

Identification and characterization of a novel ubiquitous nucleolar protein ‘NARR’ encoded by a gene overlapping the *rab34* oncogene

Alexandre Zougman^{1,2}, Matthias Mann^{1,*} and Jacek R. Wiśniewski^{1,*}

¹Max Planck Institute for Biochemistry, Department of Proteomics and Signal Transduction, Am Klopferspitz 18, Martinsried D 82152, Germany and ²Cancer Research UK Clinical Centre, St James’s University Hospital, Beckett Street, Leeds LS9 7TF, UK

Received March 11, 2011; Revised and Accepted April 8, 2011

ABSTRACT

There are only few reports on protein products originating from overlapping mammalian genes even though computational predictions suggest that an appreciable fraction of mammalian genes could potentially overlap. Mass spectrometry-based proteomics has now acquired the tools to probe proteins in an unbiased manner, providing direct evidence of the output of the genomic and gene expression machinery. In particular, proteomics can refine gene predictions and discover novel gene-processing events and gene arrangements. Here, we report the mass spectrometric discovery and biochemical validation of the novel protein encoded by a gene overlapping *rab34* oncogene. The novel protein is highly conserved in mammals. In humans, it contains 13 distinct Nine-Amino acid Residue-Repeats (NARR) with the consensus sequence PRVIV(S/T)PR in which the serine or threonine residues are phosphorylated during M-phase. NARR is ubiquitously expressed and resides in nucleoli where it colocalizes with ribosomal DNA (rDNA) gene clusters. Its distribution only partially overlaps with upstream binding factor, one of the main regulators of RNA Polymerase I activity, and is entirely uncoupled from it in mitotic cells and upon inhibition of transcription. NARR only partially colocalizes with fibrillarin, the pre-ribosomal RNA-processing protein, positioning NARR in a separate niche within the rDNA cluster.

INTRODUCTION

Mass spectrometry-based proteomics has become an indispensable protein identification and characterization

tool (1), which provides efficient means to link genomic predictions with *bona fide* proteomes. Typically, proteomics data are queried against databases of known proteins. Even though a significant portion of the spectrum queries remains unassigned to peptide sequences, these data are not routinely searched against genomic or EST expressed sequence tags (EST) databases (2). This is mainly due to processing constraints; translation in six frames creates databases of a size significantly larger than that of the conventional protein databases. Furthermore, the statistical challenges to unambiguously assign peptides to the genome require excellent data quality in order to avoid high false positive rates. Nevertheless, it is obvious that such unassigned mass spectrometric data may contain unique information about the output of the gene expression machinery that could provide key data to answer many genomic puzzles. Thus, proteomics can partially validate gene predictions and discover novel difficult to predict genes, gene-processing events and gene arrangements. Our laboratory has previously reported the MS-based discovery of new genes (3) as well as evidence for unsuspected RNA editing events leading to different protein isoforms (4).

The nucleolus is a conserved subnuclear compartment with key functions in synthesis and processing of ribosomal RNAs as well as many other cellular functions (5). It is the place where the initial steps of ribosome subunit assembly occur. The nucleolus is formed around clusters of repeated rRNA genes (rDNA) that encode ribosomal RNA (rRNA), the scaffold and catalytic heart of the eukaryotic ribosome (6). In humans, the clusters of rDNA repeats, called nucleolar organizer regions (NORs), are positioned on the short arms of the acrocentric chromosomes 3, 14, 15, 21 and 22. Extensive proteomic studies have provided insights into the nucleolar protein composition (7–9) and of its dynamics (10,11). These studies have supported the notion of the

*To whom correspondence should be addressed. Tel: +49 89 85782205; Fax: +49 89 85782219; Email: jwisniew@biochem.mpg.de
Correspondence may also be addressed to Matthias Mann. Tel: +49 89 85782557; Fax: +49 89 85782219; Email: mmann@biochem.mpg.de

plurifunctional nucleolus (5) by identifying and tracking the dynamics of proteins involved in cell cycle control, stress response and synchronization of biogenesis.

Here, we describe the discovery of a novel nucleolar protein Nine-Amino acid Residue-Repeats (NARR), which arises from the transcription of a gene overlapping the *rab34* oncogene. The protein was identified by analyzing unassigned shotgun proteomics data, verified extensively by high-resolution MS as well as by immunochemical methods. NARR is one of the very few documented examples of protein products of overlapping genes in humans.

MATERIAL AND METHODS

Cell culture

Human MCF7 and HeLa cells were cultured in dubelcco's modified eagle medium (DMEM) (GIBCO) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin. For stable isotope labelling with amino acids in cell culture (SILAC) experiments, MCF7 cells were cultured with either normal or stable isotope labeled ($^{13}\text{C}_6$ $^{15}\text{N}_2$) lysine and ($^{13}\text{C}_6$ $^{15}\text{N}_4$) arginine (12). For immunofluorescence (IF) experiments, cells were grown either on 'BD-BioCoat' collagen-coated coverslips (BD Biosciences) or on microscope slides (Superfrost Ultra Plus, Thermo Scientific). When required, MCF7 cells were treated with 100 ng/ml of actinomycin D (ActD) for 14 h.

Extraction of proteins from cells and nuclei

Nuclei were isolated by differential pelleting of MCF7 cell homogenates. Briefly, fresh cells were homogenized in a tight-fitted Potter-Elvehjem homogenizer (Sartorius) in phosphate buffered saline (PBS) containing Protease Inhibitor Cocktail (Roche Diagnostics) at 4°C. Cell debris was removed by centrifugation at 100g for 5 min and then the nuclei were collected at 800g for 10 min.

Proteins were extracted from whole cell pellets or from nuclei by three freezing (−20°C), thawing and vortexing (10 s) cycles in 0.2 M HCl and were precipitated with 25% (w/v) CF_3COOH , washed twice with 0.2% (v/v) HCl in acetone and once with pure acetone and vacuum dried. Whole SDS lysates of cultured cells and mouse tissues were prepared as described previously (13).

For pull-down experiments, the nuclei of MCF7 cells were extracted with 0.6 M NaCl in 20 mM Tris-HCl, pH 7.6, containing ethylene diamine tetraacetic acid (EDTA)-free Roche Protease Inhibitor Cocktail. The extracts were diluted eight times with PBS and clarified by centrifugation at 15000g for 10 min.

Antibody production

Antibodies against NARR were raised in rabbits using a synthetic peptide C-RQDEHSGTRAEGSR conjugated to 'Imject Maleimide Activated Ovalbumin' (Pierce). Animals were injected with 0.2 mg of the cross-linked peptide. Titer and specificity of the antisera was increased by three injections (boosts).

Mono-specific antibodies were purified on affinity columns prepared by coupling of the peptide to iodoacetate-activated gel (SulfoLink, Pierce) according to the manufacturer's protocol using 1 mg of the peptide per 1 ml of the coupling gel. The unreacted maleimide groups were quenched with 20 mM cysteine. Afterward, 1 ml of the resin was loaded onto a column and washed with 50 ml PBS. Five milliliters of the antiserum was diluted with 10 ml PBS and incubated with the resin for 4 h. After washing with 50 ml PBS, the bound antibodies were eluted with 10 ml of 0.1 M glycine-HCl, pH 2.5, and the eluates were diluted with 10 ml of PBS and concentrated in Centriprep-Ultracel YM-10 (Millipore) concentrators to a volume of 2 ml. The concentrates were dialyzed against PBS overnight and stored frozen in 50% glycerol.

