

Molecular Cloning of cDNAs Derived from a Novel Human Intestinal Mucin Gene

James R. Gum¹, James M. Hicks¹, Dallas M. Swallow², Robert L. Lagace¹, James C. Byrd¹, Derek T.A. Lamport³, Bader Siddiki¹, and Young S. Kim^{1,4}

¹Gastrointestinal Research Laboratory (151M2), Veterans Administration Medical Center, San Francisco, California 94121, the Department of Medicine, University of California, San Francisco, California 94143

²MRC Human Biochemical Genetics Unit, U.C.L., Wolfson House, 4 Stephenson Way, London NW12HE

³Department of Energy Plant Research Laboratory and Department of Biochemistry, Michigan State University, East Lansing, Michigan 48824

Received July 19, 1990

SUMMARY: A human small intestinal λ gt11 cDNA library was screened with antibodies to deglycosylated small intestinal mucin. Four partial cDNA clones were isolated that define a novel human mucin gene. These include two partial cDNA clones, SIB 124 and SIB 139, that contain 51 nucleotide tandem repeats which encode a seventeen amino acid repetitive peptide with a consensus sequence of HSTPSFTSSITTETTS. SIB 139 hybridized to messages produced by small intestine, colon, colonic tumors and also by high mucin variant LS174T colon cancer cells. The gene from which cDNAs SIB 124 and SIB 139 are derived (proposed name MUC 3) maps to chromosome 7, distinct from other known human mucin genes. ©1990 Academic Press, Inc.

Many tissues including trachea, submaxillary gland, mammary gland, pancreas, stomach, cervix and both the large and small intestine produce high molecular weight, complex glycoconjugates known as mucins (1-8). These substances, which constitute the principle glycoprotein component of the various mucus gels found in higher organisms, consist of a polypeptide core covered almost entirely by O-linked carbohydrate chains (1,2,4,8). Primarily because mucin polypeptide backbones are so heavily glycosylated, they have proven to be quite difficult to analyze by conventional peptide sequencing methodology. Recently however, cDNAs have been obtained that encode portions of the polypeptide cores of mucins found in porcine submaxillary gland (9), human mammary gland (10-12), and human small and large intestine (13). These

⁴To whom reprint requests should be addressed.

Abbreviations: SIB, deglycosylated human small intestinal mucin; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; bp, base pairs; kb, kilobase pairs.

cDNAs have revealed, in all three cases, that the mucin peptide contains extended regions of tandem repeats that contain a high content of threonine and/or serine (9-13). While all of these mucin polypeptides contain these repetitive regions, the size and amino acid composition of the tandem repeat units has been found to vary widely between mucin types.

Over the last several years, our laboratories have been interested in human intestinal mucin. We have cloned and sequenced cDNAs that encode a mucin that contains threonine and proline-rich tandem repeats from a human small intestine library (13). Mucins however, especially intestinal mucins, are heterogeneous as judged by both biochemical and immunological criteria (5-7). This led us to examine the possibility that at least one other form of human intestinal mucin exists. In this paper we report the cloning and characterization of cDNAs derived from a novel human mucin gene that is expressed at high levels in intestine.

Materials and Methods

Mucin Purification, Deglycosylation, Antibody Production, and Library Screening. Mucin was purified from normal human small intestine using gel filtration and CsCl density gradient centrifugation as described (8). The purified mucin was deglycosylated by hydrogen fluoride treatment for three hours at room temperature (14,15). This material, designated SIB, contained 20% Thr, 16% Ser, and 9% Pro, but less than 3% of its original carbohydrate. Polyclonal antibodies against this deglycosylated mucin were prepared in New Zealand white rabbits by subcutaneous injection (14). A human jejunal λ gt11 cDNA library (16) was obtained from Dr. Yvonne Edwards and screened using a 1:75 dilution of anti-SIB as described (13).

