lie below the expected regression line, whereas populations receiving more gene flow from outside will be more heterozygous and therefore will lie above the regression line.

Results

For the sake of brevity the allele-frequency tables for the 27 populations are not presented here and are being published elsewhere (Reddy et al. 2005). The following salient features of the pattern of allele-frequency distributions can be highlighted here.

All the loci were highly polymorphic in these populations, with the number of alleles ranging from 9 for D3S1358, VWA, D5S818, and D13S317 to as many as 30 for D21S11. Overall, the average heterozygosity is high (≥ 0.80), and the loci with the largest number of alleles are the ones that show the highest values of heterozygosity in different populations (Table 2). The exact test probabilities for Hardy-Weinberg equilibrium suggest significant departures in certain locus-population combinations. Except for D5S818, all the other loci show departures from Hardy-Weinberg equilibrium in some populations. Forty-one out of the 243 loci-population combinations show such departures. Both a high degree of inbreeding among these populations and a relatively small sample size in some cases may account for most of these departures, because in only two cases was the observed departure a result of excess heterozygosity.

Genetic Diversity and Relationships. The spectrums of allele-frequency distributions are fairly uniform across the 27 populations (graphs not presented),

Table 2. Locus and Average Heterozygosity and G_{ST} Values for Different Andhra Pradesh Populations

Population	D3S1358	VWA	FGA	D8S1179	D21S11	D18S51	D5S818	D13S317	D7S820	Average
Brahmin	0.744	0.847	0.884	0.857	0.903	0.876	0.725	0.813	0.774	0.825
Kshatriya	0.728	0.798	0.884	0.847	0.903	0.871	0.763	0.807	0.764	0.818
Vysya	0.724	0.754	0.794	0.858	0.918	0.865	0.709	0.815	0.795	0.804
Akuthota	0.722	0.798	0.917	0.834	0.896	0.897	0.773	0.825	0.825	0.832
Kamma	0.727	0.824	0.898	0.838	0.891	0.863	0.710	0.823	0.761	0.815
Kapu	0.718	0.773	0.879	0.779	0.853	0.867	0.708	0.824	0.804	0.801
Pokanati	0.762	0.796	0.934	0.788	0.910	0.884	0.740	0.831	0.820	0.829
Panta	0.74	0.814	0.873	0.813	0.892	0.865	0.716	0.825	0.779	0.813
Vanne	0.75	0.816	0.896	0.874	0.847	0.898	0.699	0.835	0.835	0.828
Balija	0.760	0.825	0.926	0.828	0.896	0.833	0.762	0.841	0.789	0.829
Ediga	0.778	0.838	0.851	0.792	0.813	0.888	0.695	0.815	0.797	0.807
Ekila	0.714	0.802	0.874	0.810	0.866	0.879	0.736	0.855	0.708	0.805
Gandla	0.661	0.745	0.807	0.838	0.907	0.856	0.752	0.743	0.784	0.788
Jangam	0.786	0.796	0.835	0.837	0.887	0.839	0.735	0.829	0.789	0.815
Kurava	0.699	0.763	0.924	0.852	0.866	0.850	0.705	0.795	0.797	0.806
Thogata	0.762	0.803	0.862	0.805	0.913	0.848	0.710	0.795	0.793	0.810
Yadava	0.737	0.838	0.891	0.837	0.896	0.882	0.668	0.802	0.760	0.812
Chakali	0.771	0.783	0.904	0.835	0.868	0.826	0.740	0.838	0.770	0.815
Madiga	0.752	0.813	0.891	0.853	0.865	0.838	0.748	0.811	0.785	0.817
Mala	0.732	0.809	0.913	0.84	0.904	0.852	0.740	0.822	0.777	0.821
Mangali	0.718	0.688	0.882	0.817	0.926	0.898	0.679	0.787	0.775	0.797
Vaddi	0.73	0.826	0.881	0.850	0.844	0.840	0.751	0.836	0.816	0.819
Erukula	0.794	0.783	0.927	0.816	0.896	0.814	0.669	0.822	0.782	0.812
Sugali	0.751	0.816	0.928	0.808	0.878	0.869	0.743	0.810	0.789	0.821
Yanadi	0.695	0.822	0.873	0.815	0.873	0.865	0.622	0.862	0.736	0.796
Dudekula	0.771	0.800	0.881	0.793	0.899	0.880	0.687	0.827	0.803	0.816
Sheik	0.702	0.793	0.909	0.850	0.911	0.797	0.775	0.852	0.815	0.823
Number of alleles	9	9	25	12	30	22	9	9	10	
G_{ST}	0.020	0.028	0.047	0.028	0.044	0.035	0.022	0.025	0.028	0.031 ± 0.00317

suggesting genetic homogeneity of the populations from this region. Twenty-five of the 27 populations (excluding the two Muslim groups, which do not fit strictly into the scheme of the hierarchical caste system) are grouped into 5 broad socio-economic categories, pooling the lower-middle1 and lower-middle2 and the lower1 and lower2 categories as single groups (see Table 1). The allele-frequency distributions are presented in Figure 2. The shape of the distribution of allele frequencies is fairly uniform across most socioeconomic groups for most of the loci, barring the few most polymorphic ones. The relatively greater population heterogeneity in the pattern of allele distribution is evident at the loci with the larger numbers of alleles, resulting in much greater G_{ST} values for FGA (0.047), D21S11 (0.044), and D18S51 (0.035). The G_{ST} values for the remaining loci are low and range from 0.020 for D3S1258 to 0.028 for VWA, D8S1179, and D7S820. The average G_{ST} value is low (0.031 ± 0.0032), reflecting relative homogeneity of the Andhra Pradesh populations. Among the five major socio-economic groups, upper castes ($G_{ST} = 0.015$) and tribes ($G_{ST} = 0.024$) show relatively greater population homogeneity compared to the upper-middle ($G_{ST} = 0.036$), lower-middle ($G_{ST} = 0.036$), or lower ranking caste groups ($G_{ST} = 0.032$) within them.

