

Repeat array in Epstein–Barr virus DNA is related to cell DNA sequences interspersed on human chromosomes

(cytological hybridization/Southern blot analysis/repetitive DNA/DNA·DNA hybridization)

MARK HELLER*, ANN HENDERSON†, AND ELLIOTT KIEFF*

*Departments of Medicine, Microbiology, and Virology, Kovler Viral Oncology Laboratory, University of Chicago, 910 East 58th Street, Chicago, Illinois 60637; and
†Department of Human Genetics and Development, College of Physicians and Surgeons, Columbia University, 701 West 168th Street, New York, New York 10032

Communicated by Albert Dorfman, June 28, 1982

ABSTRACT The third internal repeat (IR3) simple repeat array in Epstein–Barr virus (EBV) DNA has a high degree of homology to a reiterated component of cell DNAs. ³²P-labeled human or mouse DNAs hybridize to the IR3 sequence on Southern blots of viral DNA. EBV IR3 probe identifies many restriction enzyme fragments on Southern blots of human and mouse DNAs that have extensive homology to IR3. Cytological hybridization shows that IR3 is homologous to at least one region on each human chromosome except the Y chromosome.

Epstein–Barr virus (EBV) is the causative agent of infectious mononucleosis and is regularly found in a latent state in tumor cells of two human cancers: African Burkitt lymphoma and anaplastic nasopharyngeal carcinoma (1). Latent virus infection stimulates the growth of lymphocytes *in vitro* (2). Latently infected lymphocytes usually contain multiple copies of the viral genome (3). Most of the intracellular viral DNA is in the form of episomal closed circles (4, 5). Viral DNA is replicated concomitantly with cell DNA by cellular enzymes (6, 7). Some EBV DNA may integrate into cell DNA (8, 9). Expression of the viral genome is tightly regulated in latently infected growth-transformed cells. Three regions of the EBV genome express RNAs (10–13). Two intranuclear and one surface antigen have been detected (14–17). The ability of the EBV and cell genomes to exist and function together led us to investigate whether EBV DNA has regions that are similar to cell DNA.

We find that cell DNA has specific homology to the third internal repeat (IR3) in EBV DNA. EBV IR3 is an unusual sequence, composed of three triplets, GGG, GCA, and GGA, that are organized into three direct repeat elements: a hexanucleotide G-C-A-G-G-A and two nonanucleotides, G-C-A-G-G-A-G-G-A and G-G-G-G-C-A-G-G-A (ref. 18; Fig. 1). IR3 encodes part of a cytoplasmic polyadenylylated viral RNA detected in a latently infected growth-transformed cell line and part of two cytoplasmic polyadenylylated RNAs expressed in productively infected cells.

MATERIALS AND METHODS

Cells, Virus, and Recombinant Plasmids. B95-8, AG876, W91, and P3HR-1 are EBV-producing cell lines (25, 26). Loukes is an EBV genome-negative human B-lymphocyte cell line cultured from an American Burkitt tumor (13). Namalwa is an African Burkitt tumor cell line latently infected with EBV (10). Mouse L cells were obtained from Saul Kit. pRB103 contains the thymidine kinase (*tk*) gene *Bam*HI fragment of herpes simplex virus type 1 (HSV-1) DNA (gift of Bernard Roizman). Recombinants pDK14 (*Bam*HI V), pDK286 (*Bam*HI H), pDK39 (*Bam*HI L), pDK225 (*Bam*HI K), and pDK153 [*Sal* I F; 62–67

megadaltons (MDal) in Fig. 1a] contain the indicated B95-8 EBV DNA fragments (Fig. 1a; refs. 19 and 20). Cosmid *Eco*RI B/MUA3 contains the W91 *Eco*RI B fragment (21).

Preparation of Cellular DNA. High molecular weight DNA was isolated from human lymphocytes or mouse L cells by CsCl equilibrium sedimentation of NaDodSO₄/proteinase K cell lysates (22). Total genomic DNA was pooled from fractions with a density of 1.74–1.67 g/ml. The high-density fraction of Loukes or L-cell DNAs (1.734–1.705 g/ml) was isolated by two cycles of CsCl equilibrium sedimentation of cell lysates sheared by passage through an 18-gauge needle (average size, 10 kb).

