Structure, Expression, and Chromosomal Location of the Human c-fgr Gene

MAKOTO NISHIZAWA,¹ KENTARO SEMBA,¹ MICHIHIRO C. YOSHIDA,² TADASHI YAMAMOTO,¹ MOTOMICHI SASAKI,² and KUMAO TOYOSHIMA^{1*}

Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo, 108,¹ and Chromosome Research Unit, Faculty of Science, Hokkaido University, Kita-ku, Sapporo, 060,² Japan

Received 3 September 1985/Accepted 24 October 1985

The nucleotide sequence of seven exons of the human c-fgr gene, a cellular homolog of the oncogene of Gardner-Rasheed feline sarcoma virus, was determined. Twenty-six independent genomic clones were obtained from a human gene library with a DNA clone of Y73 avian sarcoma virus oncogene, v-yes, as a probe under relaxed hybridization conditions. Restriction mapping and partial sequence analyses revealed that two of these clones were derived from the c-fgr gene, distinct from the c-yes gene. Interestingly, the splicing points of the c-fgr gene were identical with those of the c-src gene throughout the seven exons, suggesting that the two proto-oncogenes were generated by gene duplication of an ancestral gene containing intervening sequences. On RNA blot hybridization the major transcript was found to be 2.6 kilobases long. Two additional transcripts of 3.5 and 4.7 kilobases were also detected. Furthermore, karyotype analysis of several human-mouse hybrid cells and Southern blot analyses of DNAs of the hybrids with a human c-fgr locus-specific probe showed that this gene is located on chromosome 1.

It is widely accepted that an acutely oncogenic retrovirus contains a viral oncogene that is derived from a cellular counterpart and is responsible for the initiation and maintenance of cellular transformation. In animals, cellular oncogenes have been highly conserved throughout evolution, suggesting that they play important roles in normal cells. However, the normal functions of most cellular oncogenes are still unknown (2). More than 10 oncogenes were found to be structurally related to the most well-characterized oncogene, the *src* gene of Rous sarcoma virus, and so are called the "*src* family." Most members of this gene family have been found to encode products that are associated with tyrosine kinase activity, and it is generally assumed that kinase activity plays a key role in cell transformation (2).

In the *src* family, the *yes* gene, identified as an oncogene of avian Y73 sarcoma virus (15), is the most closely related to the *src* gene (18). The oncogene product of the Y73 virus is $p90^{gag-yes}$, which shows tyrosine-specific protein kinase activity like the *src* gene product (17). The human cellular *yes*-related sequence, the proto-*yes* gene, was recently assigned to chromosome 18q21 by somatic hybrid cell analysis (33) and by in situ hybridization experiments (40), suggesting its possible involvement in the pathogenesis of follicular lymphoma, which is known frequently to be associated with translocation between chromosomes 14 and 18 (9).

Another oncogene closely related to src, the fgr gene, has been identified as an oncogene of Gardner-Rasheed feline sarcoma virus (27, 31). The amino acid sequence of the v-fgrproduct, $p70^{gag-fgr}$ (25), was deduced from nucleotide sequence data. The carboxy-terminal half of the oncogene product showed highest homology (80%) with that of the v-yes product (26). Interestingly, the kinase domain and the viral gag sequence in the amino-terminal portion were interrupted by an actin-like sequence (26). Previously we reported the isolation of a human c-fgr gene clone by using a v-yes probe and the nucleotide sequence of the two exons (28). The actin gene-like sequence of the viral fgr gene is not In this report, we present further structural analysis of the human c-fgr gene. The exon-intron structure of the c-fgr gene was found to be identical to those of the c-src genes of chickens and humans, suggesting that the fgr and src genes originated by duplication of a prototype gene. In addition, using the gene specific probe, we detected the expression and location of the gene on chromosome 1, clearly indicating that the yes gene and the fgr gene are distinct oncogenes.

