Alternative Splicing Produces Nanog Protein Variants with Different Capacities for Self-renewal and Pluripotency in Embryonic Stem Cells

Satyabrata Das, Snehalata Jena, and Dana N. Levasseur

**Background:** Nanog is a core factor that is required for the maintenance of embryonic stem (ES) cell pluripotency and self-renewal.

**Results:** Alternative splicing results in Nanog proteins with different capacities for maintaining the undifferentiated ES cell state.

**Conclusion:** The Nanog N-terminal domain is regulated by post-transcriptional modification.

**Significance:** Nanog protein variants either fully support the maintenance of pluripotency or facilitate differentiation.

Embryonic stem (ES) cells are distinguished by their ability to undergo unlimited self-renewal although retaining pluripotency, the capacity to specify cells of all germ layers. Alternative splicing contributes to these biological processes by vastly increasing the protein coding repertoire, enabling genes to code for novel variants that may confer different biological functions. The homeodomain transcription factor Nanog acts collaboratively with core factors Oct4 and Sox2 to govern the maintenance of pluripotency. We have discovered that Nanog is regulated by alternative splicing. Two novel exons and six exons have been identified that extend the known Nanog gene structure and protein coding capacity. Alternative splicing results in two novel Nanog protein variants with attenuated capacities for self-renewal and pluripotency in ES cells. Our previous results have implicated the C-terminal domain, including the tryptophan-rich (WR) domain of Nanog, to be important for the function of Nanog (Wang, J., Levasseur, D. N., and Orkin, S. H. (2008) Proc. Natl. Acad. Sci. U.S.A. 105, 6326–6331). Using point mutation analyses, serine 2 (Ser-2) of Nanog has been identified as critical for ES cell self-renewal and for stabilizing a pluripotent gene signature. An inducible conditional knock-out was created to test the ability of new Nanog variants to genetically complement Nanog null ES cells. These results reveal for the first time an expanded Nanog protein coding capacity. We further reveal that a short region of the N-terminal domain and a single phosphorylatable Ser-2 is essential for the maintenance of self-renewal and pluripotency, demonstrating that this region of the protein is highly regulated.

Embryonic stem (ES) cells, derived from the inner cell mass of the early embryo, are a vital tool for studying early developmental processes and cell-fate decisions. They have the ability to propagate indefinitely (self-renewal) in an undifferentiated state and have the potential to specify cell types of all three germ layers (pluripotency) (1, 2). Pluripotent cells therefore hold great promise for future cell replacement therapies across a broad range of diseases. Recent findings show that transcription factors form the core regulatory machinery components involved in gene expression maintenance and epigenetic regulation of pluripotency (3–9). Core pluripotency transcription factors that include Nanog (10, 11), Oct4 (12), and Sox2 (13) collaboratively form a strong self-reinforcing regulatory network that serves to govern the stable expression of self-renewal factors and repression of genes that promote differentiation. The generation of induced pluripotent stem cells from somatic cells using various mixtures of pluripotency factors like Oct4, Sox2, and Nanog coupled with accessory components such as c-Myc and Lin28 (14, 15) has created promise for the eventual use of induced pluripotent stem cells in cell therapy (16).

Alternative splicing and alternate promoter selection can significantly increase proteomic diversity in metazoans and higher eukaryotes by vastly increasing the number of distinct mRNAs produced from a single gene locus. ES cells express a particularly large diversity of splice isoforms and thus alternative splicing and alternate promoter selection likely impact lineage commitment and differentiation (17–21). The larger and more complex transcript pool in ES cells may reinforce pluripotency and self-renewal by altering protein interaction networks through the addition or removal of protein domains, or by changing the subcellular localization. Removal or inclusion of protein coding subdomains might also alter available sequences for repression by microRNAs because these small RNAs can also target regions outside of the 3’ untranslated region (22).
The homeobox protein, Nanog, was isolated as a protein capable of conferring LIF (leukemia inhibitory factor)-independent self-renewal when transfected into ES cells (10, 11). Genetic ablation of Nanog leads to the loss of pluripotency and self-renewal that characterizes the undifferentiated ES cell state (6, 11, 23, 24). A Nanog-centered protein interactome identified many known and novel protein partners essential for preservation of the pluripotent state (4). The protein domain structure of Nanog has been studied extensively to elucidate its molecular mechanism of action. The C-terminal domain, including the tryptophan-rich (WR) domain, has been deemed essential for the function of Nanog (25–29); however, study of the N-terminal domain (NTD) has been largely overlooked to date.

In the present study, we analyzed the genomic neighborhood surrounding the Nanog gene locus for evidence of an expanded Nanog gene structure. We identified novel sequences from ES cells that extend the 5' region of the known Nanog gene. Two additional new exons and 6 different subexons are differentially processed from alternative splicing. We find that this post-transcriptional regulation results in two new Nanog protein variants and we explore the function of these variants in ES cell self-renewal and pluripotency. Our studies reveal evidence that the first 25 amino acids of the NTD of Nanog are essential for the maintenance of the undifferentiated ES cell state.

**EXPERIMENTAL PROCEDURES**

**Cell Culture**—ES cell lines were maintained on gelatin-coated plates without feeders in standard ES cell media as described previously (28, 30). HEK293T cells were cultured in DMEM supplemented with 10% fetal bovine serum, 2 mM L-glutamine, and 50 units/ml of penicillin/streptomycin.

