ARTICLE
Journal of Cellular Biochemistry 114:120–133 (2013)

Journal of Cellular Biochemistry

Characterization of Human NLZ1/ZNF703 Identifies Conserved Domains Essential for Proper Subcellular Localization and Transcriptional Repression

Isabel Pereira-Castro,[1,2] Ângela M. Sousa Costa,[1,2] Maria José Oliveira,[2,3] Inês Barbosa,[1] Ana Sofia Rocha,[1] Luisa Azevedo,[1,4] and Luís Teixeira da Costa[1,2,5*]

[1]*IPATIMUP—Institute of Molecular Pathology and Immunology of the University of Porto, Porto, Portugal*
[2]*FMUP—Faculty of Medicine, University of Porto, Porto, Portugal*
[3]*INEB—Institute for Biomedical Engineering, Porto, Portugal*
[4]*FCUP—Faculty of Sciences, University of Porto, Porto, Portugal*
[5]*ICAAM—Institute of Agricultural and Environmental Mediterranean Sciences, University of Évora, Évora, Portugal*

ABSTRACT

NET family members have recently emerged as important players in the development of multiple structures, from the trachea of fly larvae to the vertebrate eye and human breast cancers. However, their mechanisms of action are still poorly understood, and we lack a detailed characterization of their functional domains, as well as gene expression patterns—particularly in adult mammals. Here, we present a characterization of human *NLZ1/ZNF703* (NocA-like zinc finger 1/Zinc finger 703), one of the two human NET family member genes. We show that the gene is ubiquitously expressed in adult human and mouse tissues, that three mRNA species with the same coding sequence are generated by alternative polyadenylation, and that the encoded protein contains six evolutionarily conserved domains, three of which are specific to NET proteins. Finally, we present functional evidence that these domains are necessary for proper subcellular distribution of and transcription repression by the NLZ1 protein, but not for its interaction with Groucho family co-repressors. J. Cell. Biochem. 114: 120–133, 2013.

KEY WORDS: ZNF703; NLZ1; NET FAMILY; CONSERVED DOMAINS; NUCLEAR LOCALIZATION

Members of the recently identified NET (for **n**ocA, Nlz, **e**lB, TLP-1) protein family [Nakamura et al., 2004] have been emerging as crucial for a variety of developmental processes. Drosophila's nocA (no ocelli) has been implicated in the development of the flies' embryonic brain [Cheah et al., 1994], larval trachea [Dorfman et al., 2002] and adult ocelli [Ashburner et al., 1982], wings and legs [Weihe et al., 2004] and its homologue elB (elbow B) has also been implicated in *Drosophila*'s leg and tracheal development [Ashburner et al., 1982; Dorfman et al., 2002]. In *C. elegans*, the single NocA/ElB known homologue, TLP-1, is involved in male tail morphogenesis, where it controls asymmetric cell fate and cell fusion [Zhao et al., 2002]. In vertebrates, the Nlz1 (also known as Znf703) and Nlz2 (Znf503) paralogues have been shown to participate in developmental processes such as brain patterning, limb formation and optic fissure closure [Andreazzoli et al., 2001; Sagerstrom et al., 2001; Chang et al., 2004; Runko and Sagerstrom, 2004; McGlinn et al., 2008; Brown et al., 2009; Ji et al., 2009].

Abbreviations: NLZ1, NocA-like zinc finger 1; NLZ2, NocA-like zinc finger 2; GRG5/AES, Groucho-related gene 5/amino-terminal enhancer of split; EST, expressed sequence tag; HDAC, histone deacetylase; UTR, untranslated region; RACE, rapid amplification of cDNA ends.

Additional supporting information may be found in the online version of this article.

Grant sponsor: Fundação para a Ciência e a Tecnologia (FCT); Grant numbers: SFRH/BD/44264/2008, SFRH/BD/24402/2005, Ciência 2007-FCT PTDC-SAU-ONC/112511/2009, Ciência 2007-IPATIMUP/AA1, Ciência 2008-ICAAM, POCTI/CBO/48218/2002.

*Correspondence to: Luís Teixeira da Costa, Instituto de Ciências Agrárias e Ambientais Mediterrânicas, Universidade de Évora, Núcleo da Mitra, Apartado 94, 7002-774 Évora, Portugal. E-mail: ltcosta@uevora.pt

Manuscript Received: 20 October 2011; Manuscript Accepted: 26 July 2012
Accepted manuscript online in Wiley Online Library (wileyonlinelibrary.com): 6 August 2012
DOI 10.1002/jcb.24309 • © 2012 Wiley Periodicals, Inc.

Computational analyses have shown that the NET family proteins are related to the Sp/Buttonhead family of transcription factors, with which they share three conserved domains: Sp, buttonhead box (Btd box) and a C_2H_2 zinc finger [Philipsen and Suske, 1999; Suske, 1999]. Apart from this similarity, several lines of evidence indicate that NET proteins regulate gene expression, primarily as transcription repressors [Dorfman et al., 2002; Runko and Sagerstrom, 2003; Nakamura et al., 2008; Ji et al., 2009; Slorach et al., 2011]: several NET proteins have been shown to interact with members of the Groucho family of co-repressors or Histone Deacetylases (HDAC) 1 and 2 [Dorfman et al., 2002; Runko and Sagerstrom, 2003; Runko and Sagerstrom, 2004; Ji et al., 2009]; Gal4-Nlz1 and Gal4-Nlz2 fusions repress transcription from Gal4-responsive promoters, both in cell lines and zebrafish embryos [Nakamura et al., 2008; Ji et al., 2009]; NLZ1 has been shown to repress *TGFBR2* [Holland et al., 2011] and E-cadherin expression, as well as TCF/LEF-dependent and TGFβ reporter activation [Slorach et al., 2011].

In humans, two NET family members, *NLZ1/ZNF703* (also known as FLJ14299, AK024361, Zeppo1 and henceforth referred to as *NLZ1* for NocA-like zinc finger 1) and *NLZ2/ZNF503* (*NLZ2*), have been identified [Dorfman et al., 2002; Zhao et al., 2002; Runko and Sagerstrom, 2003; Gerhard et al., 2004; Ota et al., 2004; Slorach et al., 2011]. They are located respectively on chromosomes 8 (8p11.23) and 10 (10q22.2) in regions which arose by duplication of the ancestral 8p11 region, around 730 million years ago [Vienne et al., 2003]. Several studies have associated gene amplification of the 8p11–12 chromosomal region with human breast cancers, predominantly of the luminal B subtype, and two recent independent studies have identified *NLZ1* as the associated oncogene [Ray et al., 2004; Garcia et al., 2005; Gelsi-Boyer et al., 2005; Yang et al., 2006; Adelaide et al., 2007; Bernard-Pierrot et al., 2008; Kwek et al., 2009; Holland et al., 2011; Sircoulomb et al., 2011]. Furthermore, overexpression of *Nlz1/Zeppo1* has been associated to breast cancer progression and metastasis in mice [Slorach et al., 2011].

Despite the growing interest in NET family members, the role of human NLZ1 in non-pathological processes is still unknown. It is therefore crucial to provide new information in order to better understand the function of this protein in normal cellular contexts. Herein, we provide the first molecular characterization of human *NLZ1*. We show that the gene is ubiquitously expressed in adult human and mouse tissues and subject to alternative polyadenylation; demonstrate that the NLZ1 protein contains six evolutionarily conserved domains, three of which have not been previously described and are specific of NET proteins. Finally, we show that human NLZ1 interacts with a Groucho family co-repressor, is a transcriptional repressor and present evidence that the NET protein-specific domains are necessary both for proper subcellular distribution of the protein and for its function as a transcription repressor.

MATERIALS AND METHODS

NLZ1 EXPRESSION STUDIES

Total RNA from human normal tissues was obtained from the FirstChoice® Human Total RNA Survey Panel (Ambion, Austin, TX). Total RNA from human cell lines and mouse normal tissues was isolated using the TriPure isolation reagent (Roche Diagnostics, Mannheim, Germany) according to the manufacturer's instructions. Complementary DNA (cDNA) was synthesized from 3 μg of total RNA using 100 U of SuperScript™ II Reverse Transcriptase (Invitrogen, San Diego, CA) and 100 ng of random hexamer primers, following the manufacturer's protocol, in a 20 μl reaction. PCR amplifications of human and mouse *Nlz1* were performed as described [Pereira-Castro et al., 2012]. Primer sequences (5′-3′) used were: human and mouse *Nlz1* ATGAGCGATTCGCCCGCTG (F) and TGCCGGTCTGCGAGCAGGTC (R); human *GAPDH* CCAGCCGAGCCACATCG (F) and GGTCATGAGTCCTTCCACG (R) and mouse *Gapdh* AATGGTGAAGGTCGGTGTG (F) and GGCAGTGATGGCATGGAC (R). Samples were amplified for 30 cycles at 95°C for 30 s, 62°C for 60 s, and 72°C for 45 s, after an initial denaturation at 95°C for 5 min and followed by a final extension step of 5 min at 72°C in a MyCycler thermocycler (Bio-Rad Laboratories, Richmond, CA).

3′-RAPID AMPLIFICATION OF cDNA ENDS (RACE) AND 3′-EXPRESSED SEQUENCE TAG (3′-EST) ANALYSES

3′-RACE cDNA synthesis was made with the SMARTer™ RACE Amplification Kit (Clontech, Palo Alto, CA) using 1 μg of human placenta total RNA (Clontech) or 1 μg of human muscle, liver, intestine or brain RNA (Ambion) according to the manufacturer's protocol. The 3′-RACE PCR was performed as described [Lopes-Marques et al., 2012] using the *NLZ1*-specific primer (5′-GGCCTAAGCCGGTACCACCCCTATG-3′) for hybridization on *NLZ1's* exon 2. Amplification conditions were as indicated in the SMARTer™ RACE Kit protocol. 3′-RACE PCR products were gel purified, cloned into an in-house developed TA cloning plasmid (pUAmp5-TA) and three clones of each were sequenced using the following vector primers TA-F (5′-GATTTTTGTGATGCTCGTCAG-3′) and TA-R (5′-GGGTTCCGCGCACATTTC-3′). We confirmed the chromosomal location of the *NLZ1* 3′-RACE transcript sequences by running a local copy of the Basic Local Alignment Search Tool (BLAST) implemented on the Geneious v5.3 program [Drummond et al., 2009] with a custom database including the reference sequence of the human chromosome 8 (NC_000008).

We also collected all 3′-Expressed Sequence Tags (3′-EST) available at the UniGene database (http://www.ncbi.nlm.nih.gov/unigene) corresponding to the *NLZ1* 3′-end (Hs.726062). From these, we selected the sequences (n = 56) that contained a polyadenylation signal or tail for 3′-ESTs analysis in order to guarantee mRNAs with the 3′-end. The location of all 3′-ESTs on NLZ1 mRNA was determined using BLAST as described above.

IDENTIFICATION OF CONSERVED DOMAINS IN NLZ1 HOMOLOGUES

The sequences of all proteins related to NLZ1 were retrieved from the Ensembl (http://www.ensembl.org) and NCBI (http://www.ncbi.nlm.nih.gov) databases. The combination of suggested orthologues in Ensembl with a BLASTP search against the NCBI database (non-redundant protein sequences), using the human NLZ1 sequence as query, allowed us to recover a total of 48 protein sequences. All complete protein sequences were aligned using the default settings of the MUSCLE 3.6 software [Edgar, 2004] running on the Geneious program v5.3.