MATERIALS AND METHODS

Screening of a human gene library. A human gene library was constructed from AluI-HaeIII partial digests of human placenta DNA and the Charon 4A EcoRI arm by the method of Maniatis et al. (22). This library was screened by hybridization to v-yes-specific pYS-2 (33) insert DNA labeled with ³²P by nick translation. Considering the divergence of the v-yes gene and human c-yes gene, in screening we used relatively relaxed conditions for hybridization (30% formamide-0.6 M NaCl-60 mM sodium citrate-10× Denhardt solution [6]-50 mM sodium HEPES [N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid] [pH 7.0]-0.2 mg of denatured salmon sperm DNA per ml-0.1 mg of yeast RNA per ml at 42°C) and for washing out the probe (0.3 M NaCl-30 mM sodium citrate-0.1% sodium dodecyl sulfate at room temperature twice and 66 mM NaCl-6.6 mM sodium citrate-0.1% sodium dodecyl sulfate at 50°C four times).

Restriction mapping and sequencing. The cloned DNAs were subjected to restriction mapping followed by Southern blot analysis to identify restriction fragments that contained possible exon sequences. The DNA sequence was determined by the dideoxynucleotide method of Messing (23) and of Sakaki et al. (personal communication) with M13mp18, M13mp19, and pUC19 (29) as cloning-sequencing vectors.

Chromosome mapping. To determine the chromosomal location of the human c-fgr gene, we needed a human c-fgr-specific probe. For this purpose, we used an EcoRI-

present in the c-fgr gene, indicating that the v-fgr gene is a tripartite gene consisting of the viral gag gene and two cellular genes, the actin gene and the c-fgr gene (28).

^{*} Corresponding author.



FIG. 1. Restriction endonuclease map of the human c-fgr locus. The human c-fgr locus was defined by two overlapping clones, $\lambda 015$ and $\lambda 010$. Restriction mapping was done by analysis of doubledigestion products. Hybridization experiments showed that only two *Eco*RI fragments of approximately 7.7 and 2.9 kb, marked by asterisks, contained the *yes*-related gene sequence, the c-fgr sequence. Sequence analysis of the *Eco*RI 7.7-kb fragment showed 7 exons (exon 4 to 10) in this region. The exon numbers correspond to these of the chicken c-*src* gene assigned by Takeya and Hanafusa (37). The positions of the exons are indicated by black boxes. Exons 11 and 12 were defined by Parker et al. (30) to be parts of the c-*src*-2 gene and were partially confirmed by sequence analysis. A 0.4-kb *Eco*RI-*ScaI* fragment (probe A) and a 0.2-kb *SacI-SacI* fragment (probe B) were used as c-fgr-specific probes.

Scal 0.4-kilobase (kb) fragment containing the exon 4 sequence that had been recloned into pUC9 (probe A in Fig. 1). We used the specific probe under stringent conditions of hybridization (50% formamide-0.45 M NaCl-45 mM NaCl-10× Denhardt solution-50 mM sodium HEPES [pH 7.0]–0.2 mg of denatured salmon sperm DNA per ml–0.1 mg of yeast RNA per ml at 42°C] and of washing out (0.3 M NaCl-30 mM sodium citrate-0.1% sodium dodecyl sulfate at room temperature twice and 30 mM NaCl-3 mM sodium citrate-0.1% sodium dodecyl sulfate at 50°C four times). Although the probe contains the exon 4 sequence of the c-fgr gene, it could detect the human c-fgr locus but could not hybridize to any mouse DNA under the stringent conditions. Somatic cell hybrids were generated by fusion of human embryo fibroblasts with mouse cells (33). DNAs extracted from human placenta, mouse cells, and 12 hybrid clones were digested with HindIII and subjected to Southern blot analysis with ³²P-labeled probe A. The human N-ras-specific probes pNP1 and pNP5 (34) were kindly supplied by K. Shimizu.

RNA isolation and blotting. Total RNAs were isolated from

GTCAGGAGGCGG—ND-

GTAACTGGGCCAGCAGCCTT

GTGCCCTGCTTCACCCCACC

8

Q

10

CACACG

GCCCAG

GCCAAG

human placenta and several cultured cells by the guanidine isothiocyanate solubilization method (4). Oligo(dT)-cellulose chromatography was used to isolate polyadenylated RNA. RNA was fractionated on 1% agarose–formamide gel (20) and transferred to nitrocellulose. Prehybridization, hybridization, and washing were carried out as described previously (33). For use as a human c-fgr specific probe for RNA blotting, we recloned the SacI-SacI 220-base-pair (bp) fragment, which contains a 154-bp exon 10 sequence, into pUC19 (probe B in Fig. 1) and labeled it by nick translation.