Radioactive Labeling of DNA and Blot Hybridization. DNAs ³²P-labeled by nick-translation had a specific activity of 1–3 × 10⁸ dpm/μg (22). Nitrocellulose filters were incubated in 0.2% polyvinylpyrrolidone/0.2% Ficoll/0.2% bovine serum albumin (10× Denhardt's solution) at 65°C and then hybridized with ³²P-labeled probes in 1× Denhardt's solution as described in the figures (22). Hybridization mixtures contained salmon sperm carrier DNA at 50 μg/ml unless otherwise indicated.

Cytological Hybridization. Metaphase plate preparations were made from phytohemagglutinin-stimulated human lymphocytes from normal donors. Slides were treated with RNase, denatured in 0.15 M NaCl/0.015 M Na citrate, pH 7.4 (1× NaCl/Cit)/95% formamide at 72°C, hybridized to ¹²⁵I-labeled IR3 DNA (8 × 10⁸ dpm/μg) (27) in 3× NaCl/Cit/50% formamide at 42°C, washed in hybridization buffer at room temperature and in 2× NaCl/Cit at 45°C, and autoradiographed (28, 29).

RESULTS

A Region in EBV DNA Is Related to Cell DNA. Sequences in EBV DNA with homology to cell DNA were identified by hybridizing ³²P-labeled cellular DNAs from Loukes cells (an EBV-negative human B-lymphocyte line) or mouse L cells to Southern blots of EBV DNA (Fig. 2). Both probes hybridize to the *Eco*RI B fragment of the B95-8 isolate of EBV DNA (map location, 55–76 MDal; Fig. 1a). Labeled Loukes or L-cell probes hybridize only to the *Bam*HI K subfragment (63–66 MDal) of cloned *Eco*RI B fragment (Fig. 2c). Similar hybridizations with labeled mixtures of cell DNA and one copy of viral DNA per cell genome indicate that the homology to the *Eco*RI B and *Bam*HI K fragments is greater than that expected for one copy of EBV DNA per cell, since sequences from the rest of the viral genome, including the tandem internal direct repeats IR1 (23), IR2, and IR4 (24) cannot be detected in the reconstruction (data not shown). Hybridization of probes nick-translated from

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: IR3, internal repeat 3; EBV, Epstein–Barr virus; bp, base pair(s); HSV-1, herpes simplex virus type 1; HSV-1 *tk*, HSV-1 thymidine kinase gene; MDal, megadalton(s); kb, kilobase(s); 1× NaCl/Cit, 0.15 M NaCl/0.015 M Na citrate, pH 7.4; Ad2, adenovirus serotype 2; SV40, simian virus 40.