PLASMID CONSTRUCTION

Plasmids were generated by subcloning the relevant human cDNA fragments (e.g., coding for residues 280–532 of NLZ1 in the case of $NLZ1_{280-532}$) into yeast two-hybrid bait (AS3, an AS2 (Clontech) derivative with a single *Eco*RI restriction site) or prey (ACT2, Clontech) plasmids and the mammalian expression vectors pCMV-Myc, pCMV-HA (Clontech), and pCMV-G4BD (generates fusion proteins with the Gal4-DNA binding domain). For the 15UTSV-Luc reporter plasmid, 15 copies of the Gal4 upstream activation sequences (UAS), generated by concatamerization of the 5 Gal4 UAS present in UAS_G-βgal [da Costa et al., 1996], were cloned into an SV40-driven luciferase reporter upstream of the SV40-promoter (from P5 [da Costa et al., 1996]). All plasmids were sequenced using the Big Dye Terminator Cycle Sequencing Kit (Applied Biosystems, Carlsbad, CA) and the ABI Prism 3130XL Sequencer (Applied Biosystems) to ensure that they had the correct sequence. Additional details of plasmid construction are available upon request.

YEAST TWO-HYBRID ASSAYS

For the yeast two-hybrid interaction assays, the *AH109* yeast reporter strain (with copies of *Ade* and *His3* reporter genes under the control of *Gal4*-dependent promoters) was transformed with pACT2-NLZ1, pACT2-$NLZ1_{341-590}$ or pACT2-$NLZ1_{1-532}$ bait plasmids and the pAS3-GRG5 prey plasmid as described previously [Gietz and Woods, 2002]. Yeast growth in synthetic complete (SC) media lacking leucine (-Leu) and tryptophan (-Trp) was used to verify the transformation efficiency. Interaction between GRG5/AES and full-length NLZ1 protein was tested using selective medium also lacking histidine (SC-His-Leu-Trp) and the strength of the interaction assessed by using increasing concentrations (up to 12 mM) of 3-Amino-1,2,4-triazole (3-AT; Sigma) in the medium. In the case of the interaction between NLZ1 domain deletion plasmids and GRG5/AES, yeast transformants were selected for their ability to grow in SC-Ade-Leu-Trp medium, determined after 5 days at 30°C.

CELL LINES

HEK293 and HeLa cell lines were maintained in DMEM (Dulbecco's modified Eagle medium; Invitrogen) and the MCF7 cell line in RPMI 1640 with GlutaMAX™ (Invitrogen), all supplemented with 10% FBS (fetal bovine serum) (Invitrogen), and 1% of a solution containing 10,000 units/ml of penicillin and 10,000 μg/ml of streptomycin. Cells were incubated at 37°C in a 5% CO_2 atmosphere and split every 3 days to maintain subconfluency.

CO-IMMUNOPRECIPITATION

Prior to transfection, HEK293 cells were grown on T25 flasks at standard culture conditions to 40–60% confluency. Cells were transiently transfected by the calcium phosphate precipitation method using 1 μg of each expression plasmid. The total amount of transfected DNA was adjusted in all cases to 10 μg by addition of the correspondent backbone vector. In each case, 200 μl of 2×HBSP (280 mM NaCl, 10 mM KCl, 50 mM Hepes, 1.6 mM $Na_2HPO_4 \cdot H_2O$ and 10 mM glucose) and 200 μl of 250 mM $CaCl_2$ were added to the plasmid DNA mixture. This solution was incubated for 40 min at room temperature and then added to each T25 flask. The medium was replaced 24 h later and the cells were harvested 48 h after transfection. Briefly, cells were washed twice with cold 1× PBS and lysed in 400 μl of lysis buffer (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 1% Triton X-100, 1% Nonidet P-40, 1 mM phenylmethylsulfonyl fluoride, 20 mM NaF, 3 mM sodium vanadate, 10 μg/ml aprotinin and 10 μg/ml leupeptin). The supernatant was collected by centrifugation at 14,000*g* for 5 min and 500 μg of total protein extracts were incubated for 2 h with rotation at 4°C with 2.5 μg of a mouse monoclonal anti-c-Myc antibody (Clontech). Immunoprecipitates were incubated for 30 min with Protein G-Sepharose beads 4 Fast Flow (GE Healthcare, Waukesha, WI). After the incubation, the beads were washed three times with 1 ml of lysis buffer and boiled in sample buffer. For the total protein lysates 20 μg of protein was loaded per case. Proteins were separated by SDS–PAGE, transferred onto a Hybond nitrocellulose membrane (GE Healthcare) and subjected to Western blot analysis. Briefly, membranes were blocked for 30 min with 5% non-fat milk in PBS + 0.5% Tween-20 (Sigma) and incubated for 2 h with a rabbit polyclonal anti-HA antibody (Clontech; 1:1,000 dilution). HA epitope-tagged GRG5/AES was detected using a horseradish peroxidase-conjugated goat anti-rabbit (Santa Cruz Biotechnology, Santa Cruz, CA) secondary antibody (1:2,000 dilution) followed by ECL detection (GE Healthcare).

IMMUNOFLUORESCENCE

Cells were plated in 24-well plates on top of glass coverslips at a density of 1×10^5 cells per well and incubated to 50% confluency. Transfections of expression plasmids were performed using 2.5 μl of Lipofectamine™ 2000 reagent (Invitrogen) and 0.5 μg of each plasmid. The total amount of transfected DNA was adjusted in all cases to 1 μg per well by addition of an irrelevant (backbone) plasmid. Lipofectamine™/DNA mixtures were incubated for 20 min at room temperature prior to addition to the cells. The media were replaced 12 h after transfection and cells were collected 48 h post-transfection. For immunofluorescence, cells were washed twice with cold PBS, fixed with 4% paraformaldehyde (Sigma–Aldrich) for 20 min, incubated in 50 mM NH_4Cl (Sigma–Aldrich) for 10 min and permeabilized for 5 min with 0.2% Triton X-100 in PBS. Thereafter, cells were blocked with 4% bovine serum albumin for 30 min and incubated with a mouse monoclonal anti-c-Myc antibody (Clontech; 1:100 dilution) for 1 h at room temperature followed by incubation for 45 min in a dark chamber with a rabbit anti-mouse IgG conjugated with FITC (fluorescein isothiocyanate; Dako, Copenhagen, Denmark) secondary antibody (1:100 dilution). Cells were mounted in Vectashield Mounting Medium with DAPI (Vector Labs, Burlingame, CA) and visualized under a Laser Scanning Confocal Microscope Leica SP2 AOBS SE (Leica, Wetzlar, Germany) using the 40× objective. Images were edited using the Leica LCS Lite image editor (version 2.61).

LUCIFERASE AND β-GALACTOSIDASE ASSAYS

For luciferase assays, cells were plated the day before transfection at a density of 1×10^5 cells per well in a 24-well plate. Transfections were done in triplicate with Lipofectamine™ 2000 (Invitrogen) as described above, using 0.65 μg of the effector plasmids and 0.25 μg of the 15UTSV-Luc reporter plasmid in the case of the Gal4-reporter assay, and 1 μg of the Myc-Nlz1 full length or Myc-Nlz1 fragments, 0.1 μg of pCMV-β-catenin and 0.1 μg of the TCF-Luc or TCF*-Luc

[da Costa et al., 1999] for the TCF/β-catenin reporter assay. In each case, 0.1 μg of the CMV-β-galactosidase plasmid (Clontech) was co-transfected to normalize for transfection efficiency. The medium was replaced 12 h after transfection and cells were collected to determine luciferase and β-galactosidase activities 24 h later. Luciferase assays were performed according to the protocol recommended by the manufacturer (Promega, Madison, WI). Briefly, cells were washed in cold PBS and lysed with 100 μl of Reporter Lysis Buffer (Promega) per well. Scraped cells were centrifuged at 14,000*g* for 2 min and the supernatant was collected. An aliquot of 20 μl of each cell lysate was added to 100 μl of Luciferase Assay Reagent (Promega), and the luciferase activities were determined using a 1450 TriLux MicroBeta luminescence counter (Wallac, PerkinElmer, MA). The β-Galactosidase assays [da Costa et al., 1996] were performed in 96-well plates using 5 μl of each cell lysate and 45 μl of Z buffer with β-mercaptoetanol. Reactions were initiated by addition of 10 μl ONPG (Sigma–Aldrich, St. Louis, MO) substrate, incubated at 37°C and stopped by addition of 25 μl Na_2CO_3. β-Galactosidase activity was measured at a wavelength of 415 nm in a Model 680 Microplate Reader (BioRad). Luciferase activities were normalized against the β-galactosidase values. To determine the statistical significance of differences between average luciferase activity values, we used a two-tailed Student's *t*-test considering a statistically significant difference when *P*-value <0.05.

RESULTS

NLZ1 IS UBIQUITOUSLY EXPRESSED IN HUMANS AND MICE

Since NET family members have generally been studied in the context of development, it is not surprising that no attempts to study their patterns of expression in adult animals have been reported. As for human genes, the only reports are those related to *NLZ1s* role in breast cancer, mainly its overexpression [Ray et al., 2004; Yang et al., 2006; Holland et al., 2011; Slorach et al., 2011; Sircoulomb et al., 2011]. As part of our effort to characterize *NLZ1*, we therefore investigated its pattern of expression in normal adult tissues and cancer cell lines. A total of 37 samples—13 human normal tissues, 10 human tumor cell lines, and 14 mouse normal tissues—were studied by RT-PCR analyses, with primers designed using human (NM_025069.1) and mouse (NM_001101502.1) *NLZ1* cDNA sequences retrieved from the NCBI database. Since there is 78.9% sequence identity between the two transcripts, primers were designed to be used for both species. We selected the *NLZ1* region between the initiation codon and the beginning of exon 2 (Fig. 1A) for amplification, in order to circumvent a stretch of repetitive sequence motifs present in the middle of the transcript in exon 2 and to avoid the potential amplification of contaminant genomic DNA. As shown in Figure 1, *NLZ1* displays a ubiquitous expression pattern in normal human tissues (Fig. 1B) and human tumor cells (Fig. 1C), as well as in adult mouse tissues (Fig. 1D).

