MOLECULAR AND CELLULAR BIOLOGY, Apr. 1987, p. 1559–1562
0270-7306/87/041559-04$02.00/0

Vol. 7, No. 4

Human Retroviral Sequences on the Y Chromosome

JONATHAN SILVER,* ARNOLD RABSON, THEODORE BRYAN, RONALD WILLEY, AND MALCOLM A. MARTIN

Laboratory of Molecular Microbiology, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland 20892

Received 24 October 1986/Accepted 12 January 1987

Novel endogenous human retroviral sequences were cloned by low-stringency hybridization, using the *pol* gene of endogenous human retrovirus 51-1. One clone, λNP-2, contained *gag*, *pol*, *env*, and long terminal repeat sequences related to the corresponding portions of clone 51-1 and the closely related full-length endogenous human retrovirus 4-1. The sequence of the *env* gene of NP-2 was 73% homologous to that of 4-1. Genomic Southern blots of male and female DNAs showed that NP-2 is located on the Y chromosome and that the Y chromosome also contains one other sequence closely related to the *env* and 3′ flanking regions of NP-2. Conservation of flanking DNA suggests that the second Y chromosome copy of the NP-2 *env* sequence arose by gene duplication rather than provirus insertion.

Several different endogenous retroviral sequences have been identified in the human genome (1, 3, 4, 11–14, 18, 19, 25). To look for new classes of endogenous human retroviruses, we screened a recombinant bacteriophage human DNA library (9) with a *pol* probe from the well-characterized human endogenous retroviral sequence 51-1 (18, 19). A *pol* probe was used because this retroviral gene is highly conserved among different retroviral families (18, 19). Recombinant bacteriophage clones were selected which hybridized at low stringency (membranes washed at 55°C in 6× SSC [1× SSC is 150 mM NaCl plus 15 mM sodium citrate]) but not at high stringency (membranes washed at 68°C in 0.1× SSC). At least 5 of 12 clones so selected appeared to have the structure of partial or complete retroviruses, since they also hybridized at low stringency to *gag*, *env*, or long terminal repeat (LTR) probes from human retroviral clones 51-1 and 4-1 (18, 25), and the relative positions of restriction fragments hybridizing to these probes were consistent, except for a few instances of apparent deletions, with the gene order LTR-*gag*-*pol*-*env*-LTR. One of these retroviral clones, λNP-2, was analyzed further because preliminary experiments indicated that the human population was highly polymorphic for NP-2-related sequences.

A restriction map of the insert in clone λNP-2 is shown in Fig. 1. The position and approximate extent of *gag*, *pol*, *env*, and LTR sequences were determined by low-stringency hybridization with subregion probes from retroviral clones 51-1 and 4-1 (18, 25). To confirm that these cross-hybridization reactions were indicative of extensive sequence homology, a 1.1-kilobase (kb) *Hin*dIII fragment from the *env* region of NP-2 (fragment a, Fig. 1) was sequenced. Overall, the sequence of this fragment was 73% homologous to that of the corresponding portion of 4-1 (Fig. 2). A small region of marked divergence (only 45% homology) was noted about one-third of the way into the *env* coding sequence of 4-1 (coordinates 6930 to 7050, Fig. 2). This region corresponds roughly to regions of hypervariability determining host range in replication-competent murine and avian retroviruses (6, 7, 15, 16). The fact that 4-1 and NP-2 differ more in this region than in other regions of the *env* gene suggests that 4-1 and NP-2 are derived from what were, at some time in the past, functional retroviruses.

When the 1.1-kb *Hin*dIII *env* fragment was used in blot hybridization analysis of human DNA under high-stringency conditions (washing at 65°C in 0.1× SSC), two sequences were detected in DNA from males, whereas DNA from females gave no, or much weaker, reactivity (Fig. 3). For example, the *env* probe detected *Bam*HI fragments of 4.9 and 0.9 kb, corresponding to the expected fragments from NP-2 (Fig. 1), and an additional *Bam*HI fragment of 9.6 kb, presumably arising from a second sequence closely related to the NP-2 *env* gene (Fig. 3, lane 6). Digestion with *Hin*dIII gave the expected 1.1-kb fragment from NP-2, and two other *env*-reactive fragments presumably because the additional NP-2 *env*-related sequence had a *Hin*dIII site within it (Fig. 3, lane 9). Digestion with *Eco*RI gave three *env*-reactive fragments of 5.8, 8.2, and 9.6 kb (Fig. 3, lanes 1 and 2). Since the DNA library from which NP-2 was cloned was constructed by partial digestion of human male DNA with *Alu*I and *Hae*III and insertion of *Eco*RI linkers (9), the *Eco*RI site forming the boundary of the NP-2 insert is an artifact of cloning. Therefore, the genomic *Eco*RI fragment containing NP-2 *env* must be larger than 7.2 kb, the size of the cloned NP-2 *Eco*RI *env* fragment (Fig. 1). Either the 8.2- or 9.6-kb *env*-reactive *Eco*RI fragment could correspond to the genomic NP-2, whereas the other fragment of this pair and the 5.8-kb fragment presumably arose from an *Eco*RI site within the NP-2 *env* homolog.