Although the neighbor-joining tree was constructed using both the D_A and the F_{ST} distances, to visualize the pattern of population relationships, given the significant similarity in the resultant trees, we present here only the F_{ST}-based neighbor-joining tree (Figure 3). Four distinct clusters of populations are apparent in the tree, but no clear separation of the populations based on social hierarchy or geographic affiliation can be discerned. For example, the first major cluster, representing 14 of the 27 Andhra Pradesh populations, consists of two distinct subclusters. One of these subclusters includes all three upper-caste groups (Brahmin, Kshatriya, and Vysya), but it also represents two upper-middle ranking castes (Akuthota and Kamma), a lower-middle ranking group (Kurava), and a tribal group (Sugali), which has a Europoid ethnic background. Similarly, the second major subcluster is a conglomeration of tribes (Yanadi and Erukela), Muslims (Sheik and Dudekula), a lower-middle ranking caste (Thogata), and two lower castes (Mangali and Madiga). The two-dimensional plot of the populations (Figure 4) based on multidimensional scaling of the F_{ST} distances also did not bring out any clear constellations of populations, based on any rational criteria, suggesting a lack of clear differentiation.

Affinities of the Andhra Pradesh Populations with Other Indian and Continental Populations. Compared to other Indian and world populations, the populations of Andhra Pradesh form a distinct cluster clearly separated from the rest (Figure 5). The other populations in the tree seem to be aligned on broad geographic, ethnic, or linguistic affiliations. For example, Asian populations from northeastern India form a distinct cluster, as do the other Asian populations from sub-Himalayan India and East Asia. Populations from western India, central
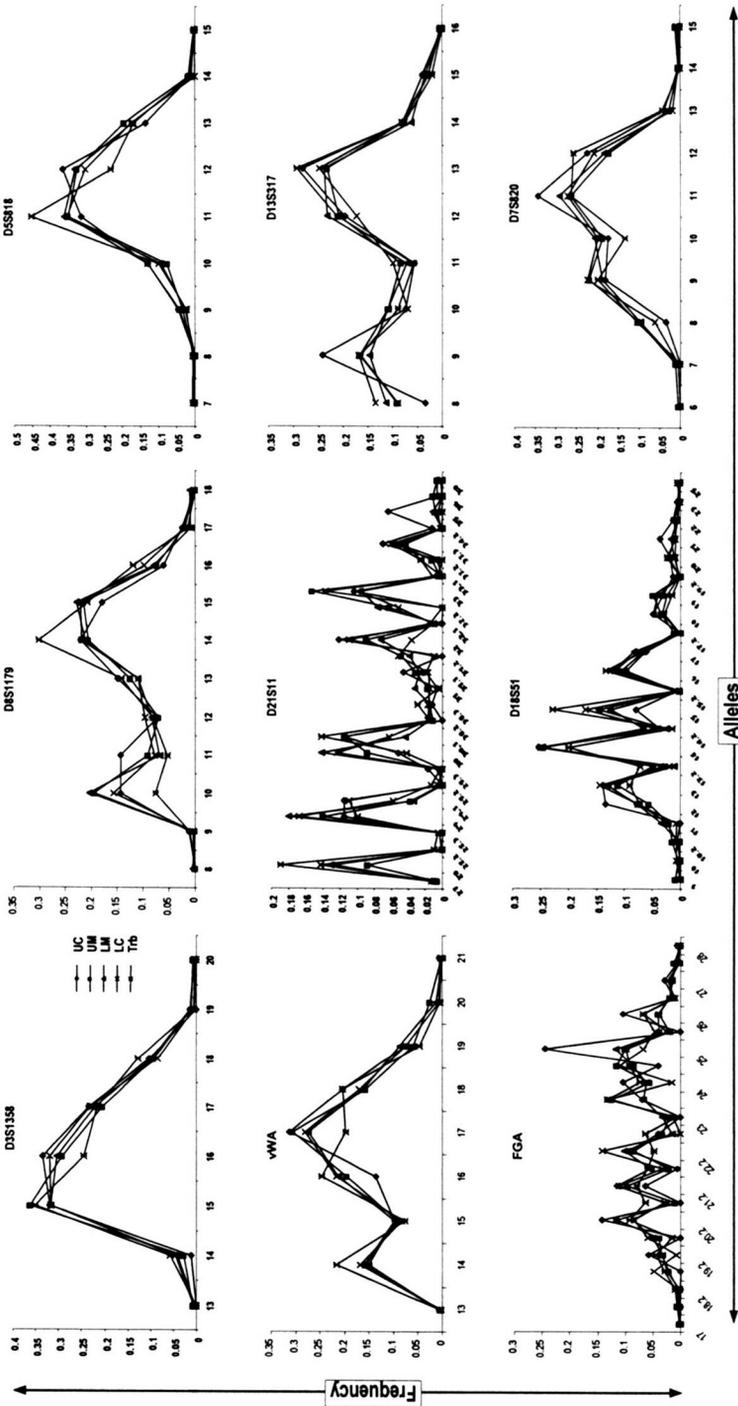
Figure 2. The frequency polygons showing the pattern of allele-frequency distributions across the five broad socioeconomic groups of Andhra Pradesh.
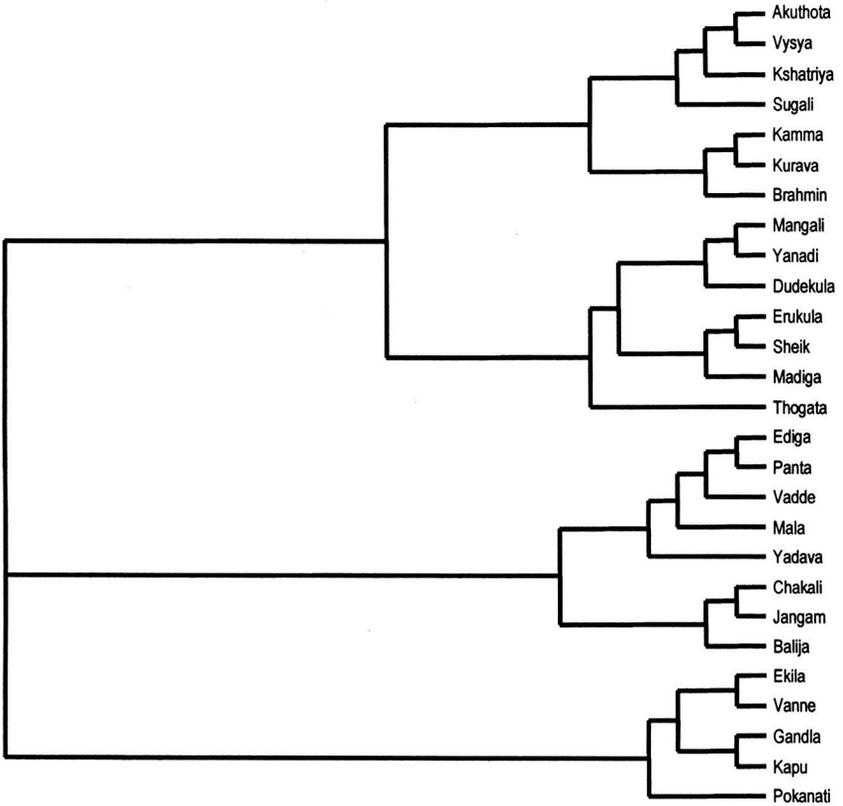
Figure 3. The neighbor-joining tree (constructed on the basis of F_{ST} distances) depicting phylogenetic relationships among the 27 populations of Andhra Pradesh.

