Figure S1: Splice variants of the *NANOG1* and *NANOG2* gene

1. NANOG1 variants

NANOG1A (E3-E4-E5-E6):

\[
\text{ATG} \text{AGTGTGAGATCCGACTGGTGTTCCGCGAGAACGACTGTGGCTTTGAAGCAGATCCGACTGTAAAGATCCATGCCTATGCCTGTGATTTGTGGGCCTGAAGAAAACTATCCATCCTTGCAAATGTCTTCTGCTGAGATGCCTCACACGGAGACTG}
\]

NANOG1Ba (E1-E2-E3(+3)-E4-E5-E6):

\[
\text{GACACAATGGGACAGGGGAGCGGGGAGGATGGGGGAATTCAGCTCAGGCTTTTATGCAAAGACCCCCTTCTGCCAAAGAACAAAGCCTTCTTGGAAGACTGGGACCTGCTGCCGACACTTGCTGCAAATGTCTTCTGCTGAGATGCCTCACACGGAGACTG}
\]

NANOG1Bb (E1-E2-E3(+6)-E4-E5-E6):

\[
\text{GACACAATGGGACAGGGGAGCGGGGAGGATGGGGGAATTCAGCTCAGGCTTTTATGCAAAGACCCCCTTCTGCCAAAGAACAAAGCCTTCTTGGAAGACTGGGACCTGCTGCCGACACTTGCTGCAAATGTCTTCTGCTGAGATGCCTCACACGGAGACTG}
\]
GAAGGCGCTTTATGTGTAATACGAGCAACACTAGGTATTTTAGTATCTCCACAAACCATGGATTTATTCCTAAACTACTCCATGAACATGCAACCTGAAGACGTG

2. NANOG2 variants

NANOG2D1: (E3-E4-E5-E6)

AGTAGTGTGGAATCTCTCTCTCTCCATAAAGCCCTGCTTCTGCTCCAAAGACATCTGAATGTGAAGACTGTTCCCCAGCATCGGTCTACGTAATCTGCTGT
CTCTGCTCTCCTCCTTCCTGTTCTCCTACGCTTCTCCTCCGCTATCCTCCACGTCTATCCTGGAGCAACCTGGAGCTGTTGTAATGACCCACAGACCCCAT
GGGACAGCTGTGTGTACTCAATGATAGATTCTAGAGACAGAAATACCTCAGCCTCAGCAGATGCAAGAACTTTCCAACATCCTGAACCTCAGCTACAAGAC
GTGAAGACCTGGTTCCAGAACCAGAGAATGAAATCTAAGAGGTGGCAGAAAAACAACTGGCTGAAGAATAGCAATGGTGTGACGCAG

GACGCAATGGGACAGGGAGCGGGGGATGGGGGAATTCAGCTCAGGCTTTTATGCAAAGTCCCCCTTCAGCAAAGAACAAAGCTTCTGGTACCTGCCCTTTGG
ACAGCTGCGGGCAAGCTCAGCCTCGAG

NANOG2D2a: (E1-E3(+3)-E4-E5-E6(+48))

AGCCACACGTGAAGGGACGGCCCTTATGTGTAATACGAGCAACACTAGGTATTTTAGTATCTCCACAAACCATGGATTTATTCCTAAACTACTCCATGAACATGC

GACACATATGGGAGACGGCGCTTTGCACTCAACATGTTGTAATACGAGCAACACTAGGTATTTTAGTATCTCCACAAACCATGGATTTATTCCTAAACTACTCC
TCAACCATGGAGGCAGCTTCTGCAATGAAATAGCAATGGTGTGACGCAG

GACCTGGTGCACCCAATCCTGGAACAATCAGGCCTGGAACAGTCCCTTCTATAACTGTGGAGAGGAATCTCTGCAGTCCTGCAATGCAGTTCCAGCCAAATTCT
CCTGCCAGTGACTTGGAGGCTGCCTTGGAAGCTGCTGGGGAAGGCCTTAATGTAATACGAGCAACACTAGGTATTTGAGTATCTCCACAAACCATGGATTTATT
CCTAAACTACTCCATGAACATGCAACCTGAAGACGTG

