Transcriptional properties of human \textit{NANOG1} and \textit{NANOG2} in acute leukemic cells

Irina Eberle, Birgit Pless, Miriam Braun, Theo Dingermann and Rolf Marschalek*

Institute of Pharmaceutical Biology/DCAL/ZAFES/CEF, Goethe-University, Biocenter, Max-von-Laue-Str. 9, 60438 Frankfurt/Main, Germany

Received September 16, 2009; Revised April 13, 2010; Accepted April 14, 2010

ABSTRACT

Transcripts of \textit{NANOG} and \textit{OCT4} have been recently identified in human t(4;11) leukemia and in a model system expressing both t(4;11) fusion proteins. Moreover, downstream target genes of \textit{NANOG/OCT4/SOX2} were shown to be transcriptionally activated. However, the \textit{NANOG1} gene belongs to a gene family, including a gene tandem duplication (named \textit{NANOG2} or \textit{NANOGP1}) and several pseudogenes (\textit{NANOGP2-P11}). Thus, it was unclear which of the \textit{NANOG} family members were transcribed in t(4;11) leukemia cells. 5’-RACE experiments revealed novel 5’-exons of \textit{NANOG1} and \textit{NANOG2}, which could give rise to the expression of two different \textit{NANOG1} and three different \textit{NANOG2} protein variants. Moreover, a novel PCR-based method was established that allows distinguishing between transcripts deriving from \textit{NANOG1}, \textit{NANOG2} and all other \textit{NANOG} pseudogenes (P2–P11). By applying this method, we were able to demonstrate that human hematopoietic stem cells and different leukemic cells transcribe \textit{NANOG2}. Furthermore, we functionally tested \textit{NANOG1} and \textit{NANOG2} protein variants by recombinant expression in 293 cells. These studies revealed that \textit{NANOG1} and \textit{NANOG2} protein variants are functionally equivalent and activate a regulatory circuit that activates specific stem cell genes. Therefore, we pose the hypothesis that the transcriptional activation of \textit{NANOG2} represents a ‘gain-of-stem cell function’ in acute leukemia.

INTRODUCTION

The \textit{NANOG} protein—in combination with \textit{OCT4} and \textit{SOX2}—was shown to be sufficient to establish an embryonic stem cell program. Since the discovery of \textit{NANOG} in 2003 (1,2), \textit{NANOG} has drawn very much attention, and the ‘core NANOG network’ has been unraveled for human and murine embryonic stem (ES) cells by ChIP-on-Chip experiments (3,4). Stem cell functions of the core NANOG network are maintained by the help of the Polycomb repressor complex II (PRC II: SUZ12, EED, EZH2) which specifically silences genes coding for transcription factors necessary for the development of all three germ layers and neuronal development (5).

Tumor research has been widely improved by the concept of cancer-initiating cells (6). Cancer-initiating cells provide features of stem cells, however, different tumors seem to use different pathways to obtain stemness (7), and little is known about the molecular mechanisms that are required to establish this unique cell population.

Recently, we have discovered that \textit{NANOG} transcription was significantly enhanced (16-fold) in murine fibroblast when stably transfected with expression constructs coding for the \textit{MLL\textbullet Af4} and \textit{Af4\textbullet MLL} fusion genes, deriving from the chromosomal translocation t(4;11)(q21;q23) (8). This genetic aberration is associated with high-risk acute lymphoblastic leukemia and very poor outcome. Subsequent analyses of the core \textit{NANOG} network provided first evidence that \textit{NANOG} downstream targets were indeed transcriptionally activated, while \textit{NANOG/PRCII}-repressed genes were transcriptionally silenced. To verify this unusual finding, leukemic cells deriving from adult and pediatric t(4;11) patients were investigated and revealed the same transcriptional profile (8). Thus, it seems that the population of t(4;11) leukemia cells—or at least a small fraction thereof—is able to turn on a stem cell program similar to the core \textit{NANOG} network identified in ES cells.

A precise analysis of \textit{NANOG} transcription, however, is hampered by the fact that \textit{NANOG1} is transcribed along with several retroposed pseudogenes of the \textit{NANOG} family, also demonstrating that \textit{NANOG2} (alias \textit{NANOGP1}) is a tandem duplication of \textit{NANOG1} (9). The comparison of human and chimpanzee genome
sequences has revealed that \textit{NANOG2} retained its intronic sequences, while \textit{NANOG2} to \textit{P11} are dispersed, intronless and reverse transcribed integrants (10). Transcription of \textit{NANOG2} and \textit{NANOG} pseudogenes (\textit{NANOGP2, P4, P5, P7 and P8}) has been demonstrated for a large variety of different solid tumors (11,12). Thus, transcripts deriving from \textit{NANOG1, NANOG2} or these pseudogene copies can only be distinguished by cloning and sequencing the resulting PCR amplimers. Based on their specific mutation spectrum (missense, frame shift or deletion) their origin can be elucidated.

Therefore, we started a detailed investigation of the \textit{NANOG} gene family and their transcriptional properties, using \textit{MLL}-rearranged leukemia cells that seem to be capable of re-activating a \textit{NANOG}-dependent stem cell program, probably as part of their pathological properties.

\section*{MATERIAL AND METHODS}

\subsection*{Cell culture}
SEM, RS4;11, KOPN8 and NOMO1 cells were cultured in RPMI 1640 containing 10\% fetal calf serum. NTERA-2, HEK-293 and HEK-293T cells were maintained in DMEM supplemented with 10\% fetal calf serum (and 5\% horse serum for NTERA-2 cells). All media were supplemented with 1\% L-glutamine and 1\% Pen/Strep.

\subsection*{Total RNA extraction and reverse transcription}
RNA was prepared from NTERA-2, SEM, RS4;11, KOPN-8 and NOMO-1 cell lines using the Qiagen RNaseq Mini Kit (Qiagen, Germany). The protocol was slightly changed to extract only cytosolic RNA (in order to prevent genomic DNA contamination for pseudogene analysis). One microgram RNA was reverse transcribed using hexamer primers in a total volume of 25\,\mu{l}. Final cDNA synthesis was diluted to 50\,\mu{l} using sterile water. In addition, all isolated RNAs were directly tested in PCR reactions to exclude any contamination with genomic DNA (data not shown). Five microliters of each cDNA was used as template in 50\,\mu{l} PCR reactions throughout all experiments.