**Mouse Blastocyst Collection and RNA Extraction**—The C57BL/6J strain mice obtained from The Jackson Laboratories were used in this study. All animals were maintained under standard laboratory conditions and handled following the institutional guide for the use and care of laboratory animals. To obtain preimplantation mouse blastocysts, 3-week-old female mice were superovulated by injecting 5 IU of human chorionic gonadotropin 45 h following 5 IU pregnant mare serum gonadotropin administration and mated with fertile male mice of the same strain. Successful mating was determined the next morning by the presence of a vaginal plug and was considered day 0.5 of development (days postcoitus). Blastocysts were flushed from uterine horns at 3.5 days postcoitus using standard procedures (31). Total RNA was isolated using TRIzol reagent (Invitrogen), and cDNA was synthesized using the SuperScript III RT with the GeneRacer oligo(dT) primer (5'-ACACGAGCAGAGGA- CACUGACAUGGACUGAGGUGAGAAA-3') and GGCCACCATAGCCTTAAGTTTACC (NgEx1A-R1) in combination with a touchdown PCR protocol: 10 cycles with 30 s at 94 °C and annealing elongation for 2 min at 65–61.4 °C (−0.4 °C per cycle); followed by 30 cycles with 30 s at 94 °C, 30 s at 58 °C, and 1.5 min at 72 °C. An aliquot of the resulting amplification product was used for nested PCR using primers 5'-GGCAGCAGACATGGACTGAGA-3' and AAAGGTTTGGTACAGGGCGCACA (NgEx1A-R2) in combination with a second PCR program (30 cycles of 30 s at 94 °C, 30 s at 58 °C, and 1.5 min at 72 °C). PCR products were run on a 2% agarose gel and different PCR products were gel-

**RNA Extraction and RT-PCR**—Total and cytosolic RNA were prepared from J1, V6.5, RF8, and E14Tg2a cell lines using the PARIS kit (Ambion) following the manufacturer's instructions. An in-column DNase digestion was performed to remove contaminating genomic DNA. Total RNA for other experiments was prepared using the Illustra RNAspin RNA extraction kit (GE Healthcare). One microgram of RNA was reverse transcribed using oligo(dT) primers in a total volume of 20 µl using GoScript reverse transcriptase (Promega). 1 µl of each cDNA was used as template in 25-µl PCR throughout all experiments. All isolated RNAs were also directly tested in PCR to exclude genomic DNA contamination.

For isolation and characterization of novel exons and cDNA sequences extending to the 5' untranslated region (UTR) of the previously known Nanog gene, PCR was performed using a forward primer 5'-ACCTCTCTGCCTGATCTT-3' (Nanog 5' UTR-F1 located 4821 bp upstream of the known TSS) with a reverse primer 5'-ATTTCGAGAAGGAAAGCCTGTGCT-3' (Nanog 3' UTR-R1). Different PCR products were gel-purified and cloned into the pCR TOPO Blunt II vector. Subsequent sequence analysis was done using a universal M13 primer 5'-CAGGAAACGCTATGACC-3'.

**5'-RACE Experiments**—5' Transcript end rapid amplification of cDNA ends (5'-RACE) was performed using the Invitrogen GeneRacer kit according to the manufacturer's instructions with J1 ES cell cytosolic RNA. Briefly, 1 µg of J1 cytosolic RNA was used for the initial dephosphorylation and subsequent decapping steps resulting in a 5'-phosphate only at full-length properly capped mRNA molecules. A 44-nucleotide long RNA oligonucleotide (5'-CGACUGGAGCAGAGGAGACUGACUGAUGGAUGAGAAA-3') was subsequently ligated to the dephosphorylated and decapped mRNA resulting in 5'-tagged RNA molecules. A first strand cDNA synthesis was performed using the SuperScript III RT with the GeneRacer oligo(dT) primer (5'-GCTGTCAACGATACGC TACGTAACGGCATGACAGTG(T)24-3), which results in cDNAs with known tags at both the 5' and 3' ends. Then the cDNA was used to perform two consecutive PCR to amplify Nanog-specific transcripts. For the first PCR the specific primers used were 5'-CGACTGGAGCAGAGGACACTG-3' and GGCACACATAGCCTTAAGTTTACC (NgEx1A-R1) in combination with a touchdown PCR protocol: 10 cycles with 30 s at 94 °C and annealing elongation for 2 min at 65–61.4 °C (−0.4 °C per cycle); followed by 30 cycles with 30 s at 94 °C, 30 s at 58 °C, and 1.5 min at 72 °C. An aliquot of the resulting amplification product was used for nested PCR using primers 5'-GGCAGCAGACATGGACTGAGA-3' and AAAGGTTTGGTACAGGGCGCACA (NgEx1A-R2) in combination with a second PCR program (30 cycles of 30 s at 94 °C, 30 s at 58 °C, and 1.5 min at 72 °C). PCR products were run on a 2% agarose gel and different PCR products were gel-

---

3 S. Das, S. Orkin, and D. Levasseur, manuscript in preparation.
purified and cloned into the pCR TOPO Blunt II vector. Subsequent sequence analysis was done using a universal M13 primer 5’-CAGGAAAACACGTATGACC-3’.

Quantitative Real-time RT-PCR (qRT-PCR)—For detection of novel Nanog variants, cDNA was synthesized from 1 μg of the J1, V6.5, RF8, and E14Tg2a cytosolic RNA using the GoScript RT (Promega). qRT-PCR was performed to quantify the relative expression of the different Nanog transcripts in the 4 different cell lines on cDNA in a 25-μl SYBR Green reaction using Bio-Rad iCycler. Expression was normalized to Ywhaz. See Table 1 for primer information. For other qRT-PCR experiments, 1 μg of total RNA was used for the cDNA synthesis. Technical replicates were represented from two independent biological replicates.

Cloning and Expression of Nanog Protein Variants—Two different identified splice variants of the mouse Nanog along with the known Nanog were amplified using specific primers exhibiting BamHI and NotI restriction sites for further cloning into the pEF6/V5 His-A expression vector (Invitrogen) enabling a C-terminal V5-His tag fusion to all Nanog coding sequences. See Table 2 for the primers used. Sequence-verified plasmids (8 μg/6-cm Petri dish) coding for different Nanog protein variants were transfected into HEK293T cells using Lipopectamine and grown for 2 days before protein or RNA extraction.

Co-immunoprecipitation (Co-IP)—HEK293T or E14T ES cells were transiently co-transfected with plasmids containing a 3× FLAG tag at the N terminus of the protein or a V5 tag at the C terminus of the protein using Lipopectamine according to the manufacturer’s protocol. 48 h after transfection, cells were washed and whole cell lystate was prepared in lysis buffer containing 20 mM Hepes-KOH (pH 7.9), 150 mM NaCl, 0.5% Nonidet P-40, 1 mM EDTA, and 10% glycerol freshly supplemented with 10 mM NaF, 1 mM Na3VO4, 1 mM PMSF, and a protease inhibitor mixture (Sigma). Antibodies directed against FLAG-M2 (Sigma), V5 (Bethyl Laboratories), or Oct4 (Santa Cruz Biotechnology, sc-9081) were added to 1 mg of whole cell lystate and incubated for 2 h with rotation at 4 °C. Pre-equilibrated Protein A (for rabbit antibody) or Protein G (for mouse antibody) Dynabeads in the cell lysis buffer were added to the whole cell lystate with antibodies and incubated overnight with rotation at 4 °C. Prior to adding the antibodies, an aliquot was removed for use as input. The next day, Dynabeads were immunomagnetically captured and washed 3 times in PBS. Pulldown proteins were then extracted in sample buffer, separated by SDS-PAGE, and Western blotted with the appropriate antibodies. All co-IPs were repeated two or more times.