GACACATATGGGAGACGGCGCTTTGCACTCAACATGTTGTAATACGAGCAACACTAGGTATTTTAGTATCTCCACAAACCATGGATTTATTCCTAAACTACTCC
TCAACCATGGAGGCAGCTTCTGCAATGAAATAGCAATGGTGTGACGCAG

GACCTGGTGCACCCAATCCTGGAACAATCAGGCCTGGAACAGTCCCTTCTATAACTGTGGAGAGGAATCTCTGCAGTCCTGCAATGCAGTTCCAGCCAAATTCT
CCTGCCAGTGACTTGGAGGCTGCCTTGGAAGCTGCTGGGGAAGGCCTTAATGTAATACGAGCAACACTAGGTATTTGAGTATCTCCACAAACCATGGATTTATT
CCTAAACTACTCCATGAACATGCAACCTGAAGACGTG

GACACATATGGGAGACGGCGCTTTGCACTCAACATGTTGTAATACGAGCAACACTAGGTATTTTAGTATCTCCACAAACCATGGATTTATTCCTAAACTACTCC
TCAACCATGGAGGCAGCTTCTGCAATGAAATAGCAATGGTGTGACGCAG

GACCTGGTGCACCCAATCCTGGAACAATCAGGCCTGGAACAGTCCCTTCTATAACTGTGGAGAGGAATCTCTGCAGTCCTGCAATGCAGTTCCAGCCAAATTCT
CCTGCCAGTGACTTGGAGGCTGCCTTGGAAGCTGCTGGGGAAGGCCTTAATGTAATACGAGCAACACTAGGTATTTGAGTATCTCCACAAACCATGGATTTATT
CCTAAACTACTCCATGAACATGCAACCTGAAGACGTG

GACACATATGGGAGACGGCGCTTTGCACTCAACATGTTGTAATACGAGCAACACTAGGTATTTTAGTATCTCCACAAACCATGGATTTATTCCTAAACTACTCC
TCAACCATGGAGGCAGCTTCTGCAATGAAATAGCAATGGTGTGACGCAG

GACCTGGTGCACCCAATCCTGGAACAATCAGGCCTGGAACAGTCCCTTCTATAACTGTGGAGAGGAATCTCTGCAGTCCTGCAATGCAGTTCCAGCCAAATTCT
CCTGCCAGTGACTTGGAGGCTGCCTTGGAAGCTGCTGGGGAAGGCCTTAATGTAATACGAGCAACACTAGGTATTTGAGTATCTCCACAAACCATGGATTTATT
CCTAAACTACTCCATGAACATGCAACCTGAAGACGTG

GACACATATGGGAGACGGCGCTTTGCACTCAACATGTTGTAATACGAGCAACACTAGGTATTTTAGTATCTCCACAAACCATGGATTTATTCCTAAACTACTCC
TCAACCATGGAGGCAGCTTCTGCAATGAAATAGCAATGGTGTGACGCAG

GACCTGGTGCACCCAATCCTGGAACAATCAGGCCTGGAACAGTCCCTTCTATAACTGTGGAGAGGAATCTCTGCAGTCCTGCAATGCAGTTCCAGCCAAATTCT
CCTGCCAGTGACTTGGAGGCTGCCTTGGAAGCTGCTGGGGAAGGCCTTAATGTAATACGAGCAACACTAGGTATTTGAGTATCTCCACAAACCATGGATTTATT
CCTAAACTACTCCATGAACATGCAACCTGAAGACGTG
All cloned and sequenced NANOG1 and NANOG2 transcript variants are summarized above. Exons are alternatively marked in red and in blue. Translation start and stop codons are indicated (underlined).
Figure S2: Identified splice sites in human *NANOG1* and *NANOG2*