\textbf{NANOG-specific RT–PCR experiments}
Initial RT–PCR experiments were carried out by using three different primer sets (a–c). Two of these pair sets (a and b) specifically bind to transcripts deriving from \textit{NANOG1, NANOG2} and all other \textit{NANOG} pseudogenes (P2–P11). Primer set c binds to the 5′-flanking UTR of \textit{NANOG1} and \textit{NANOG2} (one mismatch) and to an internal exon. Primer sets were as follows: set a (5′-catgacatctaaacagtagttgatccagctttgctgc-3′; 5′-ggattctcacaagtctctcagttgcatg-3′) results in a 938-bp PCR amplimer; set b (5′-gccgacggtctggattgacatcaggtc-3′) produces a 418-bp PCR amplimer; set c (5′-ctagacgtgttttggagactgg-3′) results in a 444-bp PCR amplimer. The \textit{NANOG2}-specific primer set d (5′-gttgaatatggtttacaaaaagttgcagc-3′; 5′-gccacccctccagttgagtgctcgtcttaa-3′) produces a 351-bp PCR amplimer. Finally, a \textit{NANOGP8}-specific primer set e (5′-cacaagcttcttctcttcatttggttaggg-3′) resulted in a 525-bp amplimer. All amplifications were carried out with a denaturation step at 94°C for 2 min followed by 35 cycles with the following profile: 30 s at 94°C, 30 s at 58°C (45°C for set c; 59°C for set e) and 30 s at 72°C; a final extension step at 72°C for 5 min was performed for all reactions. QPCR experiments were performed by using the \textit{NANOG1}-specific oligonucleotides 5′-ccttcagcaagaaacaaagcttc-3′ and 5′-tctggtagaggaagtagttaaaggg-3′ (hybridizing to \textit{NANOG1} exon 1b and exon 2) and \textit{NANOG1/2}-specific oligonucleotides 5′-caccatatgctgttagttttggtgc-3′ and 5′-ttgattttgctttggagactgg-3′ (by dridizing to \textit{NANOG1/2} exons 3 and 4).

\textbf{5′-RACE experiments}
All 5′-RACE experiments were performed by using the Invitrogen RACE Kit according to the manufacturer’s instructions. Briefly, 5\,\mu{g} of extracted RNA was used for the initial dephosphorylation step and a subsequent decapping step resulting in a 5′-phosphate only at \textit{bona fide} mRNA molecules. Next, a ligation with an 44-nt long RNA oligonucleotide was performed, leading to 5′-tagged RNA molecules. Then, a first strand cDNA synthesis was performed using \textit{NANOG}-specific oligonucleotide (5′-gtttgctccacatgtaaggg-3′) that specifically binds to the final exon. After completion of first strand cDNA synthesis, two consecutive PCR reactions were performed to amplify specifically \textit{NANOG1} and \textit{NANOG2} transcripts. For the first PCR the specific oligonucleotides were used (5′-cgactggagcagggactgactg-3′; 5′-caccaggcatcctctggctcag-3′) in combination with a touch-down PCR protocol: 10 cycles with 30 s at 94°C and annealing and elongation for 3 min in 68–64.4°C (–0.4°C per cycle); this was followed by 25 cycles with 30 s at 94°C, 30 s at 64°C and 3 min at 68°C. An aliquot of the resulting amplification products were used for a nested PCR reaction using the oligonucleotides 5′-ggacactgacatggactgaagta-3′ and 5′-gccacccccttttagttttccttggtc-3′ in combination with a second PCR program (5 cycles with 30 s at 94°C, 30 s at 64°C and 90 s at 68°C). Reactions were loaded on a 2% agarose gel and the different PCR amplimers were cut out, eluted from the gel and cloned into the pCR®4-TOPO® vector (TOPO TA Cloning® Kit For Sequencing, Invitrogen, UK).

Subsequent sequence analyses using an universal T7 oligonucleotide (5′-taatacgaactcataaggg-3′) revealed different mRNA species deriving from the \textit{NANOG1} and \textit{NANOG2} genes, respectively.

\textbf{Establishment of \textit{NANOG1} and \textit{NANOG2}-specific PCR reactions}
\textit{NANOG1}-specific oligonucleotides were 5′-cacccccaca cgagatgg-3′ (specific for novel exon 2) and 5′-caga...
Cloning and expression of NANOG1/2 protein variants

Identified splice variants of *NANOG1* (A, B3) and *NANOG2* (D1, D2c and E) were amplified using specific oligonucleotides exhibiting appropriate restriction recognition sites for further cloning into the pEXPR-IBA10-3 vector (IBA BioTAGnology, Germany), resulting in a Strep-tag fusion to all 3'-ends of the amplified *NANOG1* and *NANOG2* open reading frames. Successfully cloned open reading frames were verified by sequence analyses. Subsequently, plasmids (25 μg per 15 cm petri dishes) coding for different NANOG1 and NANOG2 protein variants were lipotransfected into 293T cells and grown for 2 days. Transfection efficiencies were monitored by co-transfecting pEGFP plasmids (25 μg per 15 cm petri dishes). GFP-positive cells were always in the range of 60–70% transfected cells. Cells were harvested and lysed for at least 1 h in a cell lysis buffer containing 150 mM NaCl, 20 mM HEPES pH 7.5, 1% TritonX-100, 0.4 mM EDTA and a phosphatase/protease inhibitor mix. The soluble fraction was added to Strep-Tactin Superflow material and the Strep-tagged proteins were purified according to the manufacturer’s instruction (IBA BioTAGnology, Germany). An aliquot of the eluted protein fractions (10 μl) were used to separate the recombinant proteins on SDS page and blotted for subsequent western blot experiments.

Western blot experiments

30 × 10^6 SEM and 1 × 10^6 NTERA cells were lysed in the above mentioned lysis buffer. Soluble fractions (S), pellet fractions (P) or Strep-tag affinity purified recombinant proteins were mixed in 2× Lämmli buffer and separated on a 12% SDS–PAGE. After blotting onto a PVDF membrane, transferred proteins were incubated in blocking solution (5% skim milk in 0.2% Triton X-100/PBS) for 1 h. The C-terminal NANOG mouse antibody (clone 2C11, Abnova) was diluted 1:1000 in blocking solution and incubated over night. The membrane was washed three times with 0.2% Triton X-100/PBS and developed by using the ECL™ Western Blotting Analysis System (GE Healthcare).