Western Blotting—Transfected cells were harvested in PBS and lysed in cell lysis buffer (0.5% Nonidet P-40). Soluble fractions were mixed in 2× Laemmll buffer and separated on an 8–14% SDS-PAGE gel and subsequently proteins were transferred onto a PVDF membrane. Membranes were incubated in blocking solution (5% skim milk in 0.1% Tween 20 containing TBS) overnight. The blocked membranes were rinsed once in washing solution (TBS-T: 0.1% Tween 20 containing TBS) and then incubated with primary antibody diluted in blocking solution overnight at 4 °C. Then the membrane was washed three times in TBS-T and incubated with HRP-conjugated secondary antibody for 2 h at room temperature. Subsequently, the membranes were washed 5 times in TBS-T and developed by using an enhanced chemiluminescent (ECL) substrate (Thermo Scientific).

LIF Withdrawal and Alkaline Phosphatase Staining—1200 Oct4-GiP cells (32) stably expressing different variants of Nanog were seeded onto a gelatin-coated 10-cm culture dish and grown with or without LIF for 8 days and then colonies were assayed for alkaline phosphatase (AP) activity with the vector blue alkaline phosphatase substrate kit (Vector Labs). Briefly, media were aspirated from the plates followed by a wash with PBS, and then cells were fixed in 1% paraformaldehyde in PBS for 2 min at room temperature. The fixative was aspirated and plates were washed with PBS. Plates were then incubated with the AP substrate solution for 20 min at room temperature in the dark. The plates were washed twice with the wash buffer (10 mM Tris-HCl, pH 8.2) for 3 min each and then AP positive and negative colonies were scored for morphology and counted.

RESULTS

An Extended Gene Structure for Nanog Is Revealed—We and others have previously detected Nanog protein as a ladder of bands in Western blots (4, 23, 28), and the prevailing assumption was that these represented degradation products or post-translational modifications. We independently replicated this laddering pattern in two independent ES cell lines (J1 and E14T) by subjecting protein lysates to Western blotting with an N-terminal Nanog antibody (Fig. 1B). Although post-translational modifications and degradation products of the protein could be plausible reasons for this observation, we reasoned that expression of multiple variants of the Nanog protein might also be a viable explanation for the ladder of bands observed.

We previously detected significant RNA polymerase II and p300 co-occupancy, coupled with enrichment of the permissive chromatin mark histone H3 lysine 4 trimethylation, at a region upstream of the Nanog TSS (30). This observation further corroborated the possibility of an expanded Nanog gene structure. We curated global databases of expressed sequence tags and other short expressed RNAs in the vicinity of the Nanog gene to search for candidate cryptic RNA sequences. Our search revealed two tags at positions 3773 and 4319 bp upstream of the known Nanog TSS and led us to hypothesize that an extension of the Nanog gene structure may exist. We performed RT-PCR with a forward primer that hybridizes upstream of the tags and

### Table 1

| Nanog transcript initiation | Primers used (5’-3’) | Amplicon size |
|----------------------------|---------------------|---------------|
| Ex3A | NgEx3A QRT-F1: ggttgaaccttctggacttt | 104 bp |
| Ex1B | NgEx1B QRT-F1: accttctggacttt | 46 bp |

---

**Nanog Protein Variants with Different Functional Capacities**

**TABLE 1**

| qPCR primers | Nanog transcript initiation | Primers used (5’-3’) | Amplicon size |
|--------------|--------------------------|---------------------|---------------|
| Ex3A | NgEx3A QRT-F1: ggttgaaccttctggacttt | 104 bp |
| Ex1B | NgEx1B QRT-F1: accttctggacttt | 46 bp |
a reverse primer that lies in the 3' UTR and discriminates the protein coding gene from all annotated retrogenes. Multiple products resulting from the PCR were purified and cloned. Subsequent sequence analysis confirmed the existence of an extension of the Nanog gene upstream of the known TSS. We identified complex splice patterns and the presence of novel 5'-exons revealed extended gene structures. As shown in Fig. 1A, the Nanog gene exhibits two additional exons that we have indicated as exons 1 and 2. Based on these findings, we propose a revised exon/intron structure for the mouse Nanog gene and a new nomenclature for the resulting splice variants as depicted in Fig. 1A. Our sequencing results reveal two alternatively used first exons (1A and 1B) that both can splice to two forms of exon 2, or exon 3.

5'-RACE was performed with cDNA obtained from polyadenylated cytosolic ES cell RNA to determine whether any of our cloned transcripts might originate from an alternate TSS. Supporting previous reports (33, 34), all cloned sequences from 5'-RACE with a reverse primer in exon 3 failed to result in reverse transcription past the known TSS. We surmised from this that an upstream TSS might exist, and 5'-RACE using reverse primers hybridizing within the novel exon 1 sequence revealed a new TSS 5 kb upstream of the previously reported TSS (27 bp upstream of our forward primer Nanog 5'-Ext-F1). This region houses a DNA regulatory signature that we and others have shown has enhancer activity (30, 35). Our results now reveal that a second Nanog promoter may contribute the full or partial RNA polymerase II recruitment activity seen previously in our analysis of DNase I hypersensitivity and chromatin conformation within this genomic region (30), and in another report analyzing chromatin interactions (36).