India, and North India also form distinct subclades, although geographic contiguity is apparent with their placement as neighboring clades. However, some of the populations from South India that have Europoid physical features (Iyenger, Lingayat, Gowda, and Muslims) form a subclade along with the three American groups (US whites, US Hispanics, and African Americans).

Genetic Structure Versus Social Structure. The analysis of molecular variance (AMOVA) considering the 27 Andhra Pradesh populations in 8 rational groups, based on socioeconomic criteria, suggests virtually no variation among the groups (Table 3). Although the variation among populations within the groups is statistically significant, it is rather small (1.37%). Most of the variation is observed to be between individuals within populations. This reiterates the relative homogeneity of the populations shown by the neighbor-joining trees based on
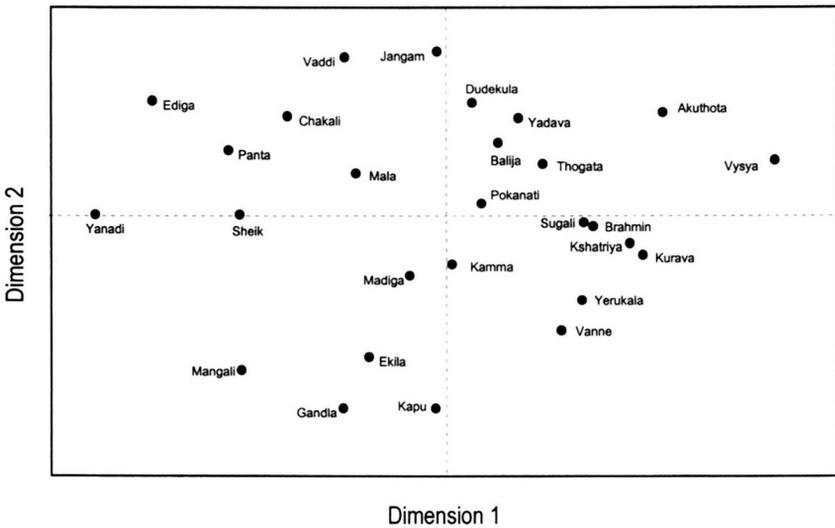
Figure 4.   Two-dimensional plot based on multidimensional scaling of the F_{ST} distance matrix of the 27 Andhra Pradesh populations.

different distance measures. The AMOVA results remained almost identical even after the number of and criteria (e.g., geographic contiguity and/or socioeconomic similarity) for constituting the rational groups was changed.