Immunohistological experiments

HeLa cells were lipotransfected with pEXPr-IBA10-3 vectors (10 μg for 10 cm petri dishes) containing open reading frames for different NANOG1 and NANOG2 protein variants. After 24 h cells were spread out on cover glass and incubated for additional 24 h. The cells were fixed with 100% Methanol (–20°C) for 20 min at –20°C. Cells were incubated with blocking solution [10% goat serum, 0.1% Triton X-100 in PBS (PBST)] for 30 min. Methanol was removed by drying the cells and subsequently rehydrating them with PBS. The C-terminal NANOG rabbit antibody (H-155, Santa Cruz, USA) was diluted 1:300 in blocking solution containing 1% goat serum and incubated over night. Cells were washed three times for 5 min with 0.1% PBST and incubated with FITC-conjugated anti-rabbit antibody (111-095-003, Dianova, USA) which was diluted 1:150 in blocking solution containing 1% goat serum and incubated over night. Cells were washed three times with 0.1% PBST and treated for 1 min with 0.1 μg/ml DAPI solution. After a second washing step, cells were embedded in 0.1% DABCO/Mowiol solution. All pictures were taken with the Axio Observer Z1 (ZEISS), the Digital CCD Camera C4742-80-12AG (Hamamatsu ORCA-ER) and the Volocity software package.

Microarray analysis

Gene expression analysis was performed using the Affymetrix HG-U133® Plus 2.0 oligonucleotide microarrays. Two μg of purified RNA from transfected cell lines expressing either GFP (mock-control) or the variants NANOG1, NANOG2D2 and NANOG2E were converted by reverse transcription into double-stranded cDNA (Roche Applied Science) and then purified using the GeneChip Sample Cleanup module (Affymetrix). Then, labeled cRNA was generated using the Microarray RNA target synthesis kit (Roche Applied Science) and an in vitro transcription labeling nucleotide mixture (Affymetrix). The cRNA was then purified using the GeneChip Sample Cleanup module (Affymetrix) and quantified using the NanoDrop spectrophotometer. Eleven micrograms of labeled cRNA were fragmented. Hybridization, washing, staining and scanning protocols were performed on Affymetrix GeneChip instruments (Hybridization Oven 640, Fluidics Station 450Dx, Scanner GCS3000Dx, respectively), following the manufacturer’s instructions. The obtained data were stringently screened by comparing against mock-control, by their *P*-values, log2-fold changes and signal intensities to obtain only very robust data sets. Presented genes were at least 4-fold deregulated.
RESULTS

Transcriptional properties of NANO G gene family members in MLL-rearranged and control cell lines

Transcripts of the NANO G gene/pseudogene family were monitored by RT–PCR experiments using different primer combinations. As shown in Figure 1A, three different primer sets (a–c) were used to analyze cDNAs obtained from four different MLL-rearranged cell lines (t(4;11) cell lines: SEM, RS4;11; the t(9;11) cell line NOMO1; the t(11;19) cell line KOPN8. The NTERA-2 (human embryonic carcinoma) cell line served as positive control. By using the exon-specific primer sets a and b, PCR amplimers were obtained for all investigated cell lines (Figure 1B, panels a and b). The forward primer of set c binds to exclusively the 5'-UTR of the NANO G1 gene, demonstrating that NANO G1 transcripts were not present in the investigated MLL-rearranged cell lines (Figure 1B, panel c). Thus, we concluded that (i) NANO G1 transcripts can be distinguished from all other family members, and (ii) that all investigated cell lines are only able to transcribe RNA species deriving from NANO G2 or other NANO G pseudogenes.

Subsequent sequence analysis of 60 individual clones that derived from NANO G amplimers of the SEM cell line revealed a first glimpse on the spectrum of activated NANO G genes/pseudogenes. Comparison to available NANO G sequences (NCBI database) demonstrated the following genes/pseudogenes to be transcribed in this MLL-rearranged leukemia cell line: NANO G2 (27%), NANO GP2 (2%), NANO GP4 (55%) and NANO GP5 (16%).

Since Booth and Holland reported for the NANO G2 gene an unusual splice event involving a cryptic exon located about 20 000-bp upstream of the NANO G2 gene, we analyzed this possibility by using specific primers (set d) in an RT–PCR experiment. As shown in Figure 1C, neither NTERA-2 cells nor the two analyzed t(4;11) cell lines SEM and RS4;11 seem to use this upstream exon that is located in intron 1 of the SLC2A14 gene (9), transcribing in the opposite direction relative to the orientation of the NANO G2 gene.

Since NANO GP8 transcripts were not identified in the 60 clones that derived from SEM cells, we wanted to rule out that NANO GP8 is indeed not transcribed in the investigated cell lines. Corresponding RT–PCR

Figure 1. NANO G transcription in MLL-rearranged and Teratocarcinoma cell lines. (A) Gene structures of NANO G1 and NANO G2. Primer sets a–c were indicated for both genes. Forward primer of primer set c binds specifically to the 5'-NTR of NANO G1, and thus, enable to monitor transcripts that derive only from the NANO G1 gene. (B) RT–PCR analysis of four different MLL-rearranged cell lines (SEM, RS4;11, NOMO1, KOPN8). The NTERA-2 cell line represents a human embryonic carcinoma cell line and served as positive control for all experiments. M: DNA size marker. N: water control. (C) Validation of transcripts starting in the upstream located SLC2A14 gene. M: DNA size marker. N: water control. (D) Validation of NANO GP8 transcription. Specific oligonucleotides that bind only to the NANO GP8 open reading frame, we were not able to detect NANO GP8 transcripts in the investigated cell lines. M: DNA size marker. N: water control. N-P8, N2, N-P5: cloned NANO GP8, NANO G2 and NANO GP5 cDNAs served as positive controls.
experiments were conducted with cDNA from NTERA-2 cells and both t(4;11) cell lines. Cloned NANO2, NANO2P8 and NANO2P5 cDNAs were used as positive control and to demonstrate specificity of the applied primer sequences (set e). As shown in Figure 1D, the investigated cell lines does not seem to transcribe NANO2P8.