**Alternative Splicing Reveals Novel Mouse Nanog Variants**—We were able to clone a variety of different Nanog splice variants from our RACE products. All of these variant mRNAs can

| Nanog ORF amplified | Primers used (5'-3') | Nanog ORF protein size |
|---------------------|----------------------|------------------------|
| Nanog a             | Nanog a V5-F1: ggtatccacatgagttggctcttggt | 305 aa |
|                     | Nanog V5-R1: ggcgggccctatttcctgcggctgcgttg | 305 aa |
| Nanog b             | Nanog b V5-F1: ggtatccacatgggctcttggtc | 305 aa |
|                     | Nanog V5-R1: ggcgggccctatttcctgcggctgcgttg | 305 aa |
| Nanog c             | Nanog c V5-F1: ggtatccacatgcctgcagttttcatcc | 280 aa |
|                     | Nanog V5-R1: ggcgggccctatttcctgcggctgcgttg | 280 aa |
Nanog Protein Variants with Different Functional Capacities

be potentially translated into 3 different Nanog proteins (Nanog a, Nanog b, and Nanog c) as shown in Fig. 1A. Splice variant Nanog a comes in multiple forms and is derived from exon 3 (the structure of the known Nanog gene) or from transcripts originating upstream of exon 3 and containing additional exons 1 and 2. Exon 2 splices with exon 3 at nucleotide +150, thereby skipping 149 nucleotides of exon 3 and resulting in exon 3B. The Nanog a variant encodes an open reading frame (ORF), which uses the first AUG start codon of exon 3 and has the potential to produce a protein with predicted molecular mass of 34.2 kDa and containing 305 amino acids, which is the known Nanog protein. For the Nanog variant b, exon 1A splices with exon 3 at nucleotide +245 of exon 3, and this splicing results in exon 3C. The Nanog b variant encodes an ORF using a start codon from exon 1A having the potential to produce a protein of 305 amino acids. However, the Nanog b protein has 9 amino acids at the beginning of the N terminus that distinguish it from the known protein, even though this novel protein has a nearly identical predicted molecular mass of 34.4 kDa. Exon 1B skips exon 2 and fuses with exon 3C to produce the variant Nanog c having an ORF using the second AUG start codon of exon 3. The Nanog c variant encodes another novel Nanog protein of 280 amino acids that has a predicted molecular mass of 31.9 kDa that can be easily distinguished by Western blotting.

Expression of Nanog Variants in HEK293T and ES Cells—To verify the predicted molecular weight of the two new splice variants identified, we expressed the variants with a V5 tag at the C terminus in both HEK293T and J1 ES cells. After transient transfection, soluble whole cell lysates were extracted and analyzed by Western blotting. Both the novel Nanog variants and the known Nanog a were successfully expressed in both HEK293T (Fig. 2A) and J1 ES (Fig. 2B) cells. For all three variants of recombinant Nanog (Nanog a (38 kDa), Nanog b (38.2 kDa), and Nanog c (35.6 kDa)) expressed in HEK293T and J1 ES cells, several protein bands were detected with bands migrating above and below the predicted Mr. This suggests that all Nanog variants are subjected to post-translational modifications. For Nanog a and Nanog b the bands detected below their predicted Mr, but of similar size to Nanog c protein bands, imply the usage of multiple start codons. Western blot band intensity suggests the usage of a putative second start codon to be minimal if compared with the first start codon. Remarkably, the migration

**FIGURE 2. Recombinant expression of the novel Nanog variants.** HEK293T (A) or J1 ES cells (B) were transiently transfected with the cloned V5-His-tagged variants of Nanog, and total lysates were prepared and subjected to Western blotting with anti-V5 antibody. Predicted molecular weight of all expression constructs with the V5-His tag is indicated. Size of protein markers are indicated on the left. Possible ATG usage with each variant is mentioned. The same blot was stripped and subjected to Western blotting with anti-Lamin B antibody to show equivalent loading.
and banding patterns for all 3 variants was identical in both HEK293T and J1 cells suggesting equivalent usage of the post-translational modification machinery in both pluripotent and somatic cells of human kidney origin.

**Novel Nanog Variants b and c Interact with Nanog a**—Nanog homodimerization is required for stem cell self-renewal and pluripotency (28). To study whether the newly identified variants are able to dimerize with Nanog a, we performed co-IP studies in both HEK293T (Fig. 3A) and E14T (Fig. 3B) ES cells. Triple FLAG-tagged Nanog a was co-transfected with each of the C-terminal V5-tagged Nanog variants and co-IP was performed in both directions (Fig. 3) to minimize the potential interference from the affinity tags. Interaction of FLAG-tagged Nanog a with itself (homodimerization) was used as a positive control. Total lysates were subjected to Western blotting with the indicated antibodies as inputs. Nanog b and c are able to interact with Nanog a.

**Nanog Variants a, b, and c Interact with Other Pluripotency Network Proteins**—Nanog interacts with a network of proteins required for maintaining the pluripotency of ES cells (4, 35, 40). We tested the interaction of novel Nanog isoforms with other known pluripotency network binding partners, by performing similar co-IPs with FLAG-tagged Oct4 and Sall4 in HEK293T and E14T cells. FLAG-Oct4 was co-transfected with each of the C-terminal V5-tagged Nanog variants and co-IP was performed in both directions in HEK293T cells (Fig. 4A) and in one direction in E14T cells (Fig. 4B). The results showed that the novel Nanog isoforms interact with Oct4 and the interaction is as strong as the interaction of Nanog a with Oct4. The interaction of Nanog variants with endogenous Oct4 was further verified by transiently transfecting E14T cells with the V5-tagged Nanog variants followed by co-IP using either anti-Oct4 or anti-V5 antibodies (Fig. 4C).

The interactions of Nanog variants with Sall4 were verified by transient co-transfection co-IPs with FLAG-Sall4 and V5-tagged Nanog in E14T cells (Fig. 4D). These observations suggest that not only are there multiple splice variants of Nanog, but also that they interact with pluripotency network proteins and possibly play a combined role in the Nanog-containing complexes.