NLZ1 IS SUBJECT TO ALTERNATIVE POLYADENYLATION

Given that no information was available about human *NLZ1*'s 3′-untranslated region (3′-UTR), we sought to map the end(s) of *NLZ1* transcript(s) by performing 3′-RACE analysis. Three different amplification products were always obtained from placenta (Fig. 2A), as well as muscle, liver, intestine and brain (not shown). Cloning and sequencing of these products showed that they correspond to three different classes of *NLZ1* mRNA 3′-ends, which were named according to their size (Fig. 2A,B). The shortest, mRNA1, has a 3′-UTR of 204 nucleotides (Fig. 2B) that matches the 3′-UTR of *NLZ1* mRNA reported at the NCBI database (NM_025069). As for

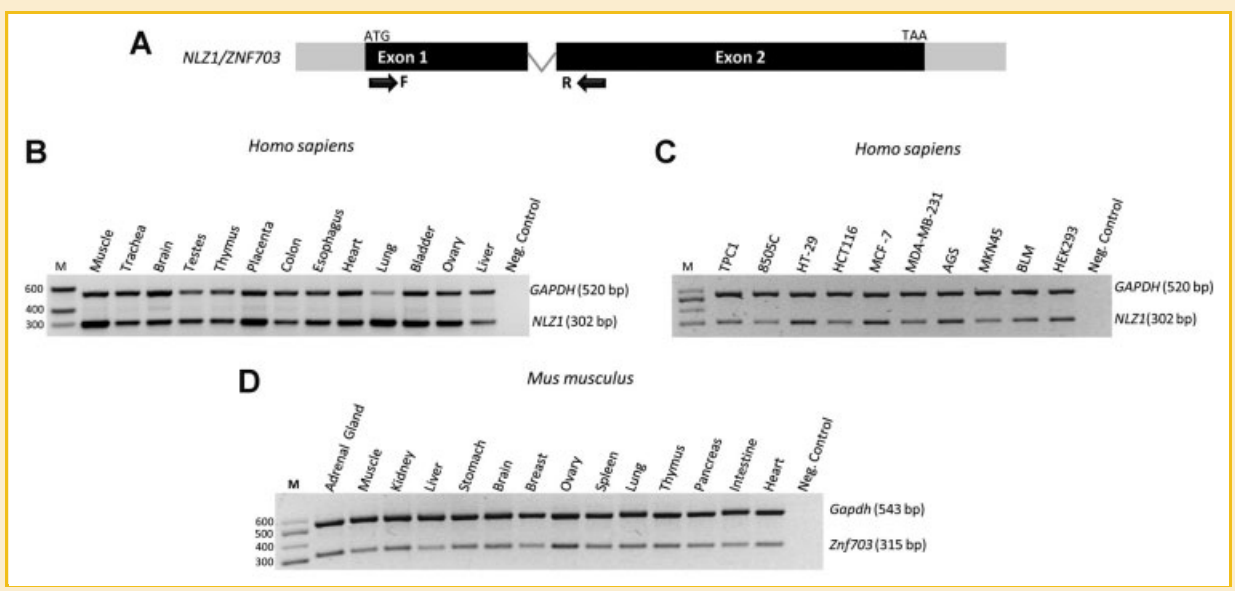

Fig. 1. *NLZ1* is ubiquitously expressed in humans and mice. A: Schematic representation of the *NLZ1* transcript. The position of the primers used in the RT-PCR is indicated by arrows. UTRs are shown in gray and coding sequence in black, F and R represent the forward and reverse primers. B: Amplification of the *NLZ1* transcript by RT-PCR (302 bp) in a panel of 13 normal human tissues and (C) 10 human cancer cell lines. D: RT-PCR analysis of *Nlz1* expression (315 bp) in 14 mouse tissues. *GAPDH* was used in the RT-PCRs as an internal control for cDNA.

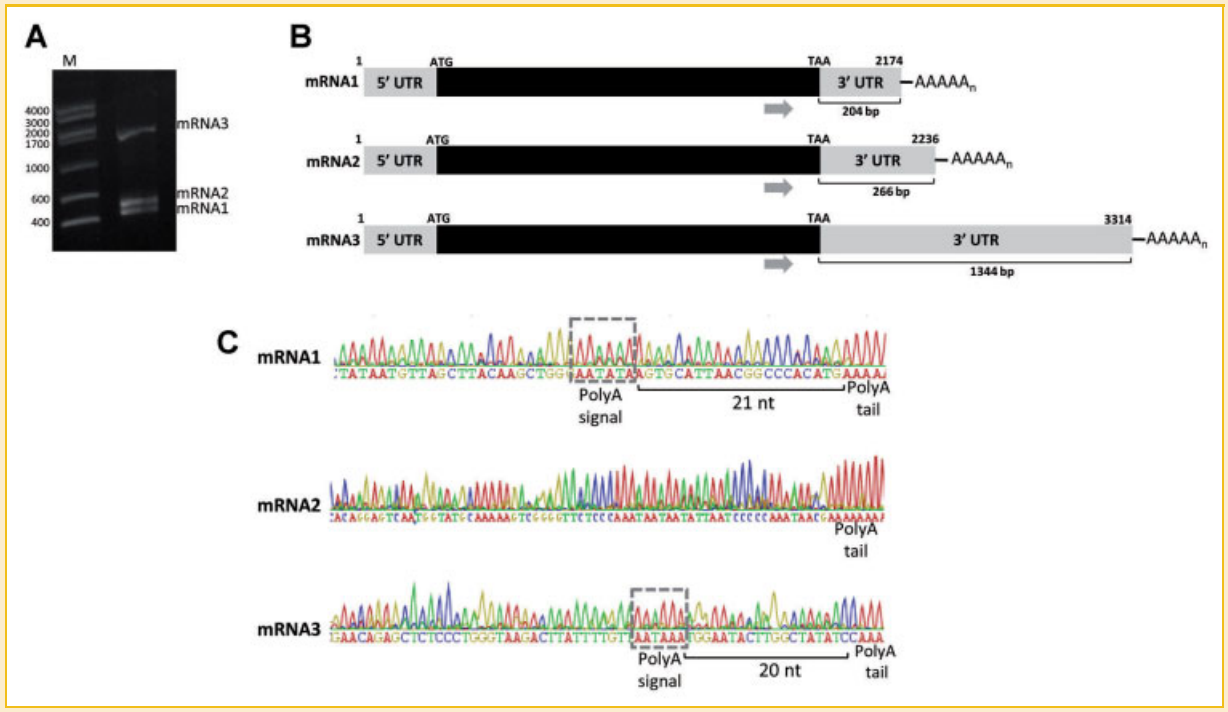

Fig. 2. *NLZ1* transcripts are subject to alternative polyadenylation. A: Three PCR products with different lengths were identified by *NLZ1* 3′-RACE analysis. The amplicons were named mRNA1, mRNA2, and mRNA3 according to their different size in a 1% agarose gel. B: Schematic representation of the three alternative polyadenylated *NLZ1* mRNAs. The arrows indicate the location of the 3′-RACE primer. The length of the different 3′-UTRs is shown below the lines. The untranslated regions are shown in gray and the coding sequence in black. C: Electropherogram depicting the alternative 3′-ends. The polyadenylation signal is shown for mRNA1 and mRNA3 sequences. The distance between the polyA signal and the polyA tail is indicated. mRNA2 does not have a conventional polyA signal.

mRNA2 and mRNA3, they also contain the full 3′-UTR sequence of mRNA1, but extend for an additional 62 and 1,140 nucleotides respectively (Fig. 2B). Curiously, it is the mRNA with the longest 3′-UTR (mRNA3) that is deposited in the Ensembl database as reference *NLZ1* mRNA (ENST00000331569). It should be noted that some heterogeneity was observed in the length of mRNA1 and mRNA2 clones, corresponding to transcripts that differ at the 3′end, by 2–3 nucleotides in mRNA1 and 10–30 nucleotides in mRNA2. This is in agreement with previous reports that heterogeneity at the cleavage site is a frequent occurrence in mammalian mRNAs [Pauws et al., 2001; Lopes-Marques et al., 2012]. Importantly, the 3′-UTR sequences are all collinear with the genomic sequence and only differ in their lengths. This indicates that the three different *NLZ1*'s 3′-UTRs are the result of alternative 3′-end formation at different positions along the pre-mRNA (alternative polyadenylation).

To confirm that the sequences identified by RACE correspond to actual mRNA species, we started by searching for the presence of polyadenylation (polyA) signals in the three mRNA species. As shown in Figure 2C, only mRNA3 contains the canonical polyadenylation signal AATAAA (known to be present in 60% of all human mRNAs 3′-ends [Beaudoing et al., 2000]), located 21 nucleotides upstream the polyA tail. The mRNA1 sequence contains a less frequent polyA signal, AATATA (which is only present in about 2% of the human mRNAs [Beaudoing et al., 2000; Tian et al., 2005]), located 20 nucleotides upstream the polyA tail. As for mRNA2 transcript, it does not have any conventional polyA signal in its 3′-end region (Fig. 2C), though it has an AT rich region around 13–29 nucleotides upstream the polyA tail, differing from the canonical AATAAA hexamer sequence only by a single nucleotide (Fig. 2C).

Additionally, we analyzed all 56 *NLZ1* 3′-EST sequences deposited at UniGene database that included a polyA tail or a polyA signal. Interestingly, 53 of them fully match the sequence of mRNA3, only three of them correspond to the mRNA1 transcript and none matches the mRNA2 sequence. Together with the sequence data, these results confirm the existence of alternative polyadenylation of *NLZ1* and suggest preferential usage of the canonical AATAAA polyA site of mRNA3 over the alternative AATATA sequence present in mRNA1, favoring formation of the longer transcript. However, tissue-specific variations in the three isoforms' relative expression levels remain a possibility (see Discussion).

THE NLZ1 PROTEIN FAMILY HAS SIX CONSERVED DOMAINS, THREE OF THEM UNIQUE TO NET PROTEINS

It is widely believed that protein motifs and domains with high level of sequence conservation among species have important structural and/or functional roles. Also, certain sequence motifs are signatures of protein families and can be used for the prediction of protein function. We therefore searched for conserved sequence block within NLZ1 by performing an in silico comparative analysis of all complete vertebrate NET protein sequences retrieved from public databases. As shown in Figure 3, six segments have a total pairwise identity of no less than 80% in the alignment of NLZ1 homologous protein sequences (accession numbers in Table S1) from 24 different