Based on these initial experiments and on the fact, that only NANO1, NANO2 and the pseudogene NANO2P8 are per se able to be translated into NANO proteins, we concluded that NTERA-2, the t(4;11) cell lines and t(4;11) patient cells most likely use NANO1 or NANO2 protein to activate the down-stream core NANO network.

**RACE experiments revealed an extended gene structure for NANO1 and NANO2**

Based on the obtained experimental results, 5’-RACE experiments were conducted with cDNA obtained from NTERA-2 and the SEM cell line by using an anchored oligonucleotide in the final exon of the NANO1/2 gene. As shown in Figure 2A, several but different PCR bands were obtained for both investigated cell lines. Therefore, all RACE–PCR amplifiers were cut out from the gel and cloned. Subsequent sequence analyses revealed the presence of NANO1 and NANO2 transcripts in NTERA-2 cells, and NANO2 transcripts in SEM cells.

The large number of analyzed clones allowed us to identify complex splice patterns in transcripts deriving from both NANO genes. In addition, the presence of novel 5’-exons for both NANO genes revealed extended gene structures. As shown in Figure 2B, the NANO1 gene exhibits three additional exons, named exon 1a, 1b/1b* and 2, while the NANO2 gene exhibits an additional exon 1b/1b*. Surprisingly, exon 2 seems to be absent, however, a careful analysis of genomic sequences revealed that there exists an NANO2 exon 2 remnant, missing most of its 3’-portion due to an ALU integration event. Based on these findings, we propose an revised exon/intron structure for both NANO genes which is depicted in Figure 2B. The yet known gene structures for both genes

---

**Figure 2.** RACE experiments reveal splice variations of NANO1 and NANO2 and novel upstream exons. (A) 5’-RACE experiment. NTERA-2 and SEM cDNA was analyzed by using an anchored primer in the final NANO exon. I–IV: bands cut out from the gel, cloned and sequenced. M: 50-bp ladder; N: water control. (B) Extended gene structure for NANO1 and NANO2. Homologous regions between both genes are depicted by grey areas. Repetitive DNA elements are indicated. Potentially transcriptional active ALU-elements are shown in light grey, while promoter-mutated ALU-elements are shown in dark grey. MER: MER-repeat. Arrows: indicate transcription initiation sites. All exons are numbered according to the nomenclature used throughout the manuscript. (C) Splice variants of NANO1 and NANO2. All cloned RACE amplimers were sequenced and revealed the additional NANO1 exons 1a, 1b, 1b*, 2 and NANO2 exons 1b, 1b*. Open reading frame analyses (marked in grey) of all cloned splice variants revealed putative open reading frames (ORFs) coding for two NANO1 (splice variants A and B) and three NANO2 protein (splice variants D, E and F). The number of isolated cDNA clones is indicated. Length of ORFs, predicted molecular weight (MW) and putative protein structures are depicted for all NANO1/2 mRNA variants.
are indicated by dotted frames. It is important to note that an earlier report (9) already described exon 1b in transcripts deriving from the \textit{NANOG1} gene.

**Detailed overview about the novel \textit{NANOG1} and \textit{NANOG2} gene structures**

\textit{NANOG2} most likely represents a tandem duplication of the \textit{NANOG1} gene located at 12p13.31. Both genes are separated by about 80 kb, a genomic region coding for the \textit{SLC2A14} gene. Based on our results, the \textit{NANOG1} gene exhibits two alternatively used first exons (1a and 1b/1b*) that both splice to the consecutive exons 2–6. For the \textit{NANOG2} gene, we identified only transcripts containing exon 1b/1b* sequences, although the presence of a \textit{NANOG2} exon 1a could be identified in the genomic DNA.

Both genes are highly saturated with repetitive ALU elements. All ALU integration events are unique for both genes, indicating that both \textit{NANOG} gene copies diverged during the evolution. Although both \textit{NANOG} genes seem to be quite diverse, the regions upstream and surrounding \textit{NANOG1}/2 exons 1a/b are highly conserved (90 and 85% identity), similar to the region upstream of \textit{NANOG1}/2 exons 3 that display 90% identical nucleotides. All other intronic sequences are much less conserved and show very little homology. This indicated that the regions used as transcriptional start sites of both \textit{NANOG} genes have been conserved during evolution.

**Alternative splice variants of \textit{NANOG1} and \textit{NANOG2}**

A series of different splice variants for \textit{NANOG1} (n = 7; named A and B\textsubscript{n–e}) and \textit{NANOG2}-derived transcripts (n = 7; named D\textsubscript{1}, D\textsubscript{2}, D\textsubscript{2*}, E, and F) were cloned (Figure 2C). All these variant mRNAs can be potentially translated into 2 different \textit{NANOG1} (predicted MW: 32 and 35 kDa) and three different \textit{NANOG2} proteins (predicted MW: 19, 27 and 29 kDa). Splice variant A is derived from transcripts starting upstream of \textit{NANOG1} exon 3 and uses the first AUG start codon of exon 3; this transcript encodes an open reading frame with the potential to produce a protein with a predicted molecular weight (MW) of 35 kDa. Splice variants B come in many different flavors, because exons 1a, 1b or 1b* are used for splice processes to consecutive exons 2–6 of \textit{NANOG1}. Moreover, splice processes between exons 2 and 3 result in exon 2 sequences fused to nt +3, +6 or +17 of exon 3, thereby skipping 2, 5 or 16 nt. All these different mRNAs lead to shorter open reading frames with the potential to produce a protein with a predicted MW of 32 kDa. In a few of these \textit{NANOG1} transcripts, alternative splicing is leading to the skipping of the first 48 nt of exon 6. This results in the loss of 16 amino acids of the CD1 domain which has previously been described (13).

