Heterogeneous nuclear ribonucleoproteins H, H', and F are members of a ubiquitously expressed subfamily of related but distinct proteins encoded by genes mapping to different chromosomes

Honoré, Bent; Rasmussen, Hanne H.; Vorum, Henrik; Dejgaard, Kurt; Liu, Xiangdang; Gromov, Pavel; Madsen, Peder; Gesser, Borbala; Tommerup, Niels; Celis, Julio E.

Published in:
Journal of Biological Chemistry

DOI (link to publication from Publisher):
10.1074/jbc.270.48.28780

Publication date:
1995

Document Version
Publisher's PDF, also known as Version of record

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 270, No. 48, Issue of December 1, pp. 28780–28789, 1995
Printed in U.S.A.

Heterogeneous Nuclear Ribonucleoproteins H, H′, and F Are Members of a Ubiquitously Expressed Subfamily of Related but Distinct Proteins Encoded by Genes Mapping to Different Chromosomes*

(Received for publication, May 17, 1995, and in revised form, August 25, 1995)

Bent Honoré‡§, Hanne H. Rasmussen‡¶, Henrik Vorum‡‖, Kurt Dejgaard‡¶, Xiangdang Liu‡, Pavel Gromov‡¶, Peder Madsen‡, Borbala Gesser‡¶‡‡, Niels Tommerup§§, and Julio E. Celis‡**

From the Danish Centre for Human Genome Research, ‡Department of Medical Biochemistry, University of Aarhus, Ole Worms Allé, Building 170, DK-8000 Aarhus C, Denmark and the §§John F. Kennedy Institute, Gl. Landevej 7, DK-2600 Glostrup, Denmark

Molecular cDNA cloning, two-dimensional gel immunoblotting, and amino acid microsequencing identified three sequence-unique and distinct proteins that constitute a subfamily of ubiquitously expressed heterogeneous nuclear ribonucleoproteins corresponding to hnRNPs H, H′, and F. These proteins share epitopes and sequence identity with two other proteins, isoelectric focusing sample spot numbers 2222 (37.6 kDa; pI 6.5) and 2326 (39.5 kDa; pI 6.6), indicating that the subfamily may contain additional members. The identity between hnRNPs H and H′ is 96%, between H and F 78%, and between H′ and F 75%, respectively. The three proteins contain three repeats, which we denote quasi-RRMs (qRRMs) since they have a remote similarity to the RNA recognition motif (RRM). The three qRRMs of hnRNP H, with a few additional NH_2-terminal amino acids, were constructed by polymerase chain reaction amplification and used for ribohomopolymer binding studies. Each qRRM repeat bound poly(rG), while only the NH_2-terminal qRRM bound poly(rC) and poly(rU). None of the repeats bound detectable amounts of poly(rA). The expression levels of hnRNPs H and F were differentially regulated in pairs of normal and transformed fibroblasts and keratinocytes. In normal human keratinocytes, the expression level of H was unaffected by treatment with several substances tested including two second messengers and seven cytokines. Likewise the expression level of F was independent of these substances, although it was strikingly down-regulated by long term treatment with 4β-phorbol 12-myristate 13-acetate, indicating that the protein kinase C signaling pathway regulates its expression. No effect of 4β-phorbol 12-myristate 13-acetate was observed on the expression of hnRNP H. The genes coding for hnRNPs H, H′, and F were chromosome-mapped to 5q35.3 (*HNRPH1*), 6q25.3-q26, and/or Xq22 (*HNRPH2*) and 10q11.21-q11.22 (*HNRPF*), respectively.

Heterogeneous nuclear ribonucleoproteins (hnRNPs)[1] constitute a set of polypeptides that bind heterogeneous nuclear RNA, the transcripts produced by RNA polymerase II (Dreyfuss *et al.*, 1993). While all of the hnRNPs are present in the nucleus some seem to shuttle between the nucleus and the cytoplasm. The full range of functions of the hnRNPs are at present unknown, although these may include pre-mRNA processing and transport as well as roles in the interaction of heterogeneous nuclear RNA with other nuclear structures (Dreyfuss *et al.*, 1993). To date, more than 20 such proteins have been described and designated with letters from A to U (Dreyfuss *et al.*, 1993). The different subclasses are characterized by their preferred binding of ribohomopolymers at 2 M NaCl; for example hnRNPs F, P, H, and M, and a subset of the E proteins bind poly(rG), hnRNP P binds poly(rA), and hnRNPs C and M bind poly(rU), while hnRNPs K and J bind poly(rC) (Swanson and Dreyfuss, 1988). The primary structure of several of these proteins, which have been deduced from their corresponding cDNAs, often exhibit a modular structure that contains one or more RNA binding domains (the RNA recognition motif, RRM also called the RNP motif; the arginine-rich motif; the RGG box; the KH motif; and other types) as well as one or more auxiliary domains (with clusters rich in a few amino acids, frequently glycines) that have been assigned a number of different putative functions (Biamonti and Riva, 1994). Isoforms of these proteins have been reported and are generated by alternative processing of pre-mRNA and by post-translational modifications such as phosphorylations of serines and threonines and methylation of arginines as well as other as yet uncharacterized modifications.

* The research was supported by grants from the Danish Medical Research Council, the Danish Biotechnology Programme, the Danish Cancer Society, the EEC Science Programme, the Velux Fund, the Danish Rheumatoid Society, Novo Nordisk, the Danish Foundation for the Advancement of Medical Science, Aarhus University Research Foundation, Fru Astrid Thaysens Legat for Lægevidenskabelig Grundforskning, and Grosserer Valdemar Foersom og Hustru Thyra Foersom, født Otto's Fond. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) L22009.

§ To whom correspondence should be addressed: Dept. of Medical Biochemistry and Danish Centre for Human Genome Research, University of Aarhus, Ole Worms Allé, Bldg. 170, DK-8000 Aarhus C, Denmark. Tel.: 45-89-42-28-88; Fax: 45-86-13-11-60; E-mail: BEH@biobase.dk.

¶ Supported by the Danish Cancer Foundation.

‖ Supported by the Michaelsen Foundation.

** Supported by the Daloon Foundation. Present address: Dept. of Pathology, University of Arizona Health Sciences Center, College of Medicine, 1501 N. Campbell Ave., Tucson, AZ.

‡‡ Present address: Dept. of Dermatology, Marselisborg Hospital, DK-8000 Aarhus C, Denmark.

[1] The abbreviations used are: hnRNP, heterogeneous nuclear ribonucleoprotein; IEF, isoelectric focusing; IL, interleukin; NEPHGE, nonequilibrium pH gradient electrophoresis; RRM, RNA recognition motif; qRRM, quasi-RRM; SSC, sodium chloride/sodium citrate; SSP, sample spot number.

As a result of a systematic analysis of human keratinocyte protein profiles using computer-aided two-dimensional gel electrophoresis we have revealed many proteins that are differentially regulated in SV40-transformed cells (Celis and Olsen, 1994). Among the highly up-regulated proteins, some are present in purified 40 S hnRNP particles, *e.g.* hnRNPs A, B, C, H, and K (Celis *et al.*, 1986; Celis and Olsen, 1994; Dejgaard *et al.*, 1994), indicating that changes in the levels of these proteins are necessary to maintain the protein expression phenotype exhibited by transformed cells. Recently, Matunis *et al.* (1994) cloned hnRNP F and showed that it is immunologically related as well as sequence related to hnRNP H, implying that these proteins are very similar. Their results, however, could not reveal whether these proteins were truly different or represented variants produced by alternative splicing of the same pre-mRNA. Using cDNA cloning, two-dimensional gel immunoblotting, and amino acid microsequencing we here show that hnRNP H has a unique primary structure and that it belongs to a subfamily of hnRNP proteins that includes hnRNP F, H′ (identical to the FTP3 gene product (Vorechovsky *et al.*, 1994)), and most likely two novel proteins with apparent molecular masses of 37.6 and 39.5 kDa, respectively. Fluorescence *in situ* hybridization revealed that the genes coding for hnRNPs H, H′, and F were localized on different chromosomes. In addition, we present ribohomopolymer binding experiments on hnRNP H as well as biochemical studies showing that hnRNPs H and F possess distinct regulatory properties.

MATERIALS AND METHODS

Cultured Cells—Primary normal human keratinocytes were obtained by plating unfractionated keratinocytes in 3-cm diameter Petri dishes coated with human dermal extract prepared as described (Madsen *et al.*, 1992) with 3 ml of serum-free keratinocyte medium supplemented with epidermal growth factor (5 ng/ml), bovine pituitary extract (50 μg/ml), and antibiotics (penicillin at 100 units/ml and streptomycin at 50 μg/ml). SV40-transformed (K14) keratinocytes (Taylor-Papadimitriou *et al.*, 1982), normal human MRC-5 fibroblasts, and SV40-transformed MRC-5 fibroblasts (MRC-5 V2) (Huschtscha and Holliday, 1983) were grown as monolayer cultures, and Molt-4 cells in suspension in Dulbecco's modified Eagle's medium that was supplemented with 10% (v/v) fetal calf serum and antibiotics (penicillin at 100 units/ml and streptomycin at 50 μg/ml).