1. NANOG1

Exon 1a intron Exon 2
GCGAAGAATGTA·gtagctggcct – cccttcccccag·AAACACACACAC

Exon 1b intron Exon 2
AGCTCAGCCTCG·gtgagtcttttgt – cccttcccccag·AAACACACACAC

Exon 1b* intron Exon 2
GTTTACTTTTCC·gtatggaagact – cccttcccccag·AAACACACACAC

Exon 2 intron Exon 3
AAGGAAACTAAG·gtaggtgctgaa – ctatactaacAT·GAGTGTTGGATC (+3)
– tataacATGAG·TGTTGATCCAGC (+6)
– GTGTGGATCCAG·CTTGTCCTACAA (+17)

Exon 3 intron Exon 4
ACACGGAGACTG·gtaagaagaa – tggcttccacag·TCTCTGCTCTCC

Exon 4 intron Exon 5
AGCTACAAACAG·gtaggtcttttt – ttttttctgctg·GTGAAGACTGG

Exon 5 intron Exon 6
GGTGTTGGCAG·gtacagtaac – cttctcttttag·AAGGCCTCAGCA
– TCCTCTCCTCTCC·GTGATCTCAGGT (+48)

2. NANOG2

Exon 1b intron Exon 3
AGCTCAGCCTCG·gtgagtcttttgt – taatgacATGAG·TGTTGATCCAGC (+6)
– GTGTGGATCCAG·CTTGTCCTACAA (+17)

Exon 1b*
CAGAGTTGGAGGT·ctgagaagaa – GGTCACCTATAC·CTGTGATTCTGT (+77)

Exon 3 intron Exon 4
ACACAGAGACTG·gtaagaagaaa – tggcttccacag·TCTCTGCTCTCC
– GTGCTATTACAG·GACAGCCATG (+39)

Exon 4 intron Exon 5
AGCTACAAACAG·gtaggtcttttt – tttttctgtgc·GTGAAGACTGG

Exon 5 intron Exon 6
GGTGTTGGCAG·gtacagtaac – cttctcttttag·AAGGCCTCAGCA
– TCCTCTCCTCTCC·GTGATCTCAGGT (+48)

Exon sequences are marked in bold letters and are shown for both the *NANOG1* and *NANOG2* gene. Intronic sequences are in lower case letters. Exonic and intronic sequences were separated by a dot. All splice sites of all identified splice variants were indicated.
Cytospinned SEM and RS4;11 cells were treated with HRP-conjugated antisera raised against NANOG, OCT4 and SOX2. Counterstainings were performed with a DAPI solution. Controls were treated with all solutions except the antisera. A few SEM cells seem to express NANOG, while OCT4 and SOX2 seem to be expressed in nearly all cells. RS4;11 cells seems to express NANOG and SOX2, while only few cells express OCT4.
Figure S5: QPCR-experiments to estimate the amount of *NANOG1* variants

Top: partial gene structure of *NANOG1* and *NANOG2* (exons 1-4). Nomenclature as in Figure 2 of the manuscript. QPCR-primer design: primer A binds to *NANOG1/2* exon 1b, primer B to *NANOG1* exon 2, primer C to *NANOG1/2* exon 3 and primer D to *NANOG1/2* exon 4. Lower left panel: Results of independent QPCR experiments made with reverse transcribed total RNA isolated from NTERA2 cells. The experiments were carried out as described in Material and methods. All experiments were carried out in parallel with log-diluted plasmid copies (1 – 10⁶ copies) that encode the *NANOG1Bb* splice variant. Lower right panel: Primers AB create an amplimer specific for the *NANOG1* gene, while primers CD are able to identify transcripts starting from both *NANOG* genes (4 different transcripts), we estimated roughly the relative amount of transcripts starting from *NANOG1* ex1b in relation to *NANOG1* ex3 about 1/50.