**Nanog Variants Have Different Capacities to Maintain LIF-independent Self-renewal**—The LIF/gp130 signaling cascade is essential for efficient preservation of self-renewal in murine ES cells (41, 42), and their human counterparts if cultured under physiological low oxygen tension (43). Nanog was originally identified as the first factor that could maintain the self-renew-
Nanog Protein Variants with Different Functional Capacities

A colony formation assay was performed with the stable Nanog variant expressing cells to provide a quantitative analysis for LIF-independent self-renewal. A limited number of stable Nanog variant expressing cells were cultured at clonal density in the presence or absence of LIF for 8 days followed by alkaline phosphatase (AP) staining, a marker of self-renewing state of ES cells in the absence of LIF (10, 11). To test if novel Nanog variants functioned equivalently, we investigated whether their enforced expression could rescue ES cells from LIF withdrawal. For this, C-terminal V5-tagged variants in the pEF6/V5-His plasmid were stably introduced into Oct4GiP ES cells that contain an Oct4-EGFP reporter.

A colony formation assay was performed with the stable Nanog variant expressing cells to provide a quantitative analysis for LIF-independent self-renewal. A limited number of stable Nanog variant expressing cells were cultured at clonal density in the presence or absence of LIF for 8 days followed by alkaline phosphatase (AP) staining, a marker of self-renewing state of ES cells in the absence of LIF (10, 11). To test if novel Nanog variants functioned equivalently, we investigated whether their enforced expression could rescue ES cells from LIF withdrawal. For this, C-terminal V5-tagged variants in the pEF6/V5-His plasmid were stably introduced into Oct4GiP ES cells that contain an Oct4-EGFP reporter.

FIGURE 4. Novel Nanog variants interact with Oct4 and Sall4. A, co-IP of FLAG-tagged Oct4 with V5-His-tagged Nanog isoforms in heterologous HEK293T cells. Total lysates were prepared from HEK293T cells transiently transfected with the constructs indicated. Co-IP was performed with anti-V5 antibody (left) or anti-FLAG antibody (right), followed by Western blotting with the antibodies indicated. Interaction of Oct4 with Nanog a was used as a positive control. Total lysates were subjected to Western blotting with the indicated antibodies as inputs. Nanog b and c are able to interact with Oct4 as well as Nanog a. B, co-IP of FLAG-tagged Oct4 with V5-His-tagged Nanog isoforms in E14T ES cells. Co-IP was repeated with anti-V5 antibody using total lysates from transiently transfected E14T cells with the indicated constructs, and the interactions of Nanog b and c with Oct4 were verified by Western blotting. C, co-IP of endogenous Oct4 in E14T ES cells with V5-His-tagged Nanog isoforms. E14T cells were transiently transfected with the constructs indicated and total lysates were prepared. Co-IP was performed using antibodies directed against V5 (top) or Oct4 (middle), followed by Western blotting with the indicated antibodies. Nontransfected E14T cell total lysates were used as a negative control in the co-IP. Total lysates were subjected to Western blotting with anti-Oct4 as inputs. * indicates nonspecific signal. D, co-IP of FLAG-tagged Sall4 with V5-His-tagged Nanog isoforms in E14T ES cells. Total lysates were prepared from E14T cells transiently transfected with the constructs indicated. Co-IP was performed with anti-V5 antibody (left) or anti-FLAG antibody (right), followed by Western blotting with the antibodies indicated. Interaction of Sall4 with Nanog a was used as a positive control. Total lysates were subjected to Western blotting with the indicated antibodies as inputs. Nanog b and c are able to interact with Sall4 as well as Nanog a.
ES cells, and differentially stained colonies were counted. When cells were grown in the presence of LIF, overexpression of both Nanog b and c resulted in a similar percentage of AP positive colonies as Nanog a (Fig. 5). However, in the absence of LIF, although the Nanog b variant expressing stable cells had a slightly higher percentage (46%) of AP positive colonies than the ES cells without any rescuing transgene (33%), this variant was not able to rescue as efficiently as the Nanog a or c variants that had about 75% AP positive colonies.

Nanog has a serine-rich motif at the N terminus (10) and regulation of Nanog by phosphorylation has been recently shown to serve a critical role in ES cell function (44). The serine positioned immediately after the starting methionine codon (Ser-2) of Nanog is conserved across all vertebrates (see alignment in supplemental Fig. S1) and is replaced by a valine residue in the Nanog b variant. Based on this observation and that regulation of this short region of the protein is mediated by alternative splicing we hypothesized that Ser-2 may be critically important for Nanog function. We mutagenized valine 2 to serine in Nanog b (V2S-Nanog b) and serine 2 to valine in Nanog a (S2V-Nanog a) to ascertain whether these mutations would be able to reflect the LIF withdrawal assay results we observed with ES cells in the absence of LIF. When the colony formation assay was performed in ES cells stably expressing mutant variants, the V2S-Nanog b mutant was better able to rescue ES cells from LIF withdrawal, resulting in colony numbers only marginally greater than in ES cells that do not contain any Nanog transgene. These results suggest that the Nanog Ser-2 residue is critically involved in the maintenance of self-renewal.

All Nanog Variants Rescue the Loss of ES Cell Self-renewal in the Absence of Endogenous Nanog—To further analyze the functional role of the individual Nanog variants in the absence of endogenous Nanog, we produced a doxycyline/tetracycline-regulatable Nanog transgene (Tet-off) containing the ES cell line, dubbed NgcKO, where the homeodomain coding exon 2 of both the endogenous Nanog alleles were excised by loxP-cre mediated recombination. This results in a null allele because exon 1–3 splicing creates a frameshift and nonsense mutations that deplete transcript message, presumably through the mRNA decay pathway. When total lysates prepared from the cell line after doxycycline (Dox, 200 ng/ml) treatment for different time periods were subjected to Western blotting, Nanog protein levels were undetectable within 12 h of Dox treatment (Fig. 6A). Nanog mRNA was also completely extinguished within 6 h of Dox treatment (data not shown).

To study the functional role of Nanog variants in the absence of Nanog a, stable NgcKO cell lines with the respective C-terminal V5-tagged Nanog variants were generated and subsequently stable cell lines were cultured in the presence of Dox and individual clones were picked for each of the variants. The clonal stable cell lines were always maintained in the presence of Dox in culture. Morphologically, these stable clonal NgcKO ES cells...
cells cultured in the presence of Dox were indistinguishable from nondepleted NgcKO cells and exhibited significant AP activity (Fig. 6C). When the control untransfected NgcKO cells were subjected to Dox treatment and subsequent Nanog depletion, they lost AP staining and differentiated into cells reminiscent of primitive endoderm (Fig. 6C), confirming the original Nanog loss of function mutation (11). This suggests that all Nanog variants tested are able to maintain self-renewal in the absence of endogenous Nanog.