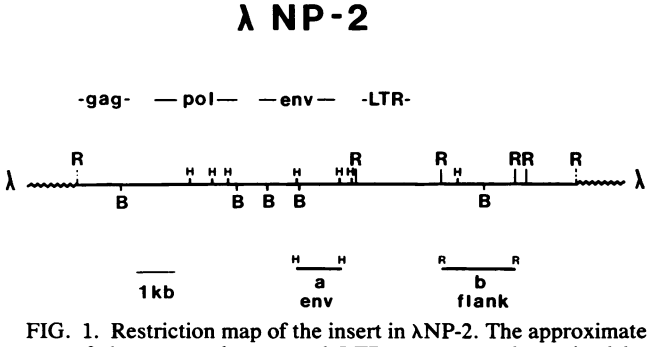

FIG. 1. Restriction map of the insert in λNP-2. The approximate extent of the *gag*, *pol*, *env*, and LTR genes was determined by low-stringency hybridization (membranes washed at 55°C in 6× SSC) of restricted DNA to retroviral probes from clones 4-1 and 51-1 (18, 25). R, *Eco*RI; B, *Bam*HI; H, *Hin*dIII.

* Corresponding author.

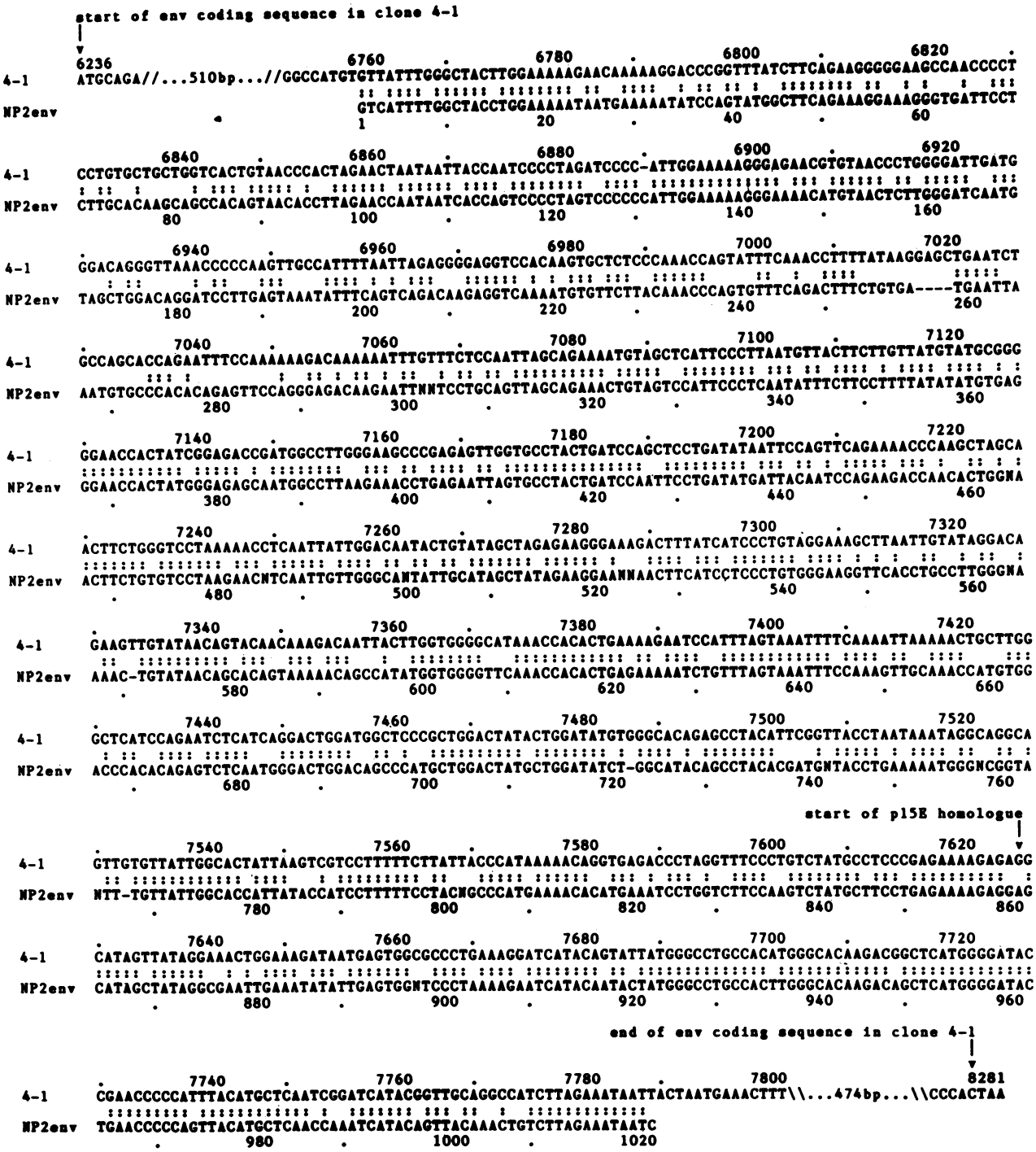

FIG. 2. Sequence comparison of a portion of the *env* genes from clones 4-1 and NP-2. The NP-2 sequence was determined by the dideoxy technique (20). Bases 1 to 370 and 652 to 1020 were sequenced in one strand only. The 4-1 sequence is from a previous study (19). Sequence alignment was performed by using NUCALN (28). The overall homology is 744 of 1,023 base pairs.

Both the NP-2 *env* gene and its homolog were detected in DNA from 25 of 26 males tested but not in DNA from any of the 20 females tested. This indicates that NP-2 and its homolog are almost certainly on the Y chromosome. The one aberrant male DNA which did not contain *env*-reactive fragments came from a cell line from a 63-year-old patient with Friedreich's ataxia. It is not known whether this cell line has lost the Y chromosome. To confirm the assignment of NP-2 and the NP-2 *env* homolog to the Y chromosome, we tested DNA from a human-Chinese hamster somatic cell hybrid in which the only human chromosome cytologically detectable is the Y chromosome (2, 21). Both sequences were detected in DNA from this hybrid (Fig. 3, lane 8) but were absent in hamster DNA (Fig. 3, lane 7). It should be noted that this somatic cell hybrid expresses a few human isozymes, indicating that it must retain a small amount of

