Proc. Natl. Acad. Sci. USA
Vol. 78, No. 2, pp. 810–814, February 1981
Biochemistry

Abundant pseudogenes for small nuclear RNAs are dispersed in the human genome

(multigene families/*Alu* family sequence)

RICHARD A. DENISON, SCOTT W. VAN ARSDELL, LAUREL B. BERNSTEIN, AND ALAN M. WEINER

Department of Molecular Biophysics and Biochemistry, Yale University, 333 Cedar Street, P.O. Box 3333, New Haven, Connecticut 06510

Communicated by Donald D. Brown, October 24, 1980

ABSTRACT We have cloned and partially characterized 24 loci from the human genome which are complementary to U1, U2, or U3, the three major species of small nuclear RNA (snRNA) in HeLa cells. When compared to the known U1 (human) and U2 (rat) snRNA sequences, the DNA sequences we report here for the complementary regions from two of the clones, U1.11 and U2.7, reveal the presence of truncated and divergent gene copies. Furthermore, most if not all of the 24 cloned loci contain gene copies that are significantly divergent from the homologous HeLa snRNA species because DNA from every recombinant phage except U1.7 and U1.15 proved unable to form snRNA·DNA hybrids which protect full-length HeLa snRNA from mild digestion with ribonuclease T1. Hence, we refer to these loci as snRNA pseudogenes. In both clones U1.11 and U2.7, an element of the dominant middle repetitive DNA sequence family in the human genome, the *Alu* family, is located upstream from the snRNA pseudogene and in the same orientation. *Alu* elements in the same location and orientation relative to bona fide genes have previously been found in the human β-globin gene cluster [Duncan, C. H., Biro, P. A., Choudary, P. V., Elder, J. T., Wang, R. C., Forget, G. B., deRiel, J. K. & Weissman, S. M. (1979) *Proc. Natl. Acad. Sci. USA* 76, 5095–5099]. We discuss the significance of these findings in relation to the nature of snRNA multigene families and other reported examples of pseudogenes.

Among the various classes of molecules found in all eukaryotic cells, the small nuclear RNAs (snRNAs) have received relatively little attention until recently. The sequences of eight of these small homogeneous RNA species from mammalian cells (U1, U2, U3A, U3B, U4, U5, U6, and 4.5S RNA) have been determined (1–6). U3 snRNA is found in the nucleolus, and other snRNA species are associated with either the nucleoskeleton or the nucleoplasm (7). At least three of the snRNAs are subject to considerable evolutionary conservation: U1 and U2 snRNA from chicken, rat, mouse, and human cells yield identical ribonuclease T1 fingerprints (1, 8, 9), and the RNA sequences of U1 from HeLa cells and U1a from rat Novikoff hepatoma differ in only 2 of 165 positions (1). We have also shown that a snRNA from the lower eukaryotic cellular slime mold *Dictyostelium discoideum* is over 40% homologous to U3 snRNA from the rat (10).

Two recent developments have prompted new interest in the intracellular packaging and function of the snRNAs. First, antibodies produced by patients with the autoimmune disease systemic lupus erythematosus have been shown to recognize discrete cellular components that contain snRNAs complexed with a defined set of proteins (8), thus opening the way for detailed structural studies of small nuclear ribonucleoprotein particles. Second, Lerner *et al.* (9) and Rogers and Wall (11) have observed a striking complementarity between the 5′ end of U1 snRNA and the sequences that span splice junctions in nuclear RNA, suggesting a possible role, at least for U1 snRNA, in the splicing of mRNA precursors.

Data concerning the organization and expression of snRNA genes are scant. Previous estimates of the reiteration frequency of snRNA genes in mammals had ranged from 100 (12) to 1000 (13) copies, based on hybridization of purified snRNA species to bulk DNA in solution. Our own recent results suggest that many of the sequences complementary to the three major snRNA species (U1, U2, and U3) in human cells are not interspersed with each other (ref. 14 and unpublished observations). Although no definitive data exist regarding transcription of the snRNAs, the inhibition of U1 and U2 biosynthesis by both α-amanitin (15) and 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (16) may imply that RNA polymerase II is responsible for the transcription of at least certain snRNAs. Furthermore, the sensitivity of U1 and U2 biosynthesis to low doses of ultraviolet irradiation suggests that these snRNA species are derived from transcription units as large as 5 kilobases (kb) (17).

We originally decided to clone the genes encoding snRNAs in the hope that a systematic investigation of their organization and expression would also open up new avenues for understanding the function of the snRNAs themselves. We report here the surprising result that most, if not all, of the genomic loci complementary to snRNAs U1, U2, and U3 that we have examined contain divergent, and in some cases truncated, gene copies when compared with the corresponding HeLa cell snRNA species.

MATERIALS AND METHODS

A library of 15-kb partial *Eco*RI fragments of human placental DNA in the λ vector Charon 4A (18) was kindly supplied by A. Biro, P. V. Choudary, J. T. Elder, and S. M. Weissman. Plaques were screened by the method of Benton and Davis (19); we used as probes U1, U2, and U3 snRNAs isolated from HeLa cells and labeled *in vitro* at the 3′ end with 5′-^{32}P-labeled pCp and T4 RNA ligase as described (14). As little as 10^6 cpm of snRNA at a specific activity of 10^6 cpm/μg was sufficient to screen six 140-mm nitrocellulose filters, each bearing 8000 plaques. With each probe, ≈0.1% of the plaques were scored as positive and repurified for further study. Small quantities of recombinant DNA were prepared from 4-ml NZY cultures (20); larger quantities were prepared from recombinant phage grown and purified as described (18). *Eco*RI fragments that hybridized with U1 or U2 were subcloned into pBR322, and the recombinant plasmids were designated pU1.11 and pU2.7.

All experiments involving recombinant molecules were performed under P2-EK2 conditions as specified by the revised National Institutes of Health guidelines for recombinant DNA research.

RESULTS

The human genomic library in Charon 4A was screened separately with U1, U2, and U3 snRNA probes to yield 13 recom-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: kb, kilobase; snRNA, small nuclear RNA.

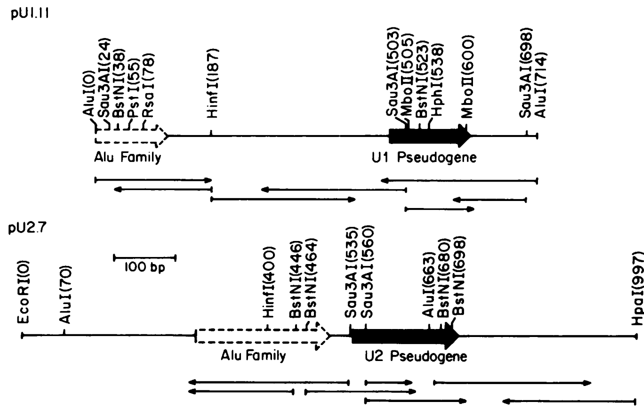

FIG. 1. Restriction maps of the region complementary to U1 and U2 snRNA in the plasmid subclones pU1.11 (*Upper*) and pU2.7 (*Lower*). The *Alu* I fragment from pU1.11 was cloned into pBR322 by using *Eco*RI linkers. The *Eco*RI/*Hpa* I fragment from U2.7 was obtained from a *Hpa* I digest of the 3-kb *Eco*RI fragment carried in pU2.7. The DNA sequencing strategy is shown below each map. bp, Base pairs.

binant phage complementary to U1 (designated U1.1, U1.2, . . .), 7 complementary to U2, and 4 complementary to U3. Each of the 24 phage contains 15 kb of human DNA derived from a distinct genomic locus because the size of the *Eco*RI restriction fragment complementary to the snRNA is different in each clone (data not shown); moreover, when hybridized to total cellular RNA labeled *in vivo*, each recombinant phage tested (U1.11, as well as all the U2 and U3 loci) selected only the expected snRNA, demonstrating that sequences complementary to one snRNA are not generally interspersed with those complementary to the other two (ref. 14 and unpublished data).

We initially chose the recombinant phage U1.11 and U2.7 for detailed characterization. The unique region complementary to the snRNA in each phage was localized as described (10) and the sequence was determined according to Maxam and Gilbert (21). Fig. 1 presents a high-resolution map of the relevant restriction fragments from each of the plasmid pBR322 subclones pU1.11 and pU2.7; Fig. 2 compares the DNA sequence of these cloned human genomic loci with the known sequence of human U1 and rat U2 snRNA (1, 2). We did not expect such striking divergence between the RNA and DNA sequences, even in the heterologous comparison of rat U2 RNA with the human genomic clone pU2.7; rat and human U2 snRNAs appear indistinguishable as judged both by RNase T1 fingerprints and by secondary analysis of T1 oligonucleotides after redigestion with pancreatic RNase A (refs. 8 and 9; J. A. Steitz, personal communication). Furthermore, the predicted T1 oligonucleotides for a snRNA transcribed from pU1.11 or

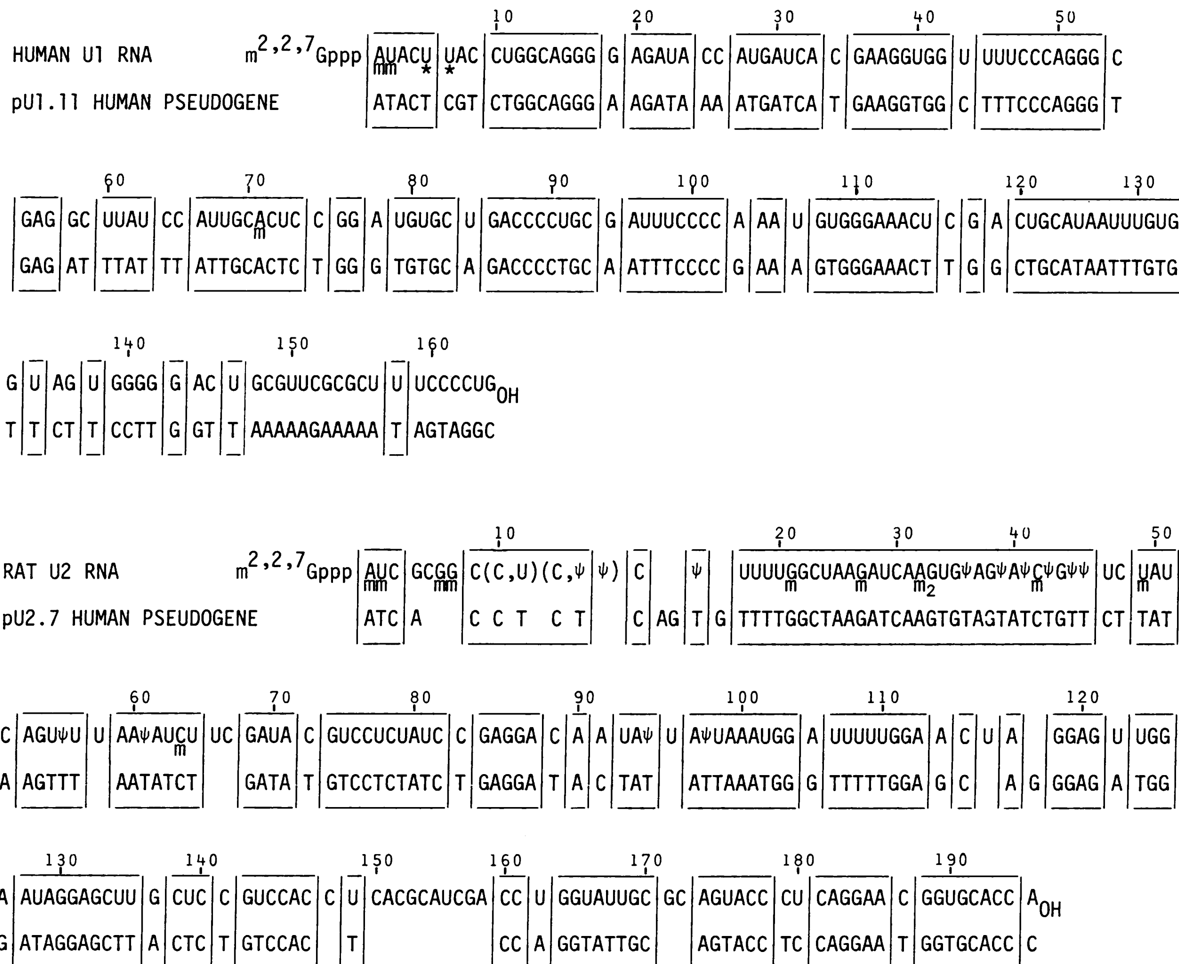

FIG. 2. Comparison of RNA sequences for U1 RNA from HeLa cells and U2 RNA from rat Novikoff hepatoma (1, 2) with the corresponding DNA sequences from human clones pU1.11 and pU2.7. Boxes denote homologous sequences. Numbers refer to positions in the RNA sequences. Ambiguities in the reported rat U2 RNA sequence (2) at positions 10–11 and 12–14 are enclosed by parentheses and aligned to give the best match to the pU2.7 DNA sequence. Asterisks (*) at positions 6 and 7 in U1 RNA denote modified uridine residues, most likely pseudouridines (1). Modified residues are abbreviated as follows: m, 2′-*O*-methyl; m_2, N^6-methyl-2′-*O*-methyl; ψ, pseudouridine; $m^{2,2,7}$, $N^{2,2,7}$-trimethyl.

pU2.7 are incompatible with the observed U1 and U2 fingerprints. We therefore refer to these loci as pseudogenes.

Both pseudogene sequences exhibit transitional and transversional single base changes, and pU2.7 also contains internal insertions and deletions relative to the rat U2 snRNA sequence. In particular, the pU1.11 sequence corresponding to the 5' end of U1 RNA contains three mismatched bases at positions 7–9 precisely in the region of U1 that is complementary to splice junction sequences in nuclear RNA (9, 11). The pU1.11 pseudogene also lacks any sequence corresponding to the last 32 nucleotides of U1 RNA, and no sequences resembling the 3' end of U1 RNA could be found in 76 additional base pairs of 3' flanking DNA (data not shown). Discounting truncation at the 3' end, the overall sequence mismatch up to position 133 in pU1.11 is 16% (21 differences in 132 nucleotides), and the RNA and DNA sequences maintain perfect registry over this entire region. The homology between pU2.7 and rat U2 RNA is somewhat better, with only 10% mismatch and no truncation, but the registry is poor (18 differences in the 173 nucleotides aligned in Fig. 2, with insertion of 4 and deletion of 21 nucleotides). The lack of registry might in part reflect sequence differences between human U2 RNA and the published rat U2 sequence which was determined by classical RNA fingerprinting techniques (ref. 2; also see *Note Added in Proof*).

T1 RNase Protection Experiments. To determine whether the other 22 cloned genomic loci also contain pseudogenes, we screened DNA from these phage for the ability to form snRNA·DNA hybrids resistant to mild digestion with T1 RNase as described (14). Recombinant DNAs were immobilized on nitrocellulose filters and hybridized with purified HeLa U1, U2, or U3 RNA uniformly labeled *in vivo* with [^{32}P]phosphate. After treatment with 10 units of T1 RNase per ml at high salt concentration under conditions that leave a perfect RNA·DNA duplex intact (14), the protected snRNA fragments were eluted from the filters and their sizes were determined on a 15% polyacrylamide gel. U2 and U3 snRNA eluted directly from duplicate filters without RNase treatment were included as a control (see Fig. 4, lanes U2 and U3).

Based on the size of the HeLa snRNA fragments protected by the pseudogenes U1.11 and U2.7 of known sequence, the T1 RNase protection experiments appeared to be a sensitive measure of homology between the various cloned loci and the corresponding snRNA. For example, the exposed 3' end of U1 snRNA in a hybrid with U1.11 DNA resulted in a shortened but otherwise intact molecule after T1 RNase digestion (Fig. 3); moreover, this protected U1 fragment was found by fingerprint analysis to have lost T1 oligonucleotides derived from the 3' end of the molecule (data not shown). Similarly, the deletion of 10 base pairs at positions 149–158 in the U2.7 pseudogene should expose one or two guanosine residues of U2 RNA to T1 RNase digestion, resulting in one large and one small fragment, as was actually observed (Fig. 4). Of the 13 U1 clones, 7 U2 clones, and 4 U3 clones analyzed (Figs. 3 and 4), only U1.7 and U1.15 appeared to have significantly protected full-length HeLa snRNA from T1 RNase digestion. These clones may represent bona fide genes (see *Note Added in Proof*). The remaining genomic loci almost certainly contain pseudogenes because the observed T1 RNase protection patterns resemble those of U1.11 and U2.7. Within the set of clones complementary to each of the snRNA species, apparently identical protected RNA fragments appear in many or all lanes; for example, six out of seven U2 clones protect a fragment of U2 RNA that migrates just ahead of tRNA (Fig. 4). We have no basis for explaining this phenomenon, although it might reflect RNA secondary structure, some degree of homology between the pseudogenes, or even the presence of intervening sequences.

Alu Family Sequences. Computer analysis (22) of DNA sequences flanking the pseudogenes produced yet another surprise. Members of the major family of interspersed middle repetitive DNA sequences in the human genome, known collectively as the *Alu* family (23, 24), lie upstream from both snRNA pseudogenes and in the same orientation as the pseudogenes themselves (Fig. 1). [The polarity of an *Alu* family sequence is defined by the direction in which the *Alu* elements of the human β-globin gene cluster are transcribed *in vitro* by RNA polymerase III (23, 25).] We have sequence data for only the second half of the *Alu* family element in U1.11, beginning at the canonical *Alu* I restriction site, but this portion of the

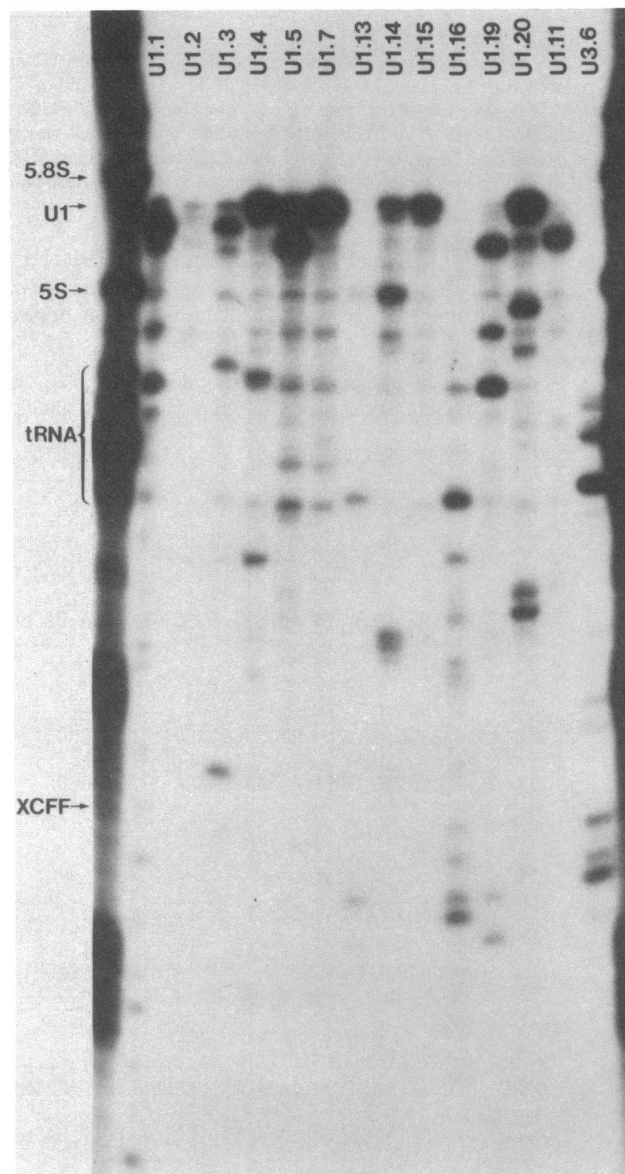

FIG. 3. Size fractionation of U1 RNA fragments remaining bound to DNA of U1 genomic clones after mild digestion with T1 RNase. Methods have been described (14). Phage DNAs from each of the U1 clones (and from one U3 clone, U3.6) were immobilized on nitrocellulose filters and hybridized with *in vivo* labeled and gel-purified U1 RNA (U3 RNA for U3.6) from HeLa cells (10^6 cpm/μg, 10^4 cpm in each hybridization). Filters were washed and treated with 10 units of T1 RNase per ml in high concentration of salt. Protected fragments were eluted and their sizes were determined on a 15% urea gel. Outer lanes are total *in vivo* labeled HeLa cell RNA. Autoradiography was for 2 weeks at −70°C with an intensifying screen. XCFF, xylene cyanol FF.

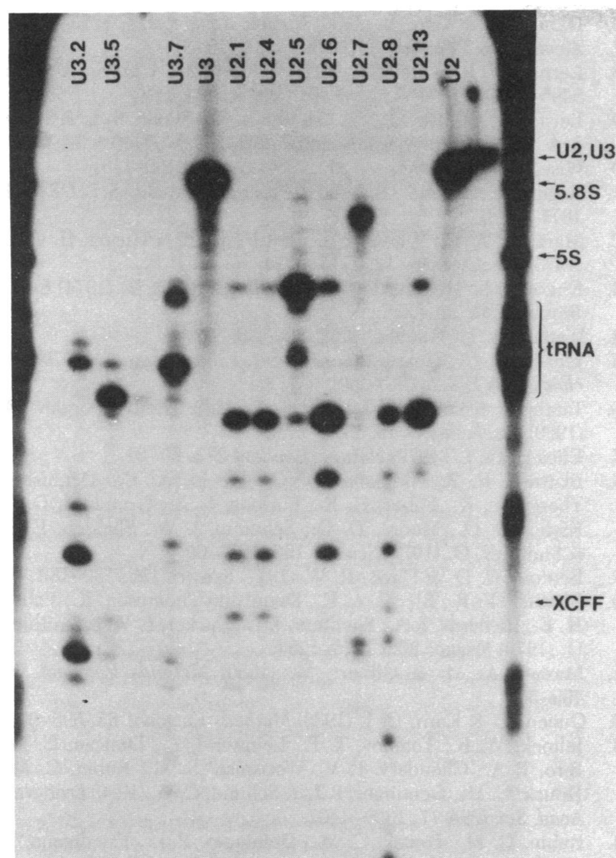

FIG. 4. Size fractionation of U2 and U3 RNA fragments remaining bound to DNA of U2 and U3 genomic clones after mild digestion with T1 RNase. See legend to Fig. 3 for details.

element maintains nearly perfect registry with the *Alu* family "consensus" sequence (24) over its entire length and is 72% homologous. In contrast, the complete DNA sequence of the *Alu* family element in U2.7 is highly divergent and scrambled relative to the *Alu* consensus sequence, with homologous tracts as long as 45 base pairs interspersed with short nonhomologous blocks (unpublished data).

DISCUSSION

From a genomic library containing 15-kb segments of human placental DNA in the λ vector Charon 4A (18), we have isolated and partially characterized 24 recombinant bacteriophage carrying distinct genomic loci complementary to one of the three major species of snRNA in HeLa cells, U1, U2, and U3. Surprisingly, complete DNA sequence data for the complementary region in two clones, U1.11 and U2.7, revealed no bona fide gene copies but rather pseudogenes that are divergent and truncated when compared to the known sequence of human U1 and rat U2 snRNAs (1, 2). The remaining 22 recombinant phage were screened for their ability to form DNA·RNA hybrids with snRNA uniformly labeled *in vivo*, and all but two of them also proved unable to protect full-length HeLa snRNA significantly from mild digestion with T1 RNase; hence, these clones are likely to contain pseudogenes as well (see *Note Added in Proof*).

Our findings raise many questions about the nature of snRNA multigene families and the significance of pseudogene sequences in general. Although snRNA genes were considered to be highly reiterated in mammals, our detailed analysis of several cloned snRNA pseudogenes implies that many or most of the genomic sequences previously detected by solution hybridization of purified snRNA to bulk DNA (12, 13) could not possibly encode the major species of snRNA. This conclusion is consistent with the earlier suggestion, based on the broad melting profile of such snRNA·DNA hybrids, that the average divergence between snRNA genes might be as high as 15% (13).

One intriguing possibility is that the pseudogene sequences we have examined actually code for snRNA species that are scarce in HeLa cells but abundant in other cell types or developmental stages. A potential analogy might be drawn to the highly divergent multigene families that encode developmentally regulated chorionic proteins in the silk moth (26). Although both U1 and U2 snRNAs appear to be homogeneous in HeLa cells (8), minor sequence heterogeneity within a single cell type has been reported for U1 in the mouse [U1a and U1b (ref. 8)] and U3 in the rat [U3A, U3B, and U3C (ref. 3)], and at least two minor sequence variants of the snRNA D2 have been identified in *D. discoideum* (10, 27). Such minor RNA sequence heterogeneity suggests that the various RNAs are actually encoded by multiple genes that belong to the extended pseudogene families described here. We have previously shown that at least two of the five genomic loci complementary to D2 snRNA in *Dictyostelium* must be genetically active (10).

Although some of the snRNA pseudogenes may actually be transcribed *in vivo*, they are unlikely to produce mature snRNAs because the strong conservation of snRNA sequence among hen liver, rat hepatoma, human cervical carcinoma, and mouse Ehrlich ascites cells (8, 9) contrasts so dramatically with the diversity of the pseudogene loci we have characterized by T1 protection experiments. A more remote possibility is that many snRNA pseudogenes are transcriptionally active and produce minor snRNA sequence variants which comigrate during electrophoresis as a single, apparently homogeneous, species. This would imply that the RNA sequences determined for U1 and U2 by classical fingerprinting techniques (2, 28) actually represent composite or "consensus" sequences derived from a large number of related RNA species. We can exclude this possibility because the very high background expected for a consensus fingerprint was not observed (compare refs. 8 and 29), and both U1 and U3B have now been shown to be homogeneous by enzymatic sequence analysis of end-labeled RNA (1, 3).

What then can be made of the finding that the human genome contains many related but clearly noncoding sequences homologous to snRNAs? Pseudogenes have been discovered in several other well-characterized gene families, including *Xenopus* 5S RNA genes (30), the α-globin (31–33) and β-globin (34–36) gene clusters in various mammals, and the actin genes of *D. discoideum* (37). Few generalizations can be drawn about the possible function of pseudogenes because the term refers to several types of unexpressed but significantly homologous sequences, ranging from a perfect mouse α-globin gene copy that has cleanly lost both intervening sequences (31, 32) to the truncated but otherwise nearly perfect 5S pseudogene of *Xenopus* (30). With the possible exception of the *Dictyostelium* actin genes (37) and the mouse α-globin genes (31, 32), pseudogene sequences reported to date seem to occur in gene families that are clustered. The snRNA pseudogenes, however, appear to be dispersed: sequences complementary to the snRNA in each of the 24 clones we have studied are confined to a single distinct *Eco*RI fragment, and the two recombinant phage U1.11 and U2.7 described in this paper as well as phage U1.7 and U3.5 each contains an isolated snRNA pseudogene copy embedded in 15 kb of human DNA.

The pseudogenes found in clustered gene families are generally thought to have arisen by a process of gene duplication or unequal crossing-over, followed by genetic drift (32–35).

Although this mechanism may also account for generation of the pseudogenes discussed here, the dispersion, high reiteration frequency, and sequence diversity of snRNA pseudogenes might have functional significance as well: the loci may serve in a regulatory capacity (32, 38), for example, by virtue of their ability to form base pairs with the corresponding snRNAs.

We were surprised to find upstream *Alu* family sequences with the same polarity as the U1.11 and U2.7 pseudogenes, and we have recently learned that two *Alu* family sequences form a hyphenated palindromic sequence surrounding a human U6 snRNA pseudogene, with the upstream *Alu* sequence also in the same orientation as the pseudogene itself (C. H. Duncan, P. V. Choudary, and S. M. Weissman, personal communication). Middle repetitive DNA sequences belonging to the *Alu* family precede both the β- and γ-globin gene cluster in humans (23, 25), and when used *in vitro* as templates for RNA polymerase III, the human *Alu* sequences produce discrete RNA transcripts with the same polarity as the globin genes themselves (25). A similar arrangement of middle repetitive sequences may also occur in the β-like globin gene cluster of the rabbit (39). One reasonable hypothesis is that in mammals, *Alu* family sequences or their equivalent (23, 40) often occur upstream from and in the same orientation as RNA polymerase II transcription units. Such a sequence arrangement may have survived whatever process generated the pseudogenes we have studied.

We have not yet identified bona fide genes encoding U1, U2, or U3 snRNA. However, we may have selected against these loci during construction, propagation, or screening of the recombinant phage library. It should be possible to isolate true snRNA genes from a more complete library by using procedures that distinguish between genes and pseudogenes. We believe that further analysis of snRNA pseudogenes will lead to a better understanding of their origin and possible function.

Note Added in Proof. The DNA sequence of clone U1.15 corresponds to that of human U1 RNA in all but 1 of 165 positions and is therefore likely to represent a bona fide gene; clone U1.7 contains a slightly divergent pseudogene. We have also obtained a revised sequence for rat U2 RNA (R. Reddy, personal communication). The U2.7 pseudogene maintains nearly perfect registry with this revised sequence except for a single deletion of $\approx$12 nucleotides centered at position 150.

We thank Steve Mount and Joan Steitz for many enlightening discussions, and Joe Hendrick for an endless supply of HeLa cells. This work was supported in part by National Science Foundation Grants PCM 76-81524 and PCM 78-21799 and National Institutes of Health Grant GM26312.

1. Branlant, C., Krol, A., Ebel, J. P., Lazar, E., Gallinaro, H., Jacob, M., Sri-Widada, J. & Jeanteur, P. (1980) *Nucleic Acids Res.* **8**, 4143–4154.
2. Shibata, H., Ro-Choi, T. S., Reddy, R., Choi, Y. C., Henning, D. & Busch, H. (1975) *J. Biol. Chem.* **250**, 3903–3920.
3. Reddy, R., Henning, D. & Busch, H. (1979) *J. Biol. Chem.* **254**, 11097–11105.
4. Reddy, R. & Busch, H., *The Cell Nucleus*, in press.
5. Epstein, P., Reddy, R., Henning, D. & Busch, H. (1980) *J. Biol. Chem.* **255**, 8901–8906.
6. Harada, F. & Kato, N. (1980) *Nucleic Acids Res.* **8**, 1273–1285.
7. Zieve, G. & Penman, S. (1976) *Cell* **8**, 19–31.
8. Lerner, M. R. & Steitz, J. A. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 5495–5499.
9. Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. & Steitz, J. A. (1980) *Nature (London)* **283**, 220–224.
10. Wise, J. A. & Weiner, A. M. (1980) *Cell* **22**, 109–118.
11. Rogers, J. & Wall, R. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1877–1879.
12. Marzluff, W. F., White, E. L., Benjamin, R. & Huang, H. C. C. (1975) *Biochemistry* **14**, 3715–3724.
13. Engberg, J., Hellung-Larsen, P. & Frederiksen, S. (1974) *Eur. J. Biochem.* **41**, 321–328.
14. Weiner, A. M. (1980) *Cell* **22**, 209–218.
15. Jensen, E. G., Hellung-Larsen, P. & Frederiksen, S. (1979) *Nucleic Acids Res.* **6**, 321–330.
16. Tamm, I., Kikuchi, T., Darnell, J. E., Jr. & Salditt-Georgiev, M. (1980) *Biochemistry* **19**, 2743–2748.
17. Elicieri, G. L. (1979) *Nature (London)* **279**, 80–81.
18. Blattner, F. R., Williams, B. G., Blechl, A. E., Denniston-Thompson, K., Faber, H. E., Furlong, L. A., Grunwald, D. J., Kiefer, D. O., Moore, D. O., Schumm, J. W., Sheldon, E. L. & Smithies, O. (1977) *Science* **196**, 161–169.
19. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
20. Blattner, F. R., Blechl, A. E., Denniston-Thompson, K., Faber, H. E., Richards, J. A., Slightom, J. L., Tucker, P. W. & Smithies, O. (1978) *Science* **202**, 1279–1284.
21. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
22. Queen, C. & Korn, C. J. (1980) *Methods Enzymol.* **65**, 595–609.
23. Jelinek, W. R., Toomey, T. P., Leinwand, L., Duncan, C. H., Biro, P. A., Choudary, P. V., Weissman, S. M., Rubin, C. M., Houck, C. M., Deininger, P. L. & Schmid, C. W. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1398–1402.
24. Rubin, C. M., Houck, C. M., Deininger, P. L., Friedmann, T. & Schmid, C. W. (1980) *Nature (London)* **284**, 372–374.
25. Duncan, C., Biro, P. A., Choudary, P. V., Elder, J. T., Wang, R. C., Forget, B. G., deRiel, J. K. & Weissman, S. M. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 5095–5099.
26. Jones, C. W., Rosenthal, N., Rodakis, G. C. & Kafatos, F. C. (1979) *Cell* **18**, 1317–1332.
27. Wise, J. A. & Weiner, A. M., *J. Biol. Chem.*, in press.
28. Reddy, R., Ro-Choi, T. S., Henning, D. & Busch, H. (1974) *J. Biol. Chem.* **249**, 6486–6494.
29. Robertson, H. D., Dickson, E. & Jelinek, W. (1977) *J. Mol. Biol.* **115**, 571–589.
30. Miller, J. R., Cartwright, E. M., Brownlee, G. G., Federoff, N. & Brown, D. D. (1978) *Cell* **13**, 717–725.
31. Nishioka, Y., Leder, A. & Leder, P. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2806–2809.
32. Vanin, E. F., Goldberg, G. I., Tucker, P. W. & Smithies, O. (1980) *Nature (London)* **286**, 222–226.
33. Proudfoot, N. & Maniatis, T. (1980) *Cell* **21**, 537–544.
34. Lacy, E. & Maniatis, T. (1980) *Cell* **21**, 545–553.
35. Jahn, C. L., Hutchinson, C. A., III, Phillips, S. A., Weaver, S., Haigwood, N. L., Voliva, C. F. & Edgell, M. H. (1980) *Cell* **21**, 159–168.
36. Cleary, M. L., Haynes, J. R., Schon, E. A. & Lingrel, J. B. (1980) *Nucleic Acids Res.* **8**, 4791–4802.
37. Firtel, R. A., Timm, R., Kimmel, A. R. & McKeown, M. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 6206–6210.
38. Davidson, E. H. & Britten, R. J. (1979) *Science* **204**, 1052–1059.
39. Shen, C.-K. J. & Maniatis, T. (1980) *Cell* **19**, 379–391.
40. Krayev, A. S., Kramerov, D. A., Skryabin, K. G., Ryskov, A. P., Bayev, A. A. & Georgiev, G. P. (1980) *Nucleic Acids Res.* **8**, 1201–1215.