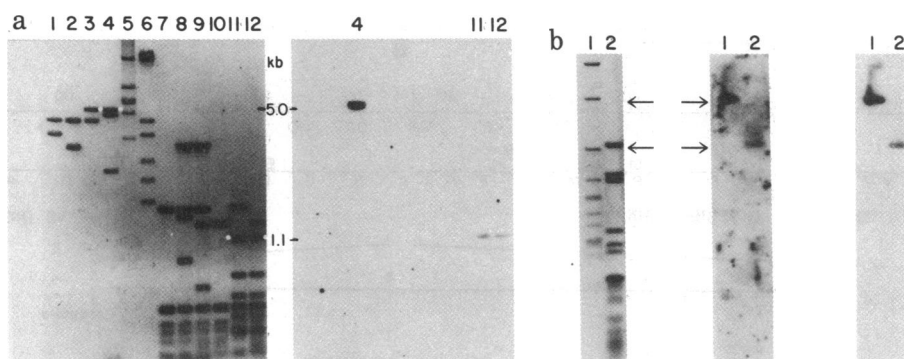


FIG. 3. Mapping the region in the *Bam*HI K fragment homologous to cell DNA. (a) Various DNAs were digested with the indicated restriction enzymes. Lanes: 1, pRB103 with *Bam*HI (HSV-1 *tk* gene); 2, pDK14 with *Bam*HI (*Bam*HI V fragment; IR1 sequence); 3, pDK39 with *Bam*HI (*Bam*HI L fragment); 4, pDK153 with *Bam*HI (*Sal* I F fragment; the *Bam*HI K fragment maps within the *Sal* I F fragment); 5, λ *Eco*RI; 6, Ad2 DNA with *Eco*RI; 7, pBR322 with *Hinf*I; 8, pDK286 with *Hinf*I (*Bam*HI H fragment; IR2 sequence); 9, pDK286 with *Bam*HI/*Hinf*I; 10, pBR322 with *Bam*HI/*Hinf*I; 11, pDK225 with *Hinf*I (*Bam*HI K fragment; IR3 repeat); 12, pDK225 with *Bam*HI/*Hinf*I. (Left) DNA fragments were separated on a 1% agarose gel, stained with ethidium bromide, and transferred to nitrocellulose. (Right) The blot was hybridized to Loukes probe (3×10^8 cpm) as described in Fig. 2. (b; Left) Ethidium bromide-stained 1.4% agarose gel of separated DNA fragments. Lanes: 1, pDK225 *Bam*HI K plasmid (1 μ g) was digested with *Hinf*I; 2, pDK225 (3 μ g) was digested with *Hinf*I/*Hpa* II. Arrows: *Hinf*I 1.15-kb fragment (upper) and *Hpa* II 0.77-kb fragment (lower). (Middle) Southern blot of the gel after hybridization to Loukes probe (2×10^8 cpm) as described in Fig. 2. (Right) The blot was further hybridized to IR3 probe (Fig. 1d).

of the cell DNA probe to other DNAs on the filters, including Ad2 DNA and DNAs with a high G+C content such as the HSV-1 *tk* gene (64%; ref. 31) and the IR1 (67%; ref. 23) and IR2 (84%; ref. 24) EBV repeat sequences.

The nucleotide sequence of the 1.15-kb pDK225 *Hinf*I fragment has been determined (18). The 735-bp *Ava* II/*Hpa* II subfragment that contains the region homologous to cell DNA is 74% G+C and consists almost entirely of the IR3 simple repeat array as well as 16 nucleotides to the left of IR3 and 11 nucleotides to the right of IR3 (Fig. 1d).

IR3 Is Homologous to a Reiterated Sequence in Cell DNA. Labeled IR3 DNA (Fig. 1d) was hybridized to nitrocellulose filters containing *Eco*RI and *Bam*HI restriction enzyme fragments of DNAs from Loukes and Namalwa human lymphocytes or mouse fibroblast L cells to identify related cellular sequences. IR3 is homologous to at least 25 fragments in these cell DNAs (Fig. 4a). IR3 probe hybridizes more extensively to

mouse DNA than to human DNA. Shorter autoradiographic exposure of the hybridization to mouse DNA (lanes 4 and 7) indicates that multiple discrete fragments are homologous to IR3. The hybridizations of IR3 to mouse and human DNAs are therefore similar in numbers of fragments identified but differ in intensity. These observations suggest that regions of IR3 homology are longer in mouse DNA, compared with human DNA, or that multiple IR3-homologous regions are clustered in mouse DNA. The intensity of hybridization of IR3 to mouse DNA fragments is similar to the extent of hybridization to a single copy of the pDK225 *Bam*HI K fragment per cell genome reconstruction (Fig. 4a) and a single copy of purified EBV DNA per cell genome reconstruction (Fig. 4c). Therefore, some regions of IR3-related DNA in the mouse genome may be near the size of viral IR3 (about 0.7 kb). Homologous regions in the human genome are not clustered or are shorter than viral IR3.