Because the AMOVA results did not conform to the population structure based on socioeconomic hierarchy, we further attempted a structural analysis of the populations without a priori information using an admixture model, and we repeated the run for different values of K ($K = 1$ to 20) to ascertain whether any genetic structure other than that based on the socioeconomic criteria exists. We found that the estimates of Pr(genotypic data | K) are similar for $K = 1$ to 11, suggesting that there is no genetic structure among these groups that could be revealed by the nine STR loci. Beyond $K = 11$ the estimates of the probability decrease substantially. Although, the estimates of the probability were similar for $K = 1$ to 11, we chose $K = 8$ as the optimum number of clusters to be inferred to compare with the results of the AMOVA. Figure 6, based on the proportion of membership of each predefined population in each cluster, suggests that none of the inferred clusters have a significant proportion of membership from any of the populations. In fact, all the populations exhibit a uniform degree of presence in each of the inferred clusters, which is consistent with the AMOVA results. Structural analysis of the data on the five subgroups of the Reddy caste (results not presented) did not bring out any genetic substructure bearing signatures of the subcaste endogamy and differentiation.
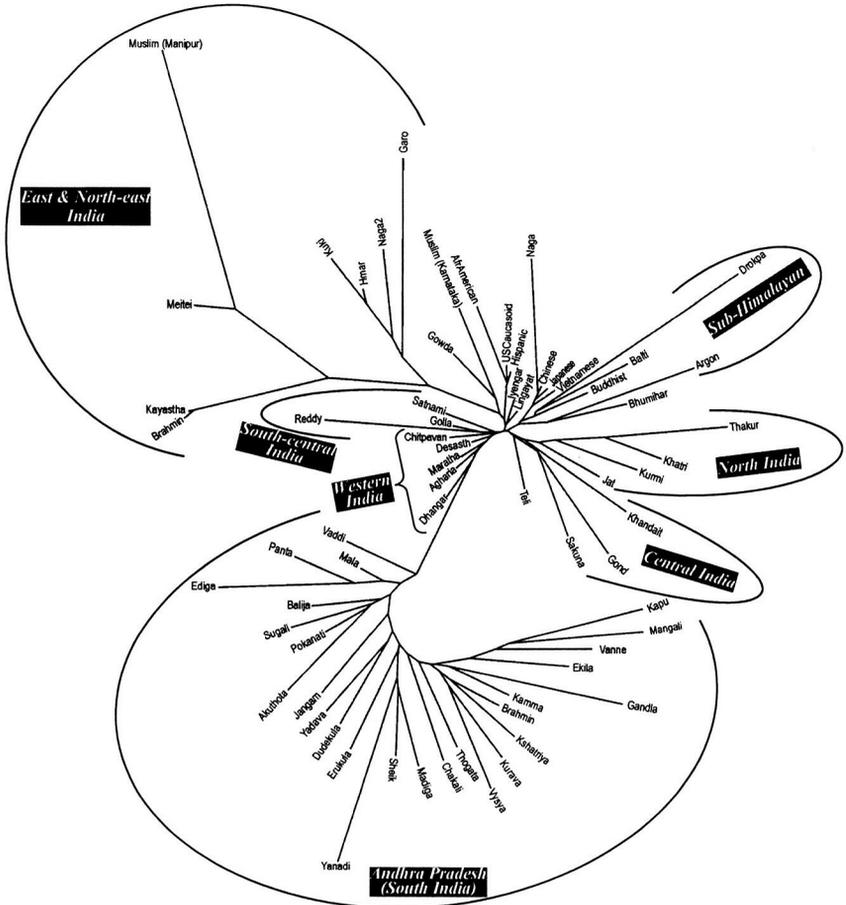
Figure 5. Cladogram depicting the relationship of the populations of Andhra Pradesh to other Indian and continental populations. Sources for comparative data: AfrAmerican (African American), USCaucasoid (US whites), and Hispanic (US Hispanics) (*AmpFlSTR Identifiler PCR Amplification Kit User's Manual* 2000); Vietnamese (Borys et al. 1999a); Japanese (Borys et al. 1999b); Chinese (Fung et al. 2001); Golla (Reddy et al. 2001b); Iyengar, Gowda, Lingayat, and Muslim (Karnataka) (Rajkumar and Kashyap 2002); Agharia, Satnami, Gond, and Teli (Sarkar and Kashyap 2002); Buddhist, Argon, Drokpa, and Balti (Trivedi et al. 2002); Maratha, Desasth Chitpavan, and Dhangar (Gaikwad and Kashyap 2002); Bhumihar, Reddy, Sakuna, Naga, and Khandait (Kashyap et al. 2002); Thakur, Khatri, Kurmi, and Jat (Tandon et al. 2002); Garo, Naga2, Kuki, and Hmar (Chattopadhyay et al. 2001); Brahmin, Kayastha, Muslim (Manipur), and Meitei (Dutta et al. 2000).

Table 3. Results of AMOVA Based on the 27 Andhra Populations Rationalized into 8 Socioeconomic Groups Depicting the Population Structure[a,b]

Source of Variation	df	Sum of Squares	Variance Component	Variation (%)
Among groups	7	51.232	−0.00057 (Va)	−0.02
Among populations within groups	19	128.751	0.04912 (Vb)	1.37
Within populations	1,869	6609.988	3.53664 (Vc)	98.65
Total	1,895	6789.97	3.58519	

a. Fixation indexes: F_{ST}: 0.01354 (significant at 0.01); F_{SC}: 0.01370 (significant at 0.01); F_{CT}: −0.00016.

b. Group 1: Brahmin, Kshatriya, Vysya; Group 2: Pokanati, Vanne, Panta, Akuthota, Kapu, Kamma; Group 3: Balija, Yadava, Kurava, Ekila, Thogata; Group 4: Jangam, Gandla, Ediga; Group 5: Chakali, Mangali, Vaddi; Group 6: Mala, Madiga; Group 7: Sugali, Erukula, Yanadi; Group 8: Sheik, Dudekula.