Chromosomal Localization. Was performed by DNA blot analysis of a panel of human/rodent somatic cell hybrids containing variable numbers of human chromosomes. DNA was digested with *Hinf*I and the fragments were separated on 0.8% agarose gels and transferred to Hybond N membranes (Amersham). Hybridization to cDNA clone SIB 124 was carried out using the protocol recommended by the manufacturer. The stringency of the final wash was 0.5X standard saline citrate, 0.1% sodium dodecyl sulfate at 65°C.

Results

Isolation of cDNAs. Previously we used antibodies against deglycosylated human colon cancer xenograft mucin to obtain cDNA clones derived from a human intestinal mucin gene (13). These clones, designated SMUC 40-42, all contain 69 bp tandem repeats that encode a threonine and proline-rich repetitive peptide. Due to the heterogeneity of intestinal mucin (5-7), we hypothesized that small intestine may contain additional forms of mucin. Thus, mucin was prepared from human small intestine, deglycosylated, and used to prepare antisera (anti-SIB). Approximately 900,000 recombinant λ gt11 plaques were screened with this antisera and fourteen positives were obtained and purified. Six of these clones were found to hybridize to SMUC 41 and were not further

characterized. Phage DNA was prepared from the remaining eight clones and digested with EcoRI. Three of the eight inserts visualized by gel electrophoresis were very small (<200 bp) and were therefore not further analyzed. Of the remaining five clones, four appeared to be derived from the same gene as they recognized common bands when used in blot analysis of human genomic DNA (data not shown). These clones were sequenced and two of them, designated SIB 124 and SIB 139, were found to contain tandem repeats characteristic of a mucin as described below. The other two clones, SIB 134 and SIB 136 (337 and 222 bp, respectively), which did not contain tandem repeats, appear to encode unique portions of the same mucin and will be described later.

Sequence of Clones SIB 124 and SIB 139. Both SIB 124 and SIB 139 were found to consist of similar 51 bp tandem repeats (Fig. 1). These tandem repeats encode a repetitive peptide that contains threonine and serine-rich clusters that are potential O-glycosylation sites such as STPS, TSS, TTT, and TTS. Unlike the SMUC-type clones isolated previously (13), these cDNAs encode a peptide that contains significant quantities of serine. Also, unlike the SMUC-type clones, SIB 124 contains a short interrupting sequence (18 bp) nested in one of its tandem repeats (Fig. 1). The structure of the SIB repetitive peptide is compared to the structures of the other two known human mucin peptides in Fig. 2. The SIB tandem repeat sequences have no significant sequence similarity to these sequences or to any other nucleic acid sequence in the GenBank Release 61.0 database.

Genomic DNA Blot Analysis. Genomic DNA was subjected to blot analysis using SIB 139 as a probe (Fig. 3A). A single band was observed with only two of the twelve enzymes used, SphI and XbaI, and in both cases, the fragment that hybridized was quite large (>23.1 kb). Multiple bands occurred using the other ten enzymes. In contrast, single major bands were observed when the same blots were probed with SMUC 41 (Fig. 3B). Furthermore, some of the fragments that hybridized to SMUC 41 were as small as 10-12 kb (NcoI, PvuII, PstI; Fig. 3B). It seems likely therefore, that the tandem repeat region of

MAMMARY MUCIN

ProAspThrArgProAlaProGlySerThrAlaProProAlaHisGlyValThrSerAla

SMUC

ProThrThrThrProIleThrThrThrThrThrValThrProThrProThrProThrGlyThrGlnThr

SIB

HisSerThrProSerPheThrSerSerIleThrThrThrGluThrThrSer

Figure 2. Comparison of the tandem repeats found in human mucins. The consensus sequences of the tandem repeats of the three known human mucin genes are shown. The sequences for mammary mucin and SMUC-type mucin are from references 12 and 13.