RESULTS

Two genomic clones derived from the c-fgr locus. We isolated 26 independent genomic clones by using the v-yes-specific probe under nonstringent conditions. Restriction maps showed that two of these clones, $\lambda 015$ and $\lambda 010$, had an overlap of about 1 kb, indicating that they were derived from the same locus. Figure 1 shows the restriction maps of the inserts of $\lambda 015$ and $\lambda 010$. The inserts of these clones were 11.0 and 15.4 kb long, respectively. Southern blot analysis of the cloned DNAs showed that two *Eco*RI fragments (7.7 and 2.9 kb) contained the v-yes-related sequence. The 7.7-kb EcoRI fragment was subcloned into plasmid pBR327, and its nucleotide sequence was analyzed by the dideoxy sequencing method.

As a result, seven putative exons were found in the 7.7-kb EcoRI fragment. The positions of the exons are shown in Fig. 1. The splicing sites were predicted from the homology of the sequence with that of the v-ves gene and by comparison with the consensus splicing site structures. The GT-AG rule (24) was maintained in the splicing sites (Table 1). In addition, the features of the surrounding nucleotide sequences (24) were also well conserved, suggesting that these regions are indeed splicing positions. The total length of the exon sequences is 920 bp, coding 306 amino acid residues. Although these two genomic clones were isolated from a genomic library by sequential screening with the v-yes probe, the exon sequence had higher homology to the v-fgr gene than to the v-yes gene, human c-src gene, or other oncogenes (Fig. 2 and 3; Table 2). The 5' extremity of the sequenced exons had no homology with the v-fgr gene corresponding to the actinlike portion (26). Thus, as discussed previously, the v-fgr gene product was considered to be a tripartite gene product consisting of the viral gag gene and two cellular genes, the actin gene and the c-fgr gene (28). The extensive homology with the v-fgr gene suggests that these genomic clones were derived from the c-fgr gene, which is distinct from the c-yes gene.

Splicing positions of c-*fgr* are identical with those of c-*src*. The coding sequence of the chicken c-*src* gene is interrupted

ACTTTCTGGCTTCTTCCCAG

TCCACACTATGGTCCCCCAG

Exon

no.

4 (99)

5 (104)

6 (150)

7 (156)

8 (180)

9 (77)

10 (154)

GCAGCT

GTAGCT

Exon Intron No. Sequence Sequence No. Sequence GTTCTGTTCTGTGCCTACAG 3 TGAAGG GTAAGTAGGGATTGGGGCAA 4 (4,100) 4 TGAAGA TTGCCTGCCTTTCCCAACAG GTGGTA 5 CCAAAG **GTAGGGGTGGTGCCACCCC** 5 (179) AAAAGTGATCCTCTCCACAG GTGCCT ACATGG 6 **GT**GAGGGCAGGGGCCTCAGA 6 (982) CTTCATGACCCCTCCCCTAG AGGTGA ACAAGACAGCCTCCGAGCAG 7 GGCTGG **GTACGGAGCTCCCGGGGGGCC** 7 (74) GCAC GT

TABLE 1. Nucleotide sequences around the splicing sites of the human c-fgr gene^a

^a Numbers in parentheses indicate lengths of exons and introns, in base pairs. Introns have the same number as the exons on their 5' side. The lengths of introns were estimated by nucleotide sequence determination and restriction mapping. ND, Not determined.

8 (480)

9 (265)

10 (900)