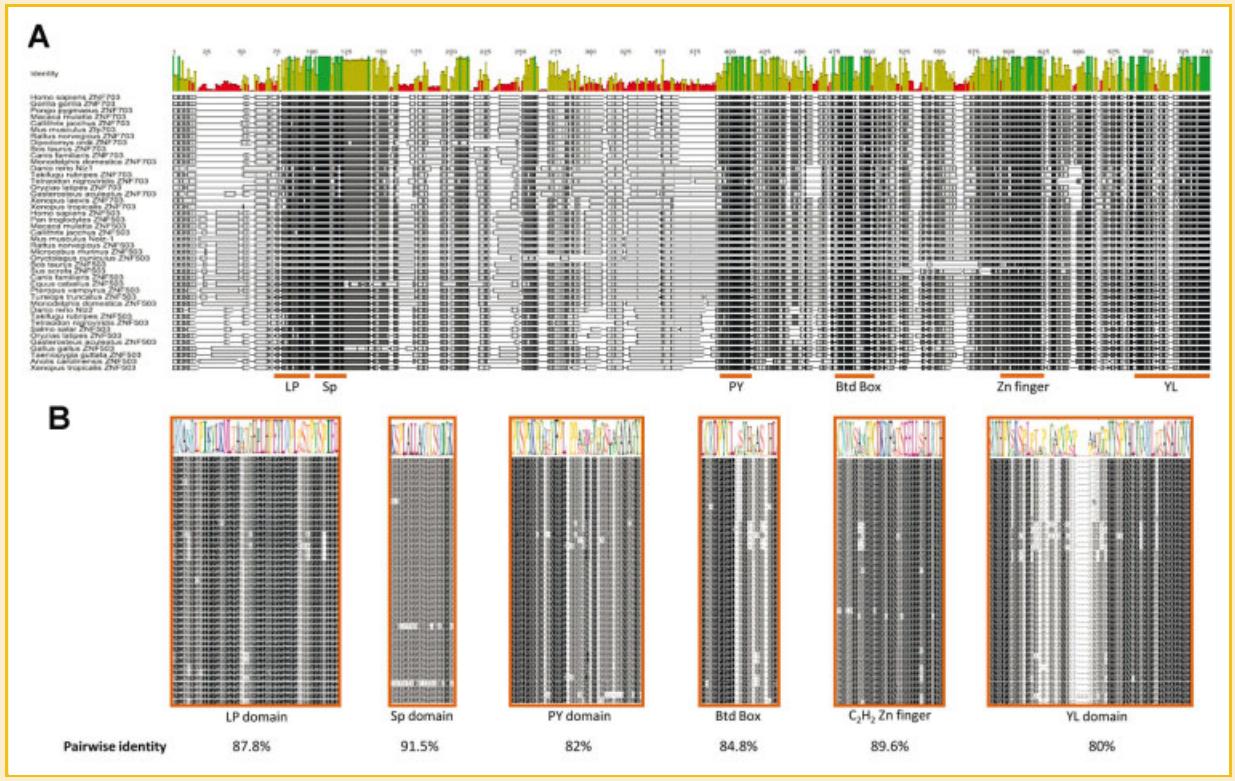

Fig. 3. NET family proteins have six conserved domains. A: General overview of the protein sequence alignment using 43 human NLZ1-related vertebrate proteins. The identity across all sequences is indicated for every position above the alignment (high and low identity values represented by green and red bars, respectively), while the names and locations of the six domains are indicated below the alignment. B: Amino acid sequences of the six conserved domains. The high degree of conservation is evident in the sequence logos and the pairwise identity values.

vertebrate species. Three of these conserved domains were previously described in NET family proteins (Sp, Btd box and C_2H_2 zinc finger domains [Nakamura et al., 2004]), and the remaining were named according to the most abundant conserved amino acids in each of them: LP, PY, and YL (Fig. 3). The Sp and Zinc finger domains show the highest levels of conservation (91.5% and 89.6%, respectively) in vertebrates. The other domains present pairwise identity values of 82% (PY), 87.8% (LP), 84.8% (Btd box domain), and 80% (YL; Fig. 3B).

The confirmation that the LP domain is specific of the NET family, and not part of the adjacent Sp domain, was accomplished by performing a BLAST analysis. By using the LP domain as query in a BLASTP search, we only retrieved proteins belonging to the NET family (data not shown). If the Sp domain was in fact larger than the previously defined consensus sequence (SPLAL/MLAA/QTC) and included what we defined here as the LP domain, we would expect to have retrieved other proteins in the search, such as those from the Sp family of transcription factors [Suske, 1999]. Searching for the LP domain in the Pfam [Finn et al., 2010] and InterProScan [Quevillon et al., 2005] databases confirmed that this domain is not present in other proteins, indicating that it is in fact specific of the NET family and not part of an extended Sp domain. Similar searches were performed using the PY and YL domains as queries and in both cases only NET proteins were recovered (not shown) showing that they are also specific of this protein family. Furthermore, all the six domains identified in vertebrate NET proteins (including the NET-specific ones) were also recognized, though with a lower degree of conservation, in invertebrate species, such as *C. elegans* (TLP-1), *D. melanogaster* (NocA and Elbow) and *Ciona intestinalis* (ZNF503; see Fig. S1).

The six domains are found both in NLZ1 orthologues and paralogues, despite the different lengths of the proteins from the two groups (e.g., 590 and 630 amino acids in human NLZ1 and NLZ2, respectively). Moreover, the high level of domain conservation was maintained even after the duplication event that originated the NLZ1 and NLZ2 paralogue groups, more than 730 million years ago, after the separation of the urochordate and vertebrate lineages [Vienne et al., 2003], suggesting functional constraints. The identification of the same domains on both the NLZ1 and NLZ2 groups is consistent with the observation that both zebrafish proteins are necessary for hindbrain rhombomere 4 formation [Runko and Sagerstrom, 2004; Nakamura et al., 2008] and in agreement with previous observations of partial functional overlap in other types of paralogues [Greer et al., 2000; Qian et al., 2010].

NEITHER OF THE NET-SPECIFIC DOMAINS IS NECESSARY FOR NLZ1 BINDING TO GROUCHOS

Having identified three NET-specific domains (LP, PY, and YL), we tried to explore what their functional roles might be. We started by considering the possibility of NET-specific domains being involved

in protein–protein interactions, as these are frequently mediated by conserved protein regions. Human NLZ1 homologues have previously been shown to bind to both class I Histone deacetylases (HDAC1 and HDAC2) [Runko and Sagerstrom, 2003, 2004; Nakamura et al., 2008], and proteins of the Groucho family [Dorfman et al., 2002; Runko and Sagerstrom, 2003, 2004; Nakamura et al., 2008; Ji et al., 2009; Slorach et al., 2011], making them attractive candidates to test this possibility. However, available data indicate that the interaction with HDACs does not require any of the NET-specific domains but instead the region between the Btd box and Zinc finger domains [Runko and Sagerstrom, 2003, 2004; Nakamura et al., 2008]. The requirements for binding to Groucho and its homologues, on the other hand, are still controversial, with different regions of the NET proteins being suggested by studies in fly [Dorfman et al., 2002], zebrafish [Runko and Sagerstrom, 2003, 2004] and chicken [Ji et al., 2009]. We therefore decided to focus on interactions with Groucho family members. First, we determined, by both yeast two-hybrid (Fig. 4A) and co-immunoprecipitation from human cells (Fig. 4B), that NLZ1 interacts with the human "short Groucho" GRG5/AES (Groucho-related gene 5/Amino-terminal enhancer of split). This interaction is strong since we observed yeast growth on a selective histidine-lacking medium even in the presence of relatively high concentrations of 3-Amino-1,2,4-triazole (3-AT), an inhibitor of histidine biosynthesis (Fig. 4A).

Next, we tested fragments of NLZ1 lacking NET-specific domains for their ability to bind human GRG5/AES in a yeast two-hybrid assay. As shown in Figure 4, deleting either the LP and PY domains ($NLZ1_{341-590}$–Fig. 4C) or the YL domain ($NLZ1_{1-532}$–Fig. 4D) has no effect on the interaction of NLZ1 with GRG5/AES, given that we observe yeast growth in the selective medium. This indicates that, like other NET proteins [Dorfman et al., 2002; Runko and Sagerstrom, 2003, 2004; Nakamura et al., 2008; Ji et al., 2009; Slorach et al., 2011], NLZ1 indeed interacts with Groucho family members, but neither of the NET-specific domains (LP, PY or YL) is required for the interactions.

THE PY AND YL DOMAINS ARE IMPORTANT FOR PROPER NLZ1 NUCLEAR LOCALIZATION

Previous studies have demonstrated that human NLZ1 homologues in zebrafish (Nlz1 and Nlz2) [Runko and Sagerstrom, 2003, 2004], fruit fly (Elbow and NocA) [Dorfman et al., 2002], rat (Zfp503) [Chang et al., 2004], and mouse (Zeppo1/Znf703) [Slorach et al., 2011] have a nuclear localization. To test whether the same held true for human NLZ1, we generated an expression vector encoding an N-terminally Myc-tagged full-length human NLZ1 ($NLZ1_{1-590}$; Fig. 5A) and used it to transfect the HEK293 human cell line. As observed in Figure 5B, human NLZ1 indeed has a nuclear distribution, with small amounts of dot-like structures dispersed throughout the nuclei (Fig. 5B—first panel). However, no nuclear localization signal (NLS) is easily discernible in NLZ1 (or its homologues) and searches using two different programs, predictNLS [Cokol et al., 2000] and NucPred [Brameier et al., 2007] failed to

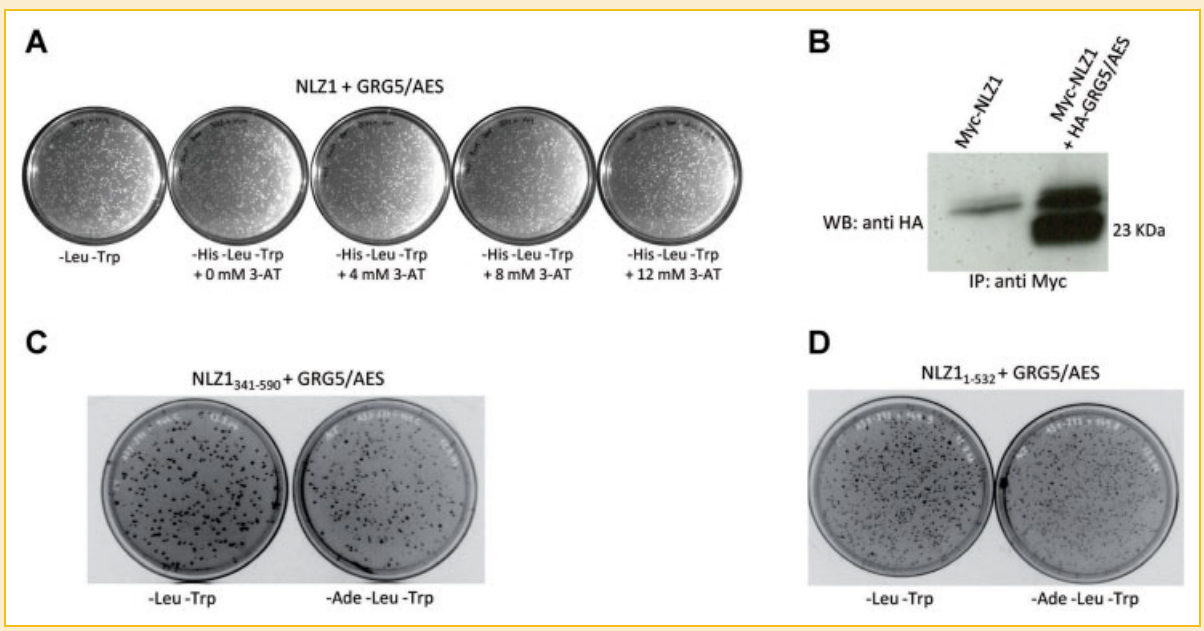

Fig. 4. The NET-specific domains LP, PY, and YL are not required for interaction between human NLZ1 and GRG5/AES. A: NLZ1 and GRG5/AES interact in a yeast two-hybrid assay. Plasmids pACT2-NLZ1 and pAS3-GRG5 were used as prey and bait vectors, respectively. Growth in SC-Leu-Trp medium was used to assess the transformation efficiency. Interaction was evaluated by growth in a His-lacking selective medium (SC-His-Leu-Trp). The strength of the interaction was evaluated using progressively high concentrations of 3-AT. B: NLZ1 and GRG5/AES interact in vivo in HEK293 human cells. Total protein extracts from HEK293 cells transfected with Myc-NLZ1 and HA-GRG5/AES expression plasmids were immunoprecipitated using a anti-Myc antibody followed by western blot analysis with an anti-HA antibody. As a negative control, total protein extracts from HEK293 cells transfected only with Myc-NLZ1 were used. C and D: The NET-specific domains LP, PY, and YL are not necessary for NLZ1-GRG5/AES interaction. Plasmids encoding fragments of NLZ1 lacking the LP and PY domains (pACT2-$NLZ1_{341-590}$) or YL domain (pACT2-$NLZ1_{1-532}$) were used as prey vectors together with the bait vector pAS3-GRG5. Selection for interaction was made using an adenine-lacking selective medium (SC-Ade-Leu-Trp). Transformant yeast colonies grew despite the absence of NET-specific domains, indicating that none of these domains is required for the interaction with GRG5/AES.