Labeling of Cells with [^{35}S]Methionine—Primary keratinocyte cultures were labeled for 14 h in serum-free keratinocyte medium lacking methionine (prepared by Life Technologies, Inc.) and containing 100 μCi of [^{35}S]methionine (SJ204, Amersham Corp.) per 0.1 ml of medium. Primary keratinocytes were treated with dibutyryl cyclic AMP, dibutyryl cyclic GMP, interferon-α, interferon-β, transforming growth factor-α, transforming growth factor-β, IL-1α, IL-1β, IL-3, IL-7, IL-8, or tumor necrosis factor-α 5 days after plating as described previously (Honoré *et al.,* 1993a). K14 keratinocytes, MRC-5, and MRC-5 V2 fibroblasts grown in microtiter wells (NUNC) were labeled for 14 h with 0.1 ml of laboratory-made Dulbecco's modified Eagle's medium (1 g/liter, $NaHCO_3$) lacking methionine and containing 10% dialyzed fetal calf serum and 100 μCi of [^{35}S]methionine (SJ204, Amersham) (Bravo *et al.*, 1982).

Two-dimensional Polyacrylamide Gel Electrophoresis and Silver Staining—The procedure for running two-dimensional gels and silver staining have been previously described in detail (Bravo *et al.*, 1982; Tunón and Johansson, 1984; Celis *et al.*, 1994b).

Microsequencing—Nuclear extracts of Molt-4 cells were separated by two-dimensional gel electrophoresis, directly or after purification on hydroxyapatite, and stained with Coomassie Brilliant Blue (Celis *et al.*, 1990). Several protein spots including IEF SSPs 4410, 4429, 5416, 6304, 7312 (previously numbered 6320), 2222, 2326, and 3415 (Celis *et al.*, 1994a) were cut from a number of gels and subjected to partial amino acid sequencing as described previously (Vandekerckhove and Rasmussen, 1994).

cDNA Libraries—The cDNA library made from MRC-5 V2 fibroblasts has previously been described in detail (Honoré *et al.*, 1992). A cDNA library from psoriatic keratinocytes was constructed by priming mRNA with $(dT)_{15}$ containing an *Xho*I and a *Sac*I site in the 5′ end. The size-fractionated cDNA (>2 kilobases) was ligated into λ Uni-ZAP™ XR cut with *Xho*I and *Eco*RI. The amplified libraries were plated on 25 × 25-cm plates containing 1.5×10^5 plaque-forming units, and overlaid with nylon filters (Hybond-N, Amersham). The DNA was covalently linked to the filters by UV light.

Screening of Libraries—A mixture of 23-mer deoxyribonucleotides was synthesized (number 2625): 5′-d(AARYTIATGGCIATGCARMGICC)-3′, where R is a purine, Y is a pyrimidine, I is deoxyinosine, and M is A or C. The oligonucleotides were labeled in the 5′-end with [γ-^{32}P]ATP and T4 polynucleotide kinase and used to screen 6×10^5 plaque-forming units from the 800-2500-base pair MRC-5 V2 library (Honoré *et al.,* 1992). Filters were hybridized at 48 °C with the tetramethylammonium chloride technique (Jacobs *et al.*, 1988; Honoré *et al.*, 1993b). The filters were sealed in plastic bags, and autoradiography was done with Kodak x-ray films and intensifying screens at −70 °C. The screening gave about six positive clones. The clones were purified by repeated platings and hybridizations. The cloned cDNAs were rescued in pBluescript according to the manual from Stratagene. Minipreps of plasmid DNA were made as described (Sambrook *et al.*, 1989). Clone 2625.5 was recloned into the *Not*I and *Sal*I sites in M13BM20 or -21 and used for large scale preparation of DNA (Sambrook *et al.*, 1989) for sequencing as described below. We then end-sequenced the other clones. Except for a few artifact clones all were found to contain parts of the same cDNA insert. Clone 2625.5 was labeled by using random primers and used to screen the psoriatic keratinocyte library as follows. The filters were first hybridized at low stringency conditions, 55 °C in 2 × SSC (1 × SSC is 0.15 M NaCl, 0.015 M Na_3citrate) overnight, then washed two times for 15 min at 55 °C in 1 × SSC and autoradiographed. A second wash was then performed at high stringency conditions, 65 °C first in 1 × SSC for 15 min and then for 15 min in 0.1 × SSC followed by autoradiography. Thirteen clones that were strongly hybridizing at low stringency and weakly at high stringency were purified, rescued in pBluescript, and sequenced in their 5′ and 3′ ends.

DNA Sequencing—DNA sequencing was performed as described previously (Honoré *et al.*, 1992, 1993a). The reported sequence of clone 2625.5 was determined from both strands. Control sequencing of clones related to 2625.5 were performed on one strand on the double-stranded plasmids with oligonucleotides for every 150 bases in the coding regions.

Computer Search for Identity and Alignments—Sequence analysis was performed with the UWGCG program package (Devereux *et al.*, 1984). A sequence identity search was carried out with the Mailfasta program in the publicly available DNA and protein data bases. Alignments of amino acid sequences were done with the Clustal V program (Higgins *et al.*, 1992) and printed with the Alscript program (Barton, 1993).

Expression of Polymerase Chain Reaction-amplified cDNA Fragments from the 2625.5-cDNA in Escherichia coli—DNA segments encoding selected amino acid parts of the protein encoded by the 2625.5-cDNA were amplified with *AmpliTaq* (Perkin Elmer) through 40 cycles (denaturation at 93 °C, annealing at 55 °C, and polymerization at 72 °C). The following peptide fragments were made: Gly^{10}-Glu^{92} (qRRM1), Gly^{111}-Glu^{190} (qRRM2), Gly^{288}-Gly^{366} (qRRM3), and the whole insert Met^{1}-Ala^{449} (TOT). The primers used in the sense direction contained a *Bam*HI site (5′-CAC*GGATCC*) followed by 18 bases encoding the first six amino acids in the fragment. The primers used in the antisense direction contained a *Hin*dIII site followed by an A (5′-TTC*AAGCTT*A) or an *Eco*RV site followed by TTA (5′-TTC-*GATATC*TTA) and then 18 bases that were complementary to the strand encoding the last six amino acids in the fragment. A stop codon was thus inserted after the last amino acid in each fragment. The polymerase chain reaction-amplified DNA-fragments were cut with the restriction enzymes *Bam*HI and *Hin*dIII or *Bam*HI and *Eco*RV and ligated into the expression vector pT7-PL (Christensen *et al.*, 1991), a derivative of the T7 promoter containing plasmid pRK172. The fusion proteins are synthesized with MGSHHHHHHGS in the NH_2 terminal. The recombinant plasmids were grown up in *E. coli* XL-1 Blue, and the three qRRM fragments were sequenced in order to check the cDNA inserts for errors introduced during the amplification process. *E. coli* XL-1 Blue or BL21(DE3) cells transformed with error-free plasmids were grown at 37 °C to an absorbance, A_{600}, of 0.6–0.7 when the recombinant proteins were expressed by adding CE6 λ-phages to XL-1 Blue cultures and isopropyl-1-thio-β-D-galactopyranoside to BL21(DE3) cultures (Studier and Moffatt, 1986). The cells were grown for an additional 4 h and then centrifuged, resuspended in buffer A (500 mM NaCl, 1 mM EDTA, 50 mM Tris-HCl, pH 8.0), and extracted with 1 volume of phenol. The cells were sonicated and the proteins precipitated by adding 2.5 volumes of ethanol. The precipitated proteins were resuspended in buffer B (6 M guanidine-HCl, 100 mM dithioerythritol, 50

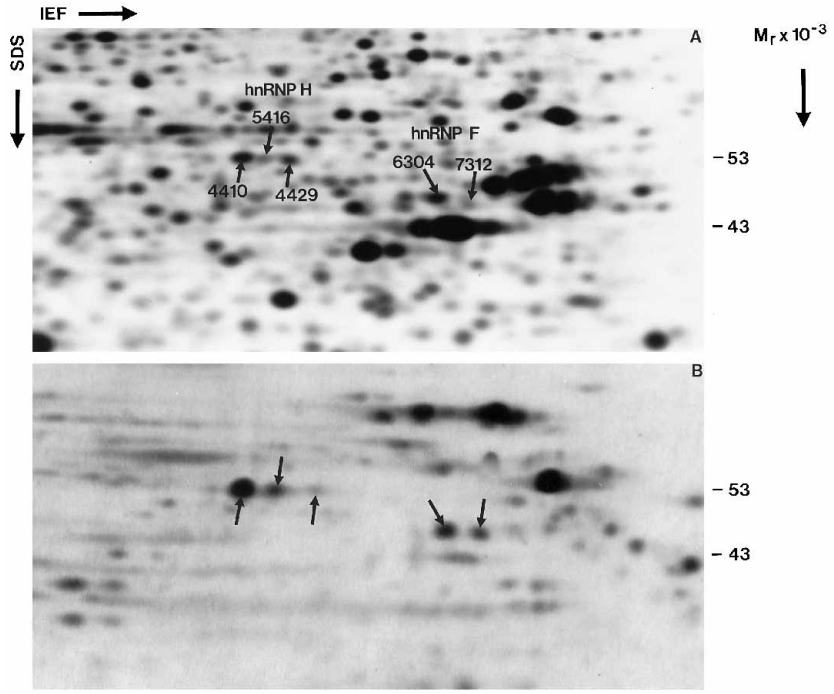

FIG. 1. **Two-dimensional IEF gels of proteins from keratinocytes (*A*) and Molt-4 cells (*B*).** *A*, autoradiogram of [^{35}S]methionine-labeled proteins from normal primary human keratinocytes; *B*, Coomassie Brilliant Blue-stained Molt-4 proteins from a fraction enriched in nuclear proteins.