Nanog b and c Variants Have Compromised Abilities to Preserve Pluripotency and Repress Differentiation—To further characterize the ability of Nanog variants to preserve pluripotency, qRT-PCR was performed to analyze the expression level of the master regulators of pluripotency and lineage specification. To ensure that the gene regulatory network governing pluripotency was not perturbed prior to analysis, cell lines were chosen that express Nanog mRNA levels between 70 and 150% of Nanog a for these studies. Western blot analysis of the selected clones with anti-V5 antibody detected similar levels of the V5-tagged transgenes in cell lysates prepared from the various clones grown in the presence or absence of Dox (Fig. 6B). Expression of polyadenylated RNA was measured from NgcKO cell lines rescued with each of the Nanog variants (Figs. 7 and 8, A and B). Among the pluripotency markers screened, the expression of Oct4, Esrrb, Zfp42, and Fgf4 was reduced below 50% (Fig. 7A), whereas minor decreases in the expression of Sox2 and Zfp281 were observed following depletion of Nanog. Transgenes harboring Nanog variants a, b, and c were largely able to maintain the expression of pluripotency markers in the absence of Nanog. We next assessed the expression of several early differentiation markers to determine whether the new Nanog variants are able to maintain repression in the absence of endogenous Nanog.

FIGURE 6. Nanog variants enable ES cells to retain self-renewal in the absence of Nanog a. A, loss of Nanog protein in Dox-treated NgcKO cells. Western blot analysis after incubation of the NgcKO cells with 200 ng/ml of Dox for the indicated time periods using anti-Nanog N-terminal antibody. There are undetectable levels of Nanog within 24 h of Dox treatment. For panels A and B, the same blot was stripped and subjected to Western blotting with anti-Lamin B antibody to demonstrate equivalent loading. Relative amounts of Nanog protein are indicated above the blot. B, Western blot analysis of stable clones of different Nanog isoforms expressing (as indicated) NgcKO cells with anti-V5 antibody to detect the V5-tagged transgenes before and after Dox treatment. All clones show similar levels of different Nanog variants. C, V5-tagged Nanog variants were transfected into the NgcKO cells as indicated and stable clones were picked for each of the variants growing in the presence of Dox and subjected to AP staining. All variants enabled morphological rescue of the NgcKO cells to retain typical ES cell morphology and stain positively for AP activity, whereas the untransfected NgcKO cells were differentiated and lost the ability for AP staining in the absence of Nanog after being cultured in the presence of Dox for 5 days (upper row middle panel). Tg, transgene.
Nanog Protein Variants with Different Functional Capacities

The undifferentiated state of ES cells is characterized by the coupling of self-renewal and pluripotency. This unique biological state is governed by a multilayered network of post-transcriptional and post-translational modifications (46). These mechanisms of regulation enable a vastly expanded usage of the proteomic space. Alternative splicing is highly pervasive in higher metazoan organisms (47). A particularly large diversity of splice isoforms may afford added functionality toward the maintenance of ES cell pluripotency and self-renewal by enabling the formation of distinct protein interactions and subnetworks involved in development of state-specific regulation. Alternative splicing affords a limited number of genes significant modularity. This can better enable the creation of multiple regulatory subnetworks that can each exhibit subtle nuance, much like that observed with functions shared among protein family members (48). The core pluripotency transcription factor human OCT4 gene has been shown to generate at least three transcripts (OCT4A, OCT4B, and OCT4B1) and four protein isoforms by alternative splicing and alternative translation initiation (49, 50). Although the long isoform OCT4A is responsible for the pluripotency of ES cells, the shorter versions OCT4B and OCT4B1 are expressed in more differentiated cell types and OCT4B may respond to cell stress (49). It is unclear whether any protein isoform other than the full-length OCT4A serves an essential role in pluripotency. Sall4 is another transcription factor that is essential for pluripotency, and exists as two isoforms, Sall4a and Sall4b. Recently, it has been shown that these two Sall4 isoforms play collaborative but distinct roles in maintenance of the pluripotent state of ES cells (40). Additionally, recent reports show that during neural and cardiac differentiation, the gamut of splice isoforms present in pluripotent versus committed lineages changes (19, 51). Because Nanog is particularly critical for facilitating and stabilizing induced pluripotency from all somatic cells analyzed, further characterization of these new isoforms may shed additional insight on this process (52, 53).

In these studies of the Nanog gene locus, we have discovered two novel 5′-exons, six different subexons, and most importantly, two new protein variants resulting from alternative splicing and alternate translational initiation. The two new Nanog proteins, dubbed Nanog b and c, respectively, differed from the previously known Nanog a in the N-terminal domain. The Nanog b variant differs from the Nanog a variant by replacement of 9 amino acids in the very N-terminal end of the protein. Nanog c contains a truncation of the initial 25 amino acids at the N terminus (see supplemental Fig. S2). Nanog has been shown to form homodimers for its functional role in ES cells (28, 54). Our immunoprecipitation studies (Fig. 3) both in heterologous HEK293T and native ES cells revealed that the novel protein variants are able to form heterodimers with the known Nanog a variant. We also demonstrated that the novel variants are able to interact with other essential pluripotency transcription factors, Oct4 and Sall4 (Fig. 4). We have previously shown that the Nanog C-terminal domain, particularly

FIGURE 7. Nanog b and c variants have compromised ability to rescue pluripotency expression markers and to fully repress certain trophodermal markers. A, qRT-PCR analysis of the expression of specific stem cell pluripotency marker genes in control NgcKO, complete Nanog knockdown, and Nanog variant rescuing NgcKO cells. Stable clones of the Nanog variants in the NgcKO cell line as mentioned were grown in the presence of Dox. Untransfected control NgcKO cells were grown in Dox for 6 days. The data shown are normalized to Gapdh, and shown along with the S.E. of technical replicates. The expression level in the control NgcKO cells is set as 1. B, qRT-PCR analysis of the expression of various cell lineage marker genes in control NgcKO, complete Nanog knockdown, and Nanog variant rescuing NgcKO cells.