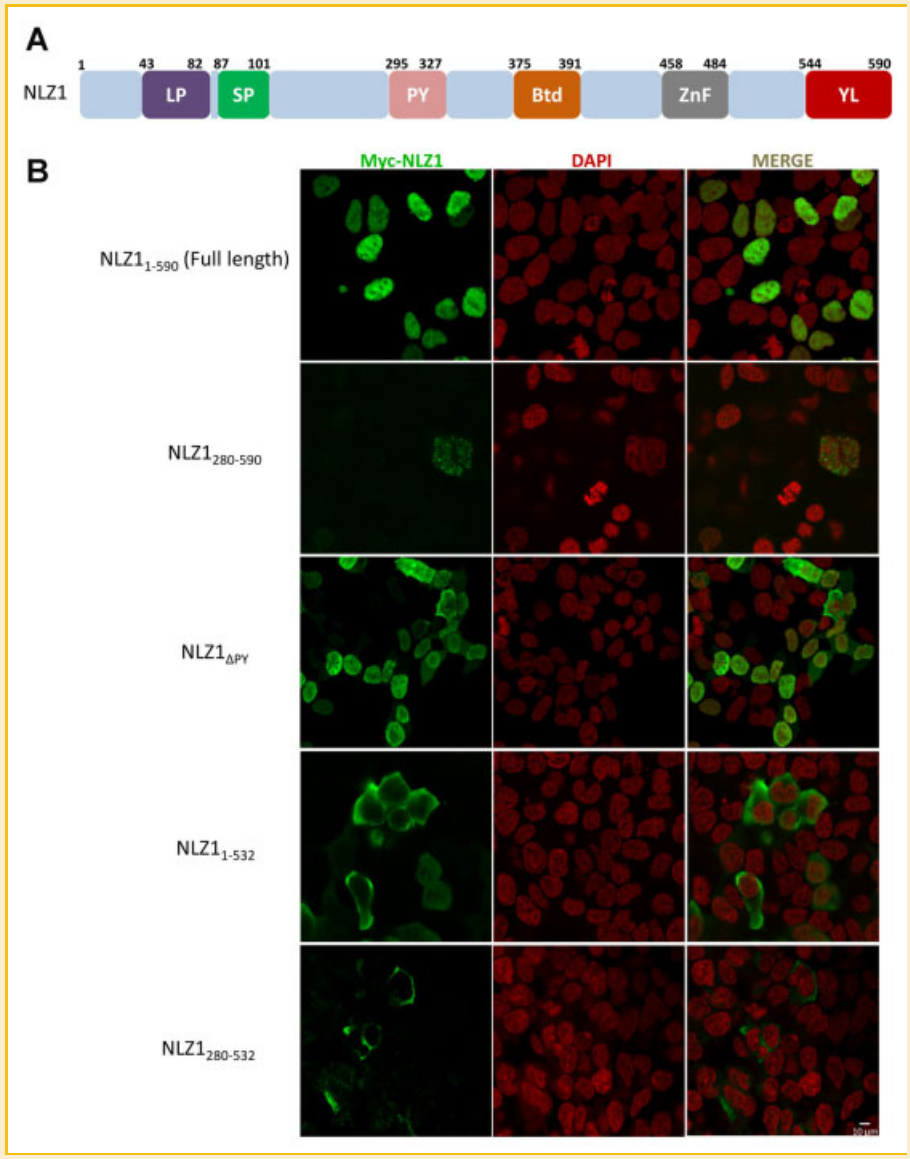

Fig. 5. The NET-specific PY and YL domains are important for NLZ1 nuclear localization. A: Schematic representation of the human NLZ1 protein showing the six conserved domains present in NET proteins: the three NET-specific domains (LP, PY, and YL) and the three domains also present in other protein families (Sp domain [Sp], Buttonhead box [Btd] and C_2H_2 Zinc finger domain [ZnF]). The numbers indicate the domains' limits in human NLZ1 protein. B: Localization of NLZ1's domain-deleted variants in the human HEK293 cell line using a confocal microscope. The localization of the different Myc-tagged NLZ1 (green) constructs was determined using an anti-Myc antibody. Nuclei were counterstained with DAPI (red). Full length NLZ1 protein ($NLZ1_{1-590}$) has a nuclear localization. PY ($NLZ1_{\triangle PY}$) or YL ($NLZ1_{1-532}$) domain deletion leads to an alteration of the normal subcellular localization of the NLZ1 protein. However, nuclear exclusion is observed only when the YL domain is absent. The results were confirmed using the DLD-1 human cell line (not shown).

detect any NLS in the human protein. This suggests that correct subcellular localization of NET proteins depends on mechanisms other than the classical nuclear import pathway [Nigg, 1997; Lange et al., 2007].

Since the NET-specific NLZ1 domains are not required for protein-protein interactions with Groucho or HDAC1 and HDAC2 (our results and [Runko and Sagerstrom, 2003, 2004; Nakamura et al., 2008]), we investigated if those domains participate in other processes, such as subcellular targeting of the protein. We therefore generated expression constructs encoding truncated forms of NLZ1 lacking the different NET-specific conserved domains and used fluorescence microscopy to evaluate their subcellular distribution. As shown in Figure 5B (second panel), deleting the first 279 amino acids of the protein ($NLZ1_{280-590}$), which includes the LP as well as the Sp domain, did not affect its nuclear localization. However, we observed a slight modification of the protein's distribution compared to that of full-length NLZ1, with less dispersion of the protein throughout the nucleus and an increase in the concentration of the dot-like structures resembling nuclear bodies. On the other hand, in cells transfected with the $NLZ1_{\triangle PY}$ construct, which express a protein lacking the PY domain, part of protein remains in the cytoplasm (Fig. 5B—third panel), though nuclear import and

distribution are maintained. This shows that the PY domain contributes for NLZ1's targeting to the nucleus, though its absence does not completely block the protein's nuclear transport, suggesting that other elements present in the protein might also be involved in the process. Indeed, when we deleted the YL domain ($NLZ1_{1-532}$), the protein remained in the cytoplasm, with complete nuclear exclusion (Fig. 5B—fourth panel). Similar results were obtained using a protein simultaneously lacking the LP, Sp and YL domains ($NLZ1_{280-532}$) (Fig. 5B—last panel). These results strongly suggest that the NET-specific conserved PY and YL domains are both required for correct targeting of human NLZ1 to the nucleus. However, the presence of the PY domain per se is not enough to promote nuclear localization, as absence of the YL domain alone results in complete nuclear exclusion of the protein (Fig. 5).

NET-SPECIFIC DOMAINS ARE REQUIRED FOR FULL TRANSCRIPTION REPRESSION BY NLZ1

Former studies in *Drosophila* [Dorfman et al., 2002], zebrafish [Runko and Sagerstrom, 2003; Nakamura et al., 2008], chicken [Ji et al., 2009], and mouse [Slorach et al., 2011] have shown that human NLZ1 homologues can act as transcription repressors. To investigate potential functional roles for NET-specific domains, other than proper nuclear localization of the protein, we therefore investigated the effects of different NLZ1 constructs on two different transcription reporter systems. First, we measured the transcriptional activity of a reporter plasmid containing 15 copies of the Gal4-binding site (UAS) upstream of a strong promoter (15UTSV-Luc; Fig. 6A) in the presence of fusions of the Gal4's DNA-binding domain (Gal4-DBD) to full length NLZ1 or NLZ1 domain-deletion fragments. In both MCF7 and HeLa cell lines, co-transfection of 15UTSV-Luc and the Gal4DBD-NLZ1 fusion constructs produces a marked reduction in the luciferase activity (~6-fold in MCF7 and ~3-fold in HeLa cells—Fig. 6B), confirming that human NLZ1 can act as a transcriptional repressor. Co-transfection with the Gal4DBD-$NLZ1_{\triangle PY}$ plasmid, lacking the NET-specific PY domain, had a similar effect, showing that the PY domain is not required for transcriptional repression. However, both the Gal4DBD-$NLZ1_{280-590}$ (lacking the LP and Sp domains) and Gal4DBD-$NLZ1_{1-532}$ (lacking the C-terminal YL domain) constructs, were unable to repress transcription as efficiently as their full-length counterpart (sixfold in FL vs. fourfold for both LP/Sp and YL domain-deletions in MCF7; and threefold vs 1.8 in HeLa, these differences being statistically significant). This suggests that these domains might be involved in NLZ1-mediated transcriptional repression. Importantly, it is unlikely that the effect observed with the YL domain-deletion construct is attributable simply to this domain's role in nuclear localization, as the fusion proteins carry an NLS in the Gal4DBD moiety.

Next, we tested whether any of the NET-specific domains is involved in Nlz1-mediated repression of the Wnt signaling pathway (which was recently reported for mouse Nlz1 [Slorach et al., 2011]), using reporter plasmids carrying three copies of a TCF/LEF-binding site upstream of a luciferase gene. As expected, co-transfection of the HEK293 cell line with β-catenin and the TCF-Luc reporter plasmid resulted in a remarkable increase of luciferase activity compared to TCF-Luc alone (Fig. 6C). Consistent with Slorach et al. results, co-transfection of NLZ1 strikingly reduced TCF/β-catenin-mediated transcriptional activation (14-fold reduction). A similar result was observed with the PY (18-fold reduction) domain-deletion construct, indicating no role in repression. However, with the YL and LP/Sp domain-deletion constructs, repression decreased to 10-fold and 3.3-fold, respectively (Fig. 6C), suggesting a possible role in Wnt signaling repression. Unsurprisingly, co-transfections with the control reporter TCF^*-Luc (with mutated TCF/LEF-binding sites) resulted in no alterations in luciferase activity (Fig. 6C).