The \textit{NANOG2} gene gives rise to splice variants D–F. Splice variant D\textsubscript{1} is similar to splice variant A of the \textit{NANOG2} gene. However, due to several missense mutations, the first two AUG start codons in exon 3 are absent, and thus, this splice variant encodes a protein with a predicted MW of 29 kDa. Splice variant D\textsubscript{2} comprises transcripts that use either exon 1b or 1b*. Exon 1b is spliced to exon 3 by fusing exon 1b sequences again with nts +3, +6 or +17 of exon 3. Presumably due to the missing exon 2 in the \textit{NANOG2} gene structure, exon 1b* splice events fuse exon 1b* with nt +77 of exon 3; these different mRNAs all encode a \textit{NANOG2} protein variant with a predicted MW of 29 kDa. Splice variant E fuses exon 1b with exon 4, resulting in an open reading frame that can be translated into a 27 kDa protein. Finally, splice variant F is nearly identical to splice variant D2, however, exon 3 is fused to nt +39 of exon 4. This results in another open reading frame that potentially encodes a protein of 19 kDa. Moreover, all cloned \textit{NANOG2} splice variants exhibit the skipping of the first 48 nt of exon 6. The number of all cloned splice variants, the length of the potentially encoded open reading frames, the predicted MW and the protein domain structure of both \textit{NANOG} genes are depicted in Figure 2C. All primary sequences and a compilation of all identified splice sites can be retrieved from the Supplementary Data (Supplementary Figures S1 and S2).

**Western blot experiments revealed different \textit{NANOG} protein variants**

Western blot experiments with soluble cell lysates (S) of NTERA-2 cells revealed the expression of several protein bands. NTERA-2 cells transcribe both the \textit{NANOG1} and \textit{NANOG2} gene in a large variety of different splice variants. This results in weaker and stronger visible protein bands in the western blot experiments. Stronger protein bands seem to have a MW of 48, 35 and 29 kDa, respectively (see Figure 3A, left panel). SEM cells, shown to transcribe only the \textit{NANOG2} gene, expressed a strong 48 kDa and a weak 29 kDa protein variant (see Figure 3A, right panel).

In order to validate that the predicted MW of the different splice variants of both genes and migration behavior in SDS-PAGE are similar, we performed recombinant protein expression in 293T cells. Several splice variants were cloned into the pEXPR-IBA10-3 vector system (A, B\textsubscript{5}, D\textsubscript{1}, D\textsubscript{2}, and E). This vector fuses a 4 kDa Strep-Tag to the C-terminus of all open reading frames. After transient transfection into 293T cells, all NANOG variants were affinity-purified and analyzed by western blot experiments. As shown in Figure 3B, two different \textit{NANOG1} and \textit{NANOG2} protein variants were successfully expressed in mammalian 293T cells. The \textit{NANOG1}B\textsubscript{5}, \textit{NANOG2}D\textsubscript{2} and \textit{NANOG2}E protein variants migrated with their expected molecular weight (32+4 = 36, 29+4 = 33 and 27+4 = 31 kDa, respectively). By contrast, the \textit{NANOG1}A variant migrated at ~50 kDa. This indicated that the \textit{NANOG1}A protein seems to be subjected to post-translational modifications (PTM). The recombinant \textit{NANOG1}A protein was subjected to mass spectrometry and verified to be the Strep-tagged \textit{NANOG1} protein (Supplementary Figure S3). However, all our attempts to uncover the potential PTM failed so far. Therefore, we have no explanation for the observed shift of about +11 kDa (35+4 = 39) in SDS-PAGE.
Subcellular localization of NANOG1/2 protein variants

To further investigate the cellular distribution of different NANOG1/2 protein variants, we performed immunohistological experiments. As shown in Figure 3C, expression and cellular localization of the recombinant NANOG1/2 protein variants was monitored by using an antiserum raised against the C-terminal portion of NANOG in combination with DAPI staining. After transient transfection of all constructs into HeLa cells, the NANOG1/2 protein variants localized in the nucleus. Thus, it can be concluded that all different NANOG1/2 protein variants are per se able to translocate into the nucleus, where these protein variants are able to provide their specific function(s).

Establishment of NANO1- and NANO2-specific RT-PCR conditions

Since the novel gene structures of NANO1 and NANO2 exhibit additional exons that are absent in all known NANO pseudogenes, they provide a perfect source to establish NANO1- and NANO2-specific PCR reactions. This may help to experimentally distinguish between bona fide NANO gene transcripts and non-functional transcripts deriving from all other NANO pseudogenes. In Figure 4 the first exons of the NANO1 and NANO2 gene are depicted. By using the specific primers A and C, we were able to establish specific conditions that specifically amplify only transcripts.
deriving from both NANOGL genes. Since the NANOGL gene encodes the additional exon 2, the resulting two different PCR amplifiers indicate transcription of NANOGL (356 bp) and NANOGL2 (195 bp). In case of using the exons 1b* splice sites, the NANOGL1-specific transcript will result in a 770-bp amplifier, while the NANOGL2-specific transcript will result in a 314-bp amplifier. All experiments were controlled by using total RNA isolated from NANOGL1Bb- and NANOGL2D2c-transfected 293T cells (Figure 4, right upper panel). The results of these experiments clearly revealed that NTERA2 cells transcribe both NANOGL genes, while SEM cells exclusively transcribe the NANOGL2 gene. The combination of primers B and C verified these results, since this PCR detected only transcripts deriving from the NANOGL1 gene (274 bp amplifier). Thus, specific conditions were established that allow to distinguish between transcripts deriving from NANOGL1 alone (B–C), NANOGL1 and NANOGL2 (A–C) from all other pseudogene transcripts of the NANOGL gene family (Figure 1B, primer

Figure 4. NANOGL1/2-specific RT–PCR experiments. (A) The two promoter regions of each NANOGL gene are highlighted. Oligonucleotides A–C are indicated. Right upper panel: RT–PCR experiments using the primer combination A/C which allowed to monitor NANOGL1- and NANOGL2-derived transcripts. Right lower panel: RT–PCR experiments using the primer combination B/C, which allowed to monitor only NANOGL1-derived transcripts. As internal control, cDNA from transfected cells were used (NANOGL1 Bb and NANOGL2 D2) and analyzed under identical conditions. N: water control. (B) Tissue cDNA panel. By using the primer combination A/C, a tissue cDNA panel was tested. With the exception of CD34+ cells, all other investigated tissue cDNAs remained negative. M: DNA size marker; neg: water control. Middle panel: Screening of t(4;11) patients, healthy volunteers and positive controls. Seven out of ten analyzed t(4;11) patients transcribed the NANOGL2 gene. Peripheral blood samples of healthy volunteers were negative. M: DNA size marker; neg: water control. Lower panel: Screening of AML patients with normal karyotype. Out of 20 samples, 16 were positive for NANOGL2 transcription, while 5 out of 20 samples were also positive for NANOGL1 transcripts. Only 17 out of 20 analyses are shown. M: DNA size marker.
set a or b). From these experiments, we concluded that MLL-rearranged cell lines are able to transcribe the NANOG2 gene.