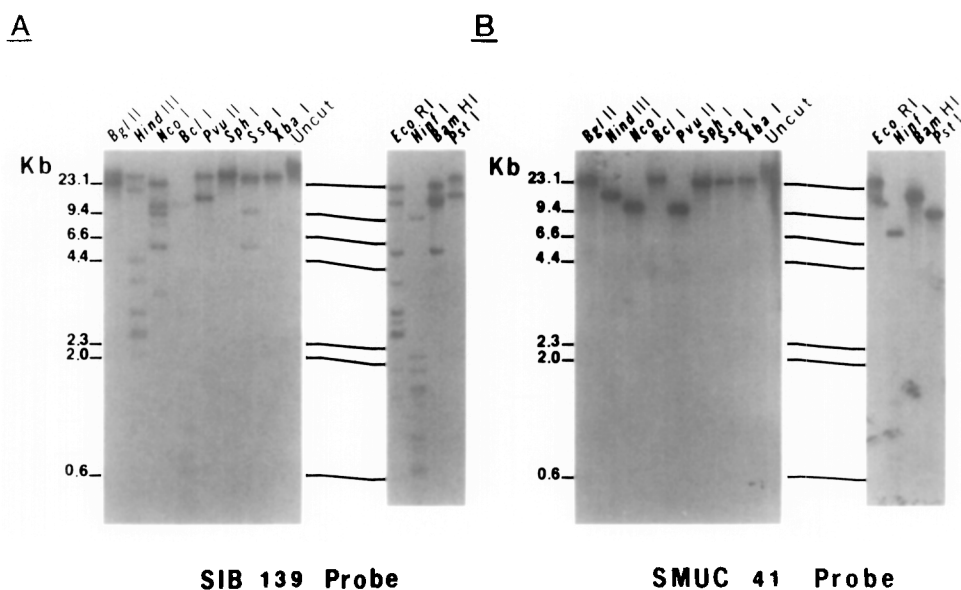


Figure 3. Genomic DNA blots using SIB and SMUC-type probes. DNA isolated from the lymphocytes of a single human donor was digested with the indicated restriction enzymes. Aliquots (8 μ g) of the digests were separated by electrophoresis on 0.8% agarose gels and transferred to nylon membranes. In Panel A, the membranes were first probed with SIB 139 as described (13, 18). In experimentation not shown, the use of clone SIB 124 as a probe gave an identical banding pattern with *Eco*RI and *Bam*HI - digested DNA as the results shown here. In Panel B, the same blots were hybridized to SMUC 41 following removal of the SIB 139 probe (13).

the gene recognized by SIB 139 is large by comparison to the tandem repeat region of the gene that contains the SMUC-type repetitive units.

RNA Blot Analysis. Poly(A)⁺ RNA isolated from human colon cancer cell line LS174T, and some quantitative mucin variants derived from this line (19), were subjected to gel electrophoresis and blot analysis (Fig. 4). The high mucin variants HM3 and HM7 both expressed mRNA that hybridized to SIB 139. This message was polydisperse in size. The most common message size in HM3, which exhibited less total polydispersity than HM7, was approximately 8 kb. In contrast, both the parental LS174T cells and the low mucin variant LM12 line produced only relatively low levels of this mucin message. Following removal of the labeled SIB 139 cDNA, this blot was probed with SMUC 41 (Fig. 4). Both high mucin variants also exhibited a polydisperse message with this probe. In contrast, when the blot was probed with a GAPDH cDNA as a control, a single, defined band was observed (Fig. 4) suggesting that there was no significant RNA degradation. A dot blot analysis of total RNA from human small intestine, colon, and colonic tumors is shown in Fig. 5. All three tissue types expressed both the SIB and SMUC-type messages. In the samples tested, small intestine and normal colon produced the largest amounts of the