To confirm the specificity of hybridization of IR3 to cell

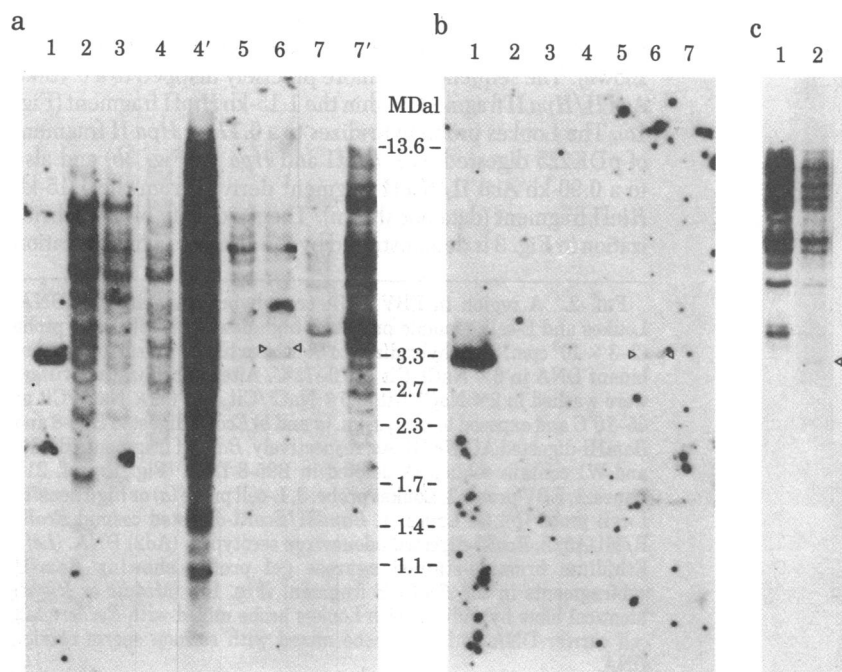


FIG. 4. IR3 is homologous to reiterated sequences in cell DNAs. (a and b) DNAs from human B-lymphocyte lines Loukes and Namalwa (a latently infected Burkitt tumor line) and mouse L cells were digested with *Eco*RI (4 μ g of DNA per lane) or *Bam*HI (2.5 μ g of DNA per lane), separated in duplicate on a 0.4% agarose gel, and blotted. Lanes: 1, size markers (1 μ g of *Eco*RI-digested Ad2 DNA and 2 pg of pDK225 5.0-kb K fragment insert); 2, *Eco*RI-cut Loukes DNA; 3, *Eco*RI-cut Namalwa DNA; 4, *Eco*RI-cut L-cell DNA; 5, *Bam*HI-cut Loukes DNA; 6, *Bam*HI-cut Namalwa DNA; 7, *Bam*HI-cut L-cell DNA. 32 P-labeled IR3 *Ava* II/*Hpa* II DNA probe (see Fig. 1d) (a) and 0.71-kb *Hinf*I DNA probe (see Fig. 1b) (b) were hybridized to identical blots in $5 \times$ NaCl/Cit/50% formamide at 50°C. The probes contained *E. coli* carrier DNA. Filters were washed after hybridization as described in Fig. 2. Lanes 4' and 7' (L-cell DNA) were exposed separately to x-ray film for the same length of time as the rest of the filter; lanes 4 and 7 were exposed for a shorter time. Open triangles indicate a *Bam*HI K fragment in Namalwa DNA (lane 6). (c) L-cell DNA (lane 1) and L-cell DNA together with one copy of P3HR-1 EBV DNA per cell genome reconstruction (lane 2) were digested with *Bam*HI, separated on a 0.3% agarose gel, blotted, hybridized to IR3 probe in $6 \times$ NaCl/Cit at 74°C, and washed as described in Fig. 2. The open triangle indicates the P3HR-1 *Bam*HI K fragment.

DNAs, a 0.71-kb *Hinf*I fragment from the *Bam*HI K fragment (2.9–3.6 kb in Fig. 1b) that does not overlap with the IR3 region was hybridized to an identical Southern blot (Fig. 4b). This probe identifies a *Bam*HI fragment in Namalwa DNA (lane 6) and the *Bam*HI K fragment in pDK225 DNA (lane 1) but does not hybridize to Loukes or L-cell DNAs. Furthermore, labeled IR3 does not hybridize to Ad2 DNA (Fig. 4), simian virus 40 (SV40) DNA, or the high G+C DNAs of HSV-1, HSV-2, and SV40 *ori* (Southern blots not shown).

Sites of IR3 Homology on Chromosomes. IR3-homologous sequences were assigned to sites in human chromosomes by cytological hybridization of 125 I-labeled IR3 DNA to metaphase chromosomes of phytohemagglutinin-stimulated lymphocytes. Regions of homology were mapped in individual chromosomes by comparing photographs of Giemsa-banded chromosomes prior to hybridization with autoradiograms of the same chromosomes after hybridization. The sequences homologous to IR3 are distributed among the chromosomes and localize to specific regions in chromosomes. Autoradiographic grains over sites in Giemsa-banded chromosomes are shown in Fig. 5. The mean distribution of silver grains over specific regions of Giemsa-banded chromosomes in 25 metaphase plates is summarized in Fig. 6. Consistent sites of IR3 homology are detected at centromeric or telomeric regions of some chromosomes. Other sites are observed at subtelomeric or midarm regions. All autosomes and the X chromosome have at least one region of homology. No IR3 homology is detected on the Y chromosome. The mean grain density over individual chromosomes is not proportional to chromosome length, indicating an unequal distribution of IR3-related DNA among chromosomes. Autoradiographic grains localize to some light and some dark Giemsa-banded regions in chromosomes (Fig. 6).