Q-PCR result:  
Primer A-B: ~10,000 copies  
Primer C-D: ~1,000,000 copies

Formula:  
A-B: transcripts starting from 
*NANOG1* ex 1b = a  
C-D: transcripts starting *NANOG1* 
ex 1b, *NANOG1* ex 3, *NANOG2* ex 1b and *NANOG2* ex 3 = 2a + 2b

Calculation:  
if a = 10,000 
and 2a + 2b = 1,000,000 
then b = 490,000
**Figure S6: RNase protection experiments to validate the existence of novel NANOG1/2 transcripts**

A. Outline of the performed RNase protection experiments. Radiolabeled antisense RNA probes were generated with the MAXIscript Kit (AMBION). For NANOG1 we used a cloned exon 2-3 fragment (Δ5 splice variant); for NANOG2 we used a cloned exon 1-3-4 fragment (Δ5 splice variant). Both fragments were cloned in pGEM-T plasmid (Promega); plasmid were digest and subjected to SP6 polymerase transcription in the presence of ³²P-UTP nucleotides. Radioactive labeled probes were separated on a 5% denaturing PAA gel. Gel slices were cut out and eluated over night at 37°C. Eluated RNA probes were then precipitated and dissolved into 40.000 cpmp/µl.
B. RNase protection experiment were carried out as recommendend by the manufacturer (AMBION). Briefly, we used 10 and 30 µg total RNA from NTERA2 cells. About 80,000 cpm were co-precipitated with these RNAs. Appropriate controls were performed by suing yeast RNA.

B. NTERA cells express both the NANOG1 and NANOG2 gene. Due to our NANOG1-probe, 3 different protected species were expected. A 277 nucleotide-long species proves that transcripts spanned exons 2 and 3, while the shorter 130 nucleotide-long species represents transcripts starting upstream of NANOG exon 3. The shortest 119 nucleotide-long species is indicative for a splice product from exon 2 to nucleotide +17 of exon 3. Thus, NANOG1 transcripts in NTERA2 cells are starting upstream of the 5' terminal nucleotide of the 277 protected fragment in exon 2 and predominantly splice to nucleotide +17 of exon 3.

Due to our NANOG2-probe, 5 different protected species were expected. The longest protected fragment is indicative for the presence of NANOG2 transcripts coding for exons 1b, 3 and 4. The 223 nucleotide-long fragment represents NANOG2 transcripts starting upstream of NANOG exon 3 and containing exon 4 sequences. The 212 nucleotide-long species represents again a splice variant from NANOG2 exon 1b to nucleotide +17 of exon 3. The 146 nucleotide-long species are again transcripts starting upstream of exon 3, but are alternatively splice to nucleotide +39 of exon 4. The shortest protected fragment with 135 nucleotides represents transcripts starting from exon 1b that alternative splice to nucleotide +17 of exon 3 and alternatively to nucleotide +39 of exon 4. Thus, transcripts starting upstream of NANOG exon 3 are predominantly used in NTERA2 cells.

In summary, this experiment validated independently the existence of NANOG1 exon 1b and 2 and NANOG2 exon 1b in transcripts deriving from both genes.
A. Tested fragments I-III for NANOG1 and NANOG2 in ChIP experiments.