To further gauge the pattern of relationships between different hierarchical caste groups and the tribes, we computed Nei's standard genetic distances (D_S) and standard errors (Nei 1972, 1978) among the five broad socioeconomic groups (upper, upper-middle, lower-middle, and lower castes and tribes) (Table 4). Although genetic distance tended to increase with increasing difference in the social hierarchy, the differences were not statistically significant. Even this meek trend
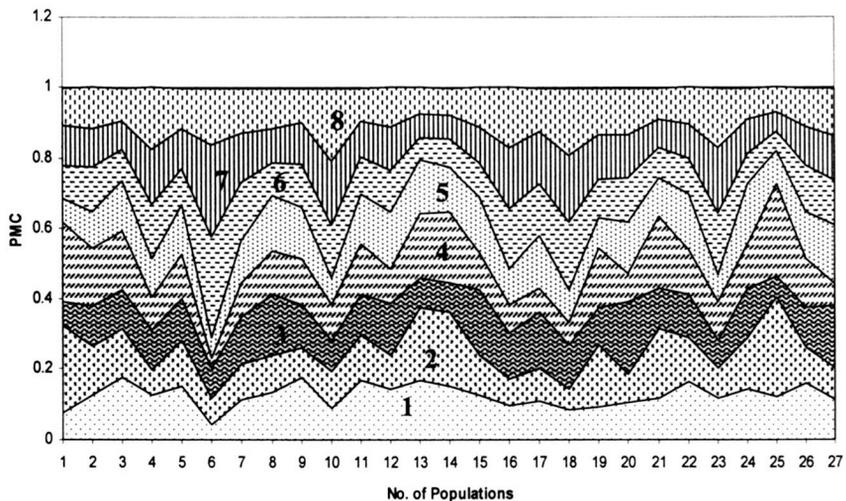
Figure 6. Proportion of membership coefficients (PMC) of the 27 Andhra Pradesh populations into the 8 inferred clusters.

Table 4. Standard Genetic Distances Between the Hierarchical Caste Groups of Andhra Pradesh and the Tribes (Below Diagonal) and Their Standard Errors (Above the Diagonal)

	Upper Castes	Upper-Middle Castes	Lower-Middle Castes	Lower Castes	Tribes
Upper castes	–	0.0129	0.0152	0.0235	0.0152
Upper-middle castes	0.0247	–	0.0051	0.0047	0.0097
Lower-middle castes	0.0295	0.0048	–	0.0040	0.0148
Lower castes	0.0365	0.0077	0.0034	–	0.0109
Tribes	0.0475	0.0304	0.0371	0.0303	–

disappears when we consider D_A distances or when we compute average distances for different pairs of populations between different hierarchical groups. Furthermore, the average distance between populations of the same socioeconomic group is not significantly different from or lower than the average distance between the populations of different groups. However, each of these hierarchical groups shows the largest genetic distance with the tribes compared to the mutual distances among them, suggesting genetic isolation and differentiation of the tribes and castes.

Gene Flow. The regression plot of mean heterozygosity versus the R_{ii} of the 27 Andhra Pradesh populations is depicted in Figure 7. It is interesting to note from the plot that the Akuthota Kapu, Vanne Kapu, and Pokanati (the three subgroups of the Reddy caste) appear as relatively more distinct outliers above the theoretical regression line, suggesting that external gene flow played a role in their differentiation. However, the position of the Akuthota Kapu as an outlier above the theoretical regression line is quite distinct, besides being far removed from the gene frequency centroid with a large value of R_{ii}. This may suggest that admixture as well as stochastic processes played a role in shaping their genetic composition. The Gandla, Yanadi, Mangali, and Kapu appear as outliers below the theoretical regression line, indicating that isolation and stochastic processes played a role in sculpting their genetic composition. The rest of the populations are scattered above and below but in the vicinity of the regression line, in conformity to the model and suggesting a uniform degree of gene flow among them.

Discussion

In recent years there has been increasing use of microsatellite loci to understand genetic relationships between closely related populations (Chu et al. 1998; Reddy et al. 2001a, 2001c). In the present study we used nine autosomal AmpF*l*-STR microsatellite markers to understand the population structure and patterns of variation of the 27 caste and tribal populations distributed in the contiguous
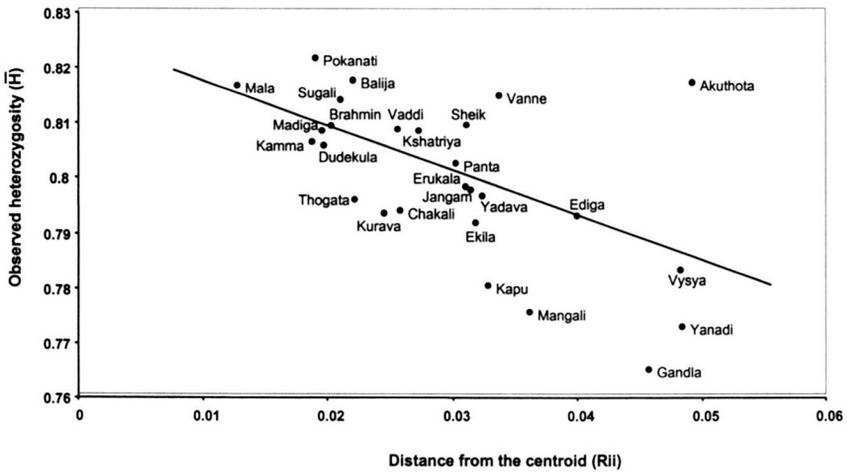
Figure 7. Regression plot of average heterozygosity (*H*) versus the distance from the centroid (R_{ii}) of the 27 Andhra Pradesh populations.

areas of the southern districts of Andhra Pradesh in India. Although most of these microsatellite loci are highly polymorphic within each of the Andhra Pradesh populations, the allele distributions are fairly uniform across the populations, suggesting relative homogeneity among them.

Concurrent with this, the coefficient of gene differentiation is low ($G_{ST} = 0.031 \pm 0.0032$) and is almost identical to the value obtained by Reddy et al. (2001c) (but using a different set of 13 STR loci) among 8 endogamous subcastes of a substructured single caste, the Golla, of Andhra Pradesh. This low coefficient of gene differentiation is reflected in the lack of clear differentiation and clustering pattern of the population based on either socioeconomic stratification or known ethnohistorical and geographic affiliations. Our analyses based on the three most informative markers (FGA, D21S11, and D18S51) did not qualitatively change the results. Consistent with the overall homogeneity among them, the 27 Andhra Pradesh populations form a single and compact cluster compared to other Indian and world populations with a high degree of support from bootstraps.

