J. Biochem. 132, 875–879 (2002)

# Characterization of Functional Regions for Nuclear Localization of NPAT[1]

Masashi Sagara,* Eri Takeda,*,†,[2] Akiyo Nishiyama,*,†,[3] Syunsaku Utsumi,† Yoshiroh Toyama,‡ Shigeki Yuasa,‡ Yasuharu Ninomiya,*,[4] and Takashi Imai*,[5]

*Frontier Research Center, National Institute of Radiological Sciences, 4-9-1 Anagawa, Inage-ku, Chiba 263-8555; †Faculty of Education, Chiba University, 1-33 Yayoi, Inage-ku, Chiba 263-8522; and ‡Graduate School of Medicine, Chiba University, 1-8-1 Inohana, Chuo-ku, Chiba 260-8670

Received January 18, 2002; accepted September 13, 2002

**NPAT plays a role in S phase entry as a substrate of cyclin E-CDK2 and activation of histone gene transcription. Although analysis of its sequence indicates that NPAT contains typical nuclear localization signals (NLS) comprising segments of positively charged amino acids, there are currently no experimental data to show that these predictive NLS are functional. To investigate whether these sequences are effective for nuclear transport of NPAT, an NPAT-green fluorescent protein fusion (NP-GFP) was constructed. After transfection of the fusion gene containing the full coding region of NPAT into cultured cells, the NP-GFP product was found exclusively in the nucleus. As expected, some deletion mutants that retained the basic amino acid clusters at the carboxyl terminus also localize the fusion protein in the nucleus. However, other fusions that lacked one of the three basic amino acid-clusters were distributed throughout the nucleus and cytoplasm. Therefore all three clusters of basic residues are necessary for localization of NPAT to the nucleus. However, another sequence outside the carboxyl terminal region functions similarly to NLS. Construction of GFP fusions with a series of truncated forms of NPAT indicated that a short peptide sequence consisting of mainly hydrophobic amino acids near the central domain of NPAT also contributes to localizing the protein in the nucleus.**

**Key words: basic amino acid cluster, GFP, hydrophobic region, NLS, NPAT.**

Ataxia telangiectasia (AT) is an autosomal recessive gene disorder characterized by a wide spectrum of defects including progressive cerebellar ataxia, oculocutaneous telangiectasia, immunological defects, and an increased incidence of cancer (*1–3*). The gene responsible for AT, ATM, was isolated from human chromosome 11q22-q23 and found to be mutated in a large number of AT patients (*4–10*). We and other groups have independently identified a housekeeping gene NPAT (E14/Cand3), which lies 0.5 kb from the 5′ end of the ATM and is transcribed divergently to it (*11–14*). Each gene may influence the expression of the other. Our nucleotide sequence analysis of NPAT revealed that the gene encodes a protein of 1,427 amino acids with a predicted molecular mass of 143 kDa. The protein contains multiple phosphorylation target sites for several types of protein kinase, for example, cyclin E–CDK2 (*12*).

NPAT plays a role in S phase entry and acts as a substrate for cyclin E–CDK2 and activation of histone gene transcription (*15–17*). Its cellular overexpression promotes S phase entry (*15*). Inactivation of NPAT by insertion of viral DNA in mice causes embryonic arrest at the eight-cell stage (*18*). Together, these data indicate that NPAT is an essential gene in both cell cycle control and early development.

Previous nucleotide and deduced amino acid sequence analysis of NPAT showed that NPAT has sequences matching the classical nuclear localization signals (NLS) consisting of clusters of four basic amino acids, $^{1368}$KKRK$^{1371}$, $^{1397}$KKKK$^{1400}$, and $^{1402}$KKKK$^{1405}$ (*12*). To verify that these charged amino acid clusters in the C-terminal domain are essential for nuclear translocation of NPAT, we constructed a chimeric gene consisting of green fluorescence protein (GFP)–NPAT fusion which was transfected into cultured mammalian cells. Analysis of several truncated NPAT fusions indicates that both the C-terminal domain and a hydrophobic region in the central domain of NPAT assist in localizing the GFP fusion protein in the nucleus.

## MATERIALS AND METHODS

*Plasmid Construction*—The eukaryotic expression construct for GFP, pCMX-SAH/Y145F (*19*), was kindly gifted

[1] This work was supported in part by Grants-in-Aid for Scientific Research on Priority Areas (Human Genome Project) from the Ministry of Education, Science, Sports and Culture of Japan.