Production of recombinant NARR

Partially purified NARR protein was produced in the Core Facility of the MPI of Biochemistry using the *Escherichia coli* BL21 expression system. *NARR* gene with its codons optimized for expression in *E. coli* was synthesized by GenScript. The gene was inserted into a modified version of the EMBL pETM vector pETM44-ccdB (N-HIS-MBP) using the SLIC cloning method (14). Specific elution from the Ni-Sepharose (GE HealthCare) beads was achieved by incubation with PreScission Protease (GE HealthCare), and subsequently protein products were purified by gel-filtration on HiLoad 26/60 Superdex 200 column (GE HealthCare).

Western blotting

For western analysis, proteins were transferred from sodium dodecylsulfate (SDS) gels onto the nitrocellulose membrane by electroblotting at 10 V/cm for 40 min. The proteins were cross-linked to the membrane by incubation in 0.5% (v/v) glutaraldehyde in PBS for 10 min. The membranes were blocked with 10% (v/v) normal goat serum (NGS) for 30 min prior to incubation with the primary antibodies together with 1% NGS in PBS containing 0.1% Tween 20 (PBS-T). Following 2-h incubation, the membrane was extensively washed with PBS-T. Primary antibodies were visualized on the membrane with peroxidase-conjugated immunoglobulins. After the detection of antibodies, blots were stained with Amido Black stain (Sigma, St Louis, MO, USA). The anti-NARR antibodies (concentration 1.0 $\mu\text{g}/\mu\text{l}$) were used in 1:4000 dilution. The following commercial antibodies were used—mouse anti-EIF5A (BD Biosciences, 611976), mouse anti-UBF (Abnova, H00007343-M01) and rabbit anti-ERF1 (Sigma, E8156)—all in 1:1000 dilutions.

Indirect IF

MCF7 cells were fixed with 4% paraformaldehyde in PBS for 15 min and permeabilized with 0.2% Triton X-100 in PBS for 20 min. Non-specific binding was blocked by incubation in 10% (v/v) NGS. The monoclonal mouse anti-EIF5A (BD Biosciences, 611976), anti-UBF (Abnova, H00007343-M01) and anti-fibrillarin antibodies (Genetex, GTX24566) were used in a 1:500, 1:1000 and

1:500 dilution, respectively. Anti-NARR antibodies were used in 1:500 dilution. Antibodies were visualized using secondary antibodies conjugated to AlexaFluor dyes 488 or 594 (Invitrogen) in 1:700 dilution. Nuclei were counterstained with 4',6-diamidino-2-phenylindole (DAPI). Coverslips were mounted with MobiGLOW (Mobitec) mounting medium. Analyses were performed using LEICA TCS SP2 and Zeiss Axioplan 2 imaging systems.

Combined fluorescence *in situ* hybridization and IF

The procedure was performed as per guidelines outlined previously (15) with minor modifications as follows. MCF7 cells were grown on microscope slides (Superfrost Ultra Plus, Thermo Scientific), washed with PBS and fixed with 4% paraformaldehyde in PBS for 20 min. The cells were permeabilized for 25 min with 0.25% Triton X-100 in PBS followed by 0.5% Saponin in PBS for 10 min. Afterwards, the slides were depurinated in 0.1 N HCl for 10 min. For denaturation, the slides were placed in 70% formamide/2× saline-sodium citrate (SSC) for 3 min at 74°C followed by 1 min in 50% formamide/2× SSC at 74°C. Acrocentric chromosome p arms red fluorophore probe (LPE NOR Aquarius probe, Cytocell), 25 ng per slide, was denatured separately for 2 min at 74°C. Subsequently, the probe was applied to the slide, covered with a coverslip, sealed with a Fixogum rubber solution which was then allowed to dry. After that, the sample and the probe were heated at 74°C for 2 min and incubation was performed for 16 h in a humidified chamber at 37°C. Following the incubation, the coverslip was removed and the slides were washed in 0.4× SSC at 72°C for 2 min, and then with 2×SSC at room temperature (RT) for 2 min, followed by three washes in PBS for 2 min at RT. Indirect IF staining was performed as described above using 1:300 dilution of the anti-NARR primary antibody. The primary antibody was visualized using 1:500 dilution of the secondary antibody conjugated to AlexaFluor 488 dye (Invitrogen). The nuclei were counterstained with DAPI. Coverslips were mounted with FluorPreserve (Calbiochem) mounting medium. The Zeiss Axioplan 2 system was used for image acquisition.

Pull-down experiments

The peptide CSPRPRNARRGTIRPRNARRGTIRPRNARRGSARAR comprising three NARR repeats and carrying N-terminal cysteine for coupling was synthesized in-house by solid phase synthesis. The peptide was coupled to SulfoLink gel (Pierce) according to the manufacturer's protocol. SulfoLink gel blocked with cysteine was utilized as a control column.

For SILAC-based pull downs, MCF7 nuclear extracts from either 'light' or 'heavy'-labeled cells were applied to the counterpart (control and peptide-bound) columns (16). The columns were washed with PBS. Elution was performed with 0.1 M glycine-HCl buffer, pH 2.2. The eluates from the control and peptide-bound columns were mixed and analyzed by liquid chromatography tandem mass spectrometry (LC-MS/MS).

For western blotting, MCF7 nuclear extracts were applied to the peptide-coupled and control columns.

The columns were washed with PBS and elution was performed in two steps—first, with 1 mM solution of NARR peptide and, secondly, with 2% SDS.

Identification of phosphorylation sites

HeLa cells at 80% confluence were treated with 0.1 µg/ml nocodazole for 16 h and then harvested and extracted with 0.2 N HCl (as described above). Proteins were precipitated with 25% (w/v) CF₃COOH as described previously (17) and the pellets were dissolved in 2% (w/v) SDS, Tris-HCl, pH 8.0. The proteins were digested with trypsin according to the filter aided sample preparation (FASP) procedure (13) using an Amicon Ultra 15 Ultracel 30 k (Millipore) device as described previously (18). Five milligrams of tryptic peptides was acidified with CF₃COOH and phosphopeptides were enriched on 10 mg TiO₂-beads according to ref. (19) with minor modifications (18). The phosphopeptide-enriched fractions were vacuum-dried and fractionated into six fractions using a pipette-based SAX column (20).

LC-MS/MS analysis

Proteins were separated by SDS-PAGE. In-gel protein tryptic digestion was performed as described (21), the peptide mixtures were desalted using in-house made C₁₈ StageTips (22), vacuum-dried and reconstituted in 0.5% acetic acid prior to analysis. The LC-MS/MS analytical setup was similar to that described before (4). Briefly, the samples were injected onto an in-house made 15 cm capillary emitter column [inner diameter 75 µm, packed with 3 µm ReproSil-Pur C18-AQ media (Dr Maisch GmbH, Ammerbuch-Entringen, Germany)], using a Proxeon EASY-nLC system (Proxeon Biosystems, Odense, Denmark, now Thermo Fisher Scientific). The LC setup was connected to an LTQ Orbitrap mass spectrometer equipped with a nanoelectrospray ion source (Proxeon Biosystems). The peptide mixtures were separated with gradients from 5% to 40% CH₃CN in 0.5% acetic acid. Data-dependent acquisition was employed. MS spectra were acquired in the Orbitrap analyzer with the resolution of 60 000 (at *m/z* of 400); MS/MS spectra were acquired in the ion trap. Data analysis was performed with the MaxQuant software (23).