The total chromosome-specific grain counts provide an in-

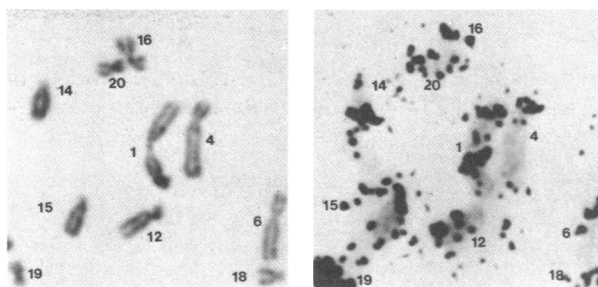


FIG. 5. Assignment of grains to chromosome regions. (Left) Photograph of Giemsa-banded chromosomes from phytohemagglutinin-stimulated lymphocytes. (Right) Autoradiogram of the same metaphase plate after cytological hybridization to 125 I-labeled IR3 probe.

dication of the amount of IR3-related DNA in cells. IR3-specific grain counts were normalized relative to the grain counts observed in similar hybridizations of human growth hormone DNA to chromosomes of rat cell lines each of which contains a different number of copies of human growth hormone DNA. The growth hormone DNA was introduced into the cells by cotransfection methods and quantitated by Southern blot analysis (27). Based on this comparison, the amount of IR3-related DNA in the human genome is estimated to be 10^8 – 10^9 daltons (28).

DISCUSSION

Each chromosome segment identified in cytological hybridization with IR3 contains at least one IR3-homologous sequence (Fig. 6). The number of such sites in human chromosomes and the number of homologous restriction endonuclease fragments in cell DNA (Fig. 4) indicate that there are at least 60 IR3-re-

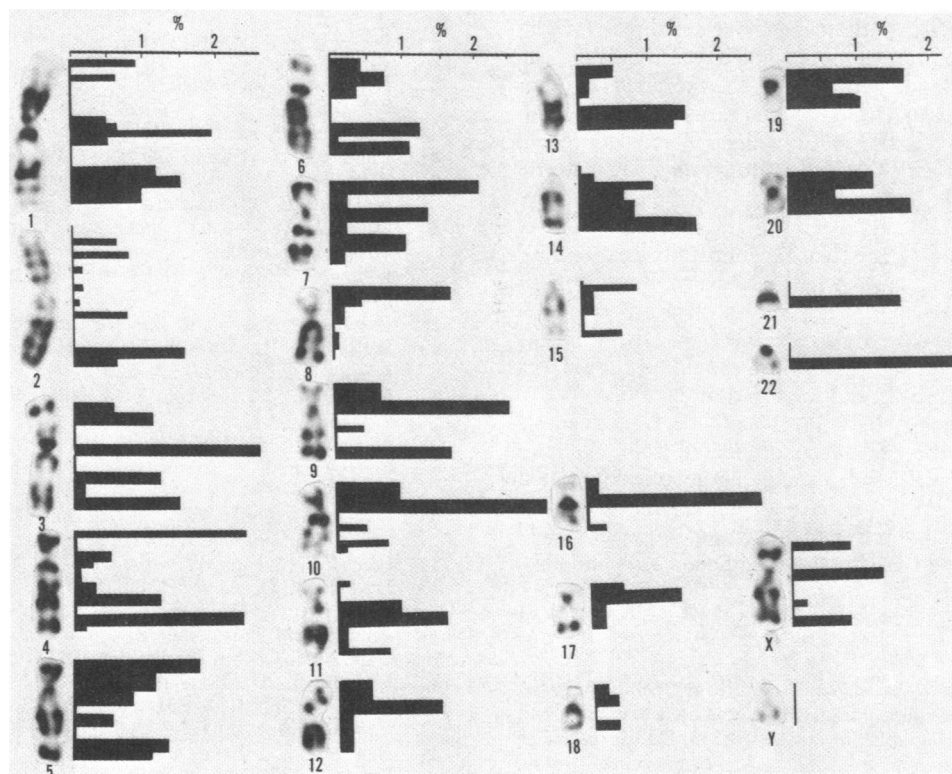


FIG. 6. Summary of grain distributions over Giemsa bands in human chromosomes. Photographs of 25 Giemsa-banded metaphase plates were compared with autoradiograms of the same plates after hybridization to 125 I-labeled IR3 DNA as shown in Fig. 5. Distribution of autoradiographic grains is plotted as a percentage of total grains over distinguishable Giemsa-banded chromosome regions. Widths of the horizontal bars reflect the widths of the light or dark reference Giemsa bands.

lated regions interspersed in the human genome. The hybridization of IR3 to many fragments of cell DNA near the level observed for the single-copy IR3 reconstruction indicates that IR3 identifies closely related cell sequences. The cellular sequences are likely to be a family of repeat arrays, since each sequence has extensive homology to IR3.