In a recent study of southwestern Indian populations based on data from 15 microsatellites, Rajkumar and Kashyap (2004) observed similar trends. Bamshad et al. (2003) and Rosenberg et al. (2002) also observed that a large number of microsatellite loci are required to differentiate populations, even on the continental scale, implying a need for a much larger number of microsatellite loci to resolve the phylogeny on the regional level. However, the nine autosomal AmpF*l*-STR markers that are validated and widely used for forensic investigations have

also been found to be useful for unraveling the local population structure and for reconstructing evolutionary relationships at the level of ethnic, geographic, and linguistic categories in India (Dutta et al. 2002; Langstieh et al. 2004) and elsewhere (Sun et al. 2003). Furthermore, on the basis of 13 STR loci, Reddy et al. (2001c) observed that these markers help to reconstruct the short evolutionary history at the level of endogamous subcastes of an Indian caste, because these loci seem to have left signatures of subcaste endogamy.

Assuming that microsatellite loci are useful for clarifying evolutionary relationships of closely related populations (Takezaki and Nei 1996), how can one explain the uniformly high degree of polymorphism and heterogeneity within the populations and the reduced diversity between the populations? Most of the genetic variation accrued in the populations of Andhra Pradesh can best be explained in terms of founder effects in the formation of the subgroups and subsequent genetic drift over the generations. Therefore, if the barriers of caste and tribal boundaries are impermeable, the observed homogeneity of the Andhra Pradesh populations may imply a relatively recent history of separation or substructuring, so that the small differences that may have accumulated are not being captured by the nine loci used in this study.

With a relatively larger number of STR markers among the regional populations of China, Chu et al. (1998) interpreted similar findings as possibly reflecting single origin, although Chu acknowledged that the lack of resolution of the STR loci used could also be one of the reasons for the observed pattern. However, Langstieh et al. (2004) recently observed that the reduced microsatellite diversity among the Meghalaya populations is due to admixture, which is perpetuated by the system of matrilineal descent and matrilocal residence, coupled with the relatively short history of separation of these tribes. Can these explanations be plausibly extended to the present situation in the Andhra Pradesh populations? The populations of southern Andhra Pradesh have clearly defined hierarchical structure, including within them the upper castes, middle-ranking castes, lower castes, and tribes. There could be two plausible scenarios that could explain the observed genetic composition and structure of these Andhra populations: (1) a recent and common origin of these populations (hence the genetic differentiation is not significant enough to show a systematic pattern of population relationships); and (2) unrecognized gene flow, albeit at a low rate, over the generations among the coexisting and closely interacting caste populations from a homogeneous area.

Even though it is interesting to note the systematic pattern of the differences in genetic distances observed between the hierarchical groups, tempting one to surmise that this is not really a case of genetic stratification being consistent with the social stratification, given the large standard errors, this stratification does not reach statistical significance. However, Bamshad et al. (2001), using a large set of autosomal markers, found this stratification to be highly significant in the caste populations of Andhra Pradesh. When we computed distances based on the nine STR loci for the hierarchical caste groups, treating the three *varna* categories in the upper castes (Brahmin, Kshatriya, and Vysya) separately, no

particular pattern of genetic distances, adhering to the implicit hierarchy, emerged between them (distance matrix not presented). This suggests a lack of strong genetic signatures consistent with the traditional *varna* system (constituting only Brahmin, Kshatriya, Vyshya, and Sudra categories), although a semblance of genetic stratification was evident with respect to socioeconomic hierarchy (i.e., upper, middle, and lower castes).

It may be pertinent to note here that Bamshad et al. (1998) did not observe genetic stratification in the Y-chromosome-based markers, whereas Bhattacharyya et al. (1999) found no evidence of male gene flow across caste and ethnic boundaries in India. On the other hand, Ramana et al. (2001) and Cordaux et al. (2004) found evidence to support bidirectional male gene flow in southern India, particularly in the caste populations of Andhra Pradesh. Furthermore, based on traditional genetic markers, Kumar et al. (2004) demonstrated a tribe-caste continuum in the genetic structure of the populations of northeast India, suggesting gene flow between them.

Bamshad et al. (1998) interpreted the observed stratification in mtDNA among the populations of Andhra Pradesh as due to movement of females, because of the practice of hypergamy in the patrilocal Hindu society, and, conversely, the lack of stratification in Y-chromosome-based markers as due to a lack of movement of males across caste boundaries. No doubt, hypergamy, which is widely acknowledged in the literature to have existed in historical times, might have been practiced only by the minority in the higher echelons of the society, merely as a mechanism to accommodate and legitimize the multiple marriages that the men of the ruling class were accustomed to. The practice of hypergamy as an institution or as a traditional practice is certainly unknown in contemporary India to have caused such a systematic and significant genetic stratification. On the contrary, given the traditional pattern of symbiotic relationships among the cohabiting and contiguously distributed hierarchical caste groups with day-to-day intimate interactions, we speculate that there might have been continuous unauthorized or unrecognized gene flow, albeit at a low rate, across the social groups, despite strict adherence to the caste endogamy, resulting in relative homogenization of at least the caste groups. This is perhaps implicit in the observed homogeneity of caste populations in southern India for Y-chromosome-based markers (Cordaux et al. 2004).