mM Tris-HCl, pH 8.0), added to a Sephadex G25 column and eluted with buffer C (8 M urea, 500 mM NaCl, 1 mM methionine, 5 mM glutathion, 50 mM Tris-HCl, pH 8.0). The collected protein fractions were pooled and added to a Ni^{2+}-nitrilotriacetic acid column (Hochuli *et al.*, 1987), washed first with buffer D (6 M guanidine-HCl, 5 mM glutathion, 50 mM Tris-HCl, pH 8.0), buffer C, and finally with buffer E (500 mM NaCl, 50 mM Tris-HCl, pH 8.0). The qRRM fragments were finally eluted with 0.2 M sodium acetate, 10 mM EDTA, pH 4. The TOT fragment was eluted with 8 M urea, 1 M NaCl, 10 mM mercaptoethanol, 10 mM Tris-HCl, pH 8. The obtained qRRM hexa-His proteins were used as such for the Northwestern dot blotting as described below. The TOT fragment on the other hand contained substantial amounts of impurities, which possibly were degradation products, and was therefore further purified by polyacrylamide gel electrophoresis, cut out from the gel, electroeluted, and finally used for injection into a mouse for the production of a polyclonal antibody.

Expression of the 2625.5-cDNA in COS-1 Cells—Clone 2625.5 was digested with *Not*I, made blunt-ended, and cut out with *Pst*I. The cDNA fragment was cloned into the plasmid pMT21 (Kaufman *et al.*, 1991), cut with *Eco*RI, made blunt-ended and finally cut with *Pst*I. The plasmid DNA was transfected into COS-1 monkey cells (Gluzman, 1981) with LipofectinAMINE (Life Technologies, Inc.) as transfection agent (Düzgünes and Felgner, 1993). COS-1 cells were resuspended in lysis solution (O'Farrell, 1975) and analyzed by two-dimensional gel electrophoresis.

Northwestern Dot Blotting—The ribohomopolymers poly(rA), poly(rC), poly(rG), and poly(rU) from Pharmacia Biotech Inc. were 5′-end labeled with ^{32}P. The peptides qRRM1, qRRM2, and qRRM3 were dot-blotted onto nitrocellulose in similar amounts as determined by visual inspection of the Coomassie Brilliant Blue-stained peptides. Hybridization and washing procedures were as described previously (Dejgaard and Celis, 1994).

Chromosome Mapping—Chromosome mapping was performed by fluorescence *in situ* hybridization as described previously (Tommerup and Vissing, 1995). Individual cDNA clones inserted in vectors, corresponding to genes *HNRPH1* (2625.5-cDNA), *HNRPH2* (4410LH.31-cDNA), and *HNRPF* (4410LH.30-cDNA) were labeled with biotin by nick-translation in the presence of bio-11-dUTP (Sigma) and hybridized to human metaphase chromosomes. For each cDNA, 50 metaphases were examined in a Zeiss Periplan epifluorescence microscope for distribution of signals. Selected metaphases were photographed on Kodak Ektachrome EPY64T, and the signal position (the relative distances from the short arm telomere to the signal (FLpter) was measured on projected slides and compared with the 4′,6-diamidino-2-phenylindole pattern and with the digitized chromosome ideogram (Francke, 1994).

RESULTS AND DISCUSSION

Isolation and Sequencing of a cDNA Encoding hnRNP H—Analysis of proteins synthesized by normal and SV40-transformed K14 keratinocytes (Taylor-Papadimitriou *et al.*, 1982) has revealed several polypeptides that are differentially regulated in K14 cells (Celis and Olsen, 1994) and that are enriched in nuclear pellets (Fig. 1) and in purified 40 S hnRNP particles (Celis *et al.*, 1986). One of these polypeptides (IEF SSP 4410 in the keratinocyte protein data base) (Celis *et al.,* 1994a) (Fig. 1*A*), which comigrated with purified hnRNP H, was selected for further studies, as this protein is 1.6-fold up-regulated in K14 cells (Celis and Olsen, 1994), and several peptide sequences were available in the keratinocyte data base (kLMAMQrPGPYDr, PGAGrGYNSIGrGAGFEr, YIEIF, and ThYDP, where residues given in lowercase could not be unambiguously assigned) (Rasmussen *et al.*, 1992). Peptide kLMAMQrP, which contained a reasonably low degree of degeneracy at codons with ambiguity, was back-translated to nucleotide sequence, and a mixture of 23-mer oligodeoxyribonucleotides (oligo number 2625) was used to screen the MRC-5 V2 cDNA library as described under "Materials and Methods." Fig. 2 shows the cDNA sequence obtained from clone 2625.5 (hnRNP H), which contains 2201 base pairs with a 19-base pair-long poly(A) tail. A multiple tissue Northern blot hybridized with this cDNA revealed a transcript length of at least 2.4 kilobases expressed at high levels in placenta and skeletal muscle, but also in brain, lung, liver, kidney, and pancreas. A number of start codons are found in the 5′-end of the cDNA, but the longest open reading frame is obtained from the start codon at position 73–75, which is preceded by an in-frame stop codon at positions 64–66. Another start codon is found at positions 76–78, immediately downstream from the first one. The observed patterns of flanking nucleotides around these two codons, **A***XX*ATG**A** and **A***XX*ATG**T** occur at a frequency of 17 and 8%, respectively, in vertebrates (Cavener and Ray, 1991). Since only 9% of vertebrate mRNAs have start codons in the 5′ noncoding region (Kozak, 1987) it is most likely that the start codon at positions 73–75 is the functional one. The stop codon, TAG, is found at positions 1420–1422, and a putative polyadenylation site,

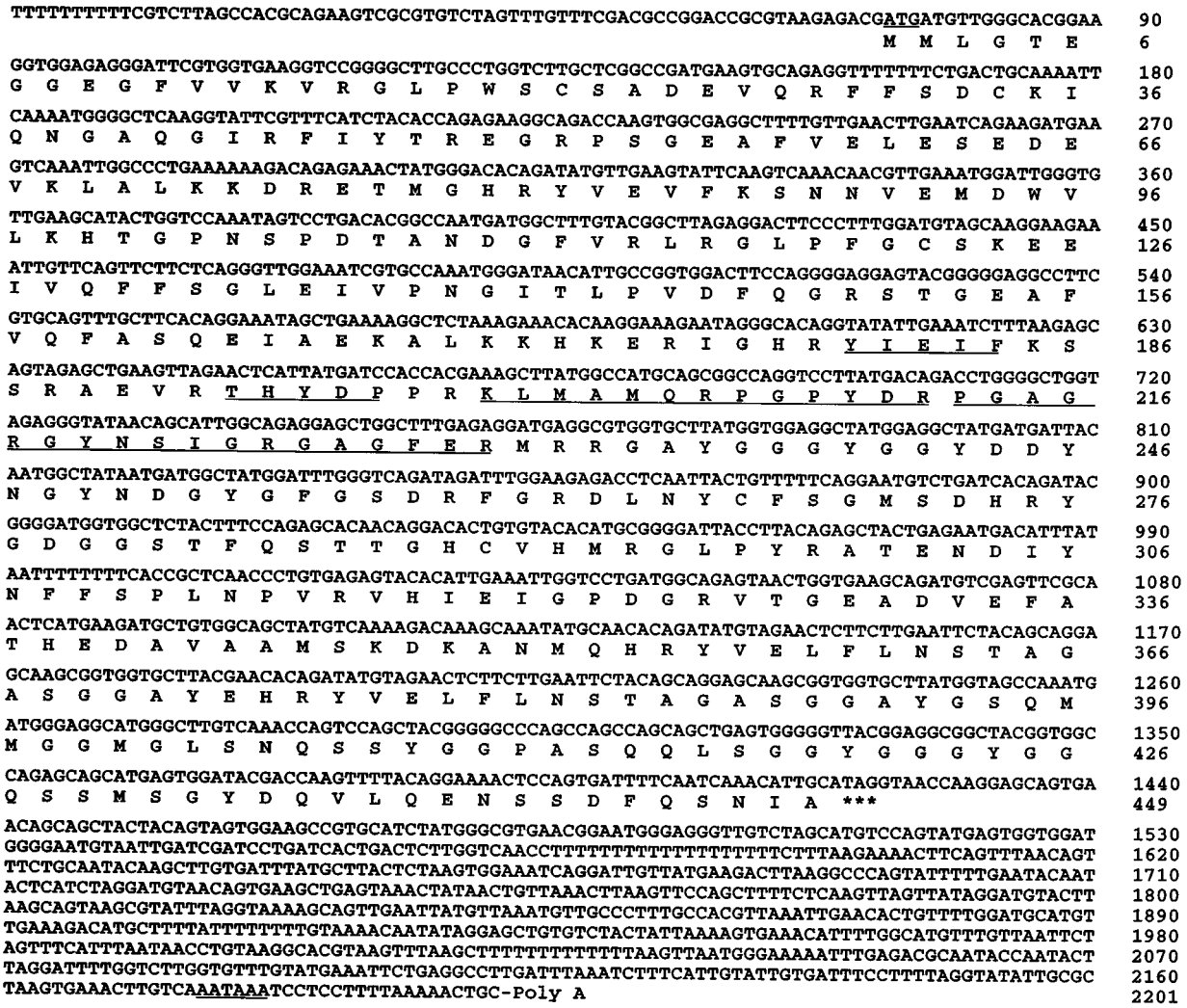

FIG. 2. **Nucleotide sequence of the 2625.5-cDNA (hnRNP H-cDNA).** The cDNA is 2201 bases long with a poly(A) tail of 19 bases. The putative polyadenylation signal, AATAAA (*underlined*), is located 19 base pairs upstream from the poly(A) tail. A 449-amino acid protein is encoded between the start codon at positions 73–75 and the stop codon at positions 1420–1422 (indicated with *asterisks*). The molecular mass of the predicted protein is 49.2 kDa, and the calculated pI is 6.26, values that are in close agreement with the experimentally observed values of keratinocyte protein IEF SSP 4410 (hnRNP H), 52.8 kDa and pI 5.9 (Celis *et al.*, 1994a). The four partial peptide sequences obtained by microsequencing are *underlined*.