Present addresses: [2] Faculties of Agriculture, Gifu University, Yanagido 1-1, Gifu, Gifu 501-1193; [3] Department of Neuromuscular Research, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Ogawahigasimachi 4-1-1, Kodaira, Tokyo 187-8502; [4] Radiation Hazards Research Group, National Institute of Radiological Sciences, Chiba 263-8555.

[5] To whom correspondence should be addressed. Tel: +81-43-206-3138, Fax: +81-43-206-4131, E-mail: imait@nirs.go.jp

Abbreviations: AT, ataxia telangiectasia; ATM, ataxia telangiectasia mutated; DAPI, 4′,6-diamidino-2-phenylindole; GFP, green fluorescent protein; NLS, nuclear localization signal; NPAT, nuclear protein mapped AT locus.

by Drs. Ogawa and Umesono (Kyoto University, Kyoto). A plasmid containing full-length NPAT cDNA, pcNPAT, was re-constructed by ligation of a *Bam*HI–*Pst*I fragment of pP3-9, a *Pst*I–*Pst*I fragment of pT4-41, and a *Pst*I–*Bam*HI fragment of pT4-81 (*12*). To generate GFP fusion constructs, a *Sph*I (nucleotide number 359 in DDBJ/EMBL/GenBank accession number D83243)–*Sca*I (nucleotide number 4822) fragment of pcNPAT was fused in-frame to pCMX-SAH/Y145F, and this construct, named NP-GFP, was used for site-directed mutagenesis.

Mutants that lacked the predicted nuclear localization signals were constructed using the QuickChange Site-Directed Mutagenesis Kit (Stratagene) following the procedure recommended by the manufacturer. PCR primers for introduction of mutations were 5′-CAAGGAGCACCACAAAAAAGCG**CTAA**ATTGAGGAATTAGATGAACG-3′ for mA ($^{1368}$KKR**stop**$^{1371}$), 5′-CATCAATACCAATGAAA**GAGC**TGAAAATTAAGAAAAAAGAAGTTCCC-3′ for mB ($^{1397}$KKKK$^{1400}$ to $^{1397}$**KELK**$^{1400}$), 5′-CCAATGAAAAAAGAAGAAAATTAAGCTTAAGAAGCTTCCCAGTTCATTTCC-3′ for mC ($^{1402}$KKKK$^{1405}$ to $^{1402}$**KLKK**$^{1405}$), and 5′-CAAGGAGCACCACAAAA**GAGCT**GAAAATTGAGGAATTAGATGAACG-3′ for mD ($^{1368}$KKRK$^{1371}$ to $^{1368}$**KELK**$^{1371}$). The sites of mutation are highlighted in bold typeface. In the cases of mA and mC, a restriction enzyme recognition site for *Eco*47III or *Afl*II, respectively, was created to confirm nucleotide changes by mutagenesis. For amplification of mutated fragments, the following temperature and time profiles were used: 18 cycles of 95°C for 30″, 55°C for 1′, 68°C for 20′.

The entire NPAT coding region was used for construction of a series of deletion fragments, D1, D5, D7, D9, and D10. These constructs were PCR-amplificated using *Pfu* Turbo polymerase (Stratagene) and pcNPAT as a template. The primer sequences used here were: 5′-TATCGGTACCATGTTGTTACCCTCGACGTAG-3′ (nucleotide number 67–90) for NP5F, 5′-TATCGGTACCTGTGGTGCTCCTTGAAGATGC-3′ (4167–4145) for NPD1R, 5′-TATCGGTACCTCCTTAATACTAGAAACAGCAC-3′ (2361–2337) for NPD5R, 5′-TATCGGTACCTGTCAAAGGCATATGAAGAACC-3′ (2988–2967) for NPD7R, 5′-TATCGGTACCTACAGTTTCTTCTGAAGATAGG-3′ (2454–2433) for NPD9R, and 5′-TATCGGTACCTGACACAGCTTGGTTGACAGC-3′ (2835–2815) for NPD10R. NP5F is a common forward primer and the other primers are reverse primers. The resulting PCR fragment was digested with *Kpn*I (indicated by the underline in the primer sequences) and then cloned in-frame into the *Kpn*I site of the pCMX-SAH/Y145F.