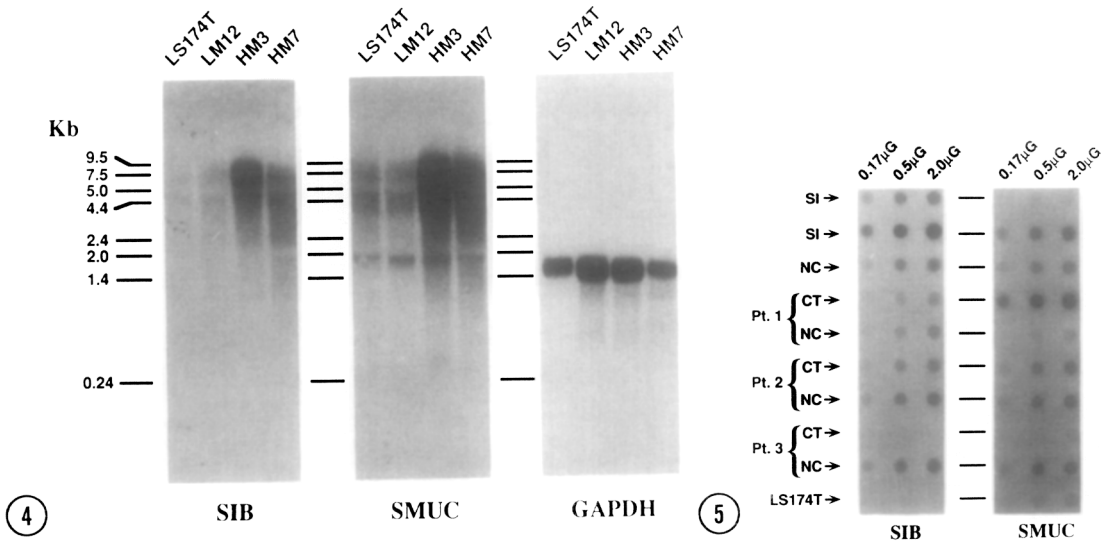


Figure 4. RNA blot analysis of human intestinal mucin messages found in LS174T cells and quantitative mucin variants. Left, 3 µg samples of poly(A)⁺ RNA from LS174T cells, and low (LM12) and high (HM3, HM7) mucin variants of this line (19) were subjected to agarose gel electrophoresis and blot analysis using SIB 139 as probe. The molecular size standards were obtained from Bethesda Research Laboratories with the exception of the 2.0 kb and 5.0 kb standards which were rRNA subunits. The autoradiograms shown on the middle and right were obtained following probe removal (13) using as probe first SMUC 41 (13) and then a GAPDH cDNA (20).

Figure 5. RNA dot blot analysis. Total RNA samples in ice-cold 10 mM NaOH and 1 mM EDTA were applied to a zeta-probe nylon membrane and hybridized to SIB 139 (13, 18). Following autoradiography, the probe was removed (13) and the blot was again probed using SMUC 41. The samples were derived from the small intestine of two individual donors (SI), from the normal colon of one individual (NC), and from matched pairs of colonic tumors (CT) and normal colons from individual patients (Pt. 1 - Pt. 3). Total RNA from LS174T cells was included as a control. The quantity of RNA used per well is given above the autoradiograms.

SIB-type message. The expression of the SMUC-type message generally followed the same pattern as the SIB-type message expression with one notable exception: the colonic tumor removed from patient 1 contained very high levels of the SMUC-type message. Thus, in this particular cancer, intestinal mucin gene expression appears to be altered.

Chromosomal Localization. In order to map the SIB-type mucin gene, we performed blot analysis of DNA extracted from human/rodent somatic cell hybrids (Table 1). It is apparent that there is complete correlation between the presence of *HinfI* hybridization bands and the presence of human chromosome 7 in all hybrids, including hybrid c121 which contains chromosome 7 as the only human chromosome present. Thus, the SIB-type mucin gene maps to human chromosome 7. We propose that this gene should be named MUC 3.