DNA from human chromosome 14 and possibly chromosomes 4 and 12 (2, 21). However, in view of the independent strong correlation of NP-2 sequences with sex, we conclude with near certainty that both NP-2 and the NP-2 *env* homolog are on the Y chromosome.

To learn more about NP-2 and its homolog, we used hybridization probes derived from the region to the right of the NP-2 *env* gene. The 2.3-kb *Eco*RI fragment located just 3′ to NP-2 *env* gave a smear when hybridized to genomic DNA and was found to contain *Alu* sequences. The next *Eco*RI fragment (labeled b and flank in Fig. 1) was free of highly repetitive sequences. In Southern blots of *Bam*HI-digested DNA, this fragment reacted with prominent 4.9- and 9.6-kb fragments present in DNA from males but not females (Fig. 4, lanes 1 and 2). A 4.9-kb *env*- and flank-reactive fragment was expected from the NP-2 clone (Fig. 1). The 9.6-kb fragment presumably represented a second Y chromosome copy of the flank sequence. The NP-2 *env* homolog also was contained in a 9.6-kb *Bam*HI fragment (Fig. 4, lanes 3 and 4). If this was the same 9.6-kb fragment, then the sequence flanking the second Y chromosome copy of the NP-2 *env* sequence is closely related to the sequence flanking the cloned NP-2 provirus. Similar results were obtained with *Hin*dIII, i.e., the NP-2 *env* and flanking probes both hybridized to a 7.8-kb fragment believed to be derived from the second NP-2 sequence on the Y chromosome (Fig. 4, lanes 5 through 8). Digestion with *Eco*RI, *Eco*RV, and *Hpa*I also resulted in fragments, believed to be derived from the second NP-2 sequence on the Y chromosome, which hybridized to both the NP-2 *env* and flanking probes (data not shown). We conclude that the sequence flanking the NP-2 provirus is very closely related to the sequence flanking the NP-2 *env* homolog.