											V-I	<u>gr</u>	acti	in-													
	exon4	•								-	118	e p	ort	Ion	ר												
	GluGl	yAsp	Tro	TrpG	luA	laAr	gS∈	erLe	uSe	Ser	GlyL	.ysT	hrGl	lyCys	lle	noS	erAs	snΤ	/rVa	alAl	aPr	oVa	lAs	pSe	rIl	eGln	(30)
c-fgr;	TGAAGC	TGAC	TGG	TGGC	AGG	CTCC	GIC	TCT	CAG	CTCC	GGAA	AAA	CTG	3CTG	CATT	CCA	GCA	ACT/	CGI	ICCC	2000	TGI	TGA	СТС	AAT	CCAA	91
v-fgr;	CCCCA4	GC	AAC	СТ	A	AGAI	A	A	GT	ATG	FICE	G	CT	Γ AA	~			-	-		-		G		C	_ G	933
v-yes;	G	С			A	AA	A	CA	TCC	I'A G			А	А	C	А		т	Т	А	Т	C	.А		C	Т	1003
		e	xon	5																							
	AlaCi	nci	- 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Ture	heC	IVIA	/s T]	ഹി	var	al.vs	٨٩٩٨	laG	יאיו	raGli	leul	euS	erPr	-oG		snPr	-0G1	nGl	VAI	aPh	eie	uIle	(60)
c-for·	ATTG/	AGAC	TGG	TACT	TTG	GAAA	GAT	TTGG	GAG	AAAG	GATC	CAG	AGA	GCA	CTG	TTT	CAC	CAG	х. ХА	ACCO	CCCA	GGG	GGC	CTT	TCT	CATI	181
v-for:	00104									G		G				С	C		G	G	C	;			c	G	1023
v-ves:	G					т	А	G	с	G			А	A T	A T	r aa	AT	т	G	Æ	AG C	Т	ТАТ	Т	СТ	AG A	1753
						ř	XOI	10	•																		
	ArgGl	luSer	Glu	Thr	hrL	ysGl	LYA]	laTy	۳Se	rLeu	Serl	lleA	rgA	spTr	oAsp(GlnT	hrAi	rgG.	lyAs	spHi	LsVa	lLy	'sHi	.sTy	rLy	sIle	(90)
c-fgr;	CGGG	AAAGC	GAG	ACCA	ACCA	AAQ	JOC	CTA	CTC	CCTG	TCCA	ATCC	:GGG/	ACTG	GAT	CAGA	CCA	GAG	3CG/	ATCA	AIGI	GAA	GCA	TTA	CAP	GAIC	2/1
v- <u>fgr</u> ;	Т	GI		Т	-			_		_			A	Т	C	jG	-		~	С,				c		A	1042
v-yes;	AA	G	A		T			Т	~	T		- A	T Y T				1		۰ <u>~</u>		c			ĉ		A	612
c-src;						-1			G	C	AC	5 61	CI	10	. 0	1 CG		40	Ų.	ICA	C			C			012
																								e	xor	17	
	AraL	/sLeu	ıAso	Met	GlvG		πT	w I l	leTh	rThr	Arg\	/al0	lnP	heAsi	' Ser	/alG	lnG	luLe	euVa	alG	lnHi	isTy	niic	etd	uVa	alAsr	(120)
c-for:	CCCA	ACTO	GAC	ATG	ccc	GTT	CT	CAT	CAO	CACA	cccc	TTC	AGT	TCAA	TCG	TOC	AGG	AGC	IGG	ICC/	ACC/	CT/	CAT	TOGA	GGI	GAAT	361
v-for;		G		С	Т	С	т					œ			Α								G	т		C	2 1203
v-yes;	ΑΑ	3	r	AT	Т	А		т		A C	ΑA	CA	Α	ΤG	A T	СС	А	т		А			C	A	ACP	ATGC	1933
c-src;		G		œ		СЛ	Г			тС	Cł	ACC			AGC	2	С			G	υT		π	ХA	ACA	AC GCC	702
	1		·····) and	T	T 1				-	110	4~+ T			o The set		1	~~~^	1 - 7 -		~ ~ ^ 1	- م			~~~		(150)
	ASPG.	y Lei	k ys	ASIL	eur	eu	IEA.		o cy	SINC	11er	Met L	JYSP	COCM		2000	TATA COC	DUA.		NCC	2'DHT	γm	r_{r}	CAT	.езе тм	2	y (150)
C-IGT;	GALG	n Ju	SIGC	AACC	TGC	ICAI	1000 7x		CIG	CALL	C	4164	ACC	CGCA	UJALU A	~100	uu	1.004	UM	400/	HUG		3000		CM		1293
v-1gr;	1 T	<u>م</u> '	`т	сти		A (^	m cr	r۵	T	ca	2	Δ	۵	<u>۸</u>	ר ג	Δ	Δ	Δ	Δ	т			Δ	πr	דא (2023
C-STC:	Ť	ĉ	1	c II	ĥ		~ A	CGI	ing i	c	CG		~	~	- п	A	~	~	~		Ť			••	ά	т	5 792
