Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochemical and Biophysical Research Communications 307 (2003) 569–577
BBRC
www.elsevier.com/locate/ybbrc

# Molecular characterization of a novel nucleolar protein, pNO40☆,☆☆

Wei-Lun Chang,[a,1] Dong-Chin Lee,[a,1] Steve Leu,[a,b] Yi-Ming Huang,[a] Ming-Chu Lu,[a] and Pin Ouyang[a,b,*]

[a] *Epithelial Biology Laboratory, Department of Anatomy, Chang Gung University Medical College, Taiwan, ROC*
[b] *Transgenic Mice Core-Laboratory, Chang Gung University Medical College, Taiwan, ROC*

Received 3 June 2003

**Abstract**

We report the discovery and characterization of a novel nucleolar protein. This protein, referred to as pNO40 based on its molecular weight on SDS–PAGE, was identified through yeast two hybrid interaction screen of a human kidney cDNA library using pinin (pnn) protein as the bait. The deduced amino acids of pNO40 derived from cDNA cloning of diverse species display high conservation; 95% identify between human and mouse and 57.3% identity for human and zebrafish. Several distinct domains are discernable in the ORF of pNO40, including a ribosomal protein S1 RNA binding region, a CCHC type zinc finger, and clusters of basic amino acid representing potential nucleolar targeting signal. Immunostaining of endogenous or transfected pNO40 indicated that it is localized to nucleoli of diverse cultured cells, with some concentration in the granular component of nucleoli. Northern blot analysis demonstrated that pNO40 message is expressed ubiquitously across all tissues examined. Characterization of human and mouse pNO40 gene revealed that mouse gene spans 44 kb in length and contains 8 exons, while that of human is 68 kb in length and displays two isoforms generated by alternative splicing of the 5′-untranslated region and differential usage of translation start site. Based on sequence features and its subcellular location, we predict that pNO40 is a novel nucleolar protein with function related to ribosome maturation and/or biogenesis.


*Keywords:* pnn; pNO40; Nucleolus; Alternative splicing; 1p34.2–34.3

Higher eukaryotic nuclei contain numerous morphologically distinct structures or nuclear bodies [1,2]. Among them the nucleolus is distinct, because it was not only one of the first intracellular structures described but it is easily visualized by light microscopy. Morphologically, the nucleolus displays three major structural components comprising, from center to periphery, the fibrillar center (FC) surrounded by the dense fibrillar component (DFC) as well as by the granular component (GC) [3,4]. It has been reported that the DFC contains newly synthesized preribosomal RNA (pre-rRNA) and a collection of proteins, while the GC is made up of nearly completed preribosomal particles destined for the cytoplasm. Therefore, it is well established that the nucleolus is involved in ribosome formation and the biosynthesis and assembly of ribosomal particles is a vectoral process, in which nascent preribosomes move from the DFC region to the peripherally located GC [5,6]. In addition to being a site dedicated to ribosome biogenesis, several lines of evidence now indicate that the nucleolus lodges the initial assembly reactions involved in the biogenesis of small RNAs transcribed by RNA polymerase III, namely 5S rRNA, tRNA, and SRP RNA [7] and acts as a sequestering compartment for regulatory complex [3,4,8]. However, the molecular mechanisms governing the assembly of ribosomes and the majority of the nucleolar proteins involved in ribosome biogenesis are still ill defined. Identification of

☆ *Abbreviations:* Pnn, pinin; ORF, open reading frame; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; FISH, fluorescence in situ hybridization.

☆☆ Sequences reported in this work have been submitted to GenBank under Accession Nos. AF247661 (human pNO40 short form, hpNO40s), AY253297 (mouse pNO40), and AY253298 (zebrafish pNO40).

* Corresponding author. Fax: +886-3211-8575.
*E-mail address:* Ouyang@mail.cgu.edu.tw (P. Ouyang).

[1] These authors contribute equally to this work.

0006-291X/03/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/S0006-291X(03)01208-7

proteins contained within nucleoli represents one of the perquisite steps to fulfill this tremendous task.

In this paper, we report the discovery and characterization of a novel nucleolar protein pNO40, which was identified through our screening of proteins interacting with pinin (pnn). As a putative tumor suppressor gene product [9], pnn was originally identified as a desmosome-associated protein [10,11] and subsequently found ubiquitously in the nuclear speckled domain of interphase cell nuclei [12,13]. Since pnn is localized both at the cell–cell junction and within the nucleus of certain cell types, it is plausible that it may form diverse protein complex at different cellular compartments [12,14] and its function may thus be regulated through its interaction partner. Identification of pnn-interacting proteins therefore is important for elucidation of cellular function of pnn. Here, we describe the molecular characterization of a novel pnn-interacting protein, pNO40. pNO40 is localized in the granular component of the nucleolus in almost every, if not all, tissues examined and it is well conserved in its sequences between human, mouse, and zebrafish. Genomic studies suggest that human pNO40 gene presents two isoforms differing in their usage of the translation initiation site. Finally, we discuss the possible involvement of pNO40 in mediating pnn cellular function.

## Materials and methods

*Reagents.* Dulbecco's modified Eagle's medium (DMEM), Hank's medium, RPMI 1640 medium, and fetal calf serum (FCS) were purchased from Gibco-BRL (Gaithersburg, MD). Glutathione–Sepharose 4B was purchased from Pharmacia Biotech (Uppsala, Sweden). Molecular biology reagents, including restriction enzymes and *Taq* polymerase, were purchased from New England Biolab (Beverly, MA). All chemicals, unless otherwise specified, were purchased from Sigma (St. Louis, MO).

*Expression and purification of recombinant proteins.* pNO40 fusion protein was produced by ligation of the *Eco*RI–*Xho*I fragment of human clone 4 (LDC4) cDNA to pGEX4T3 vector (Pharmacia Biotech, Uppsala, Sweden) and expressed in DH5α cells. Purification of recombinant proteins was performed as previously described [12]. In brief, bacteria that were successfully transformed with recombinant vector were induced with 0.1 mM IPTG in LB for 3 h at 37 °C. Purifications of expressed proteins were accomplished by absorption of fusion proteins from bacterial cell lysate to glutathione–Sepharose 4B followed by elution with 5 mM glutathione in 50 mM Tris–HCl, pH 8.0.

*Cell culture and transfection.* MDCK and Cos7 cell lines of passages 10–60 were maintained in DMEM and supplemented with 10% FCS, 2 mM glutamine, and 200 U/ml each of streptomycin and penicillin G. Cells were passed with 0.1% trypsin and 0.04% EDTA in Hank's medium. Expression plasmids encoding myc epitope conjugated pNO40 (FLDC4myc) was constructed by ligating *Eco*R–*Xho*I fragment in frame to the pCMV-Tag3c expression vector (Stratagene, La Jolla, CA). Transient transfection of Cos7 cells was performed at 37 °C for 5 h using standard calcium phosphate method and transfected cells were assayed for immunofluorescent microscopy and Western blot.

*Antibodies production.* pNO40-containing GST fusion protein (GST-LDC4) was used as immunogen and antibody generation was performed as described previously [12]. In brief, 20 μg GST-LDC4 was mixed with complete Freund's adjuvant to inject Balb/C mice. Subsequent to three or four boost injections, at 2-week intervals, mice were sacrificed and splenocyte was fused with NS1 myeloma cells. Hybridoma cells grown in HAT-containing RPMI medium were plated into 24-well dishes.

Supernatants were screened by immunofluorescence microscopy of MDCK cells cultured on coverslip and Western blot using whole cell protein extracts. Cells of positive wells were selected and cloned by limited dilution.

*Screening of cDNA and genomic libraries.* An oriented mouse kidney cDNA library constructed in UNI-ZAP XR vector or a human λFix II leukocyte genomic library (Stratagene, La Jolla, CA) or a mouse 129/svj genomic BAC library in a set of high density nylon membranes (Research Genetics, Huntsville, AL) was screened as described previously [10]. For libraries screening, a human LDC4 *Eco*RI–*Nco*I fragment was used as a probe. For BAC membrane hybridization, a mouse cDNA (mLDC4c1) corresponding to the amino half of mpNO40 was used as a probe. A total of 300,000 phage plaques were screened. Duplicate filters were prehybridized at 42 °C overnight in 50% formamide, SX SSPE, 5× Dendhardt's solution, and 0.25% SDS containing 100 μg/ml boiled salmon sperm DNA. Hybridization was performed under the same condition as prehybridization with the addition of polynucleotide kinase $^{32}$P-labeled probe. Filters were then washed for 1 h at 60 °C with 2× SSC and 0.1% SDS and were then exposed to X-ray film.

*Northern blot.* A mouse multiple tissue Northern blot was purchased from SeeGene (Soul, Korea). Filter hybridization was carried out as described before [10]. The filter was prehybridized in 50% formamide, 5× SSC, 5× Dendhardt's solution, 0.2% SDS, and 200 μg/ml salmon sperm DNA for 10 h at 42 °C. Hybridization was performed with the addition of $^{32}$P-labeled mLDC4c1 at 42 °C for 16 h. Filter was then washed twice in 2× SSC with 0.1% SDS for 30 min at room temperature and 1× SSC with 0.1% SDS at 55 °C for 2 h and was exposed to X-ray film at −80 °C with an intensifying screen.

*Immunofluorescence microscopy.* MDCK cells cultured on coverslip or Cos7 transfected with FLDC4myc were treated for indirect immunofluorescence. Primary antibodies used were monoclonal 3YC6 specific for pNO40 or a mouse anti-myc monoclonal antibody (Balbco, Berkely, CA). Cultured cells were incubated with primary antibodies for 1 h at RT. After three 5 min washes in PBS, the samples were incubated with FITC-conjugated goat anti-mouse IgG (Jackson Laboratory, West Grove, PA) 1:100 diluted for 1 h at RT. The samples were then washed extensively with PBS, mounted with 50% glycerol containing 0.4% *n*-propylgallate, and examined with a photomicroscope (Axioscope 2, Carl Zeiss) equipped with epifluorescence.

*Western blot.* Whole cell extracts of MDCK, Cos7 or their transfected counterpart were prepared by lysing cells in RIPA buffer (50 mM Tris, pH 8.0, 0.14 M NaCl, 0.2% SDS, 1% Triton X-100, and 0.5% sodium deoxycholate) containing 1 mM PMSF and 1 mg/ml each of pepstatin, leupeptin, and chemostatin. Samples were sonicated for 30 s on ice and boiled for 5 min prior to electrophoresis on a 12% polyacrylamide gel. Western blot was then performed as described previously [12]. Proteins were transferred from the gel to nitrocellulose paper. The paper was blocked with 5% non-fat dried milk in PBS followed by washing in PBS. Primary antibodies incubations, with antibodies listed above, were carried out for 1 h at room temperature. After extensive wash in PBS, the paper was incubated for 1 h with HRP conjugated goat anti-mouse IgG (diluted 1:1000 in PBS, Jackson Laboratory, West Grove, PA). The peroxidase-labeled blots were reacted with 0.5 mg/ml diaminobenzidine and color reaction was developed using 0.01% hydrogen peroxide.

*FISH.* Purified DNA from a 9 kb human pNO40 genomic subclone, hg7-1B1, was labeled with digoxigenin dUTP by nick translation (Roche Applied Science, Indianapolis, IN). Labeled probe was combined with sheared human DNA and hybridized to metaphase chromosomes derived from PHA stimulated peripheral blood lymphocytes

in a solution containing 50% formamide, 10% dextran sulfate, and 2× SSC. Specific hybridization signals were detected by incubating the hybridized slides in fluoresceinated anti-digoxigenin antibodies (Roche Applied Science, Indianapolis, IN) followed by counterstaining with DAPI. The initial experiment resulted in the specific labeling of the short arm of a group A chromosome which was believed to be chromosome 1 on the basis of size, morphology, and banding pattern. A second experiment was conducted in which a clone known to localize to 1q32 (Research Genetics, Huntsville, AL) was cohybridized with hg7-1B1. This experiment resulted in the specific labeling of both the long and short arms of chromosome 1.

Measurement of 10 specifically labeled chromosome 1 demonstrated that hg7-1B1 is located at a position which is 71% of the distance from centromere to the telomere of chromosome arm 1p, an area which corresponds to band 1p34.2–34.3. A total of 80 metaphase cells were analyzed with 68 exhibiting specific labeling.

## Results and discussion

### *Molecular cloning of pNO40*

To detect potential molecular partners of pnn, we conducted an interaction screen using the yeast two-hybrid system. We screened a Gal4 DNA activation domain-containing human kidney cDNA library with a bait plasmid, which is generated by fusing pnn C-terminal fragment (TH08LΔN, aa 462–774) in frame to the Gal4 DNA binding domain. Following yeast cotransformation and plating on selective media, we identified 8 clones that activated the reporter from 2 million independent transformants (equivalent of approximately 40 haploid genomes). To minimize false positives, the selected clones were further confirmed by cotransformation with the bait plasmid and staining for reporter expression. According to the restriction enzyme digestion and partial sequences, all mapped clones belong to the same group and we selected the longest clone, clone 4 (LDC4), for further analysis.

The LDC4 clone consists of 1479 nt (Fig. 1A), which contains a 5′-untranslated region of 90 nt, an ORF of 651 nt (nt 91–741), and a 3′ region of 738 untranslated nt (nt 742–1479) with a poly(A)$^+$ additional signal (ACTAAA) 16 bp upstream of the poly(A) tail. Blast search of the NCBI DataBank did not identify any known human sequences that bear resemblance to LDC4, suggesting that it may encode a novel human protein. However, Prosite search [15] suggests that it contains recognizable domains that may forecast its potential functions (see below). The sequence around the ATG initiator codon (GCTATGG) is compatible with Kozak sequence for efficient ribosome binding and protein translation [16]; i.e., a purine nucleotide (A or G) in position −3 or G in position +4. Using human LDC4 or restriction fragment derived from it as probes, we screened mouse kidney and zebrafish cDNA libraries for homologous sequences and tried to differentiate the functional domains among various species. In such a way, we identified mouse and zebrafish homologues of human LDC4, which generally contain identical functional domains as does human homologue (Fig. 1B). To be certain that the cDNAs we identified were bona fide protein-coding sequences, we subjected mouse and zebrafish LDC4 homologues to EST database search and identified three mouse ESTs (Accession Nos. AK012953, AJ272345, and BC018382) and two zebrafish ESTs (clone IDs: 5605033 and 5304553) which displayed >95% identify to our cDNA clones. To our surprise we found that, following sequence determination, mouse and zebrafish clones contain an additional translation start codon upstream of the originally identified ATG site found in human LDC4 (Fig. 1B). To differentiate that human LDC4 contains complete ORF, we searched for human EST database for human LDC4-related sequences. One EST clone (Accession No. AF151077) was found and sequence confirmed to have identical translation start site as those of mouse and zebrafish homologues and thus extended the human LDC4 ORF by 24 aa from 217 aa to 241 aa. Closer examination of nucleotide sequences of human LDC4 and EST AF151077, however, indicated that they did not share the 5′ extreme sequences, especially sequences upstream of the 84th nt in human LDC4 (bent arrow in Fig. 1A). Comparing these cDNAs with human genomic sequences derived from human genomic project, we can determine that human LDC4 and EST AF151077 represent two alternative splicing variants, generated by utilizing different initiation codon of the LDC4 gene (see below and Fig. 5B). Since EST AF151077 encodes a nucleolar protein with an apparent molecular weight of about 40 kDa on SDS–PAGE (see below), we named it human nucleolar protein 40 (hpNO40), while the shorter protein encoded by LDC4 was named the short form of hpNO40 (hpNO40s, Fig. 1B).

The human and mouse pNO40 ORF each encodes a polypeptide of 241 aa residues with a calculated molecular mass of 27.47 kDa and a predicted p*I* of 9.67. Sequence alignment of pNO40 conceptual translated product derived from human, mouse, and zebrafish indicated a strong conservation of this protein across different vertebrate species (Fig. 1B). Human and mouse proteins display 95% identity and 97.1% homologue in terms of conservative amino acid exchange. Human vs. zebrafish and mouse vs. zebrafish each display 57.3% identity (70.7% homologue) and 57.3% identity (70.7% homologue), respectively. Prosite search [15] of the putative translation product revealed several recognizable domains that were well conserved across the species examined (Fig. 1B). The amino terminal sequences contain a S1 RNA binding region, which is originally identified in the bacterial ribosomal protein S1 and has been suggested to play an important role in translation in *Escherichia coli* [17]. The S1 RNA binding region, composed of nearly 70 aa (aa 16–88), has been

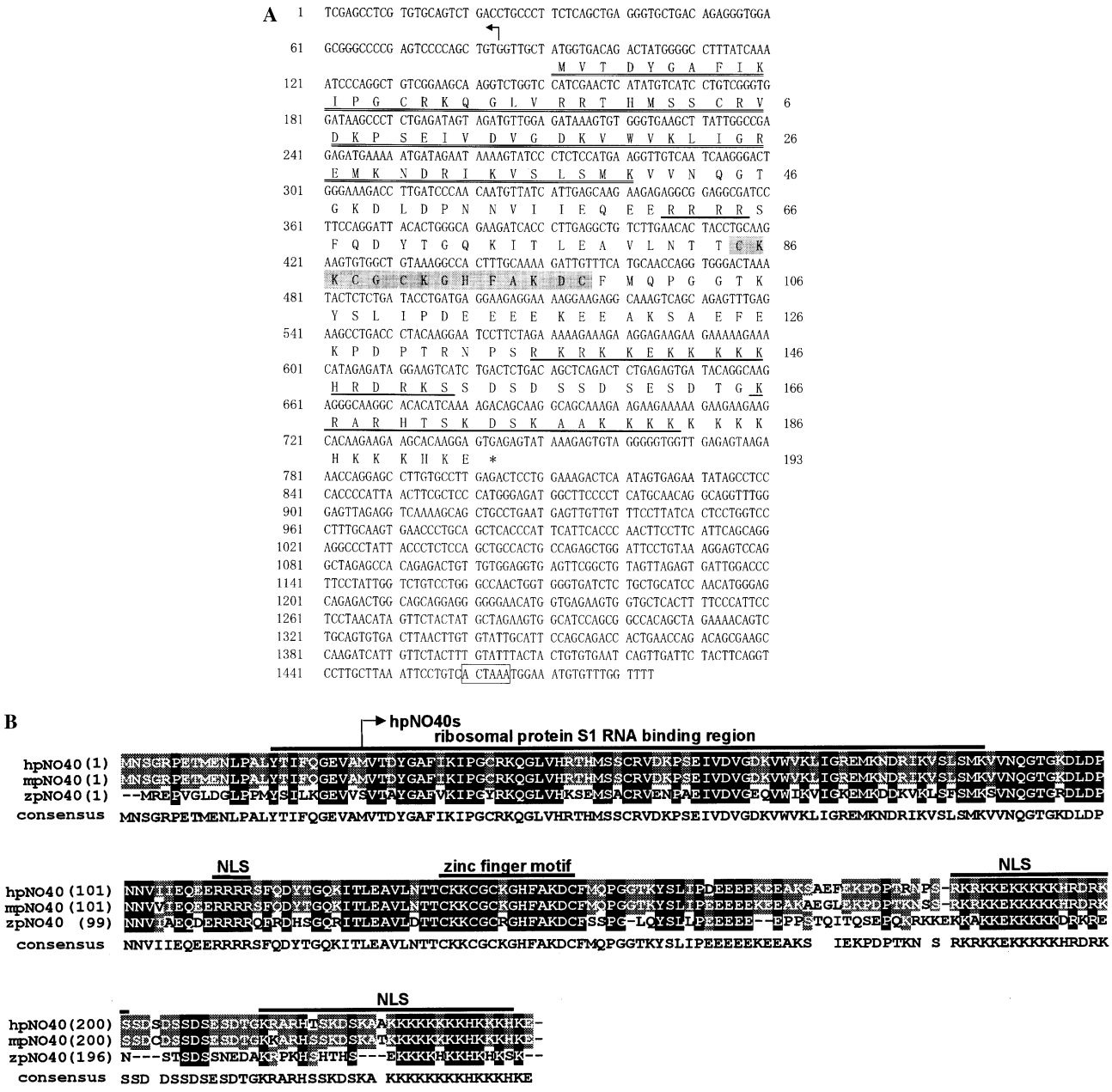


Fig. 1. Deduced amino acid sequences and domain structure of pNO40. (A) Nucleotide sequences and conceptual translation product of human pNO40 short form (hpNO40s). Sequences representing the alternative splicing isoform that differs from those of hpNO40 were denoted by a bent arrow. Deduced amino acid was displayed by single letter code and the stop codon and polyadenylation signal were represented by a star and enclosed in a rectangle, respectively. S1 RNA binding region was double underlined, while putative nuclear localization signals (NSLs) were underlined, and zinc finger motif was shadowed. (B) Alignment of pNO40 from different species. The translation start sites for human pNO40 short form (hpNO40s) was denoted by bent arrows. Human and mouse pNO40 display 95% identity, while those of human and zebrafish show 57.3% identity. Of note is that alignment of sequences among human, mouse, and zebrafish pNO40 zinc finger reveals near 100% identity. Sequences of hpNO40s, mpNO40, and zpNO40 have been submitted to GenBank under Accession Nos. AF247661, AY253297, and AY253298, respectively.

implicated in mediating RNA binding because it was found in several RNA-binding proteins such as bacteria translation initiation factor 1 (IF-1), polynucleotide phosphorylase (PNPase), and yeast PRP22 protein [18]. In the middle of the deduced amino acid sequence (aa 133–146), a zinc finger of the CCHC type is readily recognized. It is believed that zinc finger-containing proteins participate in DNA binding and there are several subfamilies of zinc finger such as CCCC type and CCHC type that have been characterized since the prototypic CCHH type is identified. Based on sequence variation, CCHC type zinc finger can further be divided into two subtypes: $CX_2CX_4HX_4C$ and $CX_2CX_{12}HX_5C$ [19]. The zinc finger motif found in pNO40 conforms to the former subtype, which was demonstrated to occur frequently in the nucleocapsid of retrovirus [20]. Since most of the

proteins containing pNO40-related CCHC zinc finger subtype, including cellular nucleic acid binding protein (CNBP), eIF-5, and 9G8, bind RNA and/or single stranded DNA with strong affinity [21,22], it is speculative that pNO40 binds DNA or RNA. Of note is that the zinc finger motifs found in pNO40 from various species display a near 100% identity (Fig. 1B), implying its important role in performing LDC4 cellular function.

Based on PSORT [23] and Prosite [15] analysis, both human and mouse pNO40 display two stretches of sequences rich in basic amino acids (aa 184–200 and aa 214–240) that may function as bipartite nuclear targeting sequences ($KRKX_{11}KPRK$, [24]). In addition, PSORT predicts that pNO40 is a nuclear protein at the probability of 82.6%. In general, a cluster of several basic amino acids residues in protein has been shown to be a signal required for nuclear import by an interaction with positively charged portions of the nuclear pore complex. Nevertheless, the second basic cluster of pNO40 (aa 214–240) appears to be too long for only nuclear transport. There are evidences that a longer basic cluster than nuclear targeting signals was necessary for targeting to nucleolus, such as KRKKEMANKSA PEAKKKK for nucleolin [25] and SKRLSSRARK RAAKRRLG for p120 nucleolar protein [26]. Therefore, pNO40 may be transported, via its carboxyl terminal basic cluster, to nucleoli in which rRNA maturation and ribosome assembly occur. Apart from nuclear targeting signals, other notable features, e.g., sequence elements involved in DNA binding and protein–protein interaction, have not been detected by Prosite or PSORT in pNO40. In fact, it is currently accepted that targeting of a specific molecule to the nucleolus results from its direct or indirect interaction with one of the nucleolar blocks, that is, rDNA or its transcripts and a shared structural feature of nucleolar proteins is the presence of an RNA recognition motif (RRM), which binds either rRNA at different stages of maturation or snoRNA [27,28]. Taken together, we predict that, based on its sequence features, pNO40 should encode a highly conserved novel nucleolar protein with a strong tendency to bind rRNA.

### *pNO40 is preferentially located in the nucleolus and displays a ubiquitous expression pattern*

To examine whether pNO40 is a nucleolar protein and its tissue distribution pattern, we performed immunofluorescent microscopy and Northern blotting. We generated monoclonal antibodies by immunizing mice with a bacterially produced, purified recombinant GST-tagged full length protein. One hybridoma, 3YC6, was selected for additional studies. Several lines of evidence demonstrated that this hybridoma indeed produces mAb that recognizes pNO40 specifically. First, on an immunoblot of total 293T and Cos7 lysates, 3YC6 recognized a single protein of ~40 kDa (Fig. 2A, lanes 3 and 4). This protein was found to be nearly identical in

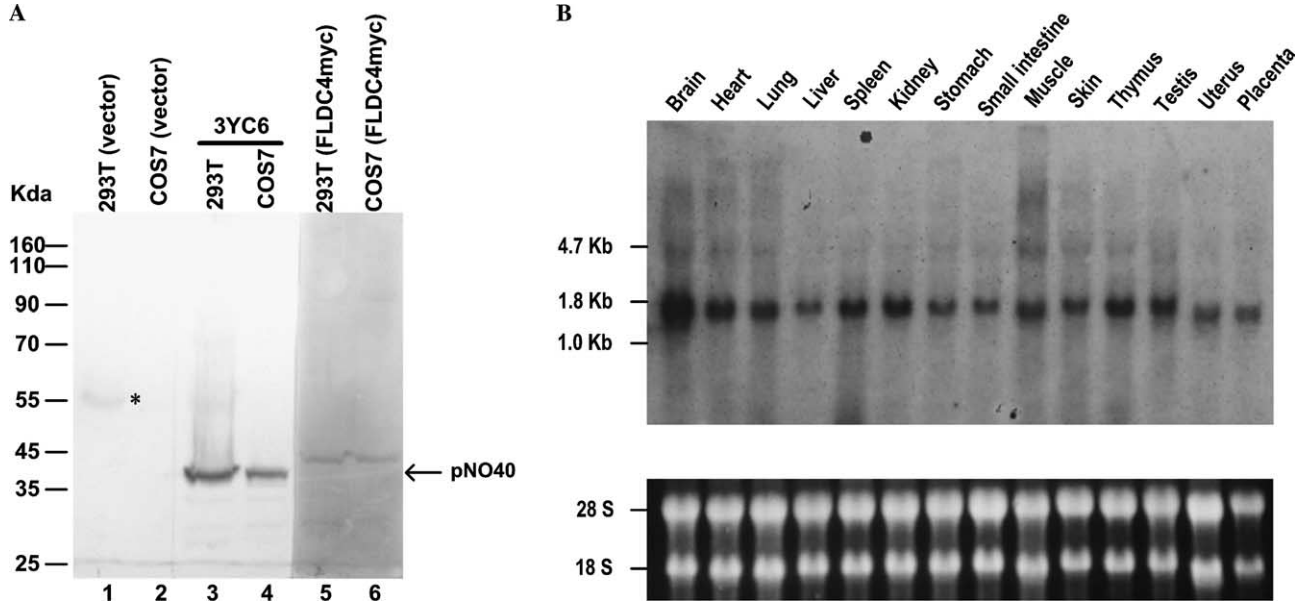


Fig. 2. Expression analysis of pNO40. (A) Identification of pNO40 as a single 40 kDa species on SDS–PAGE. Whole cell extract of 293T (lane 3) or Cos7 (lane 4) was separated by gel electrophoresis and probed with a monoclonal antibody, 3YC6, directed against pNO40. Exogenously transfected pNO40 was identified by Western blots with an anti-myc antibody on 293T or Cos7 whole cell extract containing empty vector (lanes 1 and 2) or myc-tagged full-length cDNA (FLDC4myc, lanes 5 and 6). The small (~3 kDa) difference in molecular mass between endogenous (lanes 3 and 4) and transfected pNO40 (lanes 5 and 6) is due to the tagged epitope. *: Non-specific protein cross-reacting with myc antibody. (B) Tissue expression pattern of pNO40 message. Northern blot analysis was carried out with a mouse multiple tissue blot and hybridization was performed with a mouse cDNA probe. The intensity of 28S and 18S was shown as an internal control for the assay.

electrophoretical mobility to that of a myc-tagged pNO40, albeit there is a small difference due to 3 kDa mass of myc epitope, as revealed by Western blot of whole cell 293T and Cos7 extract containing exogenously transfected full-length cDNA (Fig. 2A, lanes 5 and 6). Second, 3YC6 immunoblotted pNO40 produced by in vitro transcription and translation from full-length cDNA (data not shown). Based on the Western blot, there is a discrepancy between the predicted masses of pNO40 (27.47 kDa) and its apparent molecular masses on SDS–PAGE (40 kDa). We figure that the molecular weight difference may be due to either the basic nature of pNO40 (p*I* 9.67) or high content of potential phosphorylation sites, including 10 for protein kinase C, 6 for casein kinase II, and 2 for cAMP-dependent kinase. Presently, we do not know the actual degree of phosphorylation and thus we cannot exclude that post-translational modification might account for the observed difference.

Immunofluorescent microscopy of MDCK cell cultures using 3YC6 showed that pNO40 specifically accumulated in nucleoli, with some concentration in the granular component (arrows, Fig. 3A). The nucleolar distribution of pNO40 was also readily observed by immunostaining transfected Cos7 cells with an anti-myc antibody. In transfected cells pNO40 was detected in a unique fashion in the peripheral portion of nucleoli (arrows, Fig. 3B), corresponding to granular component, and displayed abundant distribution in nucleoplasm. To be certain that pNO40 is localized in the granular component of nucleoli, we compared the localization of pNO40 with that of other nucleolar proteins known to be enriched in the granular component of the nucleolus. Protein NO38/B23, a marker for nucleolar granular component [29], showed significant colocalization with pNO40 (data not shown). In summary, these results indicate that pNO40 is a nucleolar protein enriched preferentially in the granular component and, to some extent, disperses over the nucleoplasm.

By using a mouse multiple tissue blot, we examined tissue distribution and the size of pNO40 message (Fig. 2B). A single species of 1.6 kb was identified in all tissues examined and each tissue displays a relatively equal amount of message with brain, heart, skeletal muscle, and thymus being most prominent. The calculated message size obtained from Northern blot matches well with that from cDNA cloning (1626 bp, Accession No. AY253297), suggesting that we indeed acquire a full-length mouse pNO40 cDNA. From our cDNA cloning and EST database search, it appears that human pNO40 displays two alternative splicing variants (see below and Fig. 5B), hpNO40s and hpNO40 (EST AF151077), generated by differential usage of translation start site. However, Northern blot on multiple mouse tissues or on various human cell lines (data not shown) revealed only one reaction product upon cDNA hybridization, indicating that the two splicing variants of pNO40 may be too close in length to be differentiated with this assay. Since pNO40 exhibits high conservation between human and mouse (95%, Fig. 1B) and is expressed ubiquitously in almost, if not all, tissues examined, it may contain function which is important for all cells. Only proteins that act at intracellular key control points have a similarly high conservation between human and mouse. These proteins are found in the respiratory chain within mitochondria (e.g., cytochrome $c = 91.3\%$ identity; cytochrome *c* oxidase subunit $1 = 90.4\%$), in ribosomes (e.g., L7 = 95.9%; L12 = 98.8%; L18 = 90.4%; L19 = 99.5%; L28 = 97.8%; S2 = 98.2%; and S16 = 97.3%), or among histones (H3 = 100% and H4 = 100%). We consider pNO40 a ribosomal protein or a protein related to ribosome assembly based on several lines of evidence. First, it is a nucleolar protein and numerous proteins found within nucleoli, according to a very recent proteomic study [30], are either ribosomal proteins (15.5%) or proteins involved in ribosome biogenesis (23.5%). Second, it contains a S1 RNA binding region which was originally identified in bacterial ribosomal protein S1 [18]. Third, like some ribosomal pro-

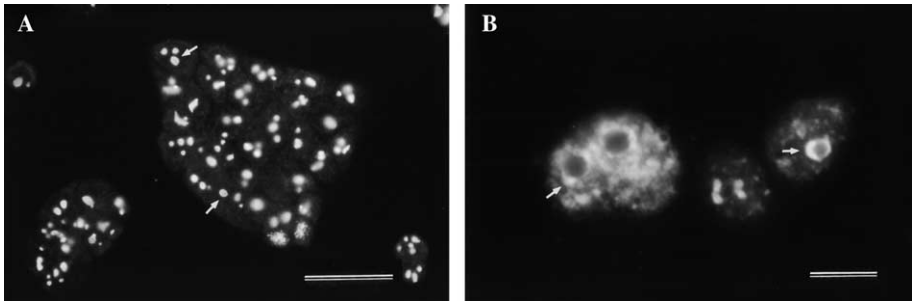


Fig. 3. Subcellular localization of pNO40. (A) MDCK cells were immunostained with pNO40 specific antibody, 3YC6, and visualized by indirect immunofluorescent microscopy. (B) Cos7 cells transfected with pNO40 full-length cDNA (FLDC4myc) were stained with myc antibody and visualized by indirect immunofluorescent microscopy. Both endogenous and transfected pNO40 showed prominent nucleolar localization, with concentration in granular component (arrows in A and B) and, in some cells, dispersed in nucleoplasm (B). Bars, 10 μm (A) and 2 μm (B).

teins, it contains a zinc finger and basic domains [31]. Fourth, it displays high conservation between human and mouse, like those of most ribosomal proteins. Recently, Andersen et al. [32] and Scherl et al. [30] each performed a detailed proteomic analysis on purified human nucleoli fraction and allowed the establishment of a list of 350 different nucleolar proteins. Of note is that hpNO40 is not identified in this list after we submitted its sequence to a BLASTp search (WWW.expasy.org) against the deduced proteins provided by both groups. Whether pNO40 is a bona fide ribosomal protein residing in nucleoli and whether pNO40 can withstand extraction methods employed by the above-mentioned proteomic analysis warrants further investigation.

*Subchromosomal localization and genomic organization of mouse and human pNO40*

To gain insight into the in vivo function and regulation of pNO40, we embarked on isolating mouse pNO40 gene and constructing targeting vector for future generation of knockout mice based on micromanipulation of engineered ES cells. To do so, we screened a 129/svj mouse BAC library with a mouse cDNA and isolated a BAC clone, BAC183. This clone spans about 97 kb in length and comprises a full-length gene encoding mpNO40 following genomic subcloning and sequence alignment with that of cDNA (Fig. 4). The whole gene of mouse pNO40 is about 44 kb in length and is composed of 8 exons. The translational start site is located in exon 2, while S1 RNA binding region is encoded by exons 2–4, and zinc finger motif encoded by exons 6 and 7 (Fig. 4). In a parallel study, we isolated human pNO40 genomic fragments from a human leukocyte λFixII genomic library and mapped its genomic locus, by fluorescent in situ hybridization (FISH), to human chromosome 1p34.2–34.3 (Fig. 5A). Compared with the genomic locus of human *pNO40*, which is 68 kb in length, mouse *pNO40* is located in chromosome 4 and in a region 61cM from the centromere based on the human–mouse synteny map. Since human pNO40 displays two protein products (Fig. 1B), we are therefore interested in figuring out at genomic level the molecular mechanism underlying this difference. Following alignment of hpNO40 and hpNO40s cDNA, respectively, with their genomic sequences obtained from human genomic project, we discerned that each represents an alternative splicing variant of hpNO40, generated by differentially utilizing the translation start codon and containing different 5′ exon (Fig. 5B). While hpNO40s starts from exon 1b followed by an ATG start codon-containing exon 3, hpNO40 starts from exon 1a, and proceeds to exon 2 followed by exon 3 (Fig. 5B). The translation initiation site of hpNO40 and hpNO40s, respectively, is located in exon 2 and exon 3. The former precedes in frame the latter by 24 aa, thus generating two forms of hpNO40, the short one with 217 aa (hpNO40s) and the long one with 241 aa (hpNO40) (Fig. 1B). It has been reported that the gene locus of

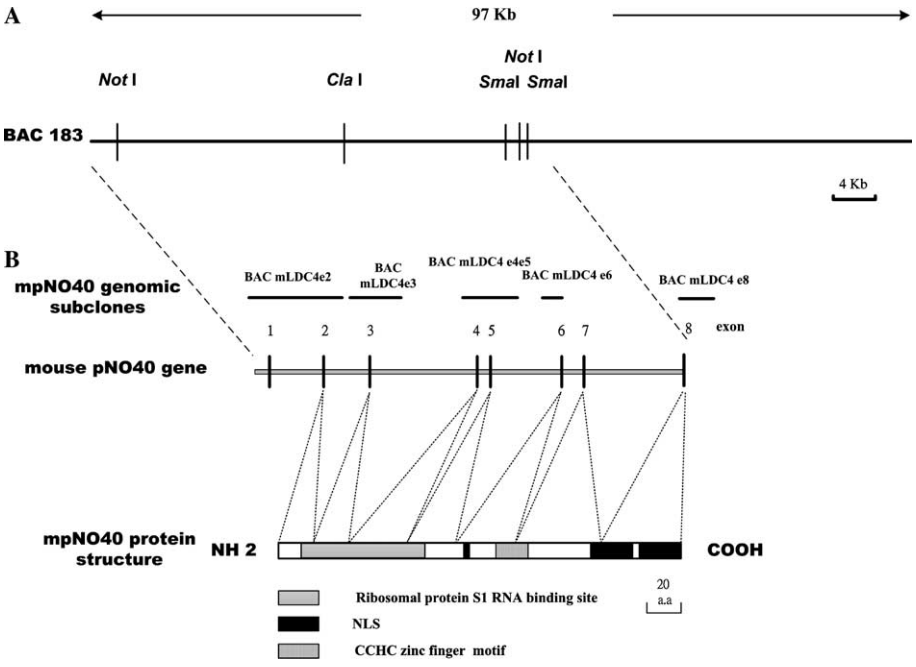


Fig. 4. Schematic diagrams showing the relationship of mouse pNO40 gene structure with its encoded protein. (A) Restriction mapping of the 129/svj derived *pNO40*-containing BAC clone, BAC 183. (B) Diagrams depicting the relative position of each mouse *pNO40* exon to its encoded protein domain. Note that *mpNO40* spans 44 kb in length and translation start site is located in second exon. NLS: nuclear localization signal.

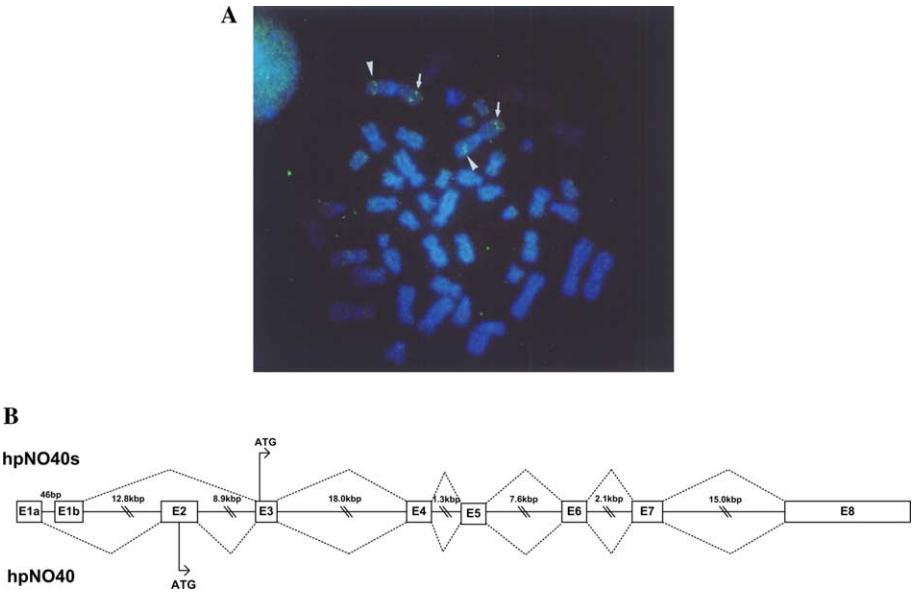


Fig. 5. Human pNO40 gene locus and genomic organization. (A) FISH analysis of *hpNO40* chromosomal localization. Human metaphase chromosomes derived from PHA stimulated peripheral blood lymphocytes were hybridized with purified *hpNO40* genomic DNA labeled with digoxigenin dUTP. Specific hybridization signals were detected by incubating samples in fluoresceinated antidigoxigenin antibodies followed by counterstaining with DAPI. Measurements of specifically labeled chromosomes indicated that *hpNO40* is located at a position corresponding to band 1p34.2–34.3 (arrows). The specificity of chromosome 1 signal was demonstrated by cohybridizing with a probe localized to 1q32 (arrowheads). (B) Genomic structure of *hpNO40* isoforms. Two isoforms of *hpNO40* were identified by cDNA screening and EST database search. Difference between the long form (*hpNO40*) and short form (*hpNO40s*) lies in the differential utilization of translation start codon (ATG) and 5′-untranslated region.

human *pNO40*, 1p34.2–34.3, is related to several human diseases. Loss of heterozygosity (LOH) studies suggested that this locus may harbor potential genes for hepatic cirrhosis [33] and malignant melanoma [34] and it may also be involved in human Schwarts–Jampel syndrome (SJS) [35]. It is therefore interesting to know whether *pNO40* is a candidate gene responsible for the said diseases.

Here, we report the identification and characterization of a novel pnn-interacting nucleolar protein, pNO40. Since nuclear pnn is localized preferentially to nuclear speckled domain, an assembly/storage site for mRNA splicing factors, it is intriguing to unravel the mechanism by which pnn interacts with pNO40. It has been reported that splicing factors for mRNA seem to follow a maturation pathway via nucleoli to their final destination. For example, Sm proteins (components of small nuclear RNPs) accumulate first in coil bodies, then in nucleoli, and finally in speckles [36]. We surmise that pnn may interact with pNO40 in nucleoli en its route to nuclear speckles. However, additional possibilities such as nucleolar location are important for optimal pnn function (preventing it from degradation) and pnn's presence in the nucleolus is one of those natural occurrences that has no real function and cannot be excluded. To summarize, pNO40 is a highly conserved nucleolar protein across different species and is expressed ubiquitously in every tissue examined. It displays two alternative splicing isoforms generated by differential usage of translation start site. Several structural features, including a S1 RNA binding region, a CCHC type zinc finger, and an extended basic domain, suggest this protein be a ribosomal protein involved in ribosome maturation/biogenesis. Currently, we are adopting the strategy of generating pNO40 deficient mice to investigate its in vivo function based on its gene structure.

## Acknowledgments

The authors thank supports from Chang Gung Memorial Hospital (CMRP 998) and National Science Council, ROC (NSC 90-2314-B-182-088 and NSC 90-2314-B-182-072).

## References

[1] A.I. Lamond, W.C. Earnshaw, Structure and function in the nucleus, Science 280 (1998) 547–553.
[2] A.G. Matera, Nuclear bodies: multifaceted subdomains of the interchromatin space, Trends Cell Biol. 9 (1999) 302–309.
[3] M. Carmo-Fonseca, L. Mendes-Soares, I. Campos, To be or not to be in the nucleolus, Nat. Cell Biol. 2 (2000) E107–E112.
[4] M.O.J. Olson, M. Dunder, A. Szebeni, The nucleolus: an old factory with unexpected capabilities, Trends Cell Biol. 10 (2000) 189–196.
[5] G. De Cárcer, F.J. Mendina, Simultaneous localization of transcription and early processing markers allow dissection of functional domains in the plant cell nucleoli, J. Struct. Biol. 128 (1999) 139–151.

[6] U. Sheer, R. Hock, Structure and function of the nucleolus, Curr. Opin. Cell Biol. 11 (1999) 385–390.
[7] T. Pederson, J.C. Politz, The nucleolus and the four ribonucleoproteins of translation, J. Cell Biol. 148 (2000) 1091–1095.
[8] R. Visintin, A. Amon, The nucleolus: the magician's hat for cell cycle tricks, Curr. Opin. Cell Biol. 12 (2000) 372–377.
[9] Y. Shi, P. Ouyang, S.P. Sugrue, Characterization of the gene encoding pinin/DRS/mem A and evidence for its potential tumor suppressor function, Oncogene 19 (2000) 289–297.
[10] P. Ouyang, S.P. Sugure, Characterization of pinin, a novel protein associated with the desmosome-intermediate filament complex, J. Cell Biol. 135 (1996) 1027–1042.
[11] P. Ouyang, S.P. Sugrue, Identification of an epithelial protein related to the desmosome and intermediate filament network, J. Cell Biol. 118 (1992) 1477–1488.
[12] P. Ouyang, Antibodies differentiate desmosome-form and nucleus-form pinin: evidence that pinin is a moonlighting protein with dual location at the desmosome and within the nucleus, Biochem. Biophys. Res. Commun. 263 (1999) 192–200.
[13] J.M. Brander, S. Reidenbach, C. Kuhn, W.W. Franke, Identification and characterization of a novel kind of nuclear protein occurring free in the nucleoplasm and in ribonucleoprotein structures of the "speckle" type, Eur. J. Cell Biol. 75 (1998) 295–308.
[14] W.G.J. Degen, M.A. Agterbos, J.P.P. Muyrers, H.P.J. Bloemers, G.M.M. Swart, memA/DRS, a putative mediator of multiprotein complexes, is overexpressed in the metastasizing human melanoma cell lines BLM and MV3, Biochim. Biophys. Acta 1444 (1999) 384–394.
[15] E.M. Zdobnov, R. Apweiler, InterProScan—an integration platform for the signature recognition methods in InterPro, Bioinformatics 17 (2001) 847–848.
[16] M. Kozak, Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes, Cell 44 (1986) 283–292.
[17] M.A. Sorensen, J. Fricke, S. Pedersen, Ribosomal protein S1 is required for translation of most, if not all natural mRNAs in *Escherichia coli* in vivo, J. Mol. Biol. 280 (1998) 561–569.
[18] M. Bycroft, T.J. Hubbard, M. Proctor, S.M. Freund, A.G. Murzin, The solution structure of the S1 RNA binding domain: a member of an ancient nucleic acid-binding fold, Cell 88 (1997) 235–242.
[19] M.F. Summers, Zinc finger motif for single-stranded nucleic acid? Investigation by nuclear magnetic resonance, J. Cell Biochem. 45 (1991) 41–45.
[20] R. Berkowitz, J. Fisher, S.P. Goff, RNA packaging, Curr. Top. Microbiol. Immunol. 214 (1996) 177–218.
[21] L. Pellizzoni, F. Lotti, B. Maras, P. Pierandrei-Amaldi, Cellular nucleic acid binding protein binds a conserved region of the 5′UTR of *Xenopus laevis* ribosomal protein mRNAs, J. Mol. Biol. 267 (1997) 264–275.
[22] T.B. Rajavashisth, A.K. Taylor, A. Andalibi, K.L. Svenson, A.J. Lusis, Identification of a zinc finger protein that binds to the sterol regulatory element, Science 245 (1989) 640–642.
[23] K. Nakai, P. Horton, PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization, Trends Biochem. Sci. 24 (1999) 34–36.
[24] C. Dingwall, R.A. Laskey, Nuclear targeting sequences—a consensus? Trends Biochem. Sci. 16 (1991) 478–481.
[25] M.S. Schmidt-Zachmann, E.A. Nigg, Protein localization to the nucleolus: a search for targeting domain in nucleolin, J. Cell Sci. 105 (1993) 799–806.
[26] B.C. Valdez, L. Perlaky, D. Henning, Y. Saijo, P.K. Chan, H. Busch, Identification of the nuclear and nucleolar localization signals of the protein p120. Interaction with translocation protein B23, J. Biol. Chem. 269 (1994) 23776–23783.
[27] P.J. Shaw, E.G. Jordan, The nucleolus, Annu. Rev. Cell Dev. Biol. 11 (1995) 93–121.
[28] M.R. Jacobson, L.-G. Cao, Y.-L. Wang, T. Pederson, Dynamic localization of RNase MRP RNA in the nucleolus observed by fluorescent RNA cytochemistry in living cells, J. Cell Biol. 131 (1995) 1649–1658.
[29] P.-K. Chan, W.-Y. Chan, B.Y.M. Yung, R.G. Cook, M.B. Aldrich, D. Ku, I.L. Goldknopf, H. Busch, Amino acid sequence of a specific antigen peptide of protein B23, J. Biol. Chem. 261 (1986) 14335–14341.
[30] A. Scherl, Y. Couté, C. Déon, A. Callé, K. Kndbeiter, J.-C. Sanchez, A. Greco, D. Hochstrasser, J.-J. Diaz, Functional proteomic analysis of human nucleolus, Mol. Biol. Cell 13 (2002) 4100–4109.
[31] I.G. Wool, Y.L. Chan, A. Gluck, Structure and evolution of mammalian ribosomal proteins, Biochem. Cell Biol. 73 (1995) 933–947.
[32] J.S. Andersen, C.E. Lyon, A.H. Fox, A.K. Leung, Y.W. Lam, H. Steen, M. Mann, A.I. Lamond, Directed proteomic analysis of the human nucleolus, Curr. Biol. 12 (2002) 1–11.
[33] M. Sun, J.R. Eshleman, L.D. Ferrell, G. Jacobs, E.C. Sudilovsky, R. Tuthill, M.R. Hussein, O. Sudilovsky, An early lesion in hepatic carcinogenesis: loss of heterozygosity in human cirrhotic livers and dysplastic nodules at the 1p36–p34 region, Hepatology 33 (2001) 1415–1424.
[34] D. Smedley, S. Sidhar, S. Birdsall, D. Bennett, M. Herlyn, C. Cooper, J. Shipley, Characterization of chromosome 1 abnormalities in malignant melanomas, Genes. Chromosomes Cancer 28 (2000) 121–125.
[35] S. Nicole, H.C. Ben, P. Beighton, S. Bakouri, S. Belal, N. Romero, D. Viljoen, G. Ponsot, A. Sammoud, J. Weissenbach, Localization of the Schwartz–Jampel syndrome (SJS) locus to chromosome 1p34–p36.1 by homozygosity mapping, Hum. Mol. Genet. 4 (1995) 1633–1636.
[36] J.E. Sleeman, A.I. Lamond, Newly assembled snRNPs associate first with coiled bodies then with speckles, suggesting a nuclear snRNP maturation pathway, Curr. Biol. 9 (1999) 1–10.