A

B

|        | Nanog KO | Nanog KO Dox | Nanog a | Nanog b | Nanog c | V2S Nanog b |
|--------|----------|--------------|---------|---------|---------|-------------|
| Oct4   | 1        | 0.7          | 0.9     | 0.7     | 1       |
| Sox2   | 1        | 0.7          | 0.9     | 0.7     | 1       |
| Zfp291 | 1        | 0.7          | 0.9     | 0.7     | 1       |
| Etno   | 1        | 0.7          | 0.9     | 0.7     | 1       |
| Tab2   | 1        | 0.7          | 0.9     | 0.7     | 1       |
| Fgfb   | 1        | 0.7          | 0.9     | 0.7     | 1       |

Pluripotency Markers

|        | Nanog KO | Nanog KO Dox | Nanog a | Nanog b | Nanog c | V2S Nanog b |
|--------|----------|--------------|---------|---------|---------|-------------|
| Tsx1   | 1        | 0.7          | 0.9     | 0.7     | 1       |
| Snai1  | 1        | 0.7          | 0.9     | 0.7     | 1       |
| Snai2  | 1        | 0.7          | 0.9     | 0.7     | 1       |
| Ectod  | 1        | 0.7          | 0.9     | 0.7     | 1       |

DISCUSSION

When it comes to the Nanog protein, it should be noted that a recent study of this transcription factor indicates it may more accurately be described as a participant in primitive endoderm formation (45). The Nanog b variant exhibited a greater impairment than Nanog c, resulting in a more significant derepression of master regulators of primitive endoderm and other critical transcription factors that govern primitive streak formation in vivo. This suggests that these novel Nanog variants, specifically Nanog b, are unable to maintain pluripotency in the absence of endogenous Nanog. Surprisingly, the V2S-Nanog b mutant behaves like the Nanog a variant in its ability to maintain the expression of pluripotency markers and suppress the differentiation genes in the absence of endogenous Nanog (Figs. 7 and 8, A and B). This result reinforces the importance of Ser-2 for the functional role of Nanog in ES cells. Taken together, these data imply that Nanog b and c variants partially compensate for the loss of Nanog but are unable to repress differentiation promoting genes, indicating a decoupling of full pluripotency from the promotion of self-renewal.
the WR subdomain is essential for homodimerization, and that the N-terminal domain was dispensable for this function (28). For this reason, the interaction of the novel isoforms with Nanog a was not unexpected. However, considering the identification of two new isoforms, it must now be considered that Nanog a can function either as a homodimer or by forming different heterodimeric combinations with Nanog b or Nanog c. Heterodimerization among different variants may afford significant combinatorial partnering and differential downstream target gene recognition. We and others have previously documented that Nanog exists primarily in the dimeric state (28, 54) and can function efficiently as an obligate dimer in self-renewal assays (28). Testing the ability of Nanog variants to rescue a complete Nanog loss of function mutation by forced homo- or heterodimerization of Nanog variants will be required to discern if different dimerization combinations enable better preservation of self-renewal and pluripotency.

Perhaps the most surprising result was the reduced ability of the Nanog b variant to maintain self-renewal in the absence of LIF (Fig. 5). The reason for this could be attributed to the subsequent finding that Nanog b alone is insufficient to fully maintain a pluripotent signature because ES cells fully depleted of endogenous Nanog but rescued with this variant were compromised in their ability to repress primitive endodermal (Gata6, Gata4, Dab2, Sox17, and Pdgfra) transcriptional regulators (Fig. 8, A and B). This would direct the ES cells toward an extraembryonic endodermal lineage. Nanog and the Gata genes (Gata6 and Gata4) are postulated to work antagonistically in the maintenance and dissolution of the pluripotent state (11, 55–57). It has been suggested that Nanog may directly repress Gata6 (11,
Nanog Protein Variants with Different Functional Capacities

28), and this may be at least partly through its binding of Gata6 proximal and distal promoter regulatory elements (4, 28). The Nanog b variant has 7 different amino acids (supplemental Fig. S1) at its N terminus compared with Nanog a and these residues are more hydrophobic than those present in Nanog a. This could cause a conformational change in the Nanog protein, altering its protein partnering repertoire. Although less likely, we cannot rule out the possibility that significant conformational changes may act in conjunction with dimerization to alter the DNA binding capacity. These scenarios may in part explain why Nanog b is less efficient at repressing Gata6 and other crucial specifiers of primitive endoderm. Once Gata6 is activated, it can directly or indirectly trigger the expression of Gata4 and Sox17 (45). Gata6 and Gata4, together with Sox17, can then consolidate the transcriptional network for primitive endoderm specification (45). The absence of 25 amino acids at the N-terminal end of the Nanog c variant may cause a reduction in its ability to repress the extraembryonic endodermal promoting Gata6, Gata4, and Sox17 proteins, as well as the trophectodermal marker Hand1 (Figs. 7B and 8, A and B). Future elaboration of the Nanog a, b, and c protein structures by crystallography or NMR will be required to determine how inclusion or exclusion of N-terminal residues may alter protein conformation and function.

Interestingly, when the valine 2 residue of Nanog b was mutated to the evolutionarily conserved serine at this position (supplemental Fig. S2), the V2S-Nanog b mutant was able to recapitulate Nanog a self-renewal capacity (Fig. 5), maintain a pluripotency marker gene signature (Fig. 7A) and keep all differentiation marker genes repressed (Figs. 7B and 8, A and B). Although, the Ser-2 position has not yet been identified as a phosphorylation site for Nanog, it has already been established that Nanog is a phosphoprotein (58) and phosphorylation is one of the post-translational mechanisms serving an important role in stabilizing Nanog (44). However, ScanSite and other phosphorylation prediction motif finders do not predict regulation at this position (data not shown); therefore, future proteomic analyses will be required to verify the existence of phosphorylation at this site. We favored the possibility that the replacement of valine with a mutant serine at position 2 might be sufficient to enable Nanog b to effectively mimic the function of Nanog a. However, when we introduced this mutation into Nanog a (S2V-Nanog a), its self-renewal ability was diminished (Fig. 5), but it was able to maintain the expression of pluripotency marker genes and to efficiently keep differentiation promoting genes suppressed (data not shown). The lone exceptions analyzed were factors in the FGF signaling cascade (Fgf5, Fgfr2) that were significantly derepressed only in this mutant but not when rescued with Nanog a, b, or c (data not shown). We note that the Nanog b variant also contains two threonine residues within the short 9-amino acid replacement that is not shared with Nanog a. Differential post-translational modification of these residues may also contribute to regulation of this alternative isoform. Collectively, an altered hydrophobic surface combined with a different repertoire of phosphorylatable residues at the N-terminal end of Nanog b could cause a conformational change that destabilizes or abolishes partnering interactions with other transcription factors essential for pluripotency. This precludes proteins other than Oct4 and Sox4 because interaction with these factors is well maintained (Fig. 4). We also cannot rule out the prospect that master regulators of primitive endodermal and primitive streak specification interact preferentially with Nanog b. Future structural and proteomic studies will be required to assess these possibilities.

The observation that the Nanog post-translational modification pattern was indistinguishable in HEK293T cells, a human embryonic kidney cell line that does not harbor the capacity for self-renewal or pluripotency, and ES cells overexpressing tagged Nanog variants, was intriguing. This implies that if the surface of the protein displays a residue that is accessible and prone to modification, then the candidate mediators of post-translational modifications are likely to be relatively ubiquitous proteins expressed across many cell types. These results also indicate that splicing regulatory proteins responsible for guiding the transcript expression choice with the alternative splicing machinery may be previously underestimated and essential mediators of the self-renewing ES cell state (59).

Our current working model is that Nanog variants may engage differential roles in early developmental fate decisions (Fig. 8C). Although Nanog a is able to maintain ES cell self-renewal by keeping critical lineage commitment regulators, especially the primitive endodermal genes, suppressed, Nanog b and c have a reduced ability to maintain repression of regulators of primitive endoderm and trophoderm. We propose that Nanog a, b, and c may be expressed differentially during their narrow developmental window of control. This would facilitate partnering with different protein regulators and subsequent recruitment to different gene promoters. Such a strategy would enable tighter control of the two lineage choices during the earliest fate decision, maintenance of a self-renewing pluripotent state or differentiation into the supportive extraembryonic layer. Stated differently, the partitioning of the Nanog gene locus into multiple protein variants enables the production of different proteins that may have altered propensities for the maintenance of self-renewal and pluripotency, active promotion of primitive endoderm formation, or a combination of both activities.

ES cells are a well accepted culture model and facsimile of the inner cell mass that forms within the blastocyst during early embryonic development. For this reason, a thorough analysis of the regulation of gene expression in both compartments of pluripotency will further our understanding of this biological state (60). As expected, expression of known and novel upstream-derived Nanog transcripts was shared between ES cells and blastocysts, although differential expression of transcripts originating from the proximal and distal promoter regions was less pronounced in blastocysts. Future studies using complementary culture models of pluripotency (30), restricted pluripotency (epiblast stem cells (61, 62)), extraembryonic endoderm (XEN (63)), and the early developmental counterparts of these cell lines will be required to further ascertain the individual contribution of Nanog variants for maintaining pluripotency. The development of antibodies that can recognize the Nanog b variant, coupled with mass spectrometric approaches to measure peptides specific for the Nanog c truncation variant, will be
Nanog Protein Variants with Different Functional Capacities

required to fully assess the role of these proteins in future studies.

The C-terminal domain of Nanog was initially found to have a modestly stronger transactivation activity than the NTD (27). The C-terminal domain was also required for Nanog homodimerization and interaction with other pluripotency transcription factors (28, 54). Conversely, the NTD was shown to be dispensable for these interactions using a heterologous reporter assay (27) and by our immunoprecipitation analyses (28). It is unclear whether previous analyses employing domain truncations, which can alter and compromise protein conformation and function, may have obscured the importance of the N-terminal portion of the protein, particularly of Ser-2. To our knowledge, our study is the first report demonstrating the importance of the NTD of Nanog in maintaining ES cell self-renewal and pluripotency, and the critical requirement of a single amino acid residue for maintaining the function of a core pluripotency factor.

Acknowledgments—One of us (D.N.L.) acknowledges Stuart Orkin for critically reading the manuscript. We are grateful to Austin Smith, Rudolf Jaenisch, and Robert Farrese for sharing ES cell lines.

REFERENCES

1. Evans, M. J., and Kaufman, M. H. (1981) Nature 292, 154–156
2. Martin, G. R. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 7634–7638
3. Orkin, S. H., Wang, J., Kim, J., Chu, J., Rao, S., Theunissen, T. W., Shen, X., and Levasseur, D. N. (2008) Cold Spring Harbor Symp. Quant. Biol. 73, 195–202
4. Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D. N., Theunissen, T. W., and Orkin, S. H. (2006) Nature 444, 364–368
5. 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., Gifford, D. K., Melton, D. A., Jaenisch, R., and Young, R. A. (2005) Cell 122, 947–956
6. Ivanova, N., Dobrin, R., Lu, R., Kotenko, I., Levorse, J. D., Coste, C., Schafer, X., Lun, Y., and Lemoschka, I. R. (2006) Nature 442, 533–538
7. Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V. B., Wong, E., Orlov, Y. L., Zhang, W., Jiang, J., Loh, Y. H., Yeo, H. C., Yeo, Z. X., Narang, V., Govindarajan, K. R., Leong, B., Shahab, A., Ruan, Y., Bourge, G., Sung, W. K., Clarke, N. D., Wei, C. L., and Ng, H. H. (2008) Cell 133, 1106–1117
8. Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S. H. (2008) Cell 132, 1049–1061
9. van den Berg, D. L., Sneson, T., Mullin, N. P., Yates, A., Zebrarostki, C., Demmers, J., Chambers, I., and Poot, R. A. (2010) Cell Stem Cell 6, 369–381
10. Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003) Cell 113, 643–655
11. Mitsui, K., Tokuwa, Y., Inoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003) Cell 113, 641–642
12. Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klim-Neubensius, D., Chambers, I., Scholer, H., and Smith, A. (1998) Cell 95, 379–391
13. Avilion, A. A., Nicolas, S. K., Pevey, L. H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003) Genes Dev. 17, 126–140
14. Takahashi, K., and Yamanaka, S. (2006) Cell 126, 663–676
15. Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R., Slukvin, I. I., and Thomson, J. A. (2007) Science 318, 1917–1920