A reflection of this can also be seen in the results of structure analyses with the admixture model, because the proportions of membership of different populations is fairly uniform across all the inferred clusters (see Figure 6). This is feasible only through male gene flow in a strictly endogamous and patrilocal society; there is no possibility of female genes transecting the caste boundaries, even in the event of unauthorized sexual interaction between caste groups. Therefore one may expect reduced Y-chromosome diversity across the populations and even stratification based on social hierarchy, given relatively greater interaction between groups of adjacent ranks. This may explain the mild gradient of genetic distances observed with the changing hierarchy of populations in the present

study. Given the small number of microsatellite loci used in this study, the conclusions remain tentative until our analysis of Y-chromosome-based markers provides more conclusive evidence on the probable processes behind the observed trend of genetic stratification vis-à-vis social stratification in the Indian populations.

Literature Cited

AmpFlSTR Identifiler PCR Amplification Kit User's Manual. 2000. Foster City, CA: Applied Biosystems.

Bamshad, M., A. E. Fraley, M. H. Crawford et al. 1996. mtDNA variation in caste populations of Andhra Pradesh, India. *Hum. Biol.* 68:1–28.

Bamshad, M. J., T. Kivisild, W. S. Watkins et al. 2001. Genetic evidence on the origin of Indian caste populations. *Genome Res.* 11:994–1004.

Bamshad, M. J., W. S. Watkins, M. E. Dixon et al. 1998. Female gene flow stratifies Hindu castes. *Nature* 395:651–652.

Bamshad, M. J., S. Wooding, W. S. Watkins et al. 2003. Human population genetic structure and inference of group membership. *Am. J. Hum. Genet.* 72:578–589.

Basu, A., N. Mukherjee, S. Roy et al. 2003. Ethnic India: A genomic view, with special reference to peopling and structure. *Genome Res.* 13:2277–2290.

Bhattacharyya, N. P., P. Basu, M. Das et al. 1999. Negligible male gene flow across ethnic boundaries in India, revealed by analysis of Y-chromosomal DNA polymorphisms. *Genome Res.* 9:711–719.

Borys, S., A. Eisenberg, G. Carmody et al. 1999a. Allele frequencies for 9 STR loci in African American, Chinese, Vietnamese, and Bangladesh populations. *J. Forensic Sci.* 44:1316–1318.

Borys, S., S. Iwaamoto, J. Miyakoshi et al. 1999b. Allele frequency distribution for 9 STR loci in the Japanese population. *J. Forensic Sci.* 44:1319.

Chattopadhyay, P., R. Dutta, and V. K. Kashyap. 2001. Allele frequency data at nine fluorescent labeled STR loci in four tribal population groups of India. *J. Forensic Sci.* 45:184–188.

Chu, J. Y., W. Huang, S. Q. Kuang et al. 1998. Genetic relationship of populations in China. *Proc. Natl. Acad. Sci. USA* 95:11,763–11,768.

Clark, V. J., S. Sivendren, N. Saha et al. 2000. The 9-bp deletion between the mitochondrial lysine tRNA and COII genes in tribal populations of India. *Hum. Biol.* 72:273–285.

Cordaux, R., R. Aunger, G. Bentley et al. 2004. Independent origins of Indian caste and tribal paternal lineages. *Curr. Biol.* 4:231–235.

Cordaux, R., N. Saha, G. R. Bentley et al. 2003. Mitochondrial DNA analysis reveals diverse histories of tribal populations from India. *Eur. J. Hum. Genet.* 11:53–64.

Dutta, R., P. Chattopadhyay, and V. K. Kashyap. 2000. STR data for the AMPFlSTR profiler plus loci among four predominant populations of eastern India. *J. Forensic Sci.* 45:1353–1357.

Dutta, R., B. M. Reddy, P. Chattopadhyay et al. 2002. Patterns of genetic diversity at the nine forensically approved STR loci in the Indian populations. *Hum. Biol.* 74:33–49.

Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* 131:479–491.

Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inferences of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* 164:1567–1587.

Fung, W. K., J. Ye, L. Hu et al. 2001. Allele frequencies for nine STR loci in Beijing Chinese. *Forensic Sci. Int.* 121:207–209.

Gadgil, M., N. V. Joshi, S. Manoharan et al. 1998. Peopling of India. In *The Indian Human Heritage*, D. Balasubramanian and N. A. Rao, eds. Hyderabad, India: University Press, 100–129.

Gaikwad, S., and V. K. Kashyap. 2002. Polymorphism at fifteen hypervariable microsatellite loci in four populations of Maharashtra, India. *Forensic Sci. Int.* 126:267–271.

Guo, S., and E. Thompson. 1992. Performing exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48:361–372.

Harpending, H. C., and R. Ward. 1982. Chemical systematics and human evolution. In *Biochemical Aspects of Evolutionary Biology*, M. Nitecki, ed. Chicago: University of Chicago Press, 213–256.

Karve, I. 1961. *Hindu Society: An Interpretation.* Poone, India: Deshmukh Prakashan.

Kashyap, V. K., N. Sarkar, and R. Trivedi. 2002. Allele frequencies for STR loci of the Powerplex 16 Multiplex system in five endogamous populations of India. *Forensic Sci. Int.* 126:178–186.

Kivisild, T., M. J. Bamshad, K. Kaldma et al. 1999a. Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Curr. Biol.* 9:1331–1334.

Kivisild, T., K. Kaldma, M. Metspalu et al. 1999b. The place of mitochondrial DNA variants in the global network of the maternal lineages and the peopling of the Old World. In *Genome Diversity: Applications in Human Population Genetics*, S. S. Papiha and R. Deka, eds. New York: Kluwer, 135–152.

Kivisild, T., S. Rootsi, M. Metspalu et al. 2003. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am. J. Hum. Genet.* 72:313–332.

Kosambi, D. D. 1991. *The Culture and Civilization of Ancient India in Historical Outline.* New Delhi, India: Vikas.

Krishnan, T., and B. M. Reddy. 1994. Geographical and ethnic variability of finger ridge counts: Biplots of male and female Indian samples. *Ann. Hum. Biol.* 21:155–169.