Table 1
Chromosomal Localization of MUC 3

		Human Chromosome																							
Hybrid.	Ref.	SIB	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X
MOG2C2	(21,22)	+	+	-	+	+	+	-	+	+	+	+	+	-	-	+	+	+	+	+	+	+	+	+	+
MOG2E5	(21,22)	+	+	-	+	+	+	+	+	+	•	+	-	+	+	+	+	+	+	+	-	-	-	+	+
CTP34B4	(23)	+	+	+	+	-	±	+	+	+	-	-	-	+	-	+	-	+	+	+	-	-	-	-	+
CRAB7	(**)	+	-	-	±	-	+	-	+	•	±	+	+	±	+	+	•	-	+	-	+	±	•	•	+
FHA7	(24)	+	-	-	+	+	+	+	+	+	•	•	-	+	•	•	+	+	+	+	•	+	-	+	+
SIF15P5	(25)	+	-	+	-	-	-	+	+	-	-	+	-	-	-	+	+	-	-	-	-	+	-	-	+
2W1R70	(26)	+	-	-	-	-	-	-	+	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	•
c121	(27)	+	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CRAB8	(**)	-	-	+	-	•	-	-	-	•	±	-	+	+	•	+	•	•	•	-	-	-	•	•	+
TWIN19F9	(28,29)	-	+	+	+	+	-	+	-	+	+	-	-	+	•	+	-	-	+	+	-	+	+	+	-
TWIN19C5	(28,29)	-	-	-	+	+	-	-	-	+	-	-	-	+	•	+	-	-	+	+	-	+	+	+	-
TWIN19F6	(28,29)	-	-	+	+	+	-	+	-	+	+	-	-	+	•	+	-	-	+	+	-	+	+	+	-
DUR4R3	(30)	-	-	-	+	-	+	-	-	•	-	-	+	+	+	+	-	-	+	+	-	+	+	+	-
SIF4A31	(25)	-	-	-	-	+	-	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	+	+
SIF4A24c1E1	(25)	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	-	+
FST9/10	(24)	-	-	-	+	+	-	+	-	+	•	+	-	+	+	+	+	-	-	+	-	+	-	+	+
HORL411B6P	(26)	-	+	-	+	-	-	-	-	-	-	-	+	-	+	-	+	-	-	+	-	-	-	•	+
SIB/chromosome	concordant	+/+	3	2	5	3	5	4	8	4	2	4	2	4	3	5	4	4	5	4	2	4	2	3	6
		-/-	7	6	3	2	8	5	9	3	5	8	6	3	2	2	6	8	3	3	9	4	3	2	4
	discordant	+/-	5	6	3	5	3	4	0	3	4	3	6	4	4	2	3	4	3	4	5	4	5	4	1
		-/+	2	3	6	6	1	4	0	4	3	1	3	6	3	7	2	0	5	6	0	5	5	5	5

Segregation of the gene encoding the SIB-type mucin in 17 human rodent somatic cell hybrids. The hybrids used have been described previously as referenced, but have been regrown and recharacterized since and the current human chromosome content is shown. They represent 13 independent hybrids and 4 related subclones. Key: human chromosome or SIB-type mucin DNA fragments present (+) or absent (-), human chromosome equivocal or not tested (.), human chromosome present in less than 20% of the cells of a clone (±). CRAB7 and CRAB8 cells (**) are previously unpublished hybrids of RAG mouse and human cells with a 9/10 translocation t(9;10)(q33;p11.2) characterized by isozyme analysis alone.

Discussion

In this work, we used antibodies against deglycosylated small intestinal mucin to identify cDNAs that define the tandem repeats of a novel human mucin. These SIB-type tandem repeats are rich in both threonine and serine. Thus, the SIB-type deduced peptide has a substantially different amino acid composition than does the threonine and proline-rich SMUC-type peptide described previously (13). Given the amino acid compositions of the two peptides, it is tempting to speculate that the SIB-type peptide predominates in the major, neutral fraction of intestinal mucin (7) while the SMUC-type peptide predominates in the acidic mucin fraction. In any event, the presence of at least two intestinal mucin genes offers a plausible explanation for the observation that purified intestinal (7) and colonic (6) mucin can be further separated into two fractions with differing amino acid content.