The nominal *m/z* (391) of the N-terminal doubly-charged VGQPQR tryptic product of NARR coincides with the *m/z* of one of the most common MS contaminants, dioctyl phthalate, in its singly-protonated form. This fact complicated identification of the N-terminal peptide of NARR. However, this peptide was nevertheless identified in the SILAC-labeled cell lines due to the mass shift introduced by 'heavy' arginine, leading to a mass devoid of interference by dioctyl phthalate.

RESULTS

Analysis of unassigned MS data reveals a novel protein

In the course of our proteomic work on the nucleus, we employed a biochemical protein extraction method aimed at small nuclear proteins. This procedure employs

extraction with diluted acids, which is a commonly used technique for isolation of highly charged nuclear proteins. In this workflow, early denaturation inhibits protein degradation and loss of post-translational modifications, such as phosphorylation, acetylation or methylation. It is not only a standard method for the isolation of histones and high-mobility group proteins (24), but also is very useful for the analysis of many other similar protein classes (17,25). A high resolution MS-based proteomics analysis of proteins extracted with 0.2N HCl from isolated MFC7 nuclei led to the identification of hundreds of proteins (data not shown); however, a portion of the MS/MS data remained unassigned to

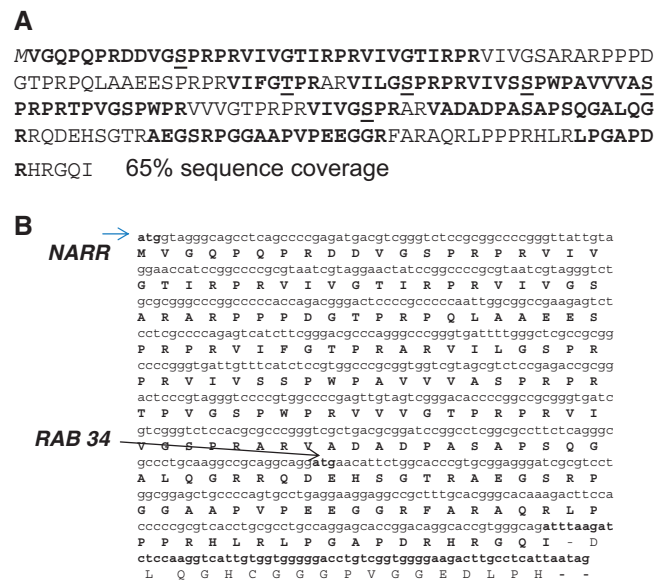


Figure 1. (A) Mass spectrometric identification of peptides matching the NARR sequence. Identified sequence (*bold*) covers 65% of the predicted protein primary structure. Experimentally identified phosphorylation sites are underlined (B) Genomic structure of the *narr* and *rab34* genes. Start of *narr* transcription is from the same exon as *rab34*. The *narr* start codon is located 443 base pairs upstream of the *rab34* start.

any known protein in the International Protein Index database. With a view to identify potential novel proteins, we then searched the unassigned data against a human EST database (see 'Materials and Methods' section). This revealed 10 peptides matching the EST CX872465 that completely covers the mRNA sequence of a so far undiscovered 198 amino acid residue proteins (Figure 1A; Supplementary Figure S1; Table 1). Peptide masses were measured with low to subparts per million (ppm) mass accuracy and the number of peptides matched to this small protein implies completely unambiguous protein identification (65% percent sequence coverage). The sequence defines a novel gene which overlaps *rab34* gene on chromosome 17 at q11.2. The transcript sequence revealed by the EST shows that the coding sequence of the protein overlaps the first 3 of the 10 exons of *rab34*. The alternative translation start site is located upstream of the *rab34* start codon (Figure 1B; Supplementary Figure S2) and the complete coding sequence of *NARR* is contained in *rab34* gene. The MS data verified the N-terminus of the protein independently of the transcript data by direct sequence information of the N-terminal peptide of the protein (Table 1).

The novel protein contains multiple copies of a nine amino acid residue repeat

Based on its amino acid sequence, the novel 198 residue protein is highly basic (pI 12.2), rich in proline (18.2%) and arginine (17.2%) but does not contain any lysines. Analysis of the primary structure revealed the presence of 13 Nine Amino acid Residue Repeats and therefore we named the protein 'NARR' (Figure 2A). Each repeat is flanked on both sides by a PR dipeptide, contains a hydrophobic core with consensus sequence VIVG and nests at S/T phosphorylation site (Figure 2B). The repeats constitute 2/3 of the N-terminal portion of the protein while the C-terminal portion does not possess any obvious structural signatures. Secondary structure analysis predicts the protein to be unstructured.

The occurrence of NARR is restricted to mammals. The protein is highly conserved across species (Figure 2A) with

Table 1. Peptides of the NARR protein identified by LC-MS/MS

Residues	Sequence	Modification ^a	Theoretical mass	Observed mass	Mass deviation (ppm)
1-7	VGQPQR ^h	-	790.4325	790.4329	-0.50
8-16	DDVGSPPR	pS-12	1077.4604	1077.4600	0.37
17-34	VIVGTIRPR	-	1009.6396	1009.6381	1.48
64-70	VIFGTPR	-	788.4545	788.4577	-4.05
64-70	VIFGTPR	pT-68	868.4208	868.4212	-0.46
73-79	VILGSPR	-	740.4545	740.4541	0.54
73-81	VILGSPRPR	-	993.6083	993.6114	-3.12
73-81	VILGSPRPR	pS-77	1073.5747	1073.5745	0.19
82-97	VIVSSPWPVVVASPR	-	1662.9457	1662.9474	-1.02
82-99	VIVSSPWPVVVASPRPR	pS-86; pS-95	2076.0323	2076.0325	-0.09
100-108	TPVGSPPWR	-	995.5189	995.5169	2.01
118-124	VIVGSPR	pS-122	806.4052	806.4046	0.74
127-144	VADADPASAPSQALQGR	-	1709.8333	1709.8357	-1.40
154-171	AEGSRPGGAAPVPEEGGR	-	1692.8179	1692.8136	2.54
186-192	LPGAPDR	-	724.3868	724.3847	2.90

^aThe phosphorylated peptides were identified only in cells treated with nocodazole; R^h - heavy-labeled arginine