• === /	•	•		-						-					-			- 0									
																	exo	nð	•								
	SerSe	erIle	Thr	Leu	GluA	rgAi	rgLe	euG]	lyTh	rGly	CysI	PheC	SlyA	spVa	lTrp	LeuG	lyT	hrT	rpA	snG.	lySe	erTt	۲Ŀ	/sVa	lAl	laVa	(180)
c-fgr;	AGCT	CAT	CACG	CTG	ACC	ccc	30C	IGGO	CAC	naar	TOT	PTCC	m.	ATTY	CTTCO	TIGG	GCA	CGT	GGA	ACG	3CA	3CAC	TA?	ACT	GGC	COGIN	541
														AIGI	0100									~~.			
v-for;				α	2			A					A	C	5100			Т		т		_	G		_		1383
v-fgr; v-yes;	GAG	тт	GG	α	AG	TTA	ΑT	A	CA	A A	T	Ţ	A T	C A	A .	A	A	T C		T T	A	2	G A	A	A	CA	1383 2113
v-fgr; v-yes; c-src;	gag gag	TT (GC (G G	α	AG G	TTA T A	АТ А	A	CA CA	A A G	т	T T	A T C	C A G	A ,	4 4	A G	T C C		T T	А (Т (2	G A C (А 3	A	CA (CA (1383 2 2113 2 882
v-fgr; v-yes; c-src;	gag Gag	TT (GC (G G	α	AG G	TTA/ TA/	АТ А	A	CA CA	A A G	T	T T	A T C	C A G	A .	4 4	A G	T C C		т	А (Т (2	G A C (A G	A	CA (CA (1383 2113 2882
v-fgr; v-yes; c-src;	GAG GAG LvsTi	TT (GC (G G CG	CC Pro(AG G	TTA TA	A T A	A	CA CA CA	A A G SAla	T	T T	A T C	C A G	A A	A A ValM	A G	T C C VSL	euL	T T	A (T (isAs	G A C (A G	A	CA C CA C	1383 22113 2882
v-fqr; v-yes; c-src; c-far;	GAG GAG Lysti AAGA	TT (GC (urLeu	G CG Lys	ProC	AG G GlyT	TTA T A	A T A etSe	A erPr	CA CA CA	A A G SAla	T Phei	T T LeuC	A T C	C A G IUAL	A aGln GCAG	A A ValM GTCA	A G letL	T C C YSL	euL	T T EuAi	A (T (rgH:	isAs	G A C (SpL)	A G VSLe	A euVa	CA C CA C alGli	1383 2113 2 882 1 (210) 5 631
v-fgr; v-yes; c-src; c-fgr; v-fgr;	GAG GAG Lystî AAGAG	TT (GC (urLeu CGCTO	G CG Lys GAAC	CC ProC	AG G GlyT GCA A	TTA T A hrife CCA	A T A etSe IGIO	A erPr	CA CA roLy CGAA	A A G SAla GGCC	T Phei TTCC	T T Leu(A T C	C A G IuAl	aGln'	A A Valm GTCA	A G letL TGA	T C C YSL	euL TGC	T T euAi TGC	A (T (rgH: GGCi	isAs ACG/	G A C (SpL)	A 3 ysLe	A euVa IQGI	CA (CA (algh IGCA	1383 2113 2882 1 (210) 5 631 1473
v-fgr; v-yes; c-src; v-fgr; v-fgr; v-yes;	GAG GAG Lysti AAGAG	TT (GC (urLeu CGCT(A :	G CG Lys GAAC	OC ProC CCGC	AG G GlyT SOCA A T	TTA T A hrife CCA A	A T A etSe IGTO	A erPr CCCC	CA CA roLy CGAA	A A G SAla GGCC	T Phei TTCC	T T LeuC CTGC	A T C SluG	G G IUAL	aGln' GCAG	A Valm GTCA A	A G letL	T C C ysL AQC A	euL TGC AAT	T T TGCO A	A (T (rgH: GGC/ A	isAs ACG/	G A C SpLy ACAV	A G YSLe	A euVa IGGI T	CA C CA C algh IGCA T C	1383 2 2113 2 882 1 (210) 5 631 1473 A 2203