AATAAA, is located 19 base pairs upstream from the poly(A) tail. The encoded protein contains 449 amino acids with a deduced molecular mass of 49.2 kDa and a pI of 6.26. These values match closely those observed experimentally for IEF SSP 4410 (Celis *et al.*, 1994a), *i.e.* 52.8 kDa and a pI of 5.9. Furthermore, the sequences of the four microsequenced peptides matched perfectly within the amino acid sequence as deduced from the cDNA (Fig. 2). As an additional control the hnRNP H-cDNA was overexpressed in COS-1 cells using the pMT21 expression vector. Subsequent two-dimensional gel electrophoresis analysis of the [^{35}S]methionine-labeled proteins indicated that the protein encoded by the hnRNP H-cDNA comigrated with endogenously synthesized hnRNP H (results not shown). Microsequencing results obtained on two other proteins migrating toward the acidic side of hnRNP H, *i.e.* IEF SSPs 5416 and 4429 (Fig. 1*A*), revealed similar sequences as those encoded by the hnRNP H-cDNA, suggesting that they correspond to modified variants of hnRNP H. The first and possibly the second variant are phosphorylated (results not shown).

Identification of Proteins Sharing Epitopes with hnRNP H—A mouse polyclonal antibody raised against the recombinant hnRNP H produced in *E. coli* was used to probe two-dimensional gel immunoblots of Molt-4 extracts enriched in nuclear proteins. As shown in Fig. 3, the antibody reacted with several proteins, which are indicated with their IEF numbers in the keratinocyte data base (Celis *et al.*, 1994a). A strong reaction was observed with hnRNP H (IEF SSP 4410) and its variants (IEF SSPs 5416 and 4429) as well as with IEF SSPs 6304, 7312, 2222, 2326, 3415, 4432, and a putative cleavage product of hnRNP H (see Fig. 3). However, since no preimmune control was available for the antibody shown in Fig. 3 we microsequenced the reacting proteins that were present in sufficient quantities, *i.e.* IEF SSPs 6304, 7312, 2222, 2326, and 3415 with the results given in Table I. The tryptic peptide maps of IEF SSPs 6304 and 7312 were identical, and the peptide sequences could be aligned to hnRNP H when very few amino acid exchanges were allowed. The five peptides obtained from 6304 showed between 80 and 100% identity to hnRNP H, and the single peptide obtained from 7312 showed 83% identity, thus confirming that both proteins are sequence-related to hnRNP H. Data base searching with the peptides identified the 6304 and 7312 proteins as hnRNP F (Fig. 3) (Matunis *et al.*, 1994).

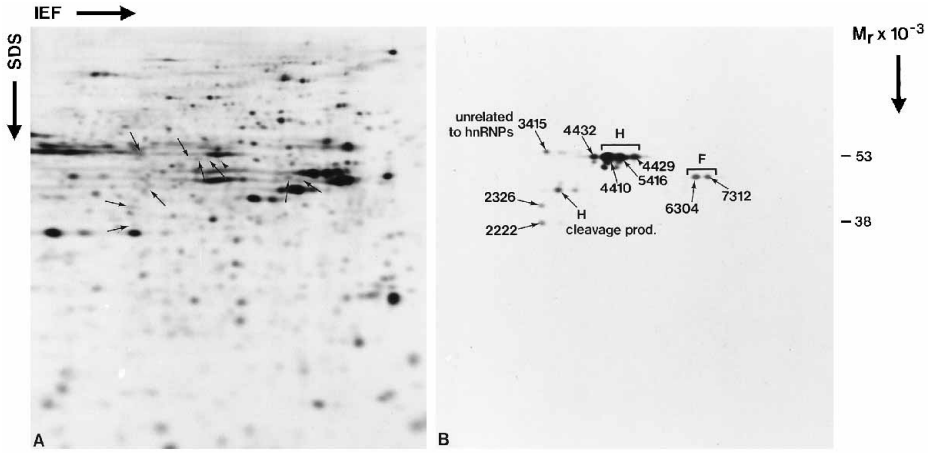

FIG. 3. **Two-dimensional IEF autoradiogram (*A*) and immunoblot (*B*) of Molt-4 nuclear pellet proteins.** The blot was reacted with a polyclonal mouse antibody raised against the recombinant protein encoded from the hnRNP H-cDNA. The proteins sharing epitopes were assigned with putative IEF SSP numbers from the data base of keratinocyte proteins (Celis *et al.*, 1994a), except for one protein (*H cleavage prod.*) that was found to be a putative cleavage product of hnRNP H. *H* indicates protein hnRNP H and its variants, while *F* indicates protein hnRNP F and a variant.

TABLE I

Microsequences of proteins sharing epitopes with hnRNP H (IEF SSP 4410)

Residues given in lowercase could not be unambiguously assigned in the microsequenced peptides. Residues that are underlined were identical with residues in hnRNP H. Numbers indicate the first and last amino acid in hnRNP H.

Protein IEF SSP number[a]	Molecular mass[b]	pI[b]	Microsequences of tryptic peptides[c]	Corresponding residues in hnRNP H	Percent identity of the microsequence with protein hnRNP H[d]
	kDa				
6304	48.1	5.2	OSGEAFVELGSE	^{53}PSGEAFVELESE^{64}	83
			YIEVFK	^{82}YVEVFK^{87}	
			or	or	83
			YIEVFK	^{180}YIEIFK^{185}	
			TEMDWVLK	^{91}VEMDWVLK^{98}	88
			GLPYK	^{295}GLPYR^{299}	80
			ANMQHR	^{350}ANMQHR^{355}	100
7312	48.3	5.1	YIEVFK	^{82}YVEVFK^{87}	
			or	or	83
			YIEVFK	^{180}YIEIFK^{185}	
2222	37.6	6.5	yIEIFR	^{180}YIEIFK^{185}	83
			sEIKGFYDPpr	^{189}AEVRTHYDPPR^{199}	55
			XXENDIANFFSPLnPV	^{300}ATENDIYNFFSPLNPV^{315}	93
			dGMDNQGGYGs	^{400}MGLSNQSSYGG^{410}	45
2326	39.5	6.6	PGPYDR	^{207}PGPYDR^{212}	100
			ATENDIANFF	^{300}ATENDIYNFF^{309}	90
3415	52.9	6.4	AVLLAGp		
			LDPSIFE	No significant identity	
			GTEDITSP		

[a] IEF SSP 7312 was previously numbered IEF SSP 6320 (Celis et al., 1994a).

[b] Molecular masses and pIs are those listed in the database of keratinocyte proteins (Celis *et al.*, 1994a).

[c] A few peptide sequences published previously (Rasmussen *et al.*, 1992) have not been included here, as they were found to originate from peaks containing more than one peptide.

[d] The tryptic peptide maps of IEF SSPs 6304 and 7312 were identical, and those of IEF SSPs 2222 and 2326 were identical. The tryptic peptide map of IEF SSP 3415 was distinct from the rest.

The tryptic peptide maps of IEF SSPs 2222 and 2326 were identical and different from the maps of 6304, 7312, and hnRNP H (4410). Sequencing results from 2222 gave identity scores between 45 and 93% to hnRNP H, while peptides from 2326 showed identities between 90 and 100% to hnRNP H. Data base searching indicated that these proteins were novel, although some similarity was found to the recently cloned poly(A)$^+$ mRNA binding protein GRSF-1 (Qian and Wilusz, 1994), which shares 46% identity to hnRNP H and 44% to hnRNP F.

One protein that shared epitopes with hnRNP H had a lower molecular mass and higher pI but revealed an almost identical tryptic peptide map to that of hnRNP H. Furthermore, microsequencing of the protein did not reveal any differences in the amino acid sequences from those predicted from the hnRNP H-cDNA. Some immunoreaction of the hnRNP antibody with this polypeptide could only be detected in nuclear proteins from Molt-4 cells and not in total cell extracts from *e.g.* AMA cells or K14 keratinocytes, so we concluded that this protein most likely corresponds to a cleavage product of hnRNP H (shown as *H cleavage prod.* in Fig. 3).

The tryptic map of polypeptide IEF 3415 was very different from that of hnRNP H, 6304, 7312, 2222, and 2326, and the peptide sequences showed no significant identity (between 29 and 43%) to hnRNP H. Thus the presence of IEF 3415 in the immunoblot may be due to a spurious cross-reactivity. The possibility, however, cannot be excluded that the microsequences are derived from a major unrelated protein and that the observed cross-reactivity is due to a minor comigrating related protein.