To analyze the characteristics of short DNA fragments encoding NLS candidates, another series of GFP fusions was constructed. The DNA fragments were synthesized by PCR as indicated above. The primer sequences were: 5′-CGCGGATCCAGTAGTACATCAAAAGTAATGGTC-3′ (3961–3984) for CA, 5′-CGCGGATCCGCATCTTCAAGGAGCACCAC-3′ (4147–4166) for CB, 5′-CGCGGATCCCTTACAAATTCATCAATACCAATG-3′ (4231–4254) for CC, 5′-CGCGGATCCCTTCCCAGTTCATTTCCAGCA-3′ (4282–4302) for CD, 5′-CGCGGATCCTTACTCATCATAATGCAATGATAAC-3′ (4350–4326) for CE, 5′-CGCGGATCCTGCTGGAAATGAACTGGGAAG-3′ (4302–4282) for CF, 5′-CGCGGATCCCATTGGTATTGATGAATTTGTAAG-3′ (4254–4231) for CG, 5′-CGCGGATCCTGTGGTGCTCCTTGAAGATGC-3′ (4167–4147) for CH, 5′-CGCGGATCCGTGTCACCAAACTTTTCAC-3′ (2830–2848) for IA, 5′-CGCGGATCCGTACTCCAAGGAATGGTAG-3′ (2887–2905) for IB, 5′-CGCGGATCCAATAACTTTTCTACTCCTCC-3′ (2941–2960) for IC, 5′-CGCGGATCCGCATACAGGTGCTGTCAAAG-3′ (3000–2981) for ID, 5′-CGCGGATCCGTTATTTCCATTCTGTCCAAC-3′ (2946–2926) for IE, and 5′-CG-CGGATCCGAGTACAGGCTGGACTGG-3′ (2892–2875) for IF. The resulting PCR fragments were digested with *Bam*HI (indicated by the underline in the primer sequences) and then cloned in-frame into the *Bam*HI site of the pCMX-SAH/Y145F. All constructs were verified by DNA sequence analysis to ensure no mistakes had been introduced during amplification.

*Cell Culture and Transient Transfection*—COS7 and HeLa cell lines were maintained at 37°C in a humidified 5% $CO_2$ incubator, and cells were grown in Dulbecco's modified Eagle medium supplemented with 10% fetal bovine serum, 2 mM L-glutamine, 100 U/ml penicillin, and 100 μg/ml streptomycin. The day prior to transfection, $5\times10^4$ cells were seeded onto glass cover slips per well of a 6-well plate. Transient transfection of 3 μg of plasmid DNA and 6 μl of FuGENE™6 Transfection Reagent (Roche) was performed as described by the manufacturer. At 48 h after transfection, cells were shifted to 30°C for a further 48 h before processing for microscopy (*19*).

*Fluorescence Microscopy*—Cells were washed twice in phosphate buffered saline (PBS), then fixed in 4% paraformaldehyde in PBS for 30 min (*19*). Following a further three washes in PBS, cells were mounted in 0.1 μg/ml DAPI in glycerol. Most images were obtained using conventional Nikon or Olympus microscopes equipped with a fluorescein isothiocyanate filter set and using a X40 objective. GFP and phase-contrast images were photographed on Fuji provia 400 film. Confocal images were obtained using a Bio-Rad MRC600 confocal system with a microscope equipped with a X60 objective. The image files were digitally processed for presentation using Adobe Photoshop.

## RESULTS AND DISCUSSION

To identify functional sequence(s) for targeting NPAT into the nucleus, we constructed a series of expression vectors for green fluorescence protein (GFP)–tagged NPAT (Fig. 1). Both the GFP-whole NPAT (NP-GFP) and GFP-NPAT lacking the first 98 amino acids (NPdN-GFP) were found exclusively in the nuclei when expressed in COS7 cells (Fig. 2, A and B). Similar results were observed when the constructs were transfected into HeLa cells (data not shown). Because the amino acid sequence of NPAT indicated that there are three clusters of basic residues at its carboxyl terminus, DNA fragments encoding the carboxyl terminal 130 amino acids (GFP-CAE) and 52 amino acids (GFP-CBF) sequences respectively were first fused to GFP, then transfected into COS7 cells. As expected, these fragments clearly directed the GFP signal into the nucleus (Fig. 2, C and E), indicating that the minimal 52 amino acids from the carboxyl terminal region of NPAT contains the nuclear localization signal(s) (NLS). Because the region contains clusters of basic amino acids, each of these was deleted or mutated to ascertain whether all of these clusters were essential as NLS.