v-fgr; v-yes; c-src; v-fgr; v-fgr; v-ges; c-src;	GAG GAG Lystî AAGAG	TT (GC (DrLeu CGCT(A C	G G G G G A A G A A G A A	ProC CCGC TT T	AG G GlyT SOCA A T	TTA T A hrife CCA G	A T A etSe IGTO A	A erPr CCC IG T	CA CA roLy CGAA CG AG	A A G SAla GGCC	Phei TTCC	T T LeuC CTGC TC	A T C	G IUAL	aGln' GCAG T	A A Valm GTCA A	A G letL TGA	T C C YSL AQC A	euLa TGC AAT A	T T TGC A A	A (T (rgH: GGC/ A	isAs ACG/ T T	G A C SipLy CAV G	A 3 ysLe	A euVa rGGI T	CA (CA (alGh IGCA(T C)	1383 2 2113 2 882 6 (210) 5 631 1473 A 2203 972
v-fgr; v-yes; c-src; v-fgr; v-fgr; v-yes; c-src;	GAG GAG Lysti AAGA(A	TT (GC (UTLEN CGCTO A C	G G G G G A A G A A G A A G A	ProC CCGC TT T	AG G G I J J J J J J J J J J J J J J J J	TTA	AT A PtSe NGTO A	A erPr CCC IG T	CA CA COLY CGAA CG AG	A A G SAla GGCC A I	Phel TTCC C	T T LeuC CTGC	A T C	G IUAL	aGln GCAG T	A A Valm GTCA A	A G letL TGA	T C C YSL AGC A A O C A O C A O C A O C A O C A O C A O C A O C A O C A O C A O C A O C A O C A O C C C C	euLa TGC AAT A	T T TGCO A A	A (T (rgH: GQC/ A	isAs ACG/ T T	G A C SpLy CAV G	A G VSLe	A NGGI T	CA C CA C algli IGCAC T C	1383 2113 882 (210) 631 1473 2203 972
v-fgr; v-yes; c-src; v-fgr; v-fgr; v-yes; c-src;	GAG GAG Lyst1 AAGAG A	TT (GC (nr Leu CGCT(A C	G G G G G G G G G G G G G G G G G G G	ProC CCGC	AG G GlyT SOCA A T	TTA	A T A PetSe IGTC A	A erPr CCC IG T	CA CA roLy CGAA CG AG	A A G SAla GGCC A I	T Phel TTC C	T T LeuC TGC TC	A T C	G IUAL	aGln GCAG T	A ValM GTCA A	A G IetL TGA	T C C YSL AQC A A C A A C A A C A A C	euL TGC AAT A	T T TGC A A	A (T (rgH: GGC/ A	isas ACG/ T T	G A C SpLy CAV G	A 3 VSLe	A euVa rGGI T	CA C CA C alGlr IGCA T C	1383 2113 882 (210) 631 1473 2203 972
v-fgr; v-yes; c-src; v-fgr; v-fgr; v-yes; c-src;	GAG GAG LysTi AAGAG A LeuT	TT (GC (urLeo CGCTO A C C	G CG Lys GAAC C A	Proc CCGC TT T Vals	GlyT GCA A T	TTA T A hrife CCA G	A T A etSe IGTO A'	A erPr CCC IG T	CA CA CA CG CG AG	A A G SAla GGCC A 1 rile	T Phel TTCC C	T T Leu CTGC	A T C SluG SAGG	G IUAL AGGC	aGln' GCAG T C	A ValM GTCA A HisO	A G IEtL: TGA	T C C ysL AGC A A C A C	euL TGC AAT A	T T TGC A A A	A (T (rgH: GGCi A	isAs ACG/ T T	G A C C SPLy ACA/ G	A 3 YSLe AGCI	A euVa rGGT T	CA C CA C algh IGCA T C	1383 2113 2882 1 (210) 3 631 1473 4 2203 972 1 (240) 2 721
v-fgr; v-ves; c-src; v-fgr; v-fgr; v-yes; c-src; v-fgr; v-yes; c-src;	GAG GAG LysTh AAGAG A LouT CTGTZ	TT (GC (UTLev CGCT(A T C	G G CG Lys GAAC F A NaVal	ProC CCGC TT T Vals	AG G G G G G G G G G G G G G G G G G G	TTA T A hrife CCA G G iluG	A T A PetSe IGTO ACCO	A erPr CCCC T T roll	CA CA COLY CGAA CG AG	A A G SAla GGCC A I rIle CATC	T Phei TTC C	T T Leu CTGC TTG (Ihr C	A T C SluG SAGG	C A G IUAL AGGC	aGln GCAG T tCys GTGT	A ValM GTCA A Hiso	A G ITGA exor IYS CCA	T C C YSL AGC A A C A C C T C C T	euL TGC AAT A euL	T T TGCO A A A EUA	A (T (T (GGC/ A spPl ATT	isAs ACG/ T T	G SpLy G G EQLy ICA	A G AGCI YSA:	A ReuVa T T SnPi ACCC	CA C CA C alght TOCAC T C/ roGht CAGAC	1383 22113 882 1 (210) 5 631 1473 4 2203 972 1 (240) 5 721 1563