The origin region of SV40 DNA also has homology to cell DNA (32, 33). The extent of homology of IR3 to cell DNA is greater than the homology of SV40 *ori* to cell DNA, since IR3 homology is detected in Southern blots of cell DNA (Fig. 4), while SV40 homology is detected only after amplification of the homologous cell sequences by cloning (32–34). The SV40-homologous monkey DNA segments contain interspersed regions shorter than 30 bp that are partially related to SV40 origin region (32, 33). The Nucleic Acid Sequence Data Base (35) was searched to determine whether an IR3-like region is part of a known sequence. Only a few 9- to 12-bp matches were found between IR3 and known interspersed, repeated, or unique cell and viral DNAs. Thus, IR3 probably identifies a new family of interspersed repeat cell DNA sequences. Sequence analysis of IR3-homologous regions isolated from clone libraries of cell DNA (36) will reveal the extent of similarity among the IR3-related cell sequences and may suggest functions of these sequences.

IR3 is unusual in that it is a simple triplet nucleotide repeat array (18). The conservation of the sequence in EBV DNA as a large array suggests function as an enhanced site for intermolecular interaction. The role of the repeat array in favoring intermolecular interaction could be at the level of DNA synthesis, recombination, transcription, RNA processing or translation, or in encoding a repetitive domain of a regulatory protein. The related cellular sequences may have similar functions and may interact with viral IR3 to regulate EBV or cell gene expression in latently infected growth-transformed cells.

At least two of these hypotheses can be tested directly. First, some aspects of the putative function of IR3 in regulation can be tested by examining how IR3 affects the expression of other EBV or cellular genes in (co)transfection experiments. Second, recombination between IR3 and a cellular homologue could lead to integration of EBV DNA into chromosomal DNA. If IR3 mediates integrative homologous recombination (37), EBV would be expected to integrate at any of the chromosomal sites containing an IR3 homologue. Our recent results localize EBV DNA to different chromosomes in the Namalwa Burkitt tumor cell line and the IB4 latently infected *in vitro* growth-transformed cell line. The close linkage of the EBV DNA domains immediately to the left and right of IR3 should be altered as a consequence of this type of IR3-mediated integration.

The extent of homology between IR3 and cell DNA suggests the possibility that IR3 could have originated from a cellular sequence. Since all EBV isolates and the genetically related baboon herpesvirus HV Papio have a colinear IR3-related sequence (18), the putative cellular progenitor sequence would have been incorporated early in the evolution of EBV and have been of selective advantage. Retrovirus transforming genes are well-described examples of cellular genes that are incorporated and expressed in viral genomes (38, 39).

This work was supported by National Institutes of Health Grants CA 17281 and CA 19264, American Cancer Society Grant MV 32 G, and National Science Foundation Grant PCM-81-04359. E.K. is the recipient of a Faculty Research Award from the American Cancer Society. A.H. is a Leukemia Society of America Scholar.