In conclusion, it appears that at least hnRNP H and variants (4410, 5416, and 4429) and hnRNP F and its variant (6304 and 7312) are closely related proteins. Based on tryptic peptide sequences IEF SSPs 2222 and 2326 are definitely related to hnRNPs H and F, while IEF 3415 apparently is unrelated.

cDNA Library Screening and Data Base Searching for cDNAs Related to the hnRNP H-cDNA—In an effort to reveal additional hnRNP H-related proteins we screened the kerati-

nocyte cDNA library with a low/high stringency hybridization technique using the hnRNP H-cDNA as probe. Among 13 clones that were found to be different from hnRNP H because they hybridized at low but not at high stringency, we found 10 containing one type of transcript and three with another type as determined by end sequencing. Data base searching revealed two transcripts that possessed very high similarity to the hnRNP H-cDNA, and these corresponded to the FTP3 transcript (82%) (Vorechovsky *et al.*, 1994) and hnRNP F mRNA (73%) (Matunis *et al.*, 1994). Sequencing on one strand of one representative clone from each of the above mentioned types of transcripts verified that clone 4410LH.31 corresponded to the FTP3 transcript, and 4410LH.30 corresponded to hnRNP F mRNA.[2] Translation of the FTP3 and hnRNP F transcripts yielded proteins with deduced molecular masses of 49.3 kDa (pI 6.26) and 45.7 kDa (pI 5.39), respectively. The coordinates of the proteins encoded by the hnRNP H-cDNA (49.2 kDa and pI 6.26) and the 4410LH.31-cDNA (FTP3 transcript) (49.3 kDa and pI 6.26) were almost identical, indicating that they may comigrate in two-dimensional gels. Alternatively, the 4410LH.31 cDNA could encode protein IEF SSP 4432 (Fig. 3), which has a slightly higher molecular mass (54.2 kDa) and a more basic pI (6.0) than hnRNP H. For convenience, we will refer to these two proteins as hnRNP H (2625.5-cDNA) and hnRNP H′ (4410LH.31-cDNA/FTP3 transcript) (see Table II). Subcloning of the cDNA encoding hnRNP H′ into pMT21 for COS-1 cell expression has repeatedly failed; therefore, we have been unable to verify the position of the encoded protein in two-dimensional gels. The deduced coordinates of hnRNP F from the 4410LH.30 cDNA (45.7 kDa and pI 5.39) indicates that this cDNA encodes protein IEF SSP 6304 and its phosphorylated variant IEF SSP 7312 (Table II).

Primary Structure of hnRNPs H, H′, and F—The alignment of the sequences of hnRNPs H, H′, and F is shown in Fig. 4*A*. The identity is 96% between hnRNPs H and H′, 78% between H and F, and 75% between H′ and F. Each of the three proteins

[2] Sequencing on one strand indicated that clone 4410LH.31 was identical with the FTP3 transcript (Vorechovsky *et al.*, 1994) from base 79 to 1428 (accession number U01923), except that we found a C instead of a T at position 687, which, however, did not change the deduced amino acid at that position. Clone 4410LH.30 was identical with the cDNA for the hnRNP F mRNA (Matunis *et al.*, 1994) from base 29 to 1276 (accession number L28010).

TABLE II
Molecular mass and pI of the hnRNPs H, H′, and F subfamily members

Numbers in parentheses give the values as deduced from their cDNAs. A question mark indicates that hnRNPH′ is only putatively the same as protein IEF SSP 4432.

Protein	IEF SSP No.	Molecular mass	pI
		kDa	
hnRNP H	4410/5416/4429	52.8/52.9/53.4 (49.2)	5.9/5.8/5.7 (6.26)
hnRNP H′	4432?	54.2? (49.3)	6.0? (6.26)
hnRNP F	6304/7312	48.1/48.3 (45.7)	5.2/5.1 (5.39)

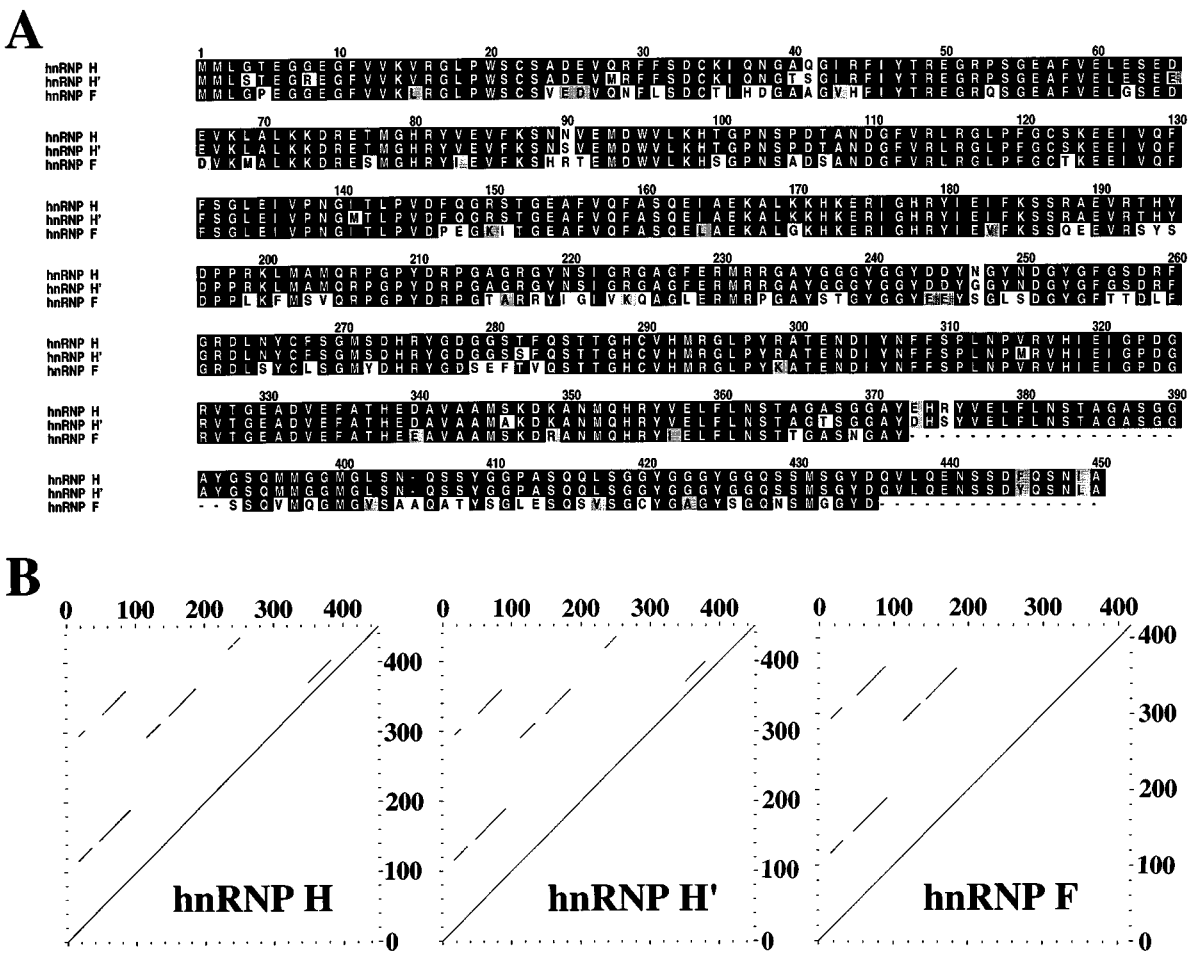

FIG. 4. **Alignment of hnRNP H, hnRNP H′, and hnRNP F (*A*) and *dot plots* showing stretches of internal duplications (*B*).** Identical amino acids are shown in *reverse print*, and conserved amino acids are *shaded* according to the following formula (A = G, I = L = V, K = R, D = E, F = Y = W, and N = Q). The *dot plots* were made with the UWGCG program COMPARE, with a window of 30 and a stringency of 18.0. Two internal regions that show a high degree of similarity will appear as a straight line, where the *x* coordinates to the ends indicate the first and last amino acid position of one region and the *y* coordinates indicates the first and last amino acid position of the other region.