DISCUSSION

NLZ1 mRNA EXPRESSION IN THE ADULT

Virtually all previously existing data on the expression of NET family genes refers to immature animals, probably as a result of the genes having primarily been identified through their roles in development [Cheah et al., 1994; Davis et al., 1997; Andreazzoli et al., 2001; Dorfman et al., 2002; Zhao et al., 2002; Runko and Sagerstrom, 2003, 2004; Chang et al., 2004; Hoyle et al., 2004; Nakamura et al., 2004, 2008; McGlinn et al., 2008; Brown et al., 2009; Ji et al., 2009]. As for the human genes, no expression studies have been reported, except for those dealing with breast cancer [Garcia et al., 2005; Gelsi-Boyer et al., 2005; Yang et al., 2006; Melchor et al., 2007; Bernard-Pierrot et al., 2008; Kwek et al., 2009; Holland et al., 2011; Slorach et al., 2011; Sircoulomb et al., 2011]. We have therefore studied *NLZ1* expression in adult human tissues as part of our effort to characterize this important protein. Our results show that *NLZ1* is ubiquitously expressed in adult tissues (Fig. 1). This pattern of expression was demonstrated by RT-PCR and RACE analyses of both human and mouse tissues, and is supported by detailed inspection of available EST data, suggesting that it might be a general phenomenon, at least in mammals. The ubiquitous expression of *NLZ1* is somewhat surprising, given the highly restricted pattern of expression displayed by NET genes during development in multiple species [Cheah et al., 1994; Davis et al., 1997; Andreazzoli et al., 2001; Dorfman et al., 2002; Zhao et al., 2002; Runko and Sagerstrom, 2003, 2004; Chang et al., 2004; Hoyle et al., 2004; Nakamura et al., 2004, 2008; McGlinn et al., 2008; Brown et al., 2009; Ji et al., 2009]. However, this might be only an apparent contradiction, as it is possible that the developmentally restricted expression of NET genes might be mirrored, in the adult, by expression in specific cell types in different organs. On the other hand, it suggests that *NLZ1* intervenes in cellular processes other than the developmental ones it has been associated with. For example, recent evidence that *NLZ1* behaves as an oncogene for the luminal B subtype of human breast cancer [Holland et al., 2011; Sircoulomb et al., 2011] points to its involvement in cell identity specification in the mammary ducts. This is supported by the findings that *NLZ1* is a Wnt signaling repressor, downregulates E-cadherin and TGFβ expression [Holland et al., 2011; Slorach et al., 2011] and is implicated in the regulation of the estrogen receptor (ER) and the E2F1 transcription factor [Sircoulomb et al., 2011].

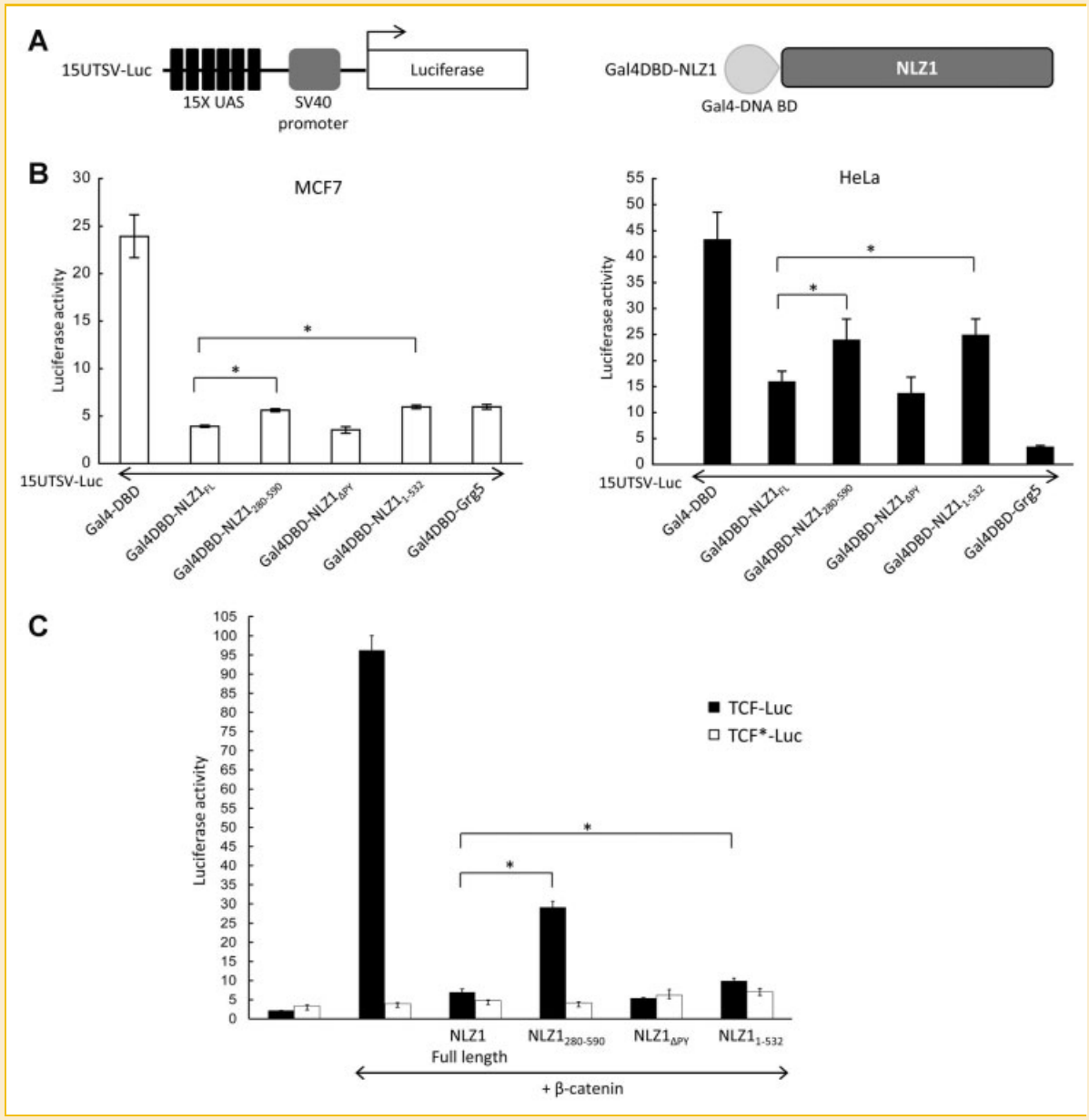

Fig. 6. Transcription repression by human NLZ1 involves NET-specific domains. A: Schematic illustration of the plasmids used in the Gal4-reporter assay. The 15UTSV-Luc reporter plasmid contains 15 GAL4 binding sites (UAS) upstream of an SV40 promoter and the luciferase reporter gene. The Gal4DBD-NLZ1 constructs were generated by fusing Gal4's DNA binding domain (DBD) to the N-termini of full length or domain-deleted human NLZ1. B: NLZ1 "intrinsic" transcriptional repression activity. The Gal4DBD-$NLZ1_{280-590}$ (lacking the LP and Sp domains) and the Gal4DBD-$NLZ1_{1-532}$ (lacking the C-terminal YL domain) fusions repress less efficiently than Gal4DBD-NLZ1 (full length) in both MCF7 and HeLa cells. These differences are statistically significant. Gal4DBD-Grg5 was used as a control for repression activity [Ren et al., 1999]. C: NLZ1 repression of TCF/β-catenin-mediated transcription. HEK293 cells were transfected with the TCF-Luc reporter plasmid (with three TCF/LEF binding sites) or the TFCF*-Luc reporter plasmid (with mutant TCF/LEF binding sites) and with a β-catenin expression plasmid to simulate activation of the Wnt signaling pathway. Addition of NLZ1 full-length or $NLZ1_{\triangle PY}$ (PY deletion) constructs resulted in repression of TCF/β-catenin-mediated transcription. Lower repression (statistically significant) was observed with the LP/Sp and YL deletion constructs, suggesting a role for these domains in Wnt signaling repression by NET-family proteins. As expected, no effect was observed with the control TCF*-Luc reporter plasmid. In (B) and (C), graphics asterisks indicate Student's *t*-test significant values (P-value < 0.05) when compared with NLZ1 full length constructs.

ALTERNATIVE POLYADENYLATION GENERATES DIFFERENT *NLZ1* ISOFORMS

In this study, three human *NLZ1* mRNA species have been identified, that we have named mRNA1 to mRNA3, according to their size (Fig. 2). Comparison of their sequences shows that they differ only in the lengths of their 3′-UTRs, indicating they are generated by alternative polyadenylation, a process estimated to occur in half of the human coding genes [Tian et al., 2005; Yan and Marr, 2005]. It has been shown previously that when 3′-UTR sequences contain more than one polyadenylation signal, there is a preferential use of one of them [Tian et al., 2005; Yan and Marr, 2005]. This seems to be the case with *NLZ1*, as 53 of the 56 human ESTs containing 3′-ends

available in the UniGene database correspond to the longest transcript (mRNA3), whereas only three of them match mRNA1 and none correspond to mRNA2. A bias towards mRNA3 is consistent with the fact that only mRNA3 contains the canonical polyadenylation signal AATAAA (21 nucleotides upstream of the cleavage site), which is a stronger signal, present in 60% of all human mRNAs 3′-ends [Beaudoing et al., 2000]. In contrast, mRNA1 seems to be generated through the use of the AATATA alternative signal (located 20 nucleotides upstream of the cleavage site), which occurs only in 2% of the human mRNAs [Beaudoing et al., 2000; Tian et al., 2005] (Fig. 2). Though mRNA2 has no recognizable polyadenylation signal within the 10–30 nucleotide interval upstream of the cleavage site, this does not completely exclude its existence in human cells. Indeed, although polyadenylation signals are particularly important as binding sites for the Cleavage and Polyadenylation Specificity Factor (CPSF), it is known that no conventional polyadenylation signal is detectable in 20 to 30% of human 3′-UTRs [Beaudoing et al., 2000; Tian et al., 2005]. Furthermore, the alignment of the human mRNAs 3′-ends with the homologous region in 12 representative mammalian species reveals a high degree of sequence conservation, including the polyadenylation signals and cleavage sites (Fig. S2). The 3′-end of mRNA2 displays a degree of conservation in mammals comparable to those found surrounding the ends of mRNA3 and mRNA1, suggesting that it might be functionally important (Supplementary Fig. S2).

The existence of multiple 3′-UTRs potentially interacting with different sets of RNA binding proteins or miRNAs, might have important implications on *NLZ1* mRNA stability, localization and translation efficiency [Mazumder et al., 2003]. Because the different mRNAs share the same coding region, this will not affect protein structure, but nevertheless might be important for the generation of developmental stage, cell-type or tissue differences in *NLZ1* mRNA species or protein levels [Yan and Marr, 2005]. In the absence of tissue-specific data on protein expression and distribution, differential regulation of NLZ1 protein levels thus remains an interesting possibility. In this regard, it is also interesting that, of the three mRNA1 EST detected, two are from pancreas and the third is from an unspecified tissue, raising the possibility of an mRNA1 bias in pancreatic tissue.

IDENTIFICATION AND FUNCTION ANALYSIS OF THE NET-SPECIFIC CONSERVED DOMAINS

A combination of functional studies with in silico evolutionary analyses was used to identify and validate conserved regions of NLZ1 homologues. We found that all NET vertebrate proteins have three additional conserved domains besides the Sp, Btd Box and C_2H_2 Zinc finger domains previously identified in both NET proteins [Dorfman et al., 2002; Nakamura et al., 2004; Runko and Sagerstrom, 2004] and members of the Sp-family of transcription factors [Athanikar et al., 1997; Suske, 1999; Schaeper et al., 2010]. The novel conserved domains identified here were named LP, PY, and YL, according to the most commonly conserved amino acids in each, and were found to be present only in NET proteins, suggesting that they might be essential elements for their specific cellular functions (Figs. 3 and 5A). Apart from the six generally conserved domains, several low complexity amino acid stretches display a certain degree of conservation within the NLZ1 or NLZ2 orthologues, or even between the two groups, though the domain sizes are not conserved. For instance, in the majority of the NLZ1 orthologues, alanine-rich and serine-rich motifs are present between the Sp and PY domains, whereas in the NLZ2 group a glycine-rich region is detectable between the same domains. In addition, between the zinc finger and the YL domains an alanine-rich motif is evident for NLZ1 and NLZ2 orthologues in the majority of the species.