The fusion protein lacking the first basic amino acid cluster, GFP-CCF, was imported into the nucleus, but the GFP

signal could also be detected in cytoplasm of the transfected cells (Fig. 2G). A similar fluorescent pattern was obtained when both the second and third basic amino acid clusters were simultaneously deleted (GFP-CBG, Fig. 2F).

When each of the basic amino acid clusters was mutated to $^{1368}$KELK$^{1371}$, $^{1397}$KELK$^{1400}$, and $^{1402}$KLKK$^{1406}$ respectively to disrupt the continuous basic amino acid stretch, the fluorescence in each of the transfected cells was detected both in the nucleus and cytoplasm (Fig. 2, I and J, K). The $^{1397}$KKKK$^{1400}$ and $^{1402}$KKKK$^{1406}$ sequences might to work cooperatively as a single nuclear localization signal, $^{1397}$KKKKIKKKK$^{1405}$, but here they appear to function independently. These data indicate that all three continuous

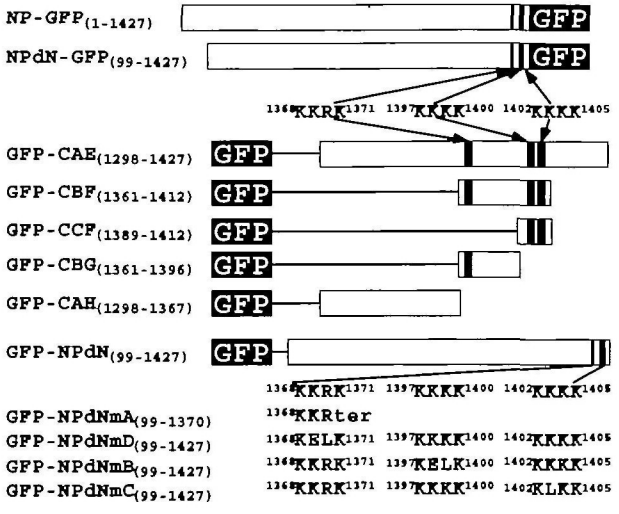



Fig. 1. **Schematic representation of green fluorescent protein (GFP)–tagged NPAT.** The clone at the top (NP-GFP) is a full-length human NPAT. The three clusters of basic residues at the carboxyl terminus of NPAT ($^{1368}$KKRK$^{1371}$, $^{1397}$KKKK$^{1400}$, and $^{1402}$KKKK$^{1406}$) are shown as filled boxes. Numbers after the clone name indicate the respective NPAT amino acids.