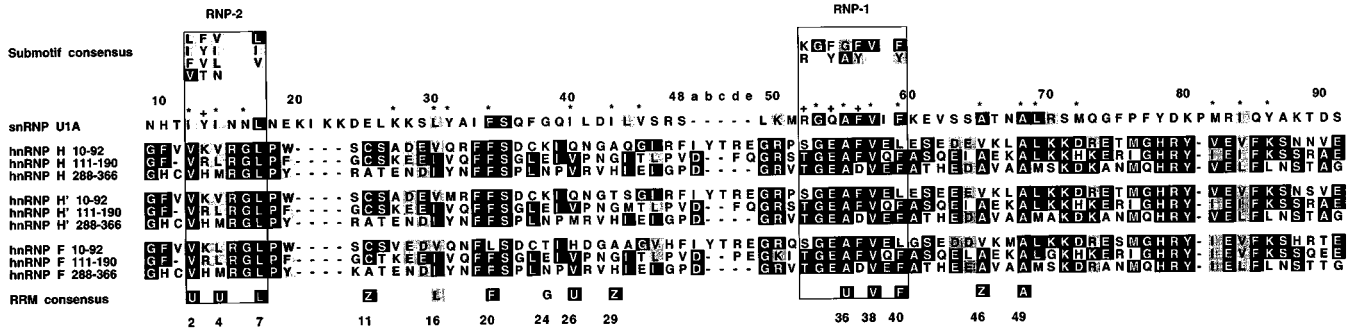

FIG. 5. **Alignment of 80 amino acid repeats in hnRNP H, hnRNP H′, and hnRNP F with the RRM in snRNP U1A and with an RRM consensus sequence.** The RNP-1 and RNP-2 submotif consensus sequences are from Dreyfuss *et al.* (1988). From similarity studies Query *et al.* (1989) have defined the RRM repeat as corresponding to residues 11–91 in snRNP U1A. Conserved hydrophobic residues that contribute to the hydrophobic core are indicated with *asterisks*, and conserved solvent-exposed positions in the RNP-1 and RNP-2 elements are indicated with *plus signs* (Kenan *et al.*, 1991). Residues that are identical to the most frequent residue in hnRNP H, *i.e.* found in at least two of three repeats, are shown in *reverse print*, while conserved residues are *shaded* according to the formula given in the legend to Fig. 4. In the RRM consensus sequence (Birney *et al.*, 1993), *U* represents uncharged residues: L, I, V, A, G, F, W, Y, C, M, and Z = U + S, T. The residues in the consensus sequence are *highlighted* when two or three residues in each protein either are identical (*reverse print*) or conserved (*shaded*). The *numbering* is according to Birney *et al.* (1993).

has a high content of Gly residues (12–15%). Internal sequence comparisons of hnRNPs H, H′, and F revealed a repeated structure of each protein (Fig. 4*B*). hnRNPs H and H′ show three repeats of about 80 residues localized at the approximate amino acid positions 10–92, 111–190, and 288–366 containing the conserved motifs; V*XX*RGLP, FFS, GR*XX*GEA*X*V, and HRY(V/I)E(V/I/L)F. hnRNP F shows similar repeats, except that one repeat contains FLS instead of FFS while another contains GK*XX* . . . instead of GR*XX*. . . . We denote these repeats qRRMs (Fig. 6, *dark gray boxes*, and see below).

Besides the three qRRMs, hnRNP H possesses a 19-residue repeat (*hatched boxes* in Fig. 6); $^{354/374}$HRYVELFLNSTAGASGGAY$^{372/392}$, where the first repeat (residues 354–372) shares amino acids with the COOH terminus of qRRM3. This repeat is also found in hnRNP H′, although it has two amino acid exchanges . . . GTSGG . . . instead of . . . GASGG . . . in the first repeat (residues 354–372) and . . . HSYVE . . . instead of . . . HRYVE . . . in the second (residues 374–392). In hnRNP F a very similar sequence is found, ^{354}HRYIELFLNSTTGASNGAY^{372}, but it is not repeated.

Finally, hnRNPs H and H′ have two 16-amino acid repeats containing many Gly residues (*open boxes* in Fig. 6): $^{234/418}$G(G/A)YGGGYGG*XXXXX*GY$^{249/433}$ with one localized between qRRM2 and qRRM3 in a region especially rich in Gly, Tyr, and Arg residues (Fig. 6, *light gray box*) but with virtually no Lys residues, while the second is localized close to the COOH terminus in a region rich in Gly and Tyr residues (Fig. 6, *light gray box*). hnRNP F has also Gly and Tyr residues more abundantly present in similar regions (Fig. 6, *light gray box*), although less pronounced than in the case of hnRNPs H and H′. hnRNP F does not contain the 16-amino acid Gly-rich repeat.

Do hnRNPs H, H′, and F Contain the RRM?—A motif analysis using the GCG program package did not identify any structure in these proteins. The most common RNA-binding motif in hnRNP proteins is the 80-residue RRM (Kenan *et al.*, 1991). Since hnRNPs H, H′, and F all contain repeats of about 80 amino acid residues and hnRNPs H and F previously have been shown to bind poly(rG) (Swanson and Dreyfuss, 1988), we searched for the presence of RRMs in these proteins. The three regions in hnRNP H (residues 10–92, 111–190, and 288–366) were noted to have a strong similarity to a variety of RRM-containing proteins such as nucleolin, eukaryotic initiation factor 4B, and hnRNP A1. We then manually aligned the three regions in hnRNPs H, H′, and F to the NH_2-terminal RRM (residues 11–91) of snRNP U1A (Sillekens *et al.*, 1987; Query *et al.*, 1989) and to the recently described RRM consensus sequence (Birney *et al.*, 1993) (Fig. 5). This alignment did not conclusively assign these regions as valid RRMs, as we found that four to five residues are conserved among seven in the RNP-1 submotif; (K/R)G(F/Y)(G/A)(F/Y)V*X*(F/Y) (Dreyfuss *et al.*, 1988) (compared with six conserved residues in snRNP U1A) and that two to three residues are conserved among four in the RNP-2 submotif; (L/I/F/V)(F/Y/V/T)(V/I/L/N)*XX*(L/I/V) (Dreyfuss *et al.*, 1988) (compared with four conserved residues in snRNP U1A). When the repeats were compared with the recently described consensus sequence (Birney *et al.*, 1993) we found 9–12 conserved residues among 14 (compared with 13 conserved residues in snRNP U1A), and it was especially noted that the highly conserved Gly-24 is not present in any of the repeats. A further profile analysis (Birney *et al.*, 1993) also did not conclusively assign these regions as valid RRMs; however, a profile constructed from an alignment of the three putative RRMs was able to retrieve a variety of RRM-containing proteins,[3] although alignments were not ideal. The hnRNP H sequence when compared with a RRM profile far outscored randomized sequences that maintained its composition, indicating that composition bias was not responsible for the observed similarity. One explanation for these observations could be that the regions are highly atypical RRMs: The current consensus (Birney *et al.*, 1993) is known not to be able to discriminate ideally, as illustrated by a variety of proteins that have a suggestive similarity but cannot be fitted to the consensus, *e.g.* U2AF35 (Birney *et al.*, 1993), although other evidence points to them as valid RNA binding proteins.[3] It may be that these proteins once had more standard RRMs, which diverged perhaps even at a structural level from their ancestral RRMs.

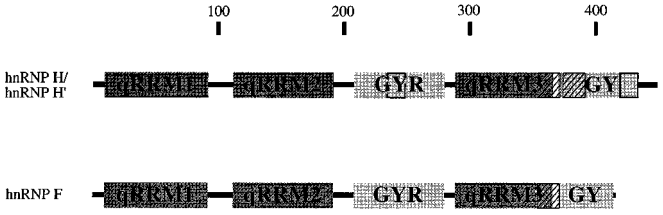

FIG. 6. **Repeats found in hnRNP H, hnRNP H′, and hnRNP F.** qRRM (quasi-RRM) indicates a domain with a remote similarity to the RRM, GYR denotes a G-, Y-, and R-rich region and GY denotes a G- and Y-rich region. *Open boxes* denote short 16-residue G-rich repeats, and hatched boxes denote 19-residue repeats.

[3] E. Birney, personal communication.

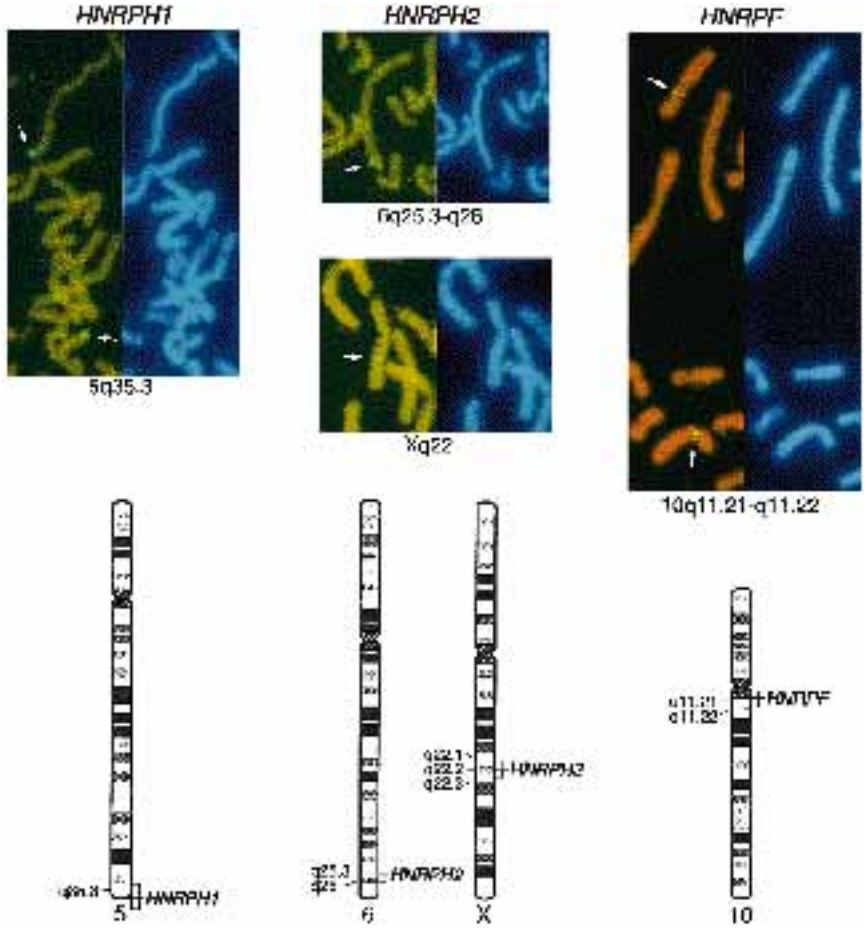

FIG. 7. **Chromosomal mapping by fluorescence *in situ* hybridization of *HNRPH1* to 5q35.3, of *HNRPH2* to 6q25.3-q26 and/or Xq22, and of *HNRPF* to 10q11.21-q11.22.** *Arrows* on the partial metaphases indicate the specific signals, with the corresponding 4′,6-diamidino-2-phenylindole-banded chromosomes displayed on the *right*. *Below* are the chromosomal idiograms (Francke, 1994) with gene localizations based on chromosomal banding and mean FLpter values. The *horizontal box* at each locus indicates the distribution of FLpter values.