Kumar, V., D. Basu, and B. M. Reddy. 2004. Genetic heterogeneity in northeastern India: Reflection of tribe-caste continuum in the genetic structure. *Am. J. Hum. Biol.* 16:334–345.

Langstieh, B. T., B. M. Reddy, K. Thangaraj et al. 2004. Genetic relationships among the tribes of Meghalaya, India: Matriliny and reduced microsatellite diversity. *Hum. Biol.* 76:569–590.

Majumder, P. P. 1998. People of India: Biological diversity and affinities. *Evol. Anthropol.* 6:100–110.

Nei, M. 1972. Genetic distances between populations. *Am. Nat.* 106:283–292.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89:583–590.

Nei, M. 1987. *Molecular Evolutionary Genetics.* New York: Columbia University Press.

Nei, M., F. Tajima, and Y. Tateno. 1983. Accuracy of estimated phylogenetic trees from molecular data. *J. Mol. Evol.* 19:153–170.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.

Rajkumar, R., and V. K. Kashyap. 2002. Distribution of alleles of 15 STR loci of the Powerplex Multiplex system in four predominant population groups of South India. *Forensic Sci. Int.* 126:173–177.

Rajkumar, R., and V. K. Kashyap. 2004. Genetic structure of four socioculturally diversified caste populations of southwest India and their affinity with related Indian and global groups. *BMC Genet.* (online) 5:23.

Ramana, G. V., B. Su, L. Jin et al. 2001. Y-chromosome SNP haplotypes suggest evidence of gene flow among caste, tribe, and the migrant Siddi populations of Andhra Pradesh, South India. *Eur. J. Hum. Genet.* 9:695–700.

Reddy, B. M. 2002. Patterns of endogamy and consanguinity in India [abstract]. Invited talk delivered at the Symposium on Community Genetics in Developing Countries, Indian Institute of Science, Bangalore.

Reddy, B. M., D. A. Demarchi, S. Bharti et al. 2004. Patterns of ethnic, linguistic, and geographic heterogeneity of palmer interdigital ridge counts in the Indian subcontinent. *Hum. Biol.* 76:211–228.

Reddy, B. M., R. Dutta, B. Langstieh et al. 2001a. Diversity at three tetrameric STR loci in a substructured Golla caste population of southern Andhra Pradesh in comparison to other Indian populations. *Int. J. Hum. Genet.* 1:1–9.

Reddy, B. M., V. M. Naidu, V. K. Madhavi et al. 2005. Data for the 9 STR loci among 27 populations of different social hierarchy from southern part of Andhra Pradesh, India. *Forensic Sci. Int.* 149:81–97.

Reddy, B. M., G. Sun, and R. Dutta. 2001b. STR data for the AmpF/STR Profiler Plus loci among Golla population of southern Andhra Pradesh, India. *J. Forensic Sci.* 46:734–735.

Reddy, B. M., G. Sun, J. R. Luis et al. 2001c. Genomic diversity at thirteen short tandem repeat loci in a substructured caste population, Golla, of southern Andhra Pradesh, India. *Hum. Biol.* 73:175–190.

Reynolds, J., B. S. Weir, and C. C. Cockerham. 1983. Estimation for coancestry coefficient: Basis for a short-term genetic distance. *Genetics* 105:767–779.

Rosenberg, N. A., J. K. Pritchard, J. L. Weber et al. 2002. Genetic structure of human populations. *Science* 298:2381–2385.

Saitou, N., and M. Nei. 1987. The neighbour-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4:406–425.

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular Cloning: A Laboratory Manual,* 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Sarkar, N., and V. K. Kashyap. 2002. Genetic diversity at two pentanucleotide STR and thirteen tetranucleotide STR loci by multiplex PCR in four predominant population groups of central India. *Forensic Sci. Int.* 128:196–201.

Schneider, S., D. Rosslie, and L. Excoffier. 1997. *Arlequin ver. 2.000: A Software for Population Genetics Data Analysis.* Geneva: Genetics and Biometry Laboratory, University of Geneva. Available at http://anthropologie.unige.ch/arlequin

Shriver, M. D., L. Jin, E. Boerwinkle et al. 1995. A novel measure of genetic distance for highly polymorphic tandem repeat loci. *Mol. Biol. Evol.* 12:914–920.

Sun, G., S. T. McGarvey, R. Bayoumi et al. 2003. Global genetic variation at nine short tandem repeat loci and implications on forensic genetics. *Eur. J. Hum. Genet.* 11:39–49.

Takezaki, N., and M. Nei. 1996. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* 144:389–399.

Tandon, M., R. Trivedi, and V. K. Kashyap. 2002. Genomic diversity at 15 fluorescent labeled short tandem repeat loci in few important populations of state of Uttar Pradesh, India. *Forensic Sci. Int.* 128:190–195.

Thapar, R. 1966. *A History of India,* v. 1. Middlesex, UK: Penguin.

Thapar, R. 1995. The first millennium B.C. in northern India (up to the end of Mauryan period). In *Recent Perspective of Early Indian History,* R. Thapar, ed. Bombay: Popular Prakshan, 80–141.

Thapar, R. 2003. *Early India: From the Origins to A.D. 1300.* Berkeley: University of California Press.

Trivedi, R., P. Chattopadhyay, B. Maity et al. 2002. Genetic polymorphism at nine microsatellite loci in four high altitude Himalayan desert human populations. *Forensic Sci. Int.* 127:150–155.

Watkins, W. S., M. Bamshad, M. E. Dixon et al. 1999. Multiple origins of the mtDNA 9-bp deletion in populations of South India. *Am. J. Phys. Anthropol.* 109:147–158.