B. ChIP experiments. Chromatin was cross-linked with 1% formaldehyde at room temperature for 10 min in PBS. Cells were then washed twice with PBS, collected in SDS Buffer (100 mM NaCl, 50 mM Tris pH 8.1, 5 mM EDTA pH 8.0, 0.5% SDS) and centrifuged for 6 min at 1,200 rpm. For sonication, cells were collected in IP Buffer (100 mM NaCl, 67 mM Tris pH 8.3, 5 mM EDTA pH 8.0, 1.7% Triton X-100, 0.3% SDS) and sonicated 4 x 20
sec with an amplitude of 30% (Branson Digital Sonifier W-250 D, Danbury, CT) followed by centrifugation for 50 min at 20,000 x g at 4°C. The chromatin was then pre-cleared with protein A/G-sepharose (25 µl of 50% slurry in TE buffer (Santa Cruz, CA), containing 2 µg of salmon sperm DNA (Trevigen USA), and 50 µg BSA) for 2-4 h at 4°C. Immunoprecipitations were performed overnight at 4°C with specific antibodies (Abcam, UK: α RNA polymerase II, α Histone H3-trimethyl K4, α Histone H3-trimethyl K9, α Sox2; Santa Cruz, USA: α Oct-3/4 C-20, α IgG). After immunoprecipitation, 20 µl protein A-Sepharose with 1,6 µg of salmon sperm DNA and 40 µg BSA were added and the incubation was continued for another 2-3 h. Precipitates were washed sequentially three times with wash buffer (WB) 1 (150 mM NaCl, 200 mM Tris pH 8.1, 50 mM EDTA pH 8.0, 5,2% sucrose, 1% Triton X-100, 0,2% SDS), two times with WB 2 (0,1% deoxycholic acid, 1 mM EDTA, 50 mM HEPES pH7,5, 500 mM NaCl, 1% Triton X-100), two times with WB 3 (0,5% deoxycholic acid, 1 mM EDTA, 250 mM LiCl, 0,5% NP-40, 10 mM Tris pH 8.0) and one time with TE Buffer. DNA was eluted by incubating the protein A-sepharose in 300 µl of 1% SDS, 0,1 M NaHCO₃ overnight at 65°C, shaking. Proteins were removed by incubation with 60 µg of Proteinase K for 60 min. DNA fragments were purified with a QIAquick Spin Kit (Qiagen, Germany). The recovered DNA was then analyzed by PCR. All PCR experiments were performed in 50 µl reactions with the following setting: initial denaturation with 2 min at 94°C, followed by 35 cycles with 30 sec at 94°C, 30 sec at 60°C and 30 sec at 72°C. A final elongation step (3 min at 72°C) was used for all reactions. The following oligonucleotides were used: NANOG1 fragment I (5'-cagggtaccatcgtactcaagtgttag-3'; 5'-cagaagcttgaaatgagctaacggct-3'); NANOG1 fragment II (5'-cagggtaccgtggccacggttag-3'; 5'-cagaagctctctgcatataaagctgtag-3'); NANOG1 fragment III (5'-cagggtactctccatcactctgtag-3'; 5'-cagaagctctggtgatctcactatgtgtag-3'); NANOG2 fragment I (5'-cagggtaccatcgtactcaagtgttag-3'; 5'-cagaagcttgaaatgagctaacggct-3'); NANOG2 fragment II (5'-cagggtaccatcgtactcaagtgttag-3'; 5'-cagaagctctgcatataaagctgtag-3'); NANOG2 fragment III (5'-cagggtaccgtggtgatctcactatgtgtag-3'; 5'-cagaagctcctctgcatataaagctgtag-3').
A. Tested NANOG1/2 DNA fragments. Different reporter plasmids containing the NANOG1-I to III and NANOG2-I to III regions were cloned into the pGL3-Luciferase plasmid. The promoter-less pGL3-Basic vector was used as negative control.

B. Luciferase reporter assays. All constructs were transiently co-transfected in $1 \times 10^5$ NTERA2 cells by using Lipofectamin-transfection. NTERA-2 cells were electroporated with 1 µg of each of the pGL-3 constructs together with 25 ng pGL3-Renilla plasmid (internal control). All experiments were performed independently 3 times in triplicates and all measurements were made 24 h after transfection.

These experiments revealed that the presence of an upstream promoter element for both tested NANOG genes.
Western blot experiments performed with leukemia patient material

Western blot experiment using patient biopsy samples from individual AML patients (2-5 x 10^6 cells), along with soluble lysates prepared from NTERA2 and SEM cells. Blots were stained with the C-terminal antibody against NANOG. Only 4 out of 10 investigated leukemia samples had enough cells to perform this experiment. The displayed patients all express NANOG protein, most likely the NANOG2 protein.