It is becoming increasingly clear that tandem repeats form a large portion of mucin protein backbones. Ligtenberg et al. (31) have recently

sequenced cDNAs that span the entire length of the coding sequence of human mammary mucin and have shown that from 60% to 80% of the mucin protein core consists of tandem repeats, depending on allele. In our own work, we have isolated a genomic DNA clone for the SMUC-type intestinal mucin, and have found that tandem repeats constitute greater than 50% of this mucin protein (Toribara, N.W., J.R.G., and Y.S.K., unpublished observations). Similarly, the genomic DNA blots presented here strongly suggest that the SIB-type tandem repeats define a major portion of the protein backbone of the second form of intestinal mucin. From a technical standpoint, it is interesting that all mucin cDNAs isolated thus far have proven to be partial cDNAs that are frequently quite short (9-13,31). The reasons for this are not clear. The cDNA library that we used to isolate both types of intestinal mucin contains larger cDNAs, as is demonstrated by its use to isolate a 2 kb sucrase-isomaltase clone (16). It may be that the short tandem repeats present in mucin clones are especially prone to undergo deletion via homologous recombination during their propagation in bacterial host strains.

The results contained in Table 1 indicate that the SIB-type mucin gene maps to human chromosome 7. This gene is thus quite distinct from MUC 1 (previously called PUM) which codes for mammary mucin and is located on chromosome 1 (32), and from MUC 2 which codes for the SMUC-type mucin and is located on chromosome 11 (33). Because this is the third human mucin gene identified, we propose that it be named MUC 3. Thus, the known human mucin genes are not clustered together in the genome. At present, clear evidence exists for at least three forms of human mucin. Other forms also could exist, perhaps even one or more that are expressed in the small intestine. Clearly, different types of tissues could produce additional forms of mucin. The identification of additional mucin genes remains a topic for further study.

Acknowledgments

We thank Dr. Sue Povey for providing the hybrid DNA samples and for help with the chromosomal mapping experiments and Dr. Neil Toribara for the human DNA samples. We also thank Frank Fearney, Beatrice Griffiths, David Matthews, Clare Harvey and Wendy Pratt for their technical assistance, and Rita Burns for manuscript preparation. This work was supported by the Veterans Administration Medical Research Service and by DOE contract #DE-AC02-76ER0-1338 (to D.T.A.L.). YSK is a Medical Investigator of the Veterans Administration.

References

1. Neutra, M., and Forstner, J.F. (1987) *In Physiology of the Digestive Tract* (L.F. Johnson, Ed.), Second Edition, pp. 975-1009, Raven Press, New York.
2. Eckhardt, A.E., Timpte, C.S., Abernethy, J.L., Toumadje, A., Johnson, W.C., and Hill, R.L. (1987) *J. Biol. Chem.* 262, 11339-11344.