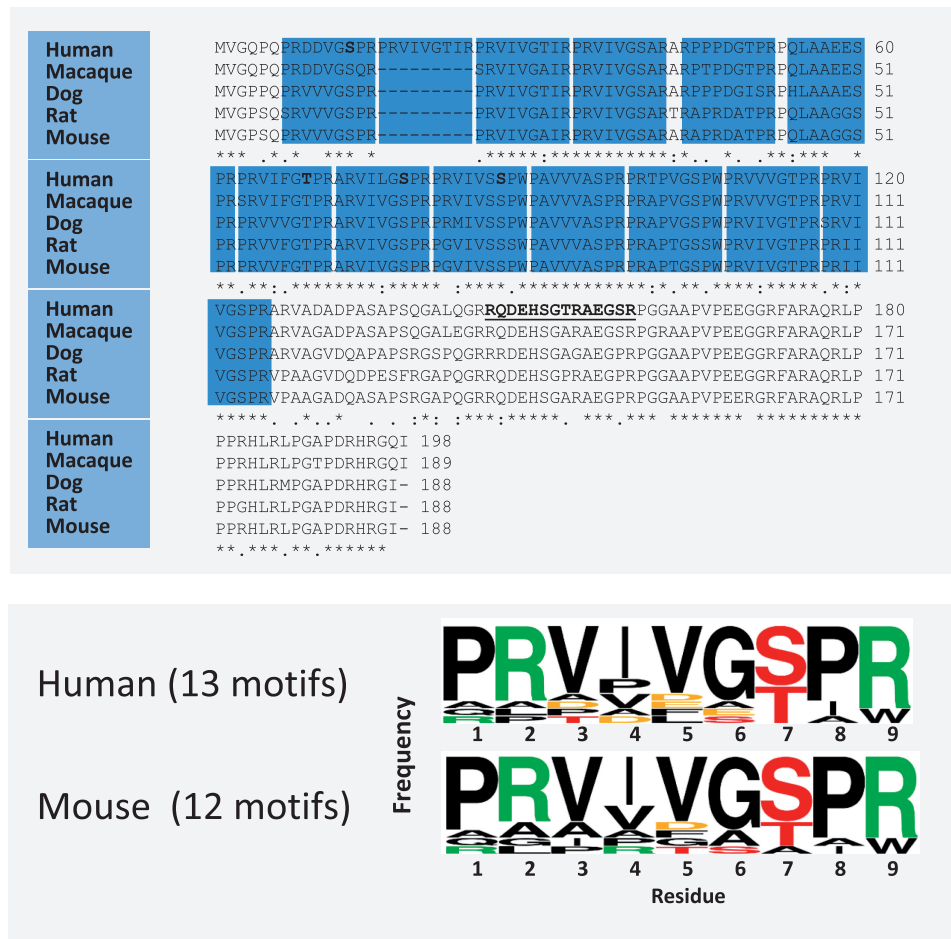


Figure 2. Primary structure of NARR. (A) Multiple alignment of NARR sequences. NAR repeats are boxed in blue. Sequence of the peptide used for antibody production is underlined. (B) WebLOGO analysis of human and mouse NARRs. S/T residues in position 7 are phosphorylated.

only one obvious difference, whereas human NARR comprises 13 repeats and predicted NARR proteins in other mammals have only 12 repeats (Figure 2A). The repeat phosphorylation site matches the consensus sequence of cyclin-dependent kinase 1 (CDK1). CDK1 is particularly active during mitosis and is a regulator of cell cycle and transcription (26). Since in normally growing cell culture only a few percent of cells undergo mitosis at any given time, we did not identify phosphorylation of NARRs in untreated cells. To specifically probe for phosphorylation of the protein during the cell cycle, we arrested cells in mitosis using nocodazole. After enrichment of phosphopeptides, we identified six phosphorylation sites—all localized to the CDK1 consensus sites in the repeats (Figure 2; Supplementary Figure S3; Table 1).

NARR protein is expressed ubiquitously and localizes to nucleoli

To further characterize the NARR protein, we raised antisera in rabbits against the peptide residues 146–159 and affinity purified the antibodies. The antibody specificity was tested on western blots. We found that the antibodies strongly and specifically reacted with the

bacterially expressed NARR protein and the endogenous protein in whole cell lysates (Figure 3A). Fractionation of cells into nuclei and cytoplasm revealed that NARR localizes to nuclei and is absent from the cytosol (Figure 3B and C). The protein is extractable from isolated nuclei with 0.8 M salt.

To shed some light on the expression of NARR in different tissues, we prepared whole lysates from five mouse organs. We found the single band of NARR in all lysates (Figure 3D and E), indicating that the protein is ubiquitously expressed in these tissues. Finally, by IF analysis we found that NARR localizes to nucleoli. The speckled appearance suggested NARR’s possible association with specific nucleolar complexes (Figure 3F–H).

NARR and UBF have dissimilar localization signatures

The distinctive nucleolar distribution pattern of NARR raised the question of colocalization with upstream binding factor (UBF) and associated proteins (27). UBF protein binds to the control elements of the rRNA promoter and is a crucial component of the RNA polymerase Pol I complex (28,29) participating in