By using the primer combination A–C, we investigated a tissue cDNA panel for NANOG transcripts. As shown in Figure 4B (upper panel), only CD34+ cells seem to express a 314-bp amplifier. This derives from the NANOG2 gene by using the exon 1b* splice site (Figure 4A). Thus CD34+ cells transcribe the NANOG2 but not the NANOG1 gene. All other tissues remained negative, as well as the investigated healthy individuals (n = 3). The same experiment was performed with leukemia biopsy material of t(4;11) patients and AML patients with normal karyotype (Figure 4B, middle and lower panels). These experiments revealed that leukemia patients weakly express predominantly NANOG2 (1b = 195 and 1b* = 314 bp), while some AML samples were also weakly transcribing NANOG1 (1b = 356 bp). Several bands were cut out from the gel and subjected to DNA sequencing analysis. In all cases, the appropriate NANOG1 or NANOG2 transcript could be confirmed. The weak amplimer production can be explained by two possibilities: (i) all cells produce only very few transcripts, or, (ii) only a minor fraction of the analyzed leukemia cells transcribe the NANOG genes. Based on immunohistochemistry experiments (Supplementary Figure S4), we assume that only a minor fraction of cells is expressing the NANOG or OCT4 protein. We also conducted QPCR experiments to analyze the absolute amount of transcripts produced by the two alternative start sites in the NANOG1 gene. For this purpose, PCR experiments were performed with total RNA isolated from NETRA-2 cells. The amount of transcripts between exon 1b/2 and exon 3/4 were measured. For quantification, we used a log-diluted NANOG1Bb expression plasmid (1–106 copies; Supplementary Figure S5). This allowed us to calculate the initiation events of transcripts starting upstream of exon 1b (100 000 transcripts in cDNA derived from 100 ng total RNA; assuming 15 fg total RNA per cell = 1 transcript in 600 cells) and exon 3 (490 000 transcripts in cDNA derived from 100 ng total RNA; assuming 15 fg total RNA per cell = 1 transcript in 12 cells). Thus, the transcriptional start site upstream of NANOG1 exon 3 is about 50-fold more used than the start site upstream of exon 1b, assuming that both NANOG genes are equally transcribed. QPCR experiments could not be performed for the NANOG2 gene, because of the multitude of different-sized PCR bands in splice variants D2, D2*, E and F did not allow reliable quantification.

Gene expression studies

To understand the biological significance of the different NANOG1/2 protein variants, we used an episomal vector system (pEPI) to express three different NANOG protein variants (1A, 2D2 and 2E) in 293 cells. The pEPI-EGFP vector encodes an additional EGFP reporter gene and the neomycin resistance gene. An empty pEPI-EGFP vector expressing only the EGFP protein served as negative control and was kept under identical conditions. After 69 days of cell culture, about 60% of the different cell populations were green fluorescent and total RNA was prepared. These RNA samples were used to perform gene expression profiling experiments using available Affymetrix chips (Figure 5A). Subsequent analysis revealed a specific activation of very few genes in all three transfected cell populations. All investigated NANOG1/2 protein variants strongly induced FOS gene transcription (NANOG1A: 39-fold; NANOG2D2: 30-fold; NANOG2E: 30-fold). The FOS protein is able to bind directly to several other transcription factors or nuclear complexes (ATF2, BCL3, COBRA1, CREBBP, CSNK2A1, CSNK2A2, DDIT3, ETS2, GTF2E2, GTF2F1, JUN, MITF, NACA, NCOA1, NCOA6, NCOR2, PML, RPS6KA4, RUNX1, RELA, RUNX2, SMAD3, SPI1, TAF1, TBP, TCF1, TSC2D3, VDR, XBP1). The second gene that was transcriptionally activated was the EGR1 gene (5- to 6-fold). This zinc finger transcription factor directly interacts with several other nuclear factors (CEBPB, CREBBP, EP300, ERBB3, NAB1, NAB2, NFATC1, PITX1, PSMA3, RELA, TP53). Most importantly, EGR1 is able to transcriptionally upregulate the CDKN1A/p21 gene. The activation of the CDKN1A/p21 gene was recently shown to be a key step for cancer stem cell quiescence and maintenance (14). Thus, ectopic activation of different NANOG protein (variants) in non-ES cells may allow to activate a specific genetic program (Figure 5B). Based on recent findings, the transcriptional activation of p21 is p33-dependent and one of the key features for tumor stem cells to acquire stem cell maintenance and drug resistance.

DISCUSSION

NANOG is a homeodomain transcription factor that is—in conjunction with OCT4 and SOX2—responsible to establish a genetic circuit to maintain the stem cell compartment at the blastocyst stage of developing embryos (3,4). This important concept has recently been verified and extended by the genetic manipulation of differentiated cells with four different transgenes (Klf4, Oct4, Sox2 and c-Myc). Transient expression of these proteins led to the activation of the NANOG gene, and subsequently to a reprogramming of the manipulated cells into induced pluripotent stem cells (iPS) (15,16).

Many reports have demonstrated transcription of ‘NANOG’ in germ cells (17,18) and other human cancer tissues (11,12). In most of these reports, the authors were only able to distinguish between transcripts deriving from NANOG1, NANOG2 (10) or NANOG pseudogenes by cloning and sequencing the obtained PCR amplimers. However, only NANOG1-, NANOG2- and NANOGP8-derived transcripts are per se able to be translated into protein, while all other pseudogenes do not exhibit intact open reading frames (9).