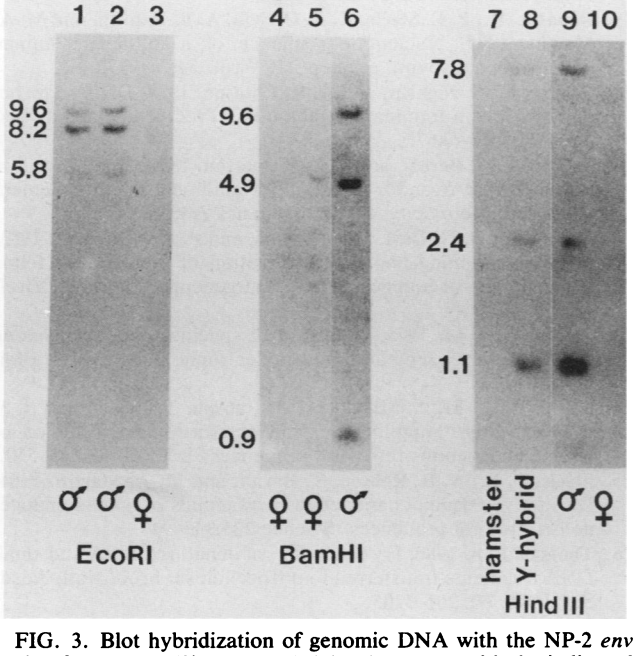

FIG. 3. Blot hybridization of genomic DNA with the NP-2 *env* probe (fragment a, Fig. 1). DNA (10 μg) was cut with the indicated restriction enzyme, electrophoresed in 0.6% agarose, transferred to nitrocellulose, hybridized to ^{32}P-labeled probe in the presence of dextran sulfate, and washed in 0.1× SSC–0.1% sodium dodecyl sulfate at 65°C (8, 23, 26). Lanes 1 through 6, 9, and 10, DNA from different individuals; lane 7, DNA from Chinese hamster; lane 8, DNA from the Y chromosome-containing somatic cell hybrid 7631 (see the text). Fragment sizes in kilobases are indicated in the margins.

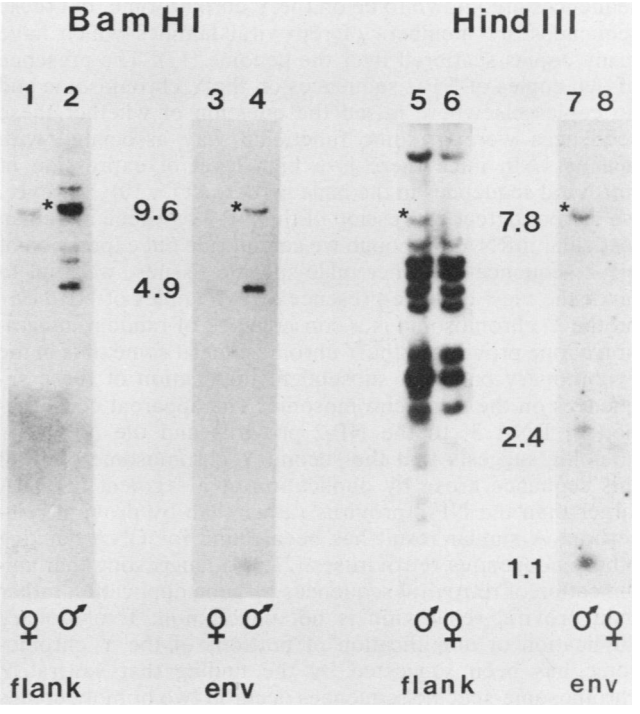

FIG. 4. Southern blot hybridization of genomic DNA with NP-2 flank and *env* probes (fragments b and a, respectively, in Fig. 1). Blot hybridizations were done as described in the legend to Fig. 3 except transfers were done bidirectionally and the membrane for lanes 5 and 6 was washed at 57°C, whereas the other membranes were washed at 65°C. Lanes 1, 2, 5, and 6, DNA from different individuals, hybridized to the flank probe; lanes 3, 4, 7, and 8, sister membranes hybridized to the *env* probe. The larger number of fragments detected by the flank probe in lanes 5 and 6 than in lanes 1 and 2 may have been the result of washing at lower stringency. *, Comigrating fragments detected with the flank and *env* probes, believed to be derived from the NP-2 *env* homolog (see the text). Fragment sizes in kilobases are indicated in the margins.

The NP-2 flanking probe reacted with several male-specific restriction fragments in addition to the fragment containing the NP-2 *env* homolog (Fig. 4, lanes 1, 2, 5, and 6). This suggests that the Y chromosome contains more than two copies of the flank sequence. In addition, the NP-2 flanking probe reacted with several restriction fragments present in females as well as males (Fig. 4, lanes 1, 2, 5, and 6), indicating that homologous sequences are located on chromosomes other than the Y chromosome.

We were interested in knowing whether the NP-2 provirus was transcribed or whether it affected transcription of adjacent sequences. We did not detect mRNA homologous to the NP-2 *env* or flanking probes in Northern blots of poly(A)$^+$ RNA from a normal human testis (removed for therapy of prostate carcinoma) or from lung, colon, or pancreatic carcinomas from human males. The mRNA samples from the last three sources contained sequences which reacted with a clone 4-1 *env* probe (data not shown).

The NP-2 "family" of endogenous human retroviral sequences is unusual in that it contains only two members, both located on the Y chromosome. To our knowledge, no other human retroviral sequences have been mapped to the Y chromosome (3, 4, 11, 13, 14, 24). In mice some retroviral

sequences are known to be on the Y chromosome, but these sequences are members of retroviral families which have many copies scattered over the genome (17). The presence of two copies of NP-2 sequences on the Y chromosome and no copies elsewhere raised the question of whether these sequences were in some functional way associated with maleness. In mice there is a high level of expression of retroviral sequences in the male germ tract (5, 10); however, we did not detect expression of the NP-2 *env* gene in human testicular mRNA. Although we cannot rule out expression of NP-2 sequences in other male-specific tissues, we tend to favor the view that the presence of two copies of NP-2 *env* on the Y chromosome is a consequence of random integration of one provirus in the Y chromosome at some time in the evolutionary past and subsequent duplication of these sequences on the same chromosome. The apparent conservation of DNA 3′ to the NP-2 provirus and the NP-2 *env* homolog suggests that the second Y chromosome copy of this sequence arose by duplication of a segment of DNA larger than the NP-2 provirus rather than by proviral reinsertion. A similar result has been found for DNA flanking other endogenous retroviruses (22, 24), suggesting that amplification of retroviral sequences by gene duplication rather than proviral reinsertion is not uncommon. Evolutionary duplication or amplification of portions of the Y chromosome has been suggested by the finding that several Y chromosome-specific sequences occur in two or more copies in different regions on the Y chromosome (27). Comparative study of NP-2-related sequences and flanking DNA in primates may provide insights into the evolutionary history of the Y chromosome.

We thank R. Cotton and D. Goldstein for providing normal human DNAs, K. Smith for a gift of DNA from the somatic cell hybrid 7631, M. Cohen for a gift of human testicular mRNA, C. Buckler for help with the sequence analysis, R. Rutledge and J. Chandler for technical assistance, and S. Rosenfeld for editorial assistance.

LITERATURE CITED

1. **Bonner, T. I., C. O'Connell, and M. Cohen.** 1982. Cloned endogenous retroviral sequences from human DNA. Proc. Natl. Acad. Sci. USA **79**:4709–4713.
2. **Burk, R. D., P. Ma, and K. D. Smith.** 1985. Characterization and evolution of a single-copy sequence from the human Y chromosome. Mol. Cell. Biol. **5**:576–581.
3. **Callahan, R., I. M. Chiu, J. F. H. Wong, S. R. Tronick, B. A. Roe, S. A. Aaronson, and J. Schlom.** 1985. A new class of endogenous human retroviral genomes. Science **228**:1208–1211.
4. **Callahan, R., W. Drohan, S. Tronick, and J. Schlom.** 1982. Detection and cloning of human DNA sequences related to the mouse mammary tumor virus genome. Proc. Natl. Acad. Sci. USA **79**:5503–5507.
5. **Del Villano, B. C., and R. A. Lerner.** 1976. Relationship between the oncornavirus gene product gp70 and major protein secretion of the mouse genital tract. Nature (London) **259**:497–499.
6. **Dorner, A. J., and J. M. Coffin.** 1986. Determinants for receptor interaction and cell killing on the avian retrovirus glycoprotein gp85. Cell **45**:365–374.
7. **Dorner, A. J., J. P. Stoye, and J. M. Coffin.** 1985. Molecular basis of host range variation in avian retroviruses. J. Virol. **53**:32–39.
8. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**:6–13.
9. **Lawn, R. M., E. F. Fritsch, R. C. Parker, G. Blake, and T. Maniatis.** 1978. The isolation and characterization of linked δ- and β-globin genes from a cloned library of human DNA. Cell **15**:1157–1174.
10. **Lerner, R. A., C. B. Wilson, B. C. Del Villano, P. J. McConahey, and F. J. Dixon.** 1976. Endogenous oncoviral gene expression in adult and fetal mice: quantitative histologic and physiologic studies of the major viral glycoprotein, gp 70. J. Exp. Med. **143**:151–166.
11. **Mager, D. L., and P. S. Henthorn.** 1984. Identification of a retrovirus-like repetitive element in human DNA. Proc. Natl. Acad. Sci. USA **81**:7510–7514.
12. **Noda, M., M. Kurihara, and T. Takano.** 1982. Retrovirus-related sequences in human DNA: detection and cloning of sequences which hybridize with the long terminal repeat of baboon endogenous virus. Nucleic Acids Res. **10**:2865–2878.
13. **O'Brien, S. J., T. I. Bonner, M. Cohen, C. O'Connell, and W. G. Nash.** 1983. Mapping of an endogenous retroviral sequence to human chromosome 18. Nature (London) **303**:74–77.
14. **O'Connell, C., S. O'Brien, W. G. Nash, and M. Cohen.** 1984. ERV3, a full-length human endogenous retrovirus: chromosomal localization and evolutionary relationships. Virology **138**:225–235.
15. **O'Neill, R. R., C. E. Buckler, T. S. Theodore, M. A. Martin, and R. Repaske.** 1985. Envelope and long terminal repeat sequences of a cloned infectious NZB xenotropic murine leukemia virus. J. Virol. **53**:100–106.
16. **O'Neill, R. R., A. S. Khan, M. D. Hoggan, J. W. Hartley, M. A. Martin, and R. Repaske.** 1986. Specific hybridization probes demonstrate fewer xenotropic than mink cell focus-forming murine leukemia virus *env*-related sequences in DNAs from inbred laboratory mice. J. Virol. **58**:359–366.
17. **Phillips, S. J., E. H. Birkenmeier, R. Callahan, and E. M. Eicher.** 1982. Male and female mouse DNAs can be discriminated using retroviral probes. Nature (London) **297**:241–243.
18. **Repaske, R., R. R. O'Neill, P. E. Steele, and M. A. Martin.** 1983. Characterization of partial nucleotide sequence of endogenous type C retrovirus segments in human chromosomal DNA. Proc. Natl. Acad. Sci. USA **80**:678–682.
19. **Repaske, R., P. E. Steele, R. R. O'Neill, A. B. Rabson, and M. A. Martin.** 1985. Nucleotide sequence of a full-length human endogenous retroviral segment. J. Virol. **54**:764–772.
20. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
21. **Shalev, A., I. Berozi, and J. L. Hamerton.** 1977. Demonstration of fluorescent X and Y bodies after antibody and complement mediated cytotoxicity. Immunogenetics **5**:405–414.
22. **Soe, L. H., B. G. Devi, J. I. Mullins, and P. Roy-Burman.** 1983. Molecular cloning and characterization of endogenous feline leukemia virus sequences from a cat genomic library. J. Virol. **46**:829–840.
23. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
24. **Steele, P. E., M. A. Martin, A. B. Rabson, T. Bryan, and S. J. O'Brien.** 1986. Amplification and chromosomal dispersion of human endogenous retroviral sequences. J. Virol. **59**:545–550.
25. **Steele, P. E., A. B. Rabson, T. Bryan, and M. A. Martin.** 1984. Distinctive termini characterize two families of human endogenous retroviral sequences. Science **225**:943–947.
26. **Thomas, P. S.** 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA **77**:5201–5205.
27. **Vergnaud, G., D. C. Page, M. C. Simmler, L. Brown, F. Rouyer, B. Noel, D. Botstein, A. De La Chapelle, and J. Weissenbach.** 1986. A deletion map of the human Y chromosome based on DNA hybridization. Am. J. Hum. Genet. **38**:109–124.
28. **Wilbur, W. J., and D. J. Lipman.** 1983. Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci. USA **80**:726–730.