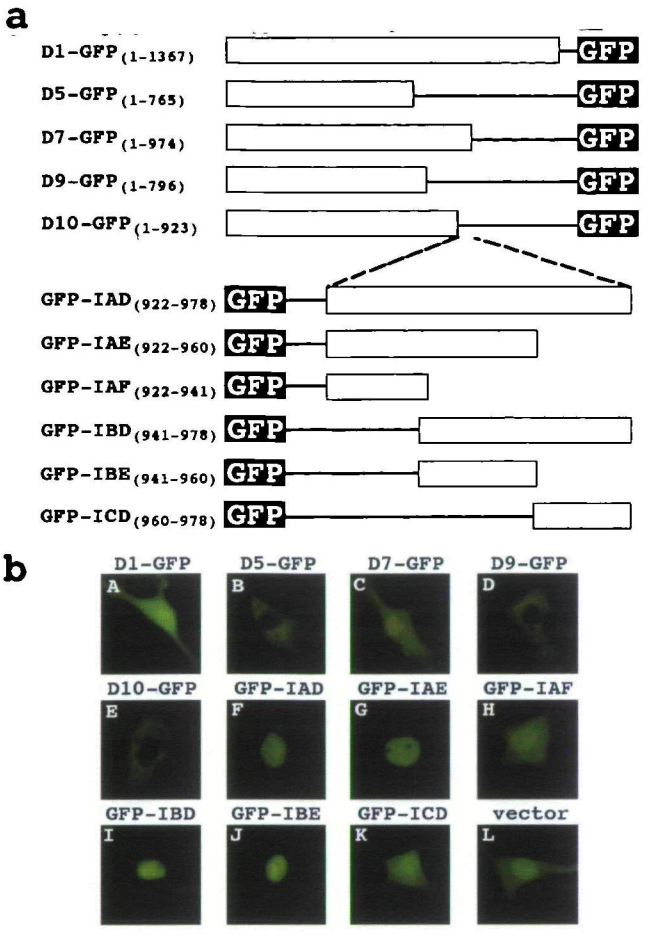
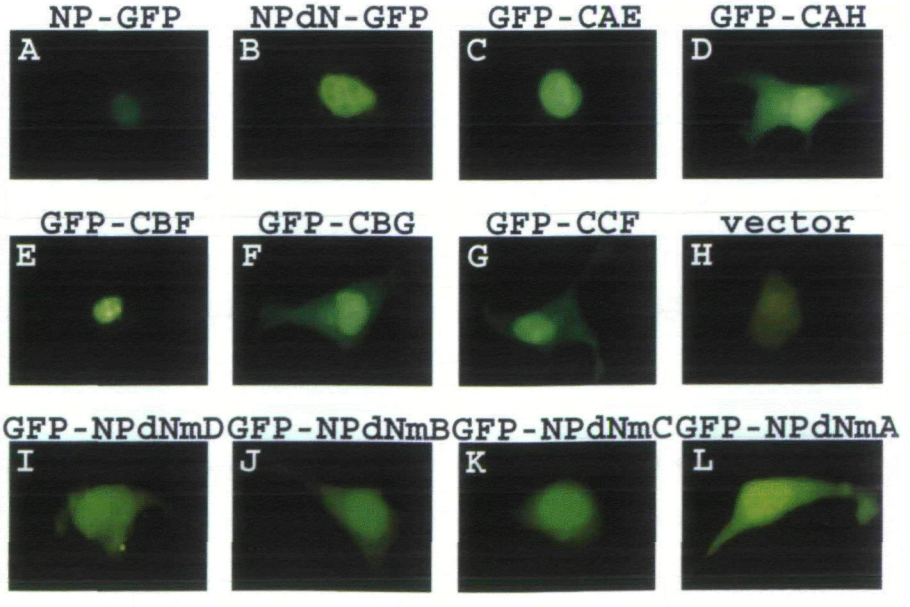



Fig. 3. **Subcellular distribution of various NPAT deletion mutants in the central region tagged with GFP.** (a) Constructs of GFP-tagged NPAT deletion mutants in the central region. (b) Fluorescent images of the COS7 cells transfected with each construct shown in (a).

Fig. 2. **Subcellular distribution of various NPAT mutants tagged with GFP in transfected COS7 cells.** The fluorescent images of the COS7 cells transfected with each construct shown in Fig. 1.



basic amino acid clusters in the carboxyl terminus of NPAT are necessary to localize the protein into nucleus completely.

To examine whether the region containing these three basic amino acid clusters is sufficient to localize NPAT into nucleus, a NPAT-GFP fusion gene lacking the basic residue clusters (D1-GFP) was constructed and transfected into the cultured cells (Fig. 3b, A). Surprisingly, the fusion protein was still detected in both the nucleus and the cytoplasm. This observation was confirmed by confocal microscopy (data not shown). Previous amino acid sequence analysis could not predict any candidate nuclear localization signal sequences in NPAT other than the carboxyl terminal sequence indicated above. Therefore, we constructed a series of deletion mutants of NPAT fused to GFP as indicated in Fig. 3a to search for additional sequence elements which may be important for the nuclear import of NPAT.

When half of the amino terminal region of NPAT, D5-GFP consisting of the amino acids 1 to 765 (Fig. 3b, B), was expressed in the cells, the fusion protein could no longer be imported into the nucleus. However, the deletion mutant, D7-GFP (amino acids 1 to 974) was still detected in both the nucleus and the cytoplasm (Fig. 3b, C). Two other deletion mutants, D9-GFP (amino acids 1 to 796) and D10-GFP (amino acids 1 to 923), were both retained solely in the cytoplasm (Fig. 3b, D and E). Therefore, we focused on the region from amino acid 923 to 974 of NPAT for further analysis. Small fragments of approximately 20 residues, corresponding to amino acids 922–941, 941–960, and 960–978, were synthesized by PCR and each was fused to GFP (GFP-IAF, GFP-IBE, and GFP-ICD in Fig. 3a). Although these fusion proteins were small enough to diffuse into the nucleus, only the GFP fusion protein containing the amino acid 941–960 of NPAT (GFP-IBE in Fig. 3b, J) was detected exclusively in the nuclei of the transfected cells. This finding suggests that the IBE region of NPAT may assist in localizing the protein in the nucleus. The fact that fusion proteins such as D7-GFP and D1-GFP could not be transported into the nucleus completely suggests that the IBE region, consisting of the sequence, VLQGMVGMIPVSVVGQNGNN, may function to localize the protein in the nucleus or may act like a number of proteins that are associated with other nuclear-targeting proteins (*20*).