This study tried to answer the question whether human cancer cells express either NANOG1 or the functional equivalents NANOG2 and NANOGP8, respectively. Expression of functional NANOG proteins may help to
establish a ‘stem cell-like’ program which in turn could be a key event to establish and maintain cancer stem cells. To answer this important question and to validate our earlier finding, namely that t(4;11) leukemia cells transcriptionally upregulate \textit{NANOG} transcripts (8), we started to investigate the transcriptional properties of \textit{NANOG} genes and pseudogenes in \textit{MLL}-rearranged cell lines.

First, the analysis of transcriptional properties of \textit{NANOG1} and \textit{NANOG2} in NTERA-2 and \textit{MLL}-rearranged cell lines led to the discovery of novel 5'-exons which reflect on more extended gene structures for \textit{NANOG1} and \textit{NANOG2}, respectively (Figure 2B). Several splice variants of both \textit{NANOG} genes were cloned which putatively encode different \textit{NANOG1}/2 protein variants (Figure 2C). The different \textit{NANOG} protein variants display a shorter N-terminal domain (ND), without affecting the important homeobox domain (HD). The ND was shown to be dispensable for the function of human \textit{NANOG} protein, because the transactivation domain is located in the C-terminal domain CD of human \textit{NANOG} protein (19). Only the \textit{NANOG2}F variant displayed a partial loss of the HD domain. Thus, all other \textit{NANOG1}/2 protein variants seem to contain all necessary domains to exhibit similar functions. Noteworthy, the existence of exon 1b as part of the \textit{NANOG1} gene has been described earlier (9), however, the authors did not further investigated their own findings.

NTERA-2 cells transcribe several splice variants deriving from the \textit{NANOG1} and \textit{NANOG2} gene. By using an antiserum specific for the C-terminal portion of \textit{NANOG} protein, different \textit{NANOG1} and \textit{NANOG2} protein variants were detected in western blot experiments. By contrast, the investigated t(4;11) cell line SEM predominantly expresses a 48 kDa protein. By using recombinant expression in HeLa cells, we could demonstrate that all tested \textit{NANOG1}/2 protein variants translocated into the nucleus. This may indicate that all

Figure 5. Gene expression profiling experiment. (A) The cell line 293 was transfected with episomal pEPI-EGFP vector containing expression cassettes for NANOG1A, NANOG2D2c or NANOG2E. After 10 weeks of culture, GFP-positive cells were sorted and used for RNA isolation. These RNA samples were used for Affymetrix chip experiments, resulting in few deregulated genes shown in the right. White boxed areas: up-regulated genes. Light grey boxed areas: down-regulated genes. Common genes identified in all three experiments using the different NANOG1/2 protein variants are highlighted. (B) Mode of action for the investigated \textit{NANOG} protein variants. \textit{NANOG} protein is able to induce transcription of \textit{EGR1} (directly or indirectly). \textit{EGR1} protein is directly binding to the promoter regions of \textit{FOS} and \textit{p21}. P21 protein is necessary for stem cell maintenance, while the FOS transcription factor is necessary for cell proliferation.
these different NANOG protein variants are able to function as transcription factors.

Based on the extended gene structures we established specific PCR conditions that allow to precisely distinguish between transcripts that derive from NANOG1 (B–C amplimer in Figure 4), NANOG2 (A–C amplimer in Figure 4) or all other NANOG pseudogene copies (primer sets a and b in Figure 1B). Thus, the analysis of transcriptional properties in stem cells, cancer cells or any other cell line will become more accurate and reliable in the future. Because of the importance of this finding, we validated independently the existence of these NANOG1/2 upstream transcripts by RNase protection experiments (Supplementary Figure S6). These experiments validated all prior experiments, thus demonstrating that the novel upstream NANOG1/2 exons are existing and no cloning artifact. Therefore, the established PCR assay could be reliably used as read-out system for the transcriptional activity of both stem cell genes without detecting any other pseudogene.

To investigate this a bit deeper, we analyzed the chromatin properties of both NANOG genes. This revealed open chromatin structures (H3K4me) in the promoter regions of both NANOG genes in NTERA2 cells, while the leukemia cell lines SEM and RS4;11 displayed open chromatin structures predominantly in the upstream region (region I and II) of the NANOG2 gene (Supplementary Figure S7). The existence of these novel upstream promoter regions was further validated by Luciferase reporter gene assays. These experiments demonstrated that both DNA sequences located directly 5′ to exon 1b function as promoter elements after transient transfection into NTERA2 cells (Supplementary Figure S8).

For the first time, the NANOG2 gene was shown to be transcribed in CD34+ cells (Figure 4B). This may indicate that the hematopoietic stem compartment may use the NANOG system to gain stem cell like properties. The hematopoietic stem compartment may use the NANOG system to gain stem cell like properties. The hematopoietic stem compartment may use the NANOG system to gain stem cell like properties. The same is true for most investigated leukemia samples that predominantly produced NANOG2D transcripts. We verified this finding by performing western blot experiments with biopsy material of acute myeloid leukemia patients (n = 10). We were able to validate NANOG protein expression in 4 out of 10 investigated leukemia patients (Supplementary Figure S9).

We further investigated the functional consequences of expressed NANOG1/2 proteins. For this purpose we used the NANOG1A, NANOG2D and NANNOG2E variants. These three NANOG variants were expressed for 69 days and remaining green fluorescent cells were used for RNA extraction. After subtracting against GFP-expressing mock-transfected cells that were maintained in parallel, the specific expression profiles were obtained. Surprisingly, only very few genes were transcriptionally deregulated. Since we expressed the NANOG variants in 293 cells, we did not see an overlap with the known core NANOG network (3,4). However, the genes FOS and EGR1 were transcriptionally activated. EGR1 directly binds to the promoter region of the FOS and CDKN1A/p21 gene (20), and serves as gatekeeper for p53 expression (21). Beside the different functions in ES cells, ectopic NANOG protein expression may exhibit additional and yet unknown functions in non-ES cells that need to be investigated further.

In conclusion, we have identified additional exons reflecting on an extended gene structure of NANOG1 and NANOG2. Moreover, we have revealed complex splice patterns for both NANOG genes that are putatively translated in a large variety of different NANOG1/2 protein variants. We have also demonstrated that the NANOG2 gene is transcribed in CD34+ and different leukemia cells. Western blot experiments performed with t(4;11) cell lines demonstrated that the NANOG2 protein is produced, potentially helping to maintain stem cell functions. We also demonstrated functional equivalence for NANOG1 and NANOG2 protein variants in terms of their downstream target genes. Further studies are ongoing, aiming to understand the molecular consequences of NANOG2 expression in leukemic cells with MLL rearrangements.