1. Epstein, M. A. & Achong, B., eds. (1979) in *The Epstein-Barr Virus* (Springer, Berlin), pp. 1–22.

2. Henle, W., Diehl, V., Kohn, G., zur Hausen, H. & Henle, G. (1967) *Science* **157**, 1064–1065.
3. zur Hausen, H., Diehl, V., Wolf, H., Schulte-Holthausen, H. & Schneider, U. (1972) *Nature (London) New Biol.* **237**, 189–190.
4. Nonoyama, M. & Pagano, J. (1972) *Nature (London) New Biol.* **238**, 169–171.
5. Lindahl, T., Adams, A., Bjursell, G., Bornkamm, S. W., Kaschka-Dietrich, E. & Jehn, U. (1976) *J. Mol. Biol.* **102**, 511–530.
6. Hampar, B., Tanaka, A., Nonoyama, M. & Derge, J. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 631–633.
7. Thorley-Lawson, D. & Strominger, J. L. (1976) *Nature (London)* **263**, 332–334.
8. Adams, A., Lindahl, T. & Klein, G. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 2888–2892.
9. Tanaka, A. & Nonoyama, M. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 4658–4661.
10. Powell, A. L. T., King, W. & Kieff, E. (1979) *J. Virol.* **29**, 261–274.
11. King, W., Powell, A. L. T., Raab-Traub, N., Hawke, M. & Kieff, E. (1980) *J. Virol.* **36**, 506–518.
12. King, W., van Santen, V. & Kieff, E. (1981) *J. Virol.* **38**, 649–660.
13. van Santen, V., Cheung, A. & Kieff, E. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 1930–1934.
14. Reedman, B. M. & Klein, G. (1973) *Int. J. Cancer* **11**, 499–520.
15. Slovin, S., Glassy, M., Dambaugh, T., Catalano, M., Ewing, R., Ferrone, S., Kieff, E., Vaughan, T. & Carson, D. (1981) *J. Immunol.* **127**, 585–590.
16. Strnad, B., Schuster, T., Hopkins, R., Neubauer, R. & Rabin, H. (1981) *J. Virol.* **38**, 996–1004.
17. Svedmyr, E. & Jondal, M. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 1622–1626.
18. Heller, M., van Santen, V. & Kieff, E. (1982) *J. Virol.* **44**, in press.
19. Dambaugh, T., Beisel, C., Hummel, M., King, W., Fennwald, S., Cheung, A., Heller, M., Raab-Traub, N. & Kieff, E. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2999–3003.
20. Dambaugh, T., Raab-Traub, N., Heller, M., Beisel, C., Hummel, M., Cheung, A., Fennwald, S., King, W. & Kieff, E. (1980) *Ann. N. Y. Acad. Sci.* **602**, 711–719.
21. Raab-Traub, N., Dambaugh, T. & Kieff, E. (1980) *Cell* **22**, 257–267.
22. Heller, M., Dambaugh, T. & Kieff, E. (1981) *J. Virol.* **38**, 632–648.
23. Cheung, A. & Kieff, E. (1982) *J. Virol.* **44**, in press.
24. Dambaugh, T. & Kieff, E. (1982) *J. Virol.*, in press.
25. Pizzo, P. A., Magrath, I. T., Chattopadhyay, S. K., Biggar, R. J. & Gerber, P. (1978) *Nature (London)* **272**, 629–631.
26. Miller, G., Coope, D., Niederman, G. & Pagano, J. (1976) *J. Virol.* **18**, 1071–1080.
27. Robins, D., Ripley, S., Henderson, A. & Axel, R. (1981) *Cell* **23**, 29–39.
28. Henderson, A. S. (1982) *Int. Rev. Cytol.*, in press.
29. Warburton, D., Atwood, K. C. & Henderson, A. S. (1976) *Cytogenet. Cell Genet.* **17**, 221–230.
30. Leiden, J. M., Frenkel, N. & Rapp, F. (1980) *J. Virol.* **33**, 272–285.
31. McKnight, S. L. (1980) *Nucleic Acids Res.* **8**, 5959–5964.
32. McCutchan, T. F. & Singer, M. F. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 95–99.
33. Queen, C., Lord, S. T., McCutchan, T. F. & Singer, M. F. (1981) *Mol. Cell Biol.* **1**, 1061–1068.
34. Botchan, M., Topp, W. & Sambrook, J. (1976) *Cell* **9**, 269–287.
35. Dayhoff, M. O., Chen, H. R., Hunt, L. T., Barker, W. C., Yeh, L.-S., George, D. G. & Orcutt, B. C. (1981) *Nucleic Acid Sequence Data Base (Natl. Biomed. Res. Found., Georgetown Univ. Med. Center, Washington, DC)*.
36. Maniatis, T., Hardison, R., Lacy, E., Laver, J., O'Connell, C., Quon, D., Sim, G. & Efstratiadis, A. (1978) *Cell* **15**, 687–701.
37. Campbell, A. M. (1962) *Adv. Genet.* **11**, 101–145.
38. Hanafusa, H., Wang, L.-H., Hanafusa, T., Anderson, S. M., Karess, R. E. & Hayward, W. S. (1980) in *Animal Virus Genetics*, eds. Fields, B., Jaenish, R. & Fox, C. F. (Academic, New York), pp. 483–497.
39. Bishop, J. (1981) *Cell* **23**, 5–6.