So far, we have not found any proteins that contain a homologous sequence to these 20 amino acids in the public protein databases. The nuclear localization signals at the carboxyl terminus of NPAT have the typical NLS sequences (*21, 22*) and activity to transport NPAT into the nucleus. The near central region of NPAT may work either by facilitating NPAT transport to the nucleus or by retaining the protein inside the nucleus. Because NPAT was implicated in binding to the upstream region of histone genes (*16, 17*), it will be interesting to examine whether this sequence is a factor in retaining the protein inside the nucleus.

We are grateful to Drs. Ogawa and Umesono (Kyoto University) for pCMX-SAH/Y145F, to Drs. Ozaki (Chiba Cancer Center), Tsuji, Inaba, and Ohyama (NIRS) for valuable suggestions for microscopic imaging and to T. Maeda for excellent technical assistance.

REFERENCES

1. Gatti, R.A., Boder, E., Vinters, H.V., Sparkes, R.S., Norman, A., and Lange, K. (1991) Ataxia-telangiectasia: An interdisciplinary approach to pathogenesis. *Medicine* **70**, 99–117
2. Harnden, D.G. (1994) The nature of ataxia-telangiectasia: Problems and perspectives. *Int. J. Radiat. Biol.* **66**, S13–S19
3. Shiloh, Y. (1995) Ataxia-telangiectasia: Closer to unraveling the mystery. *Eur. J. Hum. Genet.* **3**, 116–138
4. Savitsky, K., Bar-Shira, A., Gilad, S., Rotman, G., Ziv, Y., Vanagaite, L., Tagle, D.A., Smith, S., Uziel, T., Sfez, S., Ashkenazi, M., Pecker, I., Frydman, M., Harnik, R., Patanjali, S.R., Simons, A., Clines, G.A., Sartiel, A., Gatti, R.A., Chesa, L., Sanal, O., Lavin, M.F., Jaspers, N.G.J., Taylor, A.M.R., Arlett, C.F., Miki, T., Weossman, S.M., Lovett, T., Collins, F.S., and Shiloh, Y. (1995) A single ataxia telangiectasia gene with a product similar to PI-3 kinase. *Science* **268**, 1749–1753
5. Savitsky, K., Sfez, S., Tangle, D.A., Ziv, Y., Sartiel, A., Collins, F.S., Shiloh, Y., and Rotman, G. (1995) The complete sequence of the coding region of the ATM gene reveals similarity to cell cycle regulators in different species. *Hum. Mol. Genet.* **4**, 2025–2032
6. Gilad, S., Khosravi, R., Shekedy, D., Uziel, T., Ziv, Y., Savitsky, K., Rotman, G., Smith, S., Chessa, L., Jorgensen, T.J., Harnik, R., Frydman, M., Sanal, O., Portnoi, S., Goldwicz, Z., Jaspers, N.G.J., Gatti, R.A., Lenoir, G., Lavin, M.F., Tatsumi, K., Wegner, R.D., Shiloh, Y., and Bar-Shira, A. (1996) Predominance of null mutations in ataxia-telangiectasia. *Hum. Mol. Genet.* **5**, 433–439
7. Telatar, M., Wang, Z., Udar, N., Liang, T., Bernatowaska-Matsuszkiewicz, E., Lavin, M., Shiloh, Y., Concannon, P., Good, R.A., and Gatti, R.A. (1996) Ataxia-telangiectasia: Mutations in ATM cDNA detected by protein-truncation screening. *Am. J. Hum. Genet.* **59**, 40–44
8. McConville, C.M., Stankovic, T., Byrd, P.J., McGuire, G.M., Yao, Q.-Y., Lennox, G.G., and Taylor, A.M.R. (1996) Mutations associated with variant phenotypes in ataxia-telangiectasia. *Am. J. Hum. Genet.* **59**, 320–330
9. Byrd, P.J., McConville, C.M., Cooper, P., Parkhill, J., Stankovic, T., McGuire, G.M., Thick, J.A., and Taylor, A.M.R. (1996) Mutations revealed by sequencing the 5′ half of the gene for ataxia telangiectasia. *Hum. Mol. Genet.* **5**, 145–149
10. Wright, J., Teraoka, S., Onengut, S., Tolun, A., Gatti, R.A., Oches, H.D., and Concannon, P. (1996) A high frequency of distinct ATM gene mutations in ataxia telangiectasia. *Am. J. Hum. Genet.* **59**, 839–846
11. Byrd, P.J., Cooper, P.R., Stankovic, T., Kullar, H.S., Watts, G.D.J., Robinson, P.J., and Taylor, A.M.R. (1996) A gene transcribed from the bi-directional ATM promoter coding for a serine rich protein: amino acid sequence, structure and expression studies. *Hum. Mol. Genet.* **5**, 1785–1791
12. Imai, T., Yamauchi, M., Seki, N., Sugawara, T., Saito, T., Matsuda, Y., Ito, H., Nagase, T., Nomura, N., and Hori, T. (1996) Identification and characterization of a new gene physically linked to the ATM gene. *Genome Res.* **6**, 439–447
13. Chen, X., Yang, L., Udar, N., Liang, T., Uhrhammer, N., Xu, S., Bay, J.O., Wang, Z., Dandakar, S., Chiplunkar, S., Klisak, I., Telatar, M., Yang, H., Concannon, P., and Gatti, R.A. (1997) CAND3: a ubiquitously expressed gene immediately adjacent and in opposite transcriptional orientation to the ATM gene at 11q23.1. *Mamm. Genome* **2**, 129–133
14. Imai, T., Sugawara, T., Nishiyama, A., Shimada, R., Ohki, R., Seki, N., Sagara, M., Ito, H., Yamauchi, M., and Hori, T. (1997) The structure and organization of the human NPAT gene. *Genomics* **42**, 388–392
15. Zhao, J., Dynlacht, B., Imai, T., Hori, T., and Harlow, E. (1998) Expression of NPAT, a novel substrate of cyclin E-CDK2, promotes S-phase entry. *Genes Dev.* **12**, 456–461
16. Zhao, J., Kennedy, B.K., Lawrence, B.D., Barbie, D.A., Matera, A.G., Fletcher, J.A., and Harlow, E. (2000) NPAT links cyclin E-Cdk2 to the regulation of replication-dependent histone gene transcription. *Genes Dev.* **14**, 2283–2297