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS
The authors thank Jennifer Merkens and Silvia Bracharz for their technical assistance and Torsten Pietsch for performing the immunohistochemistry experiments with SEM and RS4;11 cells. The authors are thankful to Thomas Burmeister, Lars Bullinger, Gesine Bug and Nicola Gökbüget who provided t(4;11) and AML biopsy samples, respectively.

FUNDING
Funding for open access charge: BMBF (grant 01GS0875 to R.M.); Center of Excellence Frankfurt on Macromolecular Complexes (CEF-MC) funded by DFG (grant EXC 115 to R.M.).

Conflict of interest statement. None declared.

REFERENCES
1. Mitsui,K., Tokuzawa,Y., Itoh,H., Segawa,K., Murakami,M., Takahashi,K., Maruyama,M., Maeda,M. and Yamanaka,S. (2003) The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell, 113, 631–642.
2. Chambers,I., Colby,D., Robertson,M., Nichols,J., Lee,S., Tweedie,S. and Smith,A. (2003) Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell, 113, 643–655.
3. Boyer,L.A., Lee,T.I., Cole,M.F., Johnstone,S.E., Levine,S.S., Zucker,J.P., Guenther,M.G., Kumar,R.M., Murray,H.L., Jenner,R.G. et al. (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. Cell, 122, 947–956.
4. Loh,Y.H., Wu,Q., Chew,J.L., Vega,V.B., Zhang,W., Chen,X., Bourque,G., George,J., Leong,B., Liu,J. et al. (2006) The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat. Genet., 38, 431–440.
5. Lee,T.I., Jenner,R.G., Boyer,L.A., Guenther,M.G., Levine,S.S., Kumar,R.M., Chevalier,B., Johnstone,S.E., Cole,M.F., Isono,K,
et al. (2006) Control of developmental regulators by Polycomb in human embryonic stem cells. Cell, 125, 301–333.
6. Udomsakdi,C., Eaves,C.J., Swolin,B., Reid,D.S., Barnett,M.J. and Eaves,A.C. (1992) Rapid decline of chronic myeloid leukemic cells in long-term culture due to a defect at the leukemic stem cell level. Proc. Natl Acad. Sci. USA, 89, 6192–6196.
7. Chumsri,S., Matsui,W. and Burger,A.M. (2006) Genetic implications of leukemic stem cell pathways. Clin. Cancer Res., 13, 6549–6554.
8. Gaußmann,A., Wenger,T., Eberle,J., Bursen,A., Bracharz,S., Herrl., Dingermann,T. and Marschalek,R. (2007) Combined effects of the two reciprocal t(4;11) fusion proteins MLL-AF4 and AF4-MLL confer resistance to apoptosis, cell cycling capacity and growth transformation. Oncogene, 26, 3352–3363.
9. Booth,H.A. and Holland,P.W. (2004) Eleven daughters of NANOG. Genomics, 84, 229–238.
10. Fairbanks,D.J. and Maughan,P.J. (2006) Evolution of the NANOG pseudogene family in the human and chimpanzee genomes. BMC Evol. Biol., 6, 12.
11. Ezeh,U.I., Turek,P.J., Reije,R.A. and Clark,A.T. (2005) Human embryonic stem cell genes OCT4, NANOG, STELLAR, and GDF3 are expressed in both seminoma and breast carcinoma. Cancer, 104, 2255–2265.
12. Zhang,J., Wang,X., Li,M., Han,J., Chen,B., Wang,B. and Dai,J. (2006) NANOGP8 is a retrogene expressed in cancers. FEBS J., 273, 1723–1730.
13. Kim,J.S., Kim,J., Kim,B.S., Chung,H.Y., Lee,Y.Y., Park,C.S., Lee,Y.S., Lee,Y.H. and Chung,I.Y. (2005) Identification and functional characterization of an alternative splice variant within the fourth exon of human nanog. Exp. Mol. Med., 37, 601–607.
14. Viale,A., De Franco,F., Or leth,A., Cammachi,V., Giuliani,V., Bossi,D., Ronchini,C., Ronzoni,S., Muradore,I., Monestiroli,S. et al. (2009) Cell-cycle restriction limits DNA damage and maintains self-renewal of leukaemia stem cells. Nature, 457, 51–56.
15. Takahashi,K. and Yamanaka,S. (2006) Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell, 126, 663–676.
16. Wernig,M., Meissner,A., Foreman,R., Brambrink,T., Ku,M., Hochedlinger,K., Bernstein,B.E. and Jaenisch,R. (2007) In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature, 448, 318–324.
17. Hart,A.H., Hartley,L., Parker,K., Ibrahim,M., Looijenga,L.H., Pauchnik,M., Chow,C.W. and Robb,L. (2005) The pluripotency homeobox gene NANOG is expressed in human germ cell tumors. Cancer, 104, 2092–2098.
18. Hoei-Hansen,C.E., Almstrup,K., Nielsen,J.E., Brask Sonne,S., Graeø,N., Skakkebaek,N.E., Leffers,H. and Rajpert-De Meyts,E. (2005) Stem cell pluripotency factor NANOG is expressed in human fetal gonocytes, testicular carcinoma in situ and germ cell tumours. Histopathology, 47, 48–56.
19. Oh,J.H., Do,H.J., Yang,H.M., Moon,S.Y., Cha,K.Y., Chung,H.M. and Kim,J.H. (2005) Identification of a putative transactivation domain in human Nanog. Exp. Mol. Med., 37, 250–254.
20. Ragione,F.D., Cucciolla,V., C riniti,V., Indaco,S., Borriello,A. and Zappia,P. (2003) p21Cip1 gene expression is modulated by Egr1: a novel regulatory mechanism involved in the resveratrol antiproliferative effect. J. Biol. Chem., 278, 23360–23368.
21. Krones-Herzig,A., Adamson,E. and Mercloa,D. (2003) Early growth response 1 protein, an upstream gatekeeper of the p53 tumor suppressor, controls replicative senescence. Proc. Natl Acad. Sci. USA, 100, 3233–3238.