v-fgr; v-yes; c-src; v-fgr; v-yes; c-src; v-fgr; v-fgr; v-fgr; v-fgr; v-fgr; v-fgr; v-fgr; v-fgr; v-fgr; v-fgr; v-fgr; v-fgr; v-yes; c-src;	GAG GAG LysTi AAGAG A LeuT <u>,</u> CTGTZ	TT (GC (UTLEN CGCT(A C C VrAla ACGC(T	G G CG CG CG C AVal CGTC	ProC CCGC TT T Vals GTGT C	AG G G G G G G G G G G G G G G G G G G	TTA T A hrif CCA G G iluG	A T A etSe IGTO AC AGCO	A erPr CCCC IG T roll CCAT	CA CA COLy CGAA CG AG LeTy	A A G SAla GGCC A 1 rile CATC	T Phel TTCC C Val	T T Leu(CTGC T C T C T C T C C T C C T C C T C C T C C T C C T C C T C C T C C T C C T C C T C	A T C SluG SAGG SAGT	C A G luAl AGGC	aGln GCAG T tCys GTGT C	A ValM GTCA A A HisG CACG A A	A G ITGA ITGA IYS GCCA T	T C C YysL AQC A A C A C C A C C T	euL TGC AAT A FeuL TGC	T T TGCO A A A TGG	A (T (GGCi A SpPI ATT G	isAs ACG/ T T ITCI C	G A C C SPL S C A C A C A C A C A C A C A C A C	A G VSLe VGC1 VSAS	A ReuVa RGGT T SnPi SCCC AG	CA C CA C alght IGCAC T C/ roght CAGAC	1383 22113 882 1 (210) 5 631 1473 4 2203 972 1 (240) 5 721 1563 4 2293
v-tgr; v-ves; c-src; v-tgr; v-ves; c-src; v-tgr; v-ves; c-src; v-tgr; v-ves; c-src; v-tgr; v-ves; c-src; v-tgr; v-ves; v-	GAG GAG LysTi AAGAC A LeuT CTGT	TT (GC (In Lev CGCT(A C C VrAla C C T	G G CG DLys GAAG T A NVal CGTG T	ProC ProC CCCCC T T Vals CCCCC T T	AG G G G G G G G G G G G G G G G G G G	TTA/ T A/ hrife CCA: A G iluG: AQG/	A T A etSe IGTO A LuPo AGCO A	A PrPr CCCC IG T roll CCA9	CA CA CA CG CG AG CG AG CTA	A A G SAla GGCC A I rIle CATC A	T Phel TTCC C Val: GTG	T T LeuC CTGC TTG (Ithr(ACCC A G G	A T C Slug SAGG SAGG	C A G luAl AGGC heMe TCAT	aGln GCAG T C tCys GTGT C ACA A C	A ValM GTCA A A HisG CACG A A G	A G ITGA Exco ITSA T G G	T C C YSL AGC A A C A C A C A C T	euL TGC AAT A euL TGC A	T T TGCO A A EUA: TGG A	A (T (T (GGCi A SpPI GCC C	isAs ACG/ T T C C C	G A C C SPLy G G ULy G G	A 3 YSLe AGC1 YSA2 AGA/ G G	A CGC T SnPi ACC GGC	CA C CA C alghr IGCAC T C T C CAGAC	1383 2113 882 1 (210) 3 631 1473 972 1 (240) 3 721 1563 A 2293 4 1062
v-fgr; v-yes; c-fgr; v-fgr; c-src; c-fgr; v-ges; c-src; v-fgr; v-yes; c-src;	GAG GAG LysTf AAGA A LeuT CTGT/ T	TT (GC (Or Let CGCTC A C C V(Ala CGCCC T T T	G G CG GