Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 2435–2439, April 1984
Genetics

Novel cluster of α-interferon gene sequences in a placental cosmid DNA library

(human genomic library/multigene family/diversity)

BJÖRN LUND*, THOMAS EDLUND*, WERNER LINDENMAIER†, TOR NY*, JOHN COLLINS†, ERIK LUNDGREN*, AND ALEXANDER VON GABAIN*‡

*Unit for Applied Cell and Molecular Biology, University of Umeå, S-901 87 Umeå, Sweden; and †Gesellschaft für Biotechnologische Forschung mbH, Braunschweig-Stöckheim, Federal Republic of Germany

Communicated by L. L. Cavalli-Sforza, December 19, 1983

ABSTRACT **A human cosmid library was constructed and probed with a human α interferon (IFN-α) cDNA clone. One clone giving a strong hybridizing signal was isolated and characterized. The cosmid DNA insert represents a section of the human genome containing three regions of IFN-α-like sequences. The DNA was characterized with restriction endonuclease mapping, thereby allowing comparison to similar linkage groups reported recently and determination of homologous regions on the known physical map. The three IFN-α-like sequences were analyzed by a partial sequence analysis. Mapping and sequence data establish this section as a not-yet-described cluster of IFN-α sequences in the human genome; however, a part of the section matches to some degree to a previously described genomic region. The region described here could represent genetic polymorphism or a duplicated segment.**

Interferons (IFN) comprise a group of proteins defined by their antiviral activity. Classically, they are categorized into three groups by serological criteria: IFN-α, β, and γ (1). Recent information on the amino acid and nucleotide sequences implies that IFN-α and IFN-β have a similar genetic structure without introns and with a homology of 45% at the nucleotide level (2), while IFN-γ has an intron–exon structure and shows only remote similarities with IFN-α.

In humans, the IFN-α genes represent a multigene family. At least eight different mRNA species have been identified from a single myeloid cell line studied as cDNA copies (3). Of 14 distinct genes or gene-like sequences of IFN-α type that occur in the library constructed by Maniatis, 5 of them seem to be pseudogenes (4). A more detailed study on a chromosomal section isolated from the same gene library has given suggestive evidence for IFN-α gene evolution via a relatively recent duplication of a 4000-base-pair segment (5). Some of the biological functions have been demonstrated to differ for some of the IFN-α gene products (6).

The aim of the present study was to isolate clusters of closely linked genes in order to extend the mapping and cataloguing of the human IFN-α genes. A genomic library was used from another individual than that previously studied. Furthermore, advantage was taken of another vehicle, the cosmid cloning system. This makes it possible to clone larger DNA fragments than in the phage cloning system, increasing the probability of isolating entire regions carrying clustered genes. As the cosmid vector used also permits transfection to eukaryotic cells (7), it provides an opportunity to study further the impact of the genomic surroundings on induced expression.

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.

MATERIALS AND METHODS

The principle for the construction of the human genomic cosmid library has been described (7). As a host for cosmid propagation, the *Escherichia coli* K-12 strain HB 101 (8) was used.

Isolation of the Cosmid Hybrid Carrying IFN-α Genes. *In vivo* packaged cosmids were transduced into HB 101. The bank consisted of 5×10^5 clones, maintained on 12 nitrocellulose filters (Schleicher & Schüll, BA 85; diameter, 130 mm). Replica filters were screened by colony hybridization with a ^{32}P-labeled IFN-α cDNA (unpublished work) as a probe. A colony area giving a strong positive signal was diluted and, by repicking, a single colony was isolated.

Restriction Enzyme Digestion, DNA Ligation, Polynucleotide Kinase Treatment, and Agarose Gel Electrophoresis. The conditions for using restriction endonucleases, polynucleotide kinase, and T4 DNA ligase were those suggested by the manufacturer. Digested DNA was loaded on 0.5% (wt/vol) horizontal agarose gels in 20 mM sodium acetate/2 mM EDTA/33 mM Tris·HCl, pH 7.8, and subjected to electrophoresis at 50 V for 16 hr.

Blotting and Hybridization Procedures. Colony hybridization was performed as described (9). Southern blotting analyses were as published (10) with the following modifications. DNA was transferred to GeneScreen filters by electrophoresis as described by the manufacturer (New England Nuclear).

Filters were hybridized under selected conditions as described by Moseley *et al.* (11). IFN-α cDNA was nick-translated with [α-^{32}P]dGTP to a level of 1×10^8 cpm/μg (9). The plaque hybridization procedure was as described by Benton and Davies (12) with a synthetic oligonucleotide probe radiolabeled at the 5′ end with [γ-^{32}P]ATP by using polynucleotide kinase (13). The hybridization conditions were applied as described by Wallace *et al.* (14).

Synthesis of the Oligodeoxyribonucleotide. The oligodeoxynucleotide used as probe and primer was synthesized by the solid-phase phosphite method (15) and was supplied by KabiGen AB (Stockholm, Sweden).

Nucleotide Sequence Determination. Fragments were cloned into the phage M13 mp8 and mp9 (16). Single-stranded template DNA was isolated from the phage as described by Messing *et al.* (17). The DNA sequences were determined by the dideoxy method of Sanger *et al.* (18) with the above-mentioned oligonucleotide as a primer (see *Results*). The M13 universal primer, used to determine the orientation of the insert, was purchased from New England BioLabs.

Abbreviations: IFN, interferon; kb, kilobase(s).
‡To whom reprint requests should be addressed.

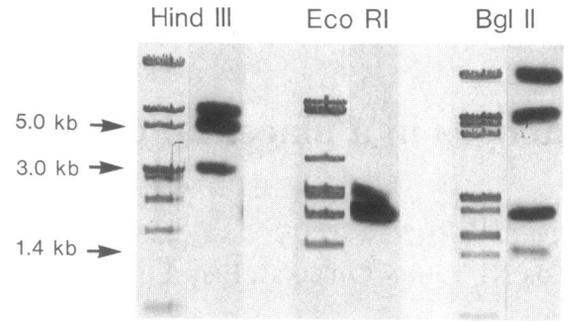

FIG. 1. Restriction enzyme analysis of cosmid DNA insert. The left-hand columns of each digestion show restriction patterns obtained after digesting the cosmid DNA with the enzymes *Hin*dIII, *Eco*RI and *Bgl* II, respectively, and fractionating the fragments on agarose gels. The right-hand columns of each digestion show which of the fragments hybridized to the IFN-α cDNA probe, used after blotting to a nitrocellulose paper.

RESULTS

Identification of a Chromosomal Section Encoding at Least Two IFN-α Sequences. Human placental DNA was isolated, partially digested with the endonuclease *Taq* I, and ligated in the *Cla* I site of the cosmid vector pHC79-2cos/tk as recently described for the construction of a library from murine liver DNA (7). The procedure for screening the library and selecting positive colonies is described in *Materials and Methods*. One colony hybridized to the probe with a stronger signal than the others. The colony was purified by replating and reprobing at low colony density.

The plasmid DNA of this candidate was digested with the restriction endonucleases *Bgl* II, *Eco*RI, and *Hin*dIII and was analyzed on Southern blots (10) with the same IFN-α cDNA probe, showing four bands for digestion with *Bgl* II, three bands with *Eco*RI, and three bands with *Hin*dIII (Fig. 1). This indicated that the identified chromosomal section might carry several regions homologous to IFN-α genes.

Preliminary Restriction Analysis and Subcloning of the Cosmid DNA. In order to simplify our further characterization, several endonucleases recognizing six-nucleotide sequences were tested on the cosmid DNA. *Bam*HI, *Sal* I and *Sph* I all cut the chromosomal insert once. *Hpa* I cut at three sites in the insert, whereas *Pvu* I cut only in the *cos* vector. By analyzing single and double digests of those enzymes, it was possible to establish a restriction map (Fig. 2). The size of the chromosomal insert was calculated to be about 29 kb, based on the individual sizes of the fragments obtained from the electrophoretic analysis (19).

The single *Bam*HI site mapping in the center of the chromosomal insert was used to subclone the cosmid. The prod-

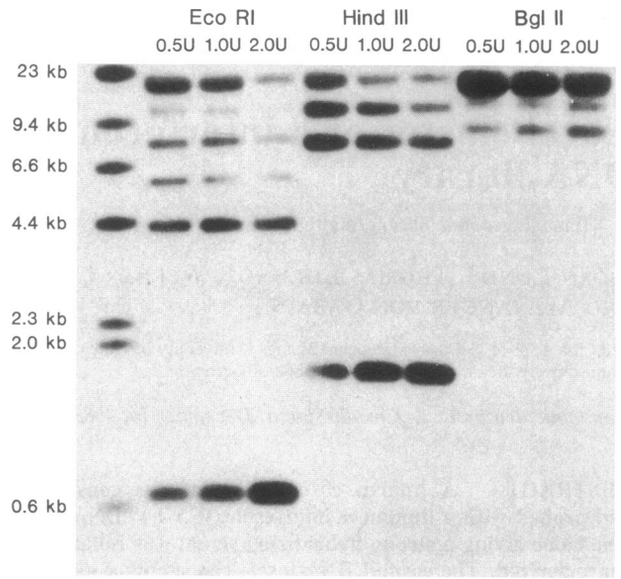

FIG. 3. Analysis of subclone 23 by Southern blot. DNA of subclone 23 was linearized with the endonuclease *Pvu* I and then subjected to limited digestion with the indicated enzymes in different concentrations. The resulting fragments were fractionated on an agarose gel and transferred to a nitrocellulose filter. The fragments were hybridized to a terminal probe as described in the text. The size distribution of the different fragments allowed the positioning of the restriction sites relative to the *Pvu* I site including their relative distances. U, units of enzyme. The left-most lane is radiolabeled *Hin*dIII fragments of phage λ used as size markers.

ucts of a complete digest with *Bam*HI were religated under conditions favoring intramolecular ligation, giving rise to a chimeric plasmid derived from a *Bam*HI fragment, which carries a part of the chromosomal insert and a part of the vector including the pBR322 replicon. This insert was designated 49 in Fig. 2.

Subsequently, the cosmid was completely digested with *Bam*HI and *Pvu* I and ligated in pBR322 at the unique *Bam*HI site. The chimeric plasmid obtained carried the other half of the chromosomal insert (designated 23 in Fig. 2).

Fine Mapping of the Cosmid with the Endonucleases *Bgl* II, *Eco*RI, and *Hin*dIII. In order to compare the DNA section to human chromosomal regions carrying IFN-α genes (4, 5, 20), we decided to establish restriction maps of the cosmid DNA, using the endonucleases *Bgl* II, *Eco*RI and *Hin*dIII because they have been used in previous publications (4, 5, 20).

A modified procedure according to Smith and Birnstiel was applied (21). The subclones (Fig. 2) were linearized by using the unique *Pvu* I site that maps in the vector. Aliquots

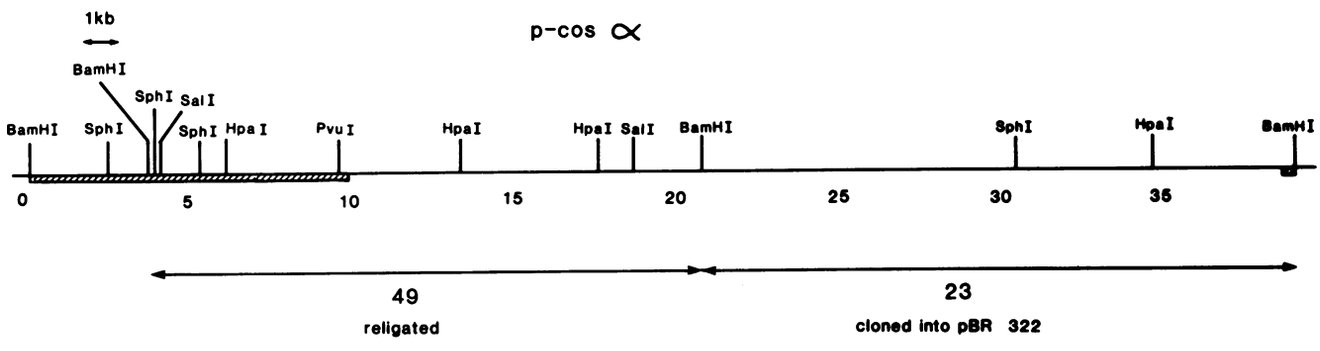

FIG. 2. Restriction map of the cosmid DNA insert. The following endonucleases were used: *Bam*HI, *Hpa* I, *Pvu* I, *Sal* I, and *Sph* I. The positions of the recognition sites for the same enzymes are indicated in the *cos* vector part (hatched box), see also Lindenmaier *et al.* (7). The terminal *Bam*HI sites are identical in the circular form of the molecule. The subclones 49 and 23 are indicated underneath the map; subcloning was done as described in the text.

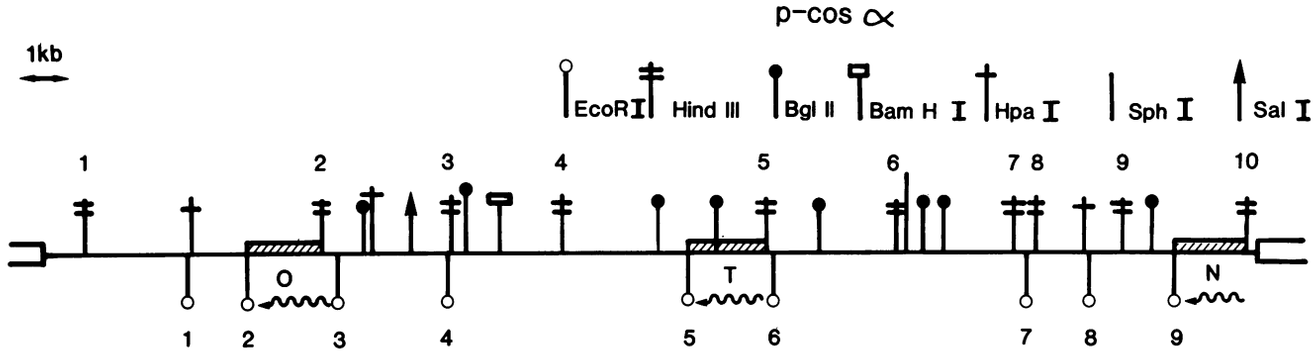

FIG. 4. Assignment of IFN-α homologous regions to the restriction map of the cosmid DNA insert. The hatched boxes indicate the regions binding to the IFN-α cDNA probe. IFN-α *O* maps on the far left, IFN-α *N* on the far right, while IFN-α *T* is positioned in the middle. The potential transcriptional direction for all three genes is from right to left.

were exposed to limited digestion with these endonucleases. Two probes were prepared by digesting pBR322 (22) either with *Eco*RI and *Pvu* I or with *Cla* I and *Bam*HI. The respective smaller fragments were isolated and nick-translated. The *Eco*RI–*Pvu* I fragment was used to probe the partial digests of both subclones from one end.

In addition, the entire cosmid was digested with *Bam*HI and probed from the other end with the *Cla* I–*Bam*HI fragment. As an example, Fig. 3 shows the result obtained for subclone 23, which allows one to read the order of the restriction sites relative to the unique *Pvu* I site.

The evaluation of the experiments revealed the complete order of the restriction sites for all three endonucleases. A complete digest of the subclones with the same enzymes as well as those previously used (Fig. 2) in single and double combinations confirmed the number and sizes of the resulting fragments. The final map is shown in Fig. 4.

Localization of the Regions Hybridizing to the IFN-α cDNA. The cosmid and its subclones 23 and 49 were digested with the same enzymes used above for restriction mapping (Figs. 1 and 4). Southern blots were prepared and probed with the IFN-α cDNA. Probing the *Eco*RI and *Hin*dIII fragments from subclone 49 revealed for both sets of fragments a unique band (Fig. 5). The sizes of these fragments [2.1 kilobases (kb) for the *Eco*RI fragment and 5.7 kb for the *Hin*dIII fragment] showed them to be between *Eco*RI sites 2 and 3 and *Hin*dIII sites 1 and 2, respectively (Fig. 4). This is confirmed by the fact that a *Hin*dIII/*Hpa* I double digest revealed hybridization of a 3.2-kb fragment. The results allowed the localization of the IFN-α region within 1.8 kb at the indicated position (Fig. 4).

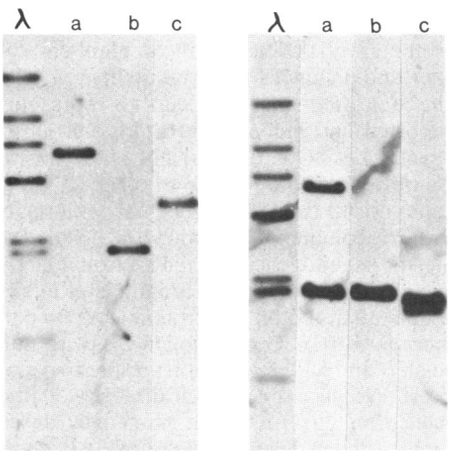

FIG. 5. Identification of restriction fragments of subclones 23 and 49. Subclone 49 (*Left*) was digested with the endonucleases *Hin*dIII (lane a), *Eco*RI (lane b), and *Hin*dIII/*Hpa* I (lane c). Subclone 23 (*Right*) was digested with *Eco*RI (lane a), *Eco*RI/*Bam*HI (lane b), and *Eco*RI/*Hin*dIII (lane c). Southern blot analysis was performed with the IFN-α cDNA as a probe, and radiolabeled *Hin*dIII fragments of phage λ DNA were used as size markers (lanes λ).

Probing the *Eco*RI fragments from subclone 23 revealed a small fragment of about 2 kb and a large fragment of about 6 kb. The size of the small fragment indicated that it lay either between *Eco*RI sites 5 and 6 or 8 and 9 (Fig. 4). The large fragment corresponded either to that between *Eco*RI sites 6 and 7 or to the hybrid fragment beginning at the *Eco*RI site 9—i.e., extending into the vector (Fig. 4).

Reprobing the fragments after digesting the subclone 23 with *Eco*RI and *Bam*HI allowed us to assign to the latter location a region homologous to IFN-α. Finally, probing the fragments obtained by cleaving the subclone with *Eco*RI and *Hin*dIII made it possible to locate the other region of homology between *Eco*RI site 5 and *Hin*dIII site 5. The same experiment refines the position of the homologous region mentioned above between *Eco*RI site 9 and *Hin*dIII site 10.

The data were independently confirmed by repeating experiments with the entire cosmid DNA. Only the central homologous region carries a *Bgl* II site, which explains why four fragments were detected when the *Bgl* II fragments were probed (Fig. 1).

Partial Sequence Analysis of IFN-α Regions. The identification of three regions on the insert binding the cDNA probe established a sequence homology to IFN-α, although the nature of the homology was not known. Therefore, we determined the sequence of a diagnostic stretch of nucleotides. As the regions of homology are limited to three *Eco*RI–*Hin*dIII fragments along the cosmid (Fig. 4), we decided to reclone these fragments into the corresponding restriction sites of the cloning vehicles M13 mp9 and mp8 (16), designed for sequence analysis as described by Sanger *et al.* (18).

The desired recombinant phages were identified by plaque hybridization to a synthetic oligonucleotide probe. This oligonucleotide (5′ C-A-A-C-C-T-C-C-C-A-G-G-C-A-C 3′) was designed to complement 15 nucleotides found to be a conserved region (positions 486–500 in ref. 3) of the majority of the IFN-α sequences characterized so far.

Twelve phages were isolated. Their inserts were screened with dideoxythymidine reactions as described by Sanger *et al.*, using the 15-mer as a primer. The DNA sequence obtained fell into three different classes, and one representative of each class was further analyzed by sequence determination. The sequence from one strand was obtained by using the synthetic 15-mer as a primer and elongating it by approximately 200 nucleotides.

The result (Fig. 6) demonstrates that, although the nucleotide sequences were very similar to segments published previously (3, 5, 20), in some positions amino acids occurred that have not been described previously in published IFN-α

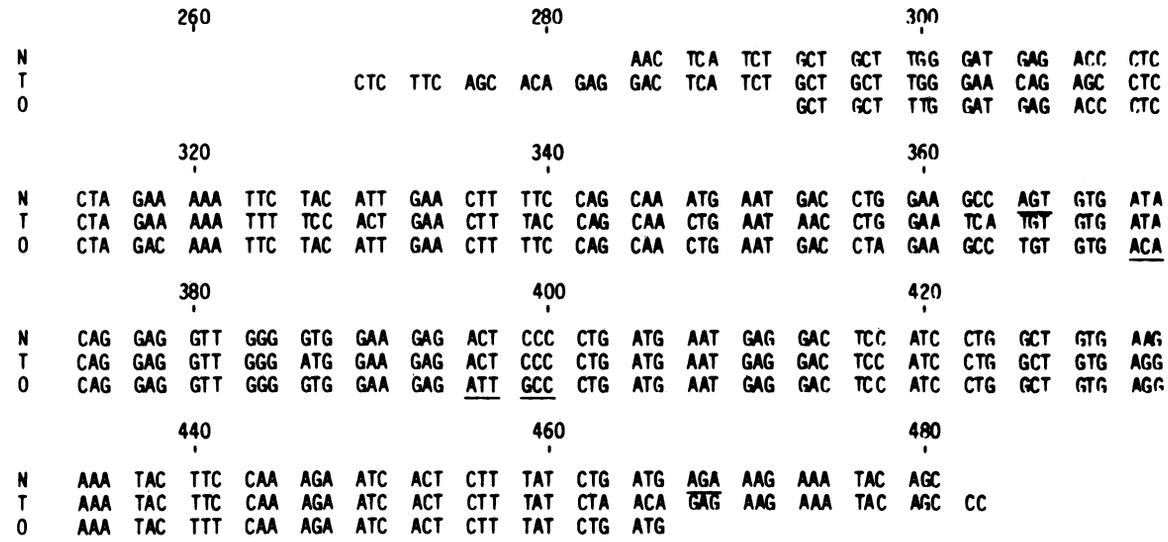

FIG. 6. Partial sequence of IFN-α gene-like sequences *O*, *T*, and *N*. The indicated positions start from the predicted translational start as proposed by Goeddel *et al.* (3), and the underlined codons represent predicted amino acids not previously demonstrated to occur in IFN-α (refs. 3 and 20).

gene sequences. These are underlined in Fig. 6. Thus, *N* has two and *O* has three such codons. Comparing them with previously described cDNA sequences revealed the following differences. The *N* and *O* sequences had the highest degree of similarity to IFN-α *H* (3), but differed in two and seven amino acid residues, respectively. Sequence of IFN-α gene *T* was to some extent similar to IFN-α *C* (3), differing in five amino acid residues. However, comparing them to recently characterized chromosomal IFN genes showed that *T* sequence is almost identical to the genomic sequence of phage λ 2c1 (20), with only one nucleotide difference changing alanine into a serine in position 98. The *N* sequence differs from the IFN-α-like sequence λ 2h (20) only with respect to two amino acids.

In order to determine the orientation of the three genes, the M13 universal primer was utilized to prime a second strand DNA synthesis reaction. This allowed us to reidentify all three genes as subclones of M13 mp9. As all the subclones hybridized to the strand-specific 15-mer, it was possible to predict the transcriptional direction for all three IFN-α genes relative to the vector part (Fig. 4). Determination of the sizes of the *Eco*RI–*Hin*dIII inserts in the M13 subclones and identification of an internal *Bgl* II site in one of them (Fig. 2) made it possible to assign them to the three regions mapping on the cosmid clone as outlined in Fig. 4.

DISCUSSION

The present paper introduces a further section of the human genome carrying three IFN-α gene-like sequences. The identified region differs from those described previously (4, 5, 20) when the following three criteria are simultaneously taken into account: (*i*) the partial sequences established for the three potential coding regions; (*ii*) the intergenic distances, and (*iii*) the pattern of the restriction sites.

The three IFN-α sequences revealed in the chromosomal DNA cosmid insert followed the same pattern of relatedness as the other members of the IFN-α genes analyzed by sequence so far (3), with a homology of 85–95% at the nucleotide level. Comparing the amino acid sequence predicted from the 200-bp nucleotide sequences of IFN-α *T*, *N*, and *O*, the following relationships were found: IFN-α *T* is most similar to IFN-α *C* (3) and phage λ 2c1 (20), while IFN-α *N* and IFN-α *O* are most similar to IFN-α *H* and λ 2h.

When one compares IFN-α *H* with IFN-α *N*, a change is seen of a cysteine to a serine, which could have structural impact. However, until the complete sequences are known, these differences are only minimal estimates, and the possibility that the three sequences represent pseudogenes cannot be ruled out.

The regions homologous to IFN-α can be limited within three restriction fragments $\approx$2 kb long, each bounded by an *Eco*RI and *Hin*dIII site. IFN-α genes have an average size of about 1 kb. Assuming that the three IFN-α-like sequences represent genuine IFN-α genes or pseudogenes allows us to map the putative genes on the cosmid DNA with a maximum error of 1 kb. The average distances between the three genes are about 10 kb. Comparison of the intergenic distances with those found for other linkage groups carrying three IFN-α genes (4, 5, 20) reveals that this chromosomal section has not been identified before. However, the genomic region described by Lawn *et al.* (20), carrying two IFN-α genes, has an intergenic distance in the same range, and the partial sequences of *N* and *T* showed striking similarities with those presented by these authors. The restriction map established for *Eco*RI might be identical to that we found between *N* and *T*, but differs outside this region. On the other hand, only a fraction of the *Hin*dIII sites might be identical concerning both number and distances between them.

Thus, the identified region appears to represent a previously unreported and independent linkage group of IFN-α gene-like sequences. Our data do not allow positioning of this section of the genome relative to other IFN-α gene-carrying regions, nor to the IFN-β gene-carrying region.

The literature contains descriptions of up to 20 different IFN-α genes or pseudogenes when data from cDNA and genomic sequences are taken into account. Still more variants have been demonstrated on the protein level by partial amino sequence data (23). Gene duplication with subsequent amino acid shift and allelic variation are the two main mechanisms considered to explain such diversity. Although the reported cluster is different from previously described regions, it shows similarities with one previously described region (20). We cannot from our data judge if the differences found in the restriction map and in the partially determined gene-like sequences could represent allelic variation—i.e., genetic polymorphism. This possibility can only be resolved by conventional genetic approaches for studies on gene segregation.

But it should also be kept in mind that the region described

here could as well be due to a recent duplication of the segment described by Lawn *et al.* (20). Then the homologies are of the same character as those demonstrated by Ullrich *et al.* (5).

It is not known how complete the different genomic libraries are, but the demonstration of new clusters of IFN-α gene sequences and also new variants of IFN-α nucleotide sequences makes it reasonable to assume that still more variants will be found. The evolvement of such a complexity of the IFN-α gene family in humans seems to be a rather recent event (24). As differences in biological properties are known to occur for the different subtypes (6), at least some of the variation on the nucleotide level is not neutral and could be the target for selective forces in evolution. This suggestion raises interesting questions about a similar heterogeneity on the IFN receptor level and points to the importance of expressing the different IFN-α genes for studies of the biological properties of the different subtypes.

This work was supported by grants from KabiGen AB, Stockholm, Sweden. B.L. was supported by a European Molecular Biology Organization short-term fellowship during his stay in Braunschweig–Stöckheim. Skillful technical assistance from Mona Andersson, Gisela Nilsson, and Helena Olsson and careful typewriting and preparation of manuscript by Elisabeth Ivarsson and Mell von Gabain are gratefully acknowledged.

1. Stewart, W. E., II, Blalock, J. E., Burke, D. C., Chany, C., Dunnick, J. K., Falcoff, E., Friedman, R. M., Galasso, G. J., Joklik, W. K., Vilcek, J. T., Youngner, J. S. & Zoon, K. C. (1980) *J. Interferon Res.* **1,** vi–vii.
2. Taniguchi, T., Mantei, N., Schwarzstein, M., Nagata, S., Muramatsu, M. & Weissmann, C. (1980) *Nature (London)* **285,** 547–549.
3. Goeddel, D. V., Leung, D. W., Dull, T. J., Gross, M., Lawn, R. M., McCandliss, R., Seeburg, P. H., Ullrich, A., Yelverton, E. & Gray, P. W. (1981) *Nature (London)* **290,** 20–26.
4. Brack, C., Nagata, S., Mantei, N. & Weissmann, C. (1981) *Gene* **15,** 379–394.
5. Ullrich, A., Gray, A., Goeddel, D. V. & Dull, T. J. (1982) *J. Mol. Biol.* **156,** 467–486.
6. Alton, K., Stabinsky, Y., Richards, R., Ferguson, B., Goldstein, L., Altrock, B., Miller, L. & Stebbing, N. (1983) in *The Biology of the Interferon System 1983*, eds. De Maeyer, E. & Schellekens, H. (Elsevier, Amsterdam), pp. 119–128.
7. Lindenmaier, W., Hauser, H., Greiser de Wilke, I. & Schütz, G. (1982) *Nucleic Acids Res.* **10,** 1243–1256.
8. Boyer, H. W. & Roulland-Dussoix, D. (1969) *J. Mol. Biol.* **41,** 459–472.
9. Gross, G., Mayr, U., Grossveld, F., Dahl, H. M., Flavell, R. A. & Collins, J. (1981) in *Molecular Biology, Pathogenicity and Ecology of Bacterial Plasmids*, eds. Levy, S. B., Clowes, R. C. & Koenig, E. L. (Plenum, New York), pp. 429–438.
10. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
11. Moseley, S. L., Huq, I., Alim, A. R. M., So, M., Samadpour-Motalebi, M. & Falkow, S. (1980) *J. Infect. Dis.* **142,** 892–898.
12. Benton, W. D. & Davies, R. W. (1977) *Science* **196,** 180–182.
13. Wallace, R. B., Schaffer, J., Murphy, R. F., Bonner, J., Hirose, T. & Itakura, K. (1979) *Nucleic Acids Res.* **6,** 3543–3557.
14. Wallace, R. B., Johnson, M. J., Hirose, T., Miyake, T., Kawashima, E. H. & Itakura, K. (1981) *Nucleic Acids Res.* **9,** 879–894.
15. Chow, F., Kempe, T. & Palm, G. (1981) *Nucleic Acids Res.* **9,** 2807–2817.
16. Messing, J. & Vieira, F. (1982) *Gene* **19,** 269–276.
17. Messing, J., Crea, R. & Seeburg, P. H. (1981) *Nucleic Acids Res.* **9,** 309–321.
18. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
19. McDonell, M. W., Simon, M. N. & Studier, F. W. (1977) *J. Mol. Biol.* **110,** 119–146.
20. Lawn, R. N., Adelman, F., Dull, T., Gross, M., Goeddel, D. & Ullrich, A. (1981) *Science* **212,** 1159–1162.
21. Smith, H. O. & Birnstiel, M. L. (1976) *Nucleic Acids Res.* **3,** 2387–2398.
22. Sutcliffe, J. G. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **43,** 77–90.
23. Allen, G. & Fantes, K. H. (1980) *Nature (London)* **287,** 408–411.
24. Weissmann, C. (1981) in *Interferon 3*, ed. Gresser, I. (Academic, New York), pp. 101–134.