Alternatively, the similarity could simply be fortuitous. Currently there is no good method of estimating the probability of a certain profile score occurring by chance; however, the sequence of hnRNP H would occur in the top 0.5% of sequences in EMBL databank when searched using an RRM profile able to place 97% of known RRM-containing sequences above an arbitrary cut-off score with no false positives.[3]

Thus, since these repeats are not conclusively identified as RRMs but in fact do bind RNA (see below), we will denote them as quasi-RRMs (qRRMs) until a more firm conclusion may be reached with the availability of a three-dimensional structure. As is the case for ordinary RRMs (Dreyfuss *et al.*, 1988) the amino acid identities between qRRMs within the same protein are lower (between 33% and 46%) than between similarly placed qRRMs in different proteins, being 72–76% for qRRM1, 86–88% for qRRM2, and 90–92% for qRRM3. Thus the lowest score was found for the NH_2-terminal domains and the highest score for the COOH-terminal domains (Fig. 6).

Chromosome Mapping of the Genes Encoding hnRNP H, H′, and F—Only one nonrandomly located signal was obtained with each of the two cDNA clones corresponding to gene *HNRPH1* (cDNA clone 2625.5 encoding protein hnRNP H) at 5q35.3 (FLpter value: 1.00 ± 0.04) and gene *HNRPF* (cDNA clone 4410LH.30 encoding protein hnRNP F) at 10q11.21-q11.22 (FLpter value: 0.36 ± 0.04) (Fig. 7). The mapping of gene *HNRPH1* to 5q35.3 is the first indication of the genomic localization of this gene. McDonald *et al.* (1992) localized the mcs94–1 clone corresponding to gene *HNRPF* to 10q11.2. The present result confirms this and refines the localization to 10q11.21-q11.22. cDNA clone 4410LH.31 (corresponding to gene *HNRPH2* encoding protein hnRNP H′) gave two specific signals, one of which was at Xq22 (FLpter value: 0.67 ± 0.02).

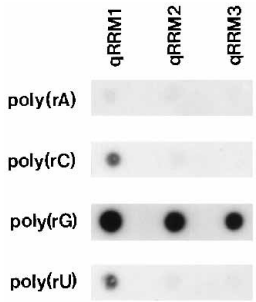

FIG. 8. **Northwestern dot blots of qRRMs in hnRNP H.** The peptide fragments were *dot-blotted* onto nitrocellulose in equal amounts, as determined by the staining intensity with Coomassie Brilliant Blue, and hybridized with labeled poly(rA), poly(rC), poly(rG), and poly(rU).

This localization is in agreement with Vorechovsky *et al.* (1994), who isolated a homologous cDNA (FTP3) by direct cDNA selection using YACs from the region Xq21.3-q22. Our fluorescence *in situ* hybridization mapping places the signal in the distal part of this segment, within Xq22 (Fig. 7). However, a specific signal was also observed at 6q25.3-q26 (FLpter value: 0.94 ± 0.02) (Fig. 7), indicating the presence of either two genes, or a pseudogene. The data presented by Vorechovsky *et al.* (1994) only imply that genomic sequences within the region Xq21.3-q22 hybridize with cDNA clones encoding hnRNP H′. If one of the two observed fluorescence *in situ* hybridization signals represents a pseudogene, it is likely that the functional *HNRPH2* gene is at 6q25.3-q26, since we observed a stronger signal at distal 6q (68 of 200 possible signals) than at Xq22 (47 of 200 possible signals). However, further genomic mapping is

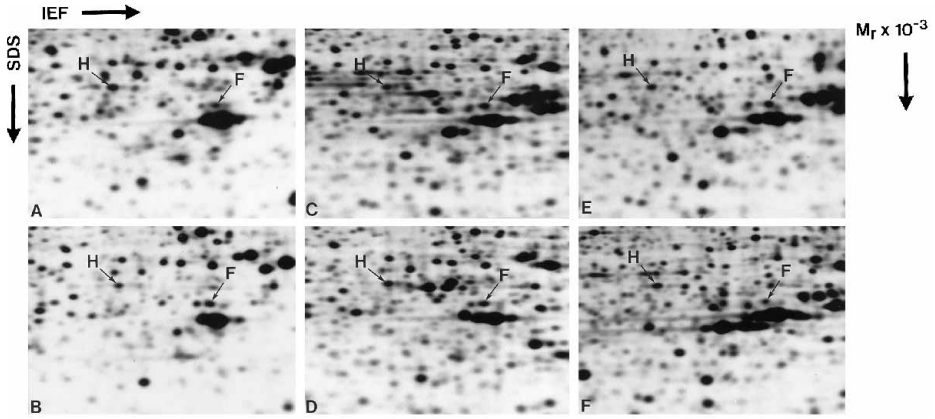

FIG. 9. **Two-dimensional IEF gels of proteins from fibroblasts (*A* and *B*) and keratinocytes (*C–F*).** Autoradiograms of normal embryonal human lung (MRC-5) fibroblasts (*A*), their SV40-transformed (MRC-5 V2) counterparts (*B*), normal human keratinocytes (*C*), their SV40-transformed (K14) counterparts (*D*), and normal human keratinocytes (*E*) as control for 4β-phorbol 12-myristate 13-acetate-treated keratinocytes (*F*) are shown. *H* indicates protein hnRNP H (IEF SSP 4410), and *F* indicates protein hnRNP F (IEF SSP 6304). hnRNPs H and F in *C* and *D* have been cut and counted by liquid scintillation relative to the total number of counts recovered from the gels (Celis and Olsen, 1994).

needed to clarify this.

RNA Binding of qRRMs in hnRNP H—To determine the polyribonucleotide binding properties of this subfamily, we selected hnRNP H and constructed peptide fragments representing each of the qRRMs by polymerase chain reaction amplification of the hnRNP H-cDNA. The peptides were dot-blotted onto nitrocellulose in approximately equal concentrations as determined by their Coomassie Brilliant Blue staining intensity. As shown in Fig. 8, each of the qRRM domains bind poly(rG) with about equal affinity. In addition, qRRM1 was able to bind poly(rC) and poly(rU) while qRRM2 and qRRM3 did not bind detectable amounts of either of these ribohomopolymers. None of the qRRMs did bind detectable amounts of poly(rA). It thus seems that the qRRMs *per se*, or at most supplemented with a few amino acids in the NH_2-terminal, are capable of binding RNA. This is interesting as most of the traditional RRMs in other hnRNPs require from 5 to 111 additional flanking residues in order to function as RNA binding domains (Kenan *et al.,* 1991). Since the identity between similarly placed qRRMs in hnRNPs H, H′, and F is higher than the similarity between qRRMs within the same protein, it may be anticipated that the binding characteristics for hnRNP H are applicable for the qRRMs in both hnRNP H′ and F. Further studies, however, are necessary to determine the affinity as well as the structure of the sequences preferred by each qRRM. The number of nucleotides involved in this binding is likely to be four to six as has been observed for RRMs in hnRNPs A1 (Shamoo *et al.*, 1994; Burd and Dreyfuss, 1994) and C (Görlach *et al.*, 1994).

Ubiquitous Tissue Expression and Differential Regulation of hnRNPs H and F—Analysis of several human fetal tissues, including the adrenal gland, brain, ear, eye, pituitary gland, liver, lung, mesonephric tissue, pancreas, smooth muscle, meninges, spleen, stomach, large intestine, thymus, tongue, and ureter revealed that hnRNPs H and F are ubiquitously expressed in these tissues (results not shown). Also, all human cell lines analyzed to date exhibit these proteins. These include HeLa, A431, AMA, Molt-4, MRC-5, and MRC-5 V2 (results not shown).

Analysis of the protein expression profiles of cultured cells showed that the expression of hnRNPs H and F are differentially regulated in different pairs of normal and transformed cells. Thus, while hnRNP F was up-regulated in SV40-transformed human embryonal lung (MRC-5) fibroblasts as compared with their normal counterparts (Fig. 9, *A* and *B*), the level of hnRNP H was down-regulated. On the other hand, when normal human keratinocytes were compared with their SV40-transformed counterparts, we found that the level of hnRNP H was up-regulated 1.6 times in the transformed cells, while the level of hnRNP F was virtually unchanged (Celis and Olsen, 1994) (Fig. 9, *C* and *D*). In addition, while the expression levels of both hnRNPs were found to be independent of exposure of normal human keratinocytes to several substances that included dibutyryl cyclic AMP, dibutyryl cyclic GMP, interferon-α, interferon-β, transforming growth factor-α, transforming growth factor-β, IL-1α, IL-1β, IL-3, IL-7, IL-8, and tumor necrosis factor-α (not shown; see "Materials and Methods"), we found that long term treatment with 4β-phorbol 12-myristate 13-acetate (Fig. 9, *E* and *F*), resulted in a strong down-regulation of the level of hnRNP F but had no effect on hnRNP H. These observations suggest that hnRNP F gene expression may be regulated through the protein kinase C signaling pathway.

Conclusion—In conclusion, we have identified three hnRNPs (H, H′, and F) and possibly two additional proteins, IEF SSPs 2222 and 2326, as sequence unique members of a ubiquitously expressed subfamily of the more than 20 known hnRNPs. Although many of the other known hnRNP variants are produced by alternative splicing (Dreyfuss *et al.*, 1993), the proteins that constitute this subfamily are encoded by different genes, localized at different chromosomes, and possess different regulatory properties (at least H and F). Even though the proteins are closely sequence-related, they exhibit obvious differences, especially in their COOH terminus, where hnRNP F is shorter and contains a deletion when compared with hnRNPs H and H′. These major differences occur outside the qRRM regions in an auxiliary domain. Auxiliary domains in hnRNPs have been assigned a number of different functions (Biamonti and Riva, 1994). Variability among the auxiliary domains could thus explain putative functional differences between various subfamily members.

Acknowledgments—We are very grateful to E. Birney for performing the profile analysis and for help with the interpretation of the result. 40 S hnRNP particles and their nuclease-resistant structures were a kind gift of Dr. W. M. LeStourgeon. The pMT21 vector was kindly supplied by Dr. R. J. Kaufmann. We thank I. Kjærgaard, E. Walbum, A. Celis, B. Basse, J. B. Lauridsen, G. Ratz, and I. Andersen for technical assistance.

REFERENCES

Barton, G. J. (1993) *Protein Eng.* **6,** 37–40
Biamonti, G., and Riva, S. (1994) *FEBS Lett.* **340,** 1–8
Birney, E., Kumar, S., and Krainer, A. R. (1993) *Nucleic Acids Res.* **21,** 5803–5816
Bravo, R., Small, J. V., Fey, S. J., Larsen, P. M., and Celis, J. E. (1982) *J. Mol. Biol.* **154,** 121–143
Burd, C. G., and Dreyfuss, G. (1994) *EMBO J.* **13,** 1197–1204
Cavener, D. R., and Ray, S. C. (1991) *Nucleic Acids Res.* **19,** 3185–3192
Celis, J. E., and Olsen, E. (1994) *Electrophoresis* **15,** 309–344
Celis, J. E., Bravo, R., Arenstorf, H. P., and LeStourgeon, W. M. (1986) *FEBS Lett.* **194,** 101–109
Celis, J. E., Gesser, B., Rasmussen, H. H., Madsen, P., Leffers, H., Dejgaard, K., Honoré, B., Olsen, E., Ratz, G., Lauridsen, J. B., Basse, B., Mouritzen, S., Hellerup, M., Andersen, A., Walbum, E., Celis, A., Bauw, G., Puype, M., Van Damme, J., and Vandekerckhove, J. (1990) *Electrophoresis* **11,** 989-1071
Celis, J. E., Rasmussen, H. H., Olsen, E., Madsen, P., Leffers, H., Honoré, B., Dejgaard, K., Gromov, P., Vorum, H., Vassilev, A., Baskin, Y., Liu, X., Celis, A., Basse, B., Lauridsen, J. B., Ratz, G. P., Andersen, A. H., Walbum, E., Kjærgaard, I., Andersen, I., Puype, M., Van Damme, J., and Vandekerckhove, J. (1994a) *Electrophoresis* **15,** 1349–1458

Celis, J. E., Ratz, G., Basse, B., Lauridsen, J. B., and Celis, A. (1994b) in *Cell Biology: A Laboratory Handbook* (Celis, J. E., ed) pp. 222–230, Academic Press, Inc., San Diego
Christensen, J. H., Hansen, P. K., Lillelund, O., and Thogersen, H. C. (1991) *FEBS Lett.* **281,** 181–184
Dejgaard, K., and Celis, J. E. (1994) in *Cell Biology: A Laboratory Handbook* (Celis, J. E., ed) pp. 339–344, Academic Press, Inc., San Diego
Dejgaard, K., Leffers, H., Rasmussen, H. H., Madsen, P., Kruse, T. A., Gesser, B., Nielsen, H., and Celis, J. E. (1994) *J. Mol. Biol.* **236,** 33–48
Devereux, J., Haeberli, P., and Smithies, O. (1984) *Nucleic Acids Res.* **12,** 387–395
Dreyfuss, G., Swanson, M. S., and Pinol Roma, S. (1988) *Trends Biochem. Sci.* **13,** 86–91
Dreyfuss, G., Matunis, M. J., Pinol-Roma, S., and Burd, C. G. (1993) *Annu. Rev. Biochem.* **62,** 289–321
Düzgünes, N., and Felgner, P. L. (1993) *Methods Enzymol.* **221,** 303–306
Francke, U. (1994) *Cytogenet. Cell Genet.* **65,** 206–219
Gluzman, Y. (1981) *Cell* **23,** 175–182
Görlach, M., Burd, C. G., and Dreyfuss, G. (1994) *J. Biol. Chem.* **269,** 23074–23078
Higgins, D. G., Bleasby, A. J., and Fuchs, R. (1992) *Comput. Appl. Biosci.* **8,** 189–191
Hochuli, E., Dobeli, H., and Schacher, A. (1987) *J. Chromatogr.* **411,** 177–184
Honoré, B., Leffers, H., Madsen, P., Rasmussen, H. H., Vandekerckhove, J., and Celis, J. E. (1992) *J. Biol. Chem.* **267,** 8485–8491
Honoré, B., Leffers, H., Madsen, P., and Celis, J. E. (1993a) *Eur. J. Biochem.* **218,** 421–430
Honoré, B., Madsen, P., and Leffers, H. (1993b) *J. Biochem. Biophys. Methods* **27,** 39–48
Huschtscha, L. I., and Holliday, R. (1983) *J. Cell Sci.* **63,** 77–99
Jacobs, K. A., Rudersdorf, R., Neill, S. D., Dougherty, J. P., Brown, E. L., and Fritsch, E. F. (1988) *Nucleic Acids Res.* **16,** 4637–4650
Kaufman, R. J., Davies, M. V., Wasley, L. C., and Michnick, D. (1991) *Nucleic Acids Res.* **19,** 4485–4490
Kenan, D. J., Query, C. C., and Keene, J. D. (1991) *Trends Biochem. Sci.* **16,** 214–220
Kozak, M. (1987) *Nucleic Acids Res.* **15,** 8125–8148
Madsen, P., Rasmussen, H. H., Leffers, H., Honoré, B., and Celis, J. E. (1992) *J. Invest. Dermatol.* **99,** 299–305
Matunis, M. J., Xing, J., and Dreyfuss, G. (1994) *Nucleic Acids Res.* **22,** 1059–1067
McDonald, H., Smailus, D., Jenkins, H., Adams, K., Simpson, N. E., and Goodfellow, P. J. (1992) *Genomics* **13,** 344–348
O'Farrell, P. H. (1975) *J. Biol. Chem.* **250,** 4007–4021
Qian, Z., and Wilusz, J. (1994) *Nucleic Acids Res.* **22,** 2334–2343
Query, C. C., Bentley, R. C., and Keene, J. D. (1989) *Cell* **57,** 89–101
Rasmussen, H. H., Van Damme, J., Puype, M., Gesser, B., Celis, J. E., and Vandekerckhove, J. (1992) *Electrophoresis* **13,** 960–969
Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
Shamoo, Y., Abdul Manan, N., Patten, A. M., Crawford, J. K., Pellegrini, M. C., and Williams, K. R. (1994) *Biochemistry* **33,** 8272–8281
Sillekens, P. T. G., Habets, W. J., Beijer, R. P., and van Venrooij, W. J. (1987) *EMBO J.* **6,** 3841–3848
Studier, F. W., and Moffatt, B. A. (1986) *J. Mol. Biol.* **189,** 113–130
Swanson, M. S., and Dreyfuss, G. (1988) *Mol. Cell. Biol.* **8,** 2237–2241
Taylor-Papadimitriou, J., Purkis, P., Lane, E. B., McKay, I. A., and Chang, S. E. (1982) *Cell Differ.* **11,** 169–180
Tommerup, N., and Vissing, H. (1995) *Genomics* **27,** 259–264
Tunón, P., and Johansson, K. E. (1984) *J. Biochem. Biophys. Methods* **9,** 171–179
Vandekerckhove, J., and Rasmussen, H. H. (1994) in *Cell Biology: A Laboratory Handbook* (Celis, J. E., ed) pp. 359–368, Academic Press, Inc., San Diego
Vorechovsky, I., Vetrie, D., Holland, J., Bentley, D. R., Thomas, K., Zhou, J.-N., Notarangelo, L. D., Plebani, A., Fontan, G., Ochs, H. D., Hammarström, L., Sideras, P., and Smith, C. I. E. (1994) *Genomics* **21,** 517–524