17. Ma, T., Van-Tine, B.A., Wei, Y., Garrett, M.D., Nelson, D., Adams, P.D., Wang, J., Qin, J., Chow, L.T., and Harper, J.W. (2000) Cell cycle-regulated phosphorylation of p220 (NPAT) by cyclin E/Cdk2 in Cajal bodies promotes histone gene transcription. *Genes Dev.* **14**, 2298–2313
18. Di-Fruscio, M., Weiher, H., Vanderhyden, B.C., Imai, T., Shiomi, T., Hori, T.A., Jaenisch, R., and Gray, D.A. (1997) Proviral inactivation of the Npat gene of Mpv 20 mice results in early embryonic arrest. *Mol. Cell. Biol.* **17**, 4080–4086
19. Ogawa, H., Inouye, S., Tsuji, F.I., Yasuda, K., and Umesono, K. (1995) Localization, trafficking, and temperature-dependence of the Aequorea green fluorescent protein in cultured vertebrate cells. *Proc. Natl. Acad. Sci. USA* **92**, 11899–11903
20. Muslin, A.J. and Xing, H. (2000) 14-3-3 proteins: regulation of subcellular localization by molecular interference. *Cell. Signal.* **12**, 703–709
21. Lanford, R.E. and Butel, J.S. (1984) Construction and characterization of an SV40 mutant defective in nuclear transport of antigen. *Cell* **37**, 801–813
22. Kalderon, D., Roberts, B.L., Richardson, W.D., and Smith, A.E. (1984) A short amino acid sequence able to specify nuclear location. *Cell* **39**, 499–509