The three novel NET specific-domains (LP, PY, and YL) display a high degree of conservation ($\geq$80%) in both paralogue groups of NET vertebrate proteins (Fig. 3), suggesting they might be important for these proteins' physiological functions, such as transcription repression [Dorfman et al., 2002; Runko and Sagerstrom, 2003; Nakamura et al., 2008; Brown et al., 2009; Ji et al., 2009; Holland et al., 2011; Slorach et al., 2011]. To explore this possibility, we started by assessing their role on binding to co-repressors of the Groucho family, since previous studies indicated different regions of the NET proteins as mediating the interactions [Runko and Sagerstrom, 2003, 2004]. However, our experiments demonstrate that neither the LP, PY or YL domains are required for human NLZ1's interaction with the Groucho family member GRG5 (Fig. 4).

Next, we tried to determine if any of the NET-specific domains are important for subcellular targeting of human NLZ1, as the protein was found to have a nuclear localization but no canonical NLS has been identified in any of the NET family members. Although NLZ1 has an essentially homogeneous nuclear distribution (Fig. 5B), a small number of dot-like structures were observed with the NLZ1 full-length construct (this study and [Sircoulomb et al., 2011]). Intriguingly, when the region that included the Sp and the LP domains was deleted ($NLZ1_{280-590}$), a considerable increase in the density of nuclear dots was observed, leading to a predominantly punctuate distribution (Fig. 5). This observation could indicate that the LP or Sp domains are necessary for the normal distribution of the protein inside the nucleus. Several nuclear proteins are located in nuclear bodies, like Promyelocytic leukaemia bodies (PML) or Oncogenic domains (PODs) [Matera, 1999] but this seems not to be the case of NLZ1, according to a recent study [Sircoulomb et al., 2011]. Even so, the punctuate nuclear distribution observed could have implications in the proteins' normal functioning, by preventing it from reaching some chromatin areas and affecting its repressor activity. More striking results were obtained with the PY and YL domain-deletion constructs: they clearly demonstrate that both domains are important for proper NLZ1 nuclear localization because deleting either of them leads to retention of NLZ1 in the cytoplasm (Fig. 5B). However, only the C-terminal YL domain seems essential for the process, as its deletion causes complete nuclear exclusion of NLZ1. These results are consistent with a previous study in zebrafish showing that the 498–566 C-terminal residues of Nlz1 are essential for nuclear localization [Runko and Sagerstrom, 2003].

Finally, we investigated potential functional roles for NET-specific domains in transcription repression, using two different assays. First, we used Gal4-DBD fusions to test the intrinsic ability of different human NLZ1 fragments to repress a promotor to which they were tethered. We confirmed that human NLZ1 is a

transcription repressor, and showed that full repression activity requires both the YL domain and a region that includes the LP and Sp domains (Fig. 6). Importantly, the effects of YL domain deletion on repression and nuclear localization are likely to be independent, as all the fusions tested contain Gal4's nuclear localization signal. Second, we assessed the effect of NLZ1 domain deletions on the protein's ability to specifically repress TCF/β-catenin-mediated transcription, which is activated by Wnt signaling. The Wnt pathway plays a crucial role in a variety of developmental processes in all animal species [Logan and Nusse, 2004] and there is evidence linking it to NET proteins in both *Drosophila* [Weihe et al., 2004] and mouse [Slorach et al., 2011]. Full-length NLZ1 was able to specifically repress transcription from a TCF/βcatenin-responsive promoter and, as in the previous experiment, full activity required the YL domain and especially the region encompassing both the LP and Sp domains. Interestingly, deletion of the YL domain had a smaller impact on repression, even though it prevents nuclear translocation of the protein. These results further suggest a role for NET-specific domains in NLZ1's activity as a transcription repressor, but also that the mechanisms of repression by NET proteins are complex, likely involving direct effects on promoters as well as cytoplasmic functions. Given recent evidence for NLZ1's involvement in breast oncogenesis [Holland et al., 2011; Slorach et al., 2011; Sircoulomb et al., 2011], our results also suggest the NET-specific domains as interesting targets for breast cancer genetic research or therapy. Our domain-deletion constructs simulate loss-of-function mutations, rather than the ''increased function'' (through amplification) observed in human tumors, but nevertheless suggest protein domains whose function might be enhanced through specific mutations. Perhaps more importantly, one can envisage that drugs capable of disrupting the function of NET-specific domains (e.g., in NLZ1 nuclear translocation or Wnt signaling) might constitute new and specific weapons against breast cancers.

In conclusion, we present here the first detailed analysis of human NLZ1 under non-pathological conditions. We show that NLZ1 is ubiquitously expressed, that its mRNA is subject to alternative polyadenylation and that NLZ1 and its NET homologues share six conserved domains, three of which are novel and NET-protein specific. Importantly, this is the first report that links nuclear localization of a NET protein to two different regions: the centrally located PY domain and the C-terminal YL domain. The mechanisms involved are unknown, but it is possible that both the PY and YL domains mediate interactions with other proteins that assist in targeting NLZ1 to the nucleus, since NET proteins lack a conventional NLS. Finally, we present evidence that NET-specific domains are also important for NLZ1's function as a transcription repressor. These results imply that further studies concerning NET proteins should take into account the repercussions that potential mutations may have in normal functioning of the newly identified domains.

Beyond the longstanding association of its homologues with embryonic development, human NLZ1 has recently gained relevance as a breast oncogene and potential key player in human breast cell differentiation. By furthering our understanding of human NLZ1, and pointing specific questions to be addressed regarding its function, our results should help shed light on its roles both in normal development and tumourigenesis.

ACKNOWLEDGMENTS

We are grateful to Bert Vogelstein for generously providing the TCF-Luc and TCF*-Luc plasmids, Fátima Gärtner for the mouse tissues and Carla Oliveira for the HeLa cell line. This work was supported by Fundação para a Ciência e a Tecnologia (FCT) through grants SFRH/BD/44264/2008 to IP-C, SFRH/BD/24402/2005 to ÂMSC, Ciência 2007-FCT and PTDC-SAU-ONC/112511/2009 to MJO, Ciência 2007-IPATIMUP/AA1 to LA, and Ciência 2008-ICAAM and POCTI/CBO/48218/2002 to LTC. IPATIMUP is an Associate Laboratory of the Portuguese Ministry of Science, Technology and Higher Education and is partially supported by FCT.

REFERENCES

Adelaide J, Finetti P, Bekhouche I, Repellini L, Geneix J, Sircoulomb F, Jauffret EC, Cervera N, Desplans J, Parzy D, Schoenmakers E, Viens P, Jacquemier J, Birnbaum D, Bertucci F, Chaffanet M. 2007. Integrated profiling of basal and luminal breast cancers. Cancer Res 67:11565–11575.

Andreazzoli M, Broccoli V, Dawid IB. 2001. Cloning and expression of noz1, a zebrafish zinc finger gene related to Drosophila nocA. Mech Dev 104:117–120.

Ashburner M, Aaron CS, Tsubota S. 1982. The genetics of a small autosomal region of Drosophila–Melanogaster, including the structural gene for alcohol-dehydrogenase .5. Characterization of X-ray-induced Adh null mutations. Genetics 102:421–435.

Athanikar JN, Sanchez HB, Osborne TF. 1997. Promoter selective transcriptional synergy mediated by sterol regulatory element binding protein and Sp1: A critical role for the Btd domain of Sp1. Mol Cell Biol 17:5193–5200.

Beaudoing E, Freier S, Wyatt JR, Claverie JM, Gautheret D. 2000. Patterns of variant polyadenylation signal usage in human genes. Genome Res 10:1001–1010.

Bernard-Pierrot I, Gruel N, Stransky N, Vincent-Salomon A, Reyal F, Raynal V, Vallot C, Pierron G, Radvanyi F, Delattre O. 2008. Characterization of the recurrent 8p11–12 amplicon identifies PPAPDC1B, a phosphatase protein, as a new therapeutic target in breast cancer. Cancer Res 68:7165–7175.

Brameier M, Krings A, MacCallum RM. 2007. NucPred—Predicting nuclear localization of proteins. Bioinformatics 23:1159–1160.

Brown JD, Dutta S, Bharti K, Bonner RF, Munson PJ, Dawid IB, Akhtar AL, Onojafe IF, Alur RP, Gross JM, Hejtmancik JF, Jiao X, Chan WY, Brooks BP. 2009. Expression profiling during ocular development identifies 2 Nlz genes with a critical role in optic fissure closure. Proc Natl Acad Sci USA 106:1462–1467.

Chang CW, Tsai CW, Wang HF, Tsai HC, Chen HY, Tsai TF, Takahashi H, Li HY, Fann MJ, Yang CW, Hayashizaki Y, Saito T, Liu FC. 2004. Identification of a developmentally regulated striatum-enriched zinc-finger gene, Nolz-1, in the mammalian brain. Proc Natl Acad Sci USA 101:2613–2618.

Cheah PY, Meng YB, Yang X, Kimbrell D, Ashburner M, Chia W. 1994. The Drosophila L(2)35Ba/Noca gene encodes a putative Zn finger protein involved in the development of the embryonic brain and the adult ocellar structures. Mol Cell Biol 14:1487–1499.

Cokol M, Nair R, Rost B. 2000. Finding nuclear localization signals. Embo Rep 1:411–415.

da Costa LT, Jen J, He TC, Chan TA, Kinzler KW, Vogelstein B. 1996. Converting cancer genes into killer genes. Proc Natl Acad Sci USA 93:4192–4196.

da Costa LT, He TC, Yu J, Sparks AB, Morin PJ, Polyak K, Laken S, Vogelstein B, Kinzler KW. 1999. CDX2 is mutated in a colorectal cancer with normal APC/beta-catenin signaling. Oncogene 18:5010–5014.

Davis T, Ashburner M, Johnson G, Gubb D, Roote J. 1997. Genetic and phenotypic analysis of the genes of the elbow-no-ocelli region of chromosome 2L of Drosophila melanogaster. Hereditas 126:67–75.

Dorfman R, Glazer L, Weihe U, Wernet MF, Shilo BZ. 2002. Elbow and Noc define a family of zinc finger proteins controlling morphogenesis of specific tracheal branches. Development 129:3585–3596.

Drummond A, Ashton B, Cheung M, Heled J, Kearse M, Moir R, Stones-Havas S, Thierer T, Wilson A. 2009. Geneious v5.3. Available from http://www.geneious.com

Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797.

Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer EL, Eddy SR, Bateman A. 2010. The Pfam protein families database. Nucleic Acids Res 38:D211–D222.

Garcia MJ, Pole JCM, Chin SF, Teschendorff A, Naderi A, Ozdag H, Vias M, Kranjac T, Subkhankulova T, Paish C, Ellis I, Brenton JD, Edwards PAW, Caldas C. 2005. A 1 Mb minimal amplicon at 8p11–12 in breast cancer identifies new candidate oncogenes. Oncogene 24:5235–5245.