3. Burchell, J., Durbin, H., and Taylor-Papadimitriou, J. (1983) *J. Immunol.* 131, 508-513.
4. Nardelli, J., Byrd, J.C., Ho, J.J.L., Fearney, F.J., Tasman-Jones, C., and Kim, Y.S. (1988) *Pancreas* 3, 631-641.
5. Gold, D.V., and Shochat, D. (1989) *Molec. Immunol.* 26, 769-777.
6. Gold, D.V., Shochat, D., and Miller, F. (1981) *J. Biol. Chem.* 256, 6354-6358.
7. Wesley, A., Mantle, M., Man, D., Qureshi, R., Forstner, G., and Forstner, J. (1985) *J. Biol. Chem.* 260, 7955-7959.
8. Byrd, J.C., Nardelli, J., Siddiqui, B., and Kim, Y.S. (1988) *Cancer Res.* 48, 6678-6685.
9. Timppe, C.S., Eckhardt, A.E., Abernethy, J.L., and Hill, R.L. (1988) *J. Biol. Chem.* 263, 1081-1088.
10. Gendler, S.J., Burchell, J.M., Duhig, T., Lampport, D., White, R., Parker, M., and Taylor-Papadimitriou, J. (1987) *Proc. Natl. Acad. Sci. USA* 84, 6060-6064.
11. Siddiqui, J., Abe, M., Hayes, D., Shani, E., Yunis, E., and Kufe, D. (1988) *Proc. Natl. Acad. Sci. USA* 85, 2320-2323.
12. Gendler, S., Taylor-Papadimitriou, J., Duhig, T., Rothbard, J., and Burchell, J. (1988) *J. Biol. Chem.* 263, 12820-12823.
13. Gum, J.R., Byrd, J.C., Hicks, J.W., Toribara, N.W., Lampport, D.T.A., and Kim, Y.S. (1989) *J. Biol. Chem.* 264, 6480-6487.
14. Byrd, J.C., Lampport, D.T.A., Siddiqui, B., Kuan, S.F., Erickson, R.H., Itzkowitz, S.H., and Kim, Y.S. (1989) *Biochem. J.* 261, 617-625.
15. Mort, A.J., and Lampport, D.T.A. (1977) *Anal. Biochem.* 82, 289-309.
16. Green, F., Edwards, Y., Hauri, H., Povey, S., Ho, M.W., Pinto, M., and Swallow, D. (1987) *Gene (Amst.)* 57, 101-110.
17. Gum, J.R., Hicks, J.W., Sack, T.L., and Kim, Y.S. (1990) *Cancer Res.* 50, 1085-1091.
18. Gum, J.R., Kam, W.K., Byrd, J.C., Hicks, J.W., Sleisenger, M.H., and Kim, Y.S. (1987) *J. Biol. Chem.* 262, 1092-1097.
19. Kuan, S.-F., Byrd, J.C., Basbaum, C.B., and Kim, Y.S. (1987) *Cancer Res.* 47, 5715-5724.
20. Allen, R.W., Trach, K.A., and Hoch, J.A. (1987) *J. Biol. Chem.* 262, 649-653.
21. Solomon, E., Swallow, D., Burgess, S., and Evans, L. (1979) *Ann. Hum. Genet.* 42, 273-281.
22. Whitehead, A.S., Solomon, E., Chambers, S., Bodmer, W.F., Povey, S., and Fey, G. (1982) *Proc. Natl. Acad. Sci. USA* 79, 5021-5025.
23. Jones, E.A., Goodfellow, P.W., Kennett, R.H., and Bodmer, W.F. (1976) *Somatic Cell Genet.* 2, 483-496.
24. Kielty, C.M., Povey, S., and Hopkinson, D.A. (1982) *Ann. Hum. Genet.* 46, 307-327.
25. Edwards, T.H., Parkar, M., Povey, S., West, L.F., Parrington, J.M., and Solomon, E. (1985) *Ann. Hum. Genet.* 49, 101-109.
26. Van Heyningen, V., Bobrow, M., Bodmer, W.F., Gardiner, S.E., Povey, S., and Hopkinson, D.A. (1975) *Ann. Hum. Genet.* 38, 295-303.
27. Croce, C.M., and Koprowski, H. (1974) *J. Exp. Med.* 140, 1221-1229.
28. Phillips, I.A., Shephard, E.A., Povey, S., Davis, M.B., Kelsey, G., Monteiro, M., West, L.F., and Cowell, J. (1985) *Ann. Hum. Genet.* 49, 267-274.
29. Swallow, D.M., Povey, S., Parkar, M., Andrews, P.W., Harris, H., Pym, B., and Goodfellow, P. (1986) *Ann. Hum. Genet.* 50, 229-235.
30. Solomon, E., Bobrow, M., Goodfellow, P.N., Bodmer, W.F., Swallow, D.M., Povey, S., and Noel, B. (1976) *Somatic Cell Genet.* 2, 125-140.
31. Ligtenberg, M.J.L., Vos, H.L., Gennissen, A.M.C., and Hilkens, J. (1990) *J. Biol. Chem.* 265, 5573-5578.
32. Swallow, D.M., Gendler, S., Griffiths, B., Kearney, A., Povey, S., Sheer, D., Palmer, R.W., and Taylor-Papadimitriou, J. (1987) *Ann. Hum. Genet.* 51, 289-294.
33. Griffiths, B., Gum, J., West, L.F., Povey, S., Swallow, D.M., and Kim, Y.S. (1989) *Cytogenet. Cell Genet.* 51, 1008.