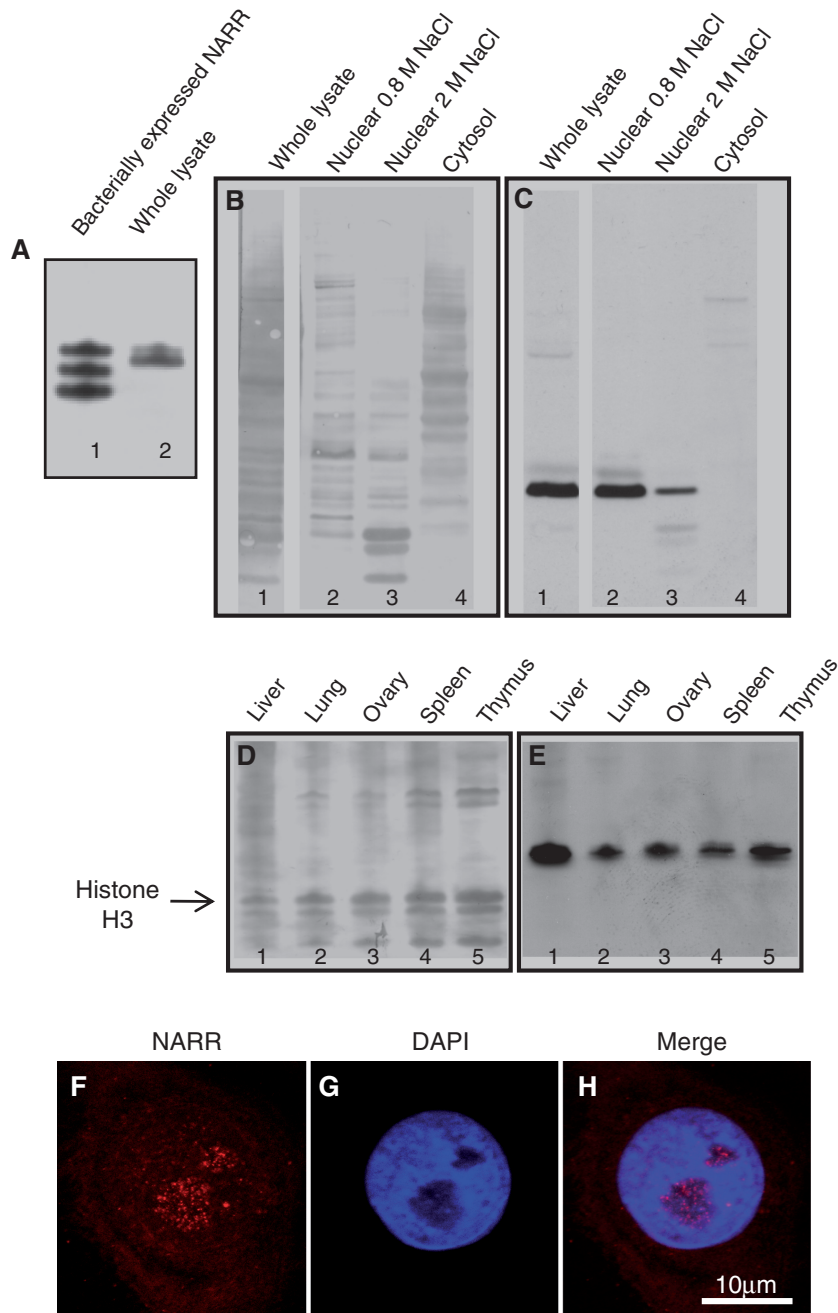


Figure 3. Occurrence of NARR in human cells and mouse tissues. Western blot of the partially purified and truncated recombinant NARR (lane 1) and of whole MCF7 lysate (lane 2) using polyclonal anti-NARR antibody (A). Whole HeLa cell lysates (1), their nuclear extracts (2,3) and cytosol were separated on SDS-PAGE (B), blotted to nitrocellulose, and probed with polyclonal anti-NARR antibody (C). Whole lysates of mouse tissues were separated on SDS-PAGE (D), blotted to nitrocellulose, and probed with polyclonal anti-NARR antibody (E). (F–H) Indirect IF confocal microscopy images of HeLa cells probed with anti-NARR antibodies (F), DAPI staining (G), merged F and G (H).

pre-ribosomal RNA (pre-rRNA) production. Our analysis showed that NARR and UBF only partially colocalize during interphase (Figure 4A). In mitotic cells, NARR seems to disappear or disperse, whereas UBF localizes in distinct areas (Figure 4B). Actinomycin D (ActD), a potent inhibitor of rRNA production, causes segregation of different nucleolar regions. Upon treatment with ActD, UBF is known to migrate to the nucleolar caps (27). It turned out that NARR does not follow the UBF path into

the nucleolar caps after the ActD treatment (Figure 4C). The results suggest that NARR disengages from the UBF-associated machinery upon inhibition of rDNA transcription.

NARR only partially colocalizes with fibrillar

UBF is known to reside both in the nucleolar fibrillar (FC) and dense fibrillar (DFC) centers. The transcription of the rRNA genes by RNA polymerase I is believed to take

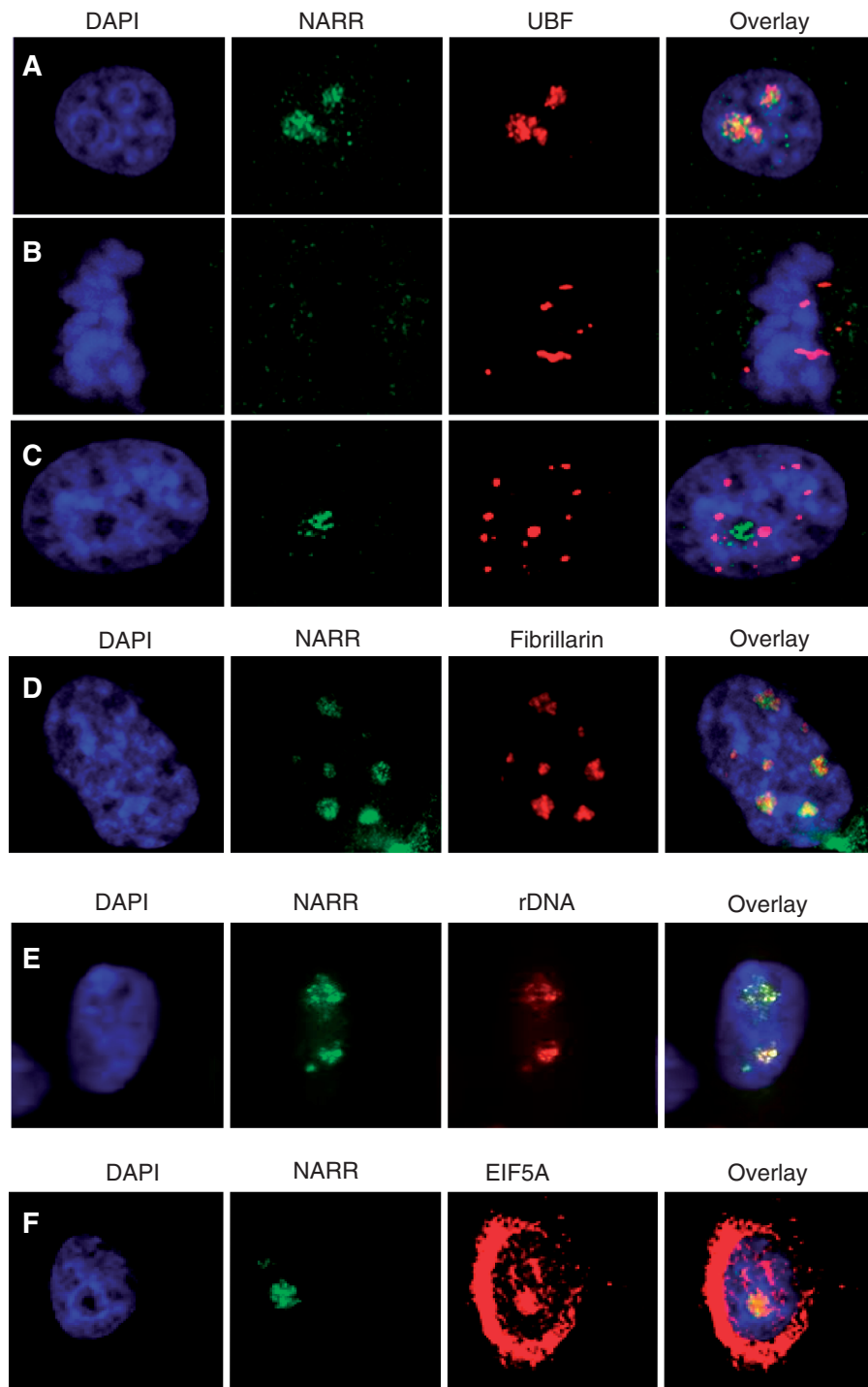


Figure 4. Indirect IF confocal microscopy images of MCF7 cells probed with anti-NARR, anti-UBF, anti-fibrillarin and anti-EIF5A antibodies. Visualization is with Alexa-488 (NARR, green) and Alexa-594 (UBF, fibrillarin and EIF5A, red)-coupled secondary antibodies (A–F). (A) In interphase, NARR only partially colocalizes with UBF in nucleoli; (B) in mitosis, NARR disappears or disperses, whereas UBF is restricted to distinct areas; (C) upon ActD treatment UBF moves to the nucleolar caps but NARR's localization pattern is unaffected; (D) NARR only partially colocalizes with fibrillarin in interphase nucleoli; (E) in interphase NARR and rDNA clusters colocalize; visualization in (E) is with combined FISH (LPE NOR probe, red) and IF (anti-NARR primary, Alexa-488 secondary antibodies, green); (F) EIF5A is visible both in the cytoplasm and nucleus; in the nucleus EIF5A is concentrated in nucleoli where it is diffusely distributed.

place either in DFC or on the border between FC and DFC (30). While trying to correctly position NARR within nucleolus, we used antibodies against fibrillarin, the RNA-2'-O-methyltransferase present solely in the

dense fibrillar centers of nucleolus. Our analysis shows only partial colocalization of NARR with fibrillarin in interphase cells (Figure 4D), suggesting that NARR may occupy a special niche near the rDNA areas.

NARR colocalizes with rDNA gene clusters

Because it is known that UBF only associates with active rDNA clusters in NORs (6), the sites of rRNA transcription, our findings suggested that NARR localization may be linked to rDNA. Combined fluorescence *in situ* hybridization (FISH) and immunofluorescence assay (immuno-FISH) was performed using a probe against acrocentric chromosome p arms and anti-NARR antibody. This showed that NARR and rDNA clusters indeed colocalize in interphase cells (Figure 4E) providing evidence for NARR's association with NORs.

NARR interacts with eukaryotic translation initiation factor 5A and ERF1

The presence of NARR repeats is a striking unique property of the protein and we hypothesized that the repeats may constitute protein interaction sites. To identify interacting proteins, we performed SILAC-based pull-down experiments (31,32) using a synthetic bait-peptide representing a fragment of NARR (residues 13–45) containing three repeats (Figure 5A). The consequent analysis highlighted two proteins, ERF1 and EIF5A, as potentially interesting NARR interactors

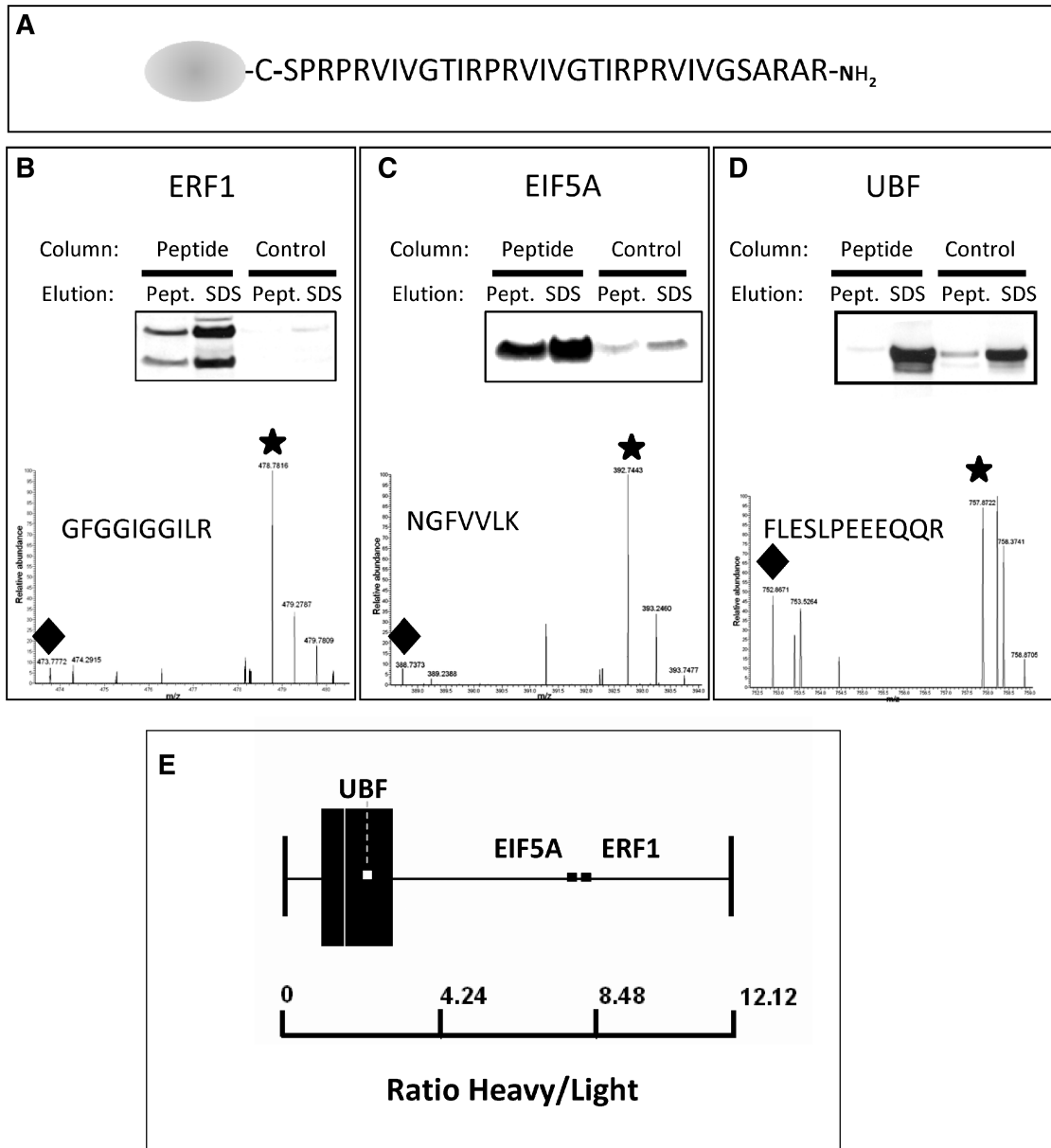


Figure 5. SILAC-based pull-down experiments with immobilized peptide containing three NARR repeats. (A) Sequence of the peptide used for the pull-down. (B) ERF1 and (C) EIF5A were found with SILAC ratios indicating specific binding. Representative spectra of EIF5A, ERF1 and UBF peptides obtained by the SILAC-based profiling are shown in corresponding boxes. Asterisks and diamonds indicate relative amount of proteins bound to the NARR-peptide column and control column, respectively. The interactions were confirmed by western blotting. (D) UBF does not specifically interact with the NARR peptide. (E) Boxplot of the results of the SILAC-based pull-down experiment indicates the interactions of EIF5A and ERF1 with the NARR peptide.

(Figure 5B, C and E). They showed SILAC heavy to light ratios greater than three and are also present in the nucleolar proteome database (7). To verify these interactions, we analyzed the eluates from NARR peptide and control by western blotting using antibodies against these proteins (inserts in Figure 5B and C). Indeed, this experiment showed that ERF1 and EIF5A were bound only to the NARR peptide column and that they were specifically eluted with the free NARR peptide but not with the control. Conversely, UBF did not show a significant SILAC ratio and western blots also indicated non-specific binding nature of UBF to the NARR peptide column (Figure 5D).

Although its name suggests a specific role in translation, published reports of functional roles of eukaryotic translation initiation factor 5A (EIF5A) are ambiguous (33). This ubiquitously expressed protein contains a nuclear localization signal (34) and is the only protein shown to carry hypusine—the post-translationally modified lysine residue, N(epsilon)-(4-amino-2-hydroxy-butyl)lysine. This modification was found to be essential for cell viability (35,36). Interestingly, EIF5A is known to be associated with HIV-1 Rev protein which partly resides in nucleoli and orchestrates the export of the viral mRNAs out of the nucleus (37). Unlike the ERF1 antibody which functioned only in western blotting, the EIF5A antibody proved to work successfully both for western blotting and IF. We used it to probe EIF5A cellular distribution in MCF7 cells and found that even though EIF5A is detected both in cytoplasmic and nuclear cellular compartments, in the nucleus it is concentrated in the nucleolar regions (Figure 4F). This link to nucleolar processes may be a clue to the essential role EIF5A plays in cell viability.

DISCUSSION

Correct identification of translational start sites is vital for understanding of protein structure, function and transcriptional regulation. Until recently, the most accepted ‘scanning’ hypothesis of translational site initiation in eukaryotic mRNAs suggested that proximity to the 5′ mRNA end plays a dominant role in identifying the start codon. This ‘position effect’ becomes apparent in cases where a mutation creates the AUG codon upstream of the normal start site of translation (38). Such a monocistronic model (one mRNA—one protein) has recently come under scrutiny and, nowadays it is acknowledged that an mRNA molecule can carry multiple sites of translation recognizable by the ribosomal machinery (39). It has also been suggested that the diversity of translation start sites may define increased complexity of the human short ORFeome (40).

The annotated translational start sites contained in genome databases are often predicted using bioinformatics and rarely verified experimentally, and hence it is not known how accurate they are (41). In general, computational gene predictions can have high false positive rates and in principle require experimental verification as they do not by themselves establish that a predicted gene is

definitely going to be translated into a protein (42). Proteomics identifies protein products of the genes and therefore reports on the final output of the translation machinery. Hence, in addition to confirming predicted genes, proteomics can supply unique information potentially leading to discovery of novel genes and gene arrangements. Our results do not yet establish the mechanism of gene expression of the two proteins. While we have no data supporting the existence of two different mRNAs, this cannot be formally ruled out at this stage. This and the mode of translation initiation of *NARR* are interesting open questions for the future.

Overlapping genes are made possible by the degeneracy of the genetic code and are most common in viruses where they are necessitated by the compactness of viral genomes. It is also often observed in organisms with a limited genome size (43). There are not many reports on the protein products of mammalian overlapping genes, notwithstanding the fact that a considerable fraction of mammalian genes is predicted to have overlaps (44). The most well-known example of genuine protein products of overlapping genes in humans is the case of mitochondrial ATPase-6 and ATPase-8 subunits translated from the same mRNA transcript and in the same direction (although the mitochondrial genome is very different to the nuclear genome) (45). Other examples include Tenascin-X, the extracellular matrix protein encoded by the gene overlapping *P450c21B* Steroid 21-hydroxylase gene (46); the tumor suppressors ARF and p16 encoded by overlapping genes transcribed in the same direction (47); very-long-chain acyl-CoA dehydrogenase (VLCAD) and discs-large-related 4 (DLG4) proteins whose overlapping genes are transcribed in opposite directions (48) and neuropeptide Y receptor subtypes Y1 and Y5 whose overlapping genes are also transcribed in opposite directions (49). Despite these well-established examples, the extent of overlap of various genomic sequence messages, in particular in the genomes of higher eukaryotes, is an open issue for evolutionary theoreticians, geneticists and bioinformaticians.

Whereas different methods have been developed for prediction of overlapping genes (44), so far, proteomics methods using unassigned peptide/protein sequences have hardly been employed for the experimental verification of products of such genes. Given the rapid evolution of MS-based proteomics, we suggest that these methods should now be used for detecting overlapping genes in a systematic manner. The possibility of additional ‘protein output’ of the same gene, with potentially different functions to its conventional protein product raises interesting functional questions that should be taken into account. For example, when performing knock-down experiments on *rab34*, our results suggest that one would have to consider not only the effects of such a treatment with regards to *RAB34*, a member of the Rab GTPase family associated with the Golgi apparatus (50), but also on the nucleolar protein *NARR*. While difficult for the above reasons, we believe that further research into *NARR*’s intriguing link to rDNA clusters may provide interesting clues about its roles in the nucleolar apparatus.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr Stephan Uebel for the expert peptide synthesis and Dr Brian McStay for providing immuno-FISH protocol. We are grateful to Dr Sabine Suppmann and Mrs Claudia Franke for the recombinant protein production.

FUNDING

Max-Planck Society for the Advancement of Science, by the European Commission's 7th Framework Program (grant agreement HEALTH-F4-2008-201648/PROSPECTS); Munich Center for Integrated Protein Science (CIPSM). Funding for open access charge: Munich Center for Integrated Protein Science (CIPSM).

Conflict of interest statement. None declared.

REFERENCES

- Aebersold, R. and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature*, **422**, 198–207.
- Boguski, M.S., Lowe, T.M. and Tolstoshev, C.M. (1993) dbEST—database for “expressed sequence tags”. *Nat. Genet.*, **4**, 332–333.
- Kuster, B., Mortensen, P., Andersen, J.S. and Mann, M. (2001) Mass spectrometry allows direct identification of proteins in large genomes. *Proteomics*, **1**, 641–650.
- Zougman, A., Ziolkowski, P., Mann, M. and Wisniewski, J.R. (2008) Evidence for insertional RNA editing in humans. *Curr. Biol.*, **18**, 1760–1765.
- Pederson, T. (1998) The plurifunctional nucleolus. *Nucleic Acids Res.*, **26**, 3871–3876.
- McStay, B. and Grummt, I. (2008) The epigenetics of rRNA genes: from molecular to chromosome biology. *Annu. Rev. Cell. Dev. Biol.*, **24**, 131–157.
- Andersen, J.S., Lyon, C.E., Fox, A.H., Leung, A.K., Lam, Y.W., Steen, H., Mann, M. and Lamond, A.I. (2002) Directed proteomic analysis of the human nucleolus. *Curr. Biol.*, **12**, 1–11.
- Pendle, A.F., Clark, G.P., Boon, R., Lewandowska, D., Lam, Y.W., Andersen, J., Mann, M., Lamond, A.I., Brown, J.W. and Shaw, P.J. (2005) Proteomic analysis of the Arabidopsis nucleolus suggests novel nucleolar functions. *Mol. Biol. Cell*, **16**, 260–269.
- Scherl, A., Coute, Y., Deon, C., Calle, A., Kindbeiter, K., Sanchez, J.C., Greco, A., Hochstrasser, D. and Diaz, J.J. (2002) Functional proteomic analysis of human nucleolus. *Mol. Biol. Cell*, **13**, 4100–4109.
- Andersen, J.S., Lam, Y.W., Leung, A.K., Ong, S.E., Lyon, C.E., Lamond, A.I. and Mann, M. (2005) Nucleolar proteome dynamics. *Nature*, **433**, 77–83.
- Boulton, S., Westman, B.J., Hutten, S., Boisvert, F.M. and Lamond, A.I. The nucleolus under stress. *Mol. Cell*, **40**, 216–227.
- Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A. and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell Proteomics*, **1**, 376–386.
- Wisniewski, J.R., Zougman, A., Nagaraj, N. and Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nat. Methods*, **6**, 359–362.
- Li, M.Z. and Elledge, S.J. (2007) Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat. Methods*, **4**, 251–256.
- Bridger, J.M., Kill, I.R. and Lichter, P. (1998) Association of pKi-67 with satellite DNA of the human genome in early G1 cells. *Chromosome Res.*, **6**, 13–24.
- Schulze, W.X. and Mann, M. (2004) A novel proteomic screen for peptide-protein interactions. *J. Biol. Chem.*, **279**, 10756–10764.
- Wisniewski, J.R., Zougman, A., Kruger, S., Ziolkowski, P., Pudelko, M., Bebenek, M. and Mann, M. (2008) Constitutive and dynamic phosphorylation and acetylation sites on NUCKS, a hypermodified nuclear protein, studied by quantitative proteomics. *Proteins*, **73**, 710–718.
- Wisniewski, J.R., Nagaraj, N., Zougman, A., Gnäd, F. and Mann, M. Brain phosphoproteome obtained by a FASP-based method reveals plasma membrane protein topology. *J. Proteome Res.*, **9**, 3280–3289.
- Larsen, M.R., Thingholm, T.E., Jensen, O.N., Roepstorff, P. and Jorgensen, T.J. (2005) Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol. Cell Proteomics*, **4**, 873–886.
- Wisniewski, J.R., Zougman, A. and Mann, M. (2009) Combination of FASP and stageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *J. Proteome Res.*, **8**, 5674–5678.
- Shevchenko, A., Wilm, M., Vorm, O. and Mann, M. (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.*, **68**, 850–858.
- Rappsilber, J., Ishihama, Y. and Mann, M. (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.*, **75**, 663–670.
- Cox, J., Matic, I., Hilger, M., Nagaraj, N., Selbach, M., Olsen, J.V. and Mann, M. (2009) A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. *Nat. Protoc.*, **4**, 698–705.
- Johns, E.W. and Butler, J.A. (1962) Further fractionations of histones from calf thymus. *Biochem. J.*, **82**, 15–18.
- Zougman, A. and Wisniewski, J.R. (2006) Beyond linker histones and high mobility group proteins: global profiling of perchloric acid soluble proteins. *J. Proteome Res.*, **5**, 925–934.
- Malumbres, M. and Barbacid, M. (2005) Mammalian cyclin-dependent kinases. *Trends Biochem. Sci.*, **30**, 630–641.
- Valdez, B.C., Henning, D., So, R.B., Dixon, J. and Dixon, M.J. (2004) The Treacher Collins syndrome (TCOF1) gene product is involved in ribosomal DNA gene transcription by interacting with upstream binding factor. *Proc. Natl Acad. Sci. USA*, **101**, 10709–10714.
- Grummt, I. (1999) Regulation of mammalian ribosomal gene transcription by RNA polymerase I. *Prog. Nucleic Acid Res. Mol. Biol.*, **62**, 109–154.
- Chen, D., Belmont, A.S. and Huang, S. (2004) Upstream binding factor association induces large-scale chromatin decondensation. *Proc. Natl Acad. Sci. USA*, **101**, 15106–15111.
- Cheutin, T., O'Donohue, M.F., Beorchia, A., Vandelaer, M., Kaplan, H., Defever, B., Ploton, D. and Thiry, M. (2002) Three-dimensional organization of active rRNA genes within the nucleolus. *J. Cell Sci.*, **115**, 3297–3307.
- Mann, M. (2006) Functional and quantitative proteomics using SILAC. *Nat. Rev. Mol. Cell Biol.*, **7**, 952–958.
- Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W., van Schaik, F.M., Varier, R.A., Baltissen, M.P., Stunnenberg, H.G., Mann, M. and Timmers, H.T. (2007) Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell*, **131**, 58–69.
- Zanelli, C.F. and Valentini, S.R. (2007) Is there a role for eIF5A in translation? *Amino Acids*, **33**, 351–358.
- Parreiras, E.S.L.T., Gomes, M.D., Oliveira, E.B. and Costa-Neto, C.M. (2007) The N-terminal region of eukaryotic translation initiation factor 5A signals to nuclear localization of the protein. *Biochem. Biophys. Res. Commun.*, **362**, 393–398.
- Schnier, J., Schwelberger, H.G., Smit-McBride, Z., Kang, H.A. and Hershey, J.W. (1991) Translation initiation factor 5A and its hypusine modification are essential for cell viability in the yeast *Saccharomyces cerevisiae*. *Mol. Cell Biol.*, **11**, 3105–3114.
- Park, M.H., Lee, Y.B. and Joe, Y.A. (1997) Hypusine is essential for eukaryotic cell proliferation. *Biol. Signals*, **6**, 115–123.
- Ruhl, M., Himmelspach, M., Bahr, G.M., Hammerschmid, F., Jaksche, H., Wolff, B., Aschauer, H., Farrington, G.K., Probst, H.,

- Bevec, D. *et al.* (1993) Eukaryotic initiation factor 5A is a cellular target of the human immunodeficiency virus type 1 Rev activation domain mediating trans-activation. *J. Cell Biol.*, **123**, 1309–1320.
38. Kozak, M. (2002) Pushing the limits of the scanning mechanism for initiation of translation. *Gene*, **299**, 1–34.
39. Kochetov, A.V. (2008) Alternative translation start sites and hidden coding potential of eukaryotic mRNAs. *Bioessays*, **30**, 683–691.
40. Oyama, M., Kozuka-Hata, H., Suzuki, Y., Semba, K., Yamamoto, T. and Sugano, S. (2007) Diversity of translation start sites may define increased complexity of the human short ORFeome. *Mol. Cell Proteomics*, **6**, 1000–1006.
41. Smollett, K.L., Fivian-Hughes, A.S., Smith, J.E., Chang, A., Rao, T. and Davis, E.O. (2009) Experimental determination of translational start sites resolves uncertainties in genomic open reading frame predictions - application to *Mycobacterium tuberculosis*. *Microbiology*, **155**, 186–197.
42. Bork, P. (2000) Powers and pitfalls in sequence analysis: the 70% hurdle. *Genome Res.*, **10**, 398–400.
43. Peleg, O., Kirzhner, V., Trifonov, E. and Bolshoy, A. (2004) Overlapping messages and survivability. *J. Mol. Evol.*, **59**, 520–527.
44. Makalowska, I., Lin, C.F. and Makalowski, W. (2005) Overlapping genes in vertebrate genomes. *Comput. Biol. Chem.*, **29**, 1–12.
45. Fearnley, I.M. and Walker, J.E. (1986) Two overlapping genes in bovine mitochondrial DNA encode membrane components of ATP synthase. *EMBO J.*, **5**, 2003–2008.
46. Bristow, J., Tee, M.K., Gitelman, S.E., Mellon, S.H. and Miller, W.L. (1993) Tenascin-X: a novel extracellular matrix protein encoded by the human XB gene overlapping P450c21B. *J. Cell Biol.*, **122**, 265–278.
47. Freedberg, D.E., Rigas, S.H., Russak, J., Gai, W., Kaplow, M., Osman, I., Turner, F., Randerson-Moor, J.A., Houghton, A., Busam, K. *et al.* (2008) Frequent p16-independent inactivation of p14ARF in human melanoma. *J. Natl Cancer Inst.*, **100**, 784–795.
48. Zhou, C. and Blumberg, B. (2003) Overlapping gene structure of human VLCAD and DLG4. *Gene*, **305**, 161–166.
49. Herzog, H., Darby, K., Ball, H., Hort, Y., Beck-Sickinger, A. and Shine, J. (1997) Overlapping gene structure of the human neuropeptide Y receptor subtypes Y1 and Y5 suggests coordinate transcriptional regulation. *Genomics*, **41**, 315–319.
50. Wang, T. and Hong, W. (2002) Interorganellar regulation of lysosome positioning by the Golgi apparatus through Rab34 interaction with Rab-interacting lysosomal protein. *Mol. Biol. Cell*, **13**, 4317–4332.