Gelsi-Boyer W, Orsetti B, Cervera N, Finetti P, Sircoulomb F, Rouge C, Lasorsa L, Letessier A, Ginestier C, Monville F, Esteyries S, Adelaide J, Esterni B, Henry C, Ethier SP, Bibeau F, Mozziconacci MJ, Charafe-Jauffret E, Jacquemier J, Bertucci F, Birnbaum D, Theillet C, Chaffanet M. 2005. Comprehensive profiling of 8p11–12 amplification in breast cancer. Mol Cancer Res 3:655–667.

Gerhard DS, Wagner L, Feingold EA, Shenmen CM, Grouse LH, Schuler G, Klein SL, Old S, Rasooly R, Good P, Guyer M, Peck AM, Derge JG, Lipman D, Collins FS. 2004. The status, quality, and expansion of the NIH full-length cDNA project: The Mammalian Gene Collection (MGC). Genome Res 14:2121–2127.

Gietz RD, Woods RA. 2002. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. Methods Enzymol 350:87–96.

Greer JM, Puetz J, Thomas KR, Capecchi MR. 2000. Maintenance of functional equivalence during paralogous Hox gene evolution. Nature 403:661–665.

Holland DG, Burleigh A, Git A, Goldgraben MA, Perez-Mancera PA, Chin SF, Hurtado A, Bruna A, Ali HR, Greenwood W, Dunning MJ, Samarajiwa S, Menon S, Rueda OM, Lynch AG, McKinney S, Ellis IO, Eaves CJ, Carroll JS, Curtis C, Aparicio S, Caldas C. 2011. ZNF703 is a common Luminal B breast cancer oncogene that differentially regulates luminal and basal progenitors in human mammary epithelium. EMBO Mol Med 3(3):167–180.

Hoyle J, Tang YP, Wiellette EL, Wardle FC, Sive H. 2004. nlz gene family is required for hindbrain patterning in the zebrafish. Dev Dyn 229:835–846.

Ji SJ, Periz G, Sockanathan S. 2009. Nolz1 is induced by retinoid signals and controls motoneuron subtype identity through distinct repressor activities. Development 136:231–240.

Kwek SS, Roy R, Zhou H, Climent J, Martinez-Climent JA, Fridlyand J, Albertson DG. 2009. Co-amplified genes at 8p12 and 11q13 in breast tumors cooperate with two major pathways in oncogenesis. Oncogene 28:1892–1903.

Lange A, Mills RE, Lange CJ, Stewart M, Devine SE, Corbett AH. 2007. Classical nuclear localization signals: Definition, function, and interaction with importin alpha. J Biol Chem 282:5101–5105.

Logan CY, Nusse R. 2004. The Wnt signaling pathway in development and disease. Annu Rev Cell Dev Biol 20:781–810.

Lopes-Marques M, Pereira-Castro I, Amorim A, Azevedo L. 2012. Characterization of the human ornithine transcarbamylase 3' untranslated regulatory region. DNA Cell Biol 31:427–433.

Matera AG. 1999. Nuclear bodies: Multifaceted subdomains of the interchromatin space. Trends Cell Biol 9:302–309.

Mazumder B, Seshadri V, Fox PL. 2003. Translational control by the 3′-UTR: The ends specify the means. Trends Biochem Sci 28:91–98.

McGlinn E, Richman JM, Metzis V, Town L, Butterfield NC, Wainwright BJ, Wicking C. 2008. Expression of the NET family member Zfp503 is regulated by hedgehog and BMP signaling in the limb. Dev Dyn 237:1172–1182.

Melchor L, Garcia MJ, Honrado E, Pole JCM, Alvarez S, Edwards PAW, Caldas C, Brenton JD, Benitez J. 2007. Genomic analysis of the 8p11–12 amplicon in familial breast cancer. Int J Cancer 120:714–717.

Nakamura M, Runko AP, Sagerstrom CG. 2004. A novel subfamily of zinc finger genes involved in embryonic development. J Cell Biochem 93:887–895.

Nakamura M, Choe SK, Runko AP, Gardner PD, Sagerstrom CG. 2008. Nlz1/Znf703 acts as a repressor of transcription. Bmc Dev Biol 8:108.

Nigg EA. 1997. Nucleocytoplasmic transport: Signals, mechanisms and regulation. Nature 386:779–787.

Ota T, Suzuki Y, Nishikawa T, Otsuki T, Sugiyama T, Irie R, Wakamatsu A, Hayashi K, Sato H, Nagai K, Kimura K, Makita H, Sekine M, Obayashi M, Nishi T, Shibahara T, Tanaka T, Ishii S, Yamamoto J, Saito K, Kawai Y, Isono Y, Nakamura Y, Nagahari K, Murakami K, Yasuda T, Iwayanagi T, Wagatsuma M, Shiratori A, Sudo H, Hosoiri T, Kaku Y, Kodaira H, Kondo H, Sugawara M, Takahashi M, Kanda K, Yokoi T, Furuya T, Kikkawa E, Omura Y, Abe K, Kamihara K, Katsuta N, Sato K, Tanikawa M, Yamazaki M, Ninomiya K, Ishibashi T, Yamashita H, Murakawa K, Fujimori K, Tanai H, Kimata M, Watanabe M, Hiraoka S, Chiba Y, Ishida S, Ono Y, Takiguchi S, Watanabe S, Yosida M, Hotuta T, Kusano J, Kanehori K, Takahashi-Fujii A, Hara H, Tanase T, Nomura Y, Togiya S, Komai F, Hara R, Takeuchi K, Arita M, Imose N, Musashino K, Yuuki H, Oshima A, Sasaki N, Aotsuka S, Yoshikawa Y, Matsunawa H, Ichihara T, Shiohata N, Sano S, Moriya S, Momiyama H, Satoh N, Takami S, Terashima Y, Suzuki O, Nakagawa S, Senoh A, Mizoguchi H, Goto Y, Shimizu F, Wakebe H, Hishigaki H, Watanabe T, Sugiyama A, Takemoto M, Kawakami B, Yamazaki M, Watanabe K, Kumagai A, Itakura S, Fukuzumi Y, Fujimori Y, Komiyama M, Tashiro H, Tanigami A, Fujiwara T, Ono T, Yamada K, Fujii Y, Ozaki K, Hirao M, Ohmori Y, Kawabata A, Hikiji T, Kobatake N, Inagaki H, Ikema Y, Okamoto S, Okitani R, Kawakami T, Noguchi S, Itoh T, Shigeta K, Senba T, Matsumura K, Nakajima Y, Mizuno T, Morinaga M, Sasaki M, Togashi T, Oyama M, Hata H, Watanabe M, Komatsu T, Mizushima-Sugano J, Satoh T, Shirai Y, Takahashi Y, Nakagawa K, Okumura K, Nagase T, Nomura N, Kikuchi H, Masuho Y, Yamashita R, Nakai K, Yada T, Nakamura Y, Ohara O, Isogai T, Sugano S. 2004. Complete sequencing and characterization of 21,243 full-length human cDNAs. Nat Genet 36:40–45.

Pauws E, van Kampen AH, van de Graaf SA, de Vijlder JJ, Ris-Stalpers C. 2001. Heterogeneity in polyadenylation cleavage sites in mammalian mRNA sequences: Implications for SAGE analysis. Nucleic Acids Res 29:1690–1694.

Pereira-Castro I, Quental R, da Costa LT, Amorim A, Azevedo L. 2012. Successful COG8 and PDF overlap is mediated by alterations in splicing and polyadenylation signals. Hum Genet 131:265–274.

Philipsen S, Suske G. 1999. A tale of three fingers: The family of mammalian Sp/XKLF transcription factors. Nucleic Acids Res 27:2991–3000.

Qian W, Liao BY, Chang AY-F, Zhang J. 2010. Maintenance of duplicate genes and their functional redundancy by reduced expression. Trends Genet 26:425–430.

Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. 2005. InterProScan: Protein domains identifier. Nucleic Acids Res 33:W116–W120.

Ray ME, Yang ZQ, Albertson D, Kleer CG, Washburn JG, Macoska JA, Ethier SP. 2004. Genomic and expression analysis of the 8p11–12 amplicon in human breast cancer cell lines. Cancer Res 64:40–47.

Ren B, Chee KJ, Kim TH, Maniatis T. 1999. PRDI-BF1/Blimp-1 repression is mediated by corepressors of the Groucho family of proteins. Genes Dev 13:125–137.

Runko AP, Sagerstrom CG. 2003. Nlz belongs to a family of zinc-finger-containing repressors and controls segmental gene expression in the zebrafish hindbrain. Dev Biol 262:254–267.

Runko AP, Sagerstrom CG. 2004. Isolation of nlz2 and characterization of essential domains in Nlz family proteins. J Biol Chem 279:11917–11925.

Sagerstrom GC, Kao BK, Lane ME, Sive H. 2001. Isolation and characterization of posteriorly restricted genes in the zebrafish gastrula. Dev Dyn 220:402–408.

Schaeper ND, Prpic NM, Wimmer EA. 2010. A clustered set of three Sp-family genes is ancestral in the Metazoa: Evidence from sequence analysis, protein domain structure, developmental expression patterns and chromosomal location. BMC Evol Biol 10:88.

Sircoulomb F, Nicolas N, Ferrari A, Finetti P, Bekhouche I, Rousselet E, Lonigro A, Adelaide J, Baudelet E, Esteyries S, Wicinski J, Audebert S, Charafe-Jauffret E, Jacquemier J, Lopez M, Borg JP, Sotiriou C, Popovici C, Bertucci F, Birnbaum D, Chaffanet M, Ginestier C. 2011. ZNF703 gene amplification at 8p12 specifies luminal B breast cancer. EMBO Mol Med 3(3):153–166.

Slorach EM, Chou J, Werb Z. 2011. Zeppo1 is a novel metastasis promoter that represses E-cadherin expression and regulates p120-catenin isoform expression and localization. Genes Dev 25:471–484.

Suske G. 1999. The Sp-family of transcription factors. Gene 238:291–300.

Tian B, Hu J, Zhang H, Lutz CS. 2005. A large-scale analysis of mRNA polyadenylation of human and mouse genes. Nucleic Acids Res 33:201–212.

Vienne A, Rasmussen J, bi-Rached L, Pontarotti P, Gilles A. 2003. Systematic phylogenomic evidence of en bloc duplication of the ancestral 8p11.21–8p21.3-like region. Mol Biol Evol 20:1290–1298.

Weihe U, Dorfman R, Wernet MF, Cohen SM, Milan M. 2004. Proximodistal subdivision of Drosophila legs and wings: The elbow-no ocelli gene complex. Development 131:767–774.

Yan J, Marr TG. 2005. Computational analysis of 3'-ends of ESTs shows four classes of alternative polyadenylation in human, mouse, and rat. Genome Res 15:369–375.

Yang ZQ, Streicher KL, Ray ME, Abrams J, Ethier SP. 2006. Multiple interacting oncogenes on the 8p11–p12 amplicon in human breast cancer. Cancer Res 66:11632–11643.

Zhao XJ, Yang Y, Fitch DHA, Herman MA. 2002. TLP-1 is an asymmetric cell fate determinant that responds to Wnt signals and controls male tail tip morphogenesis in C-elegans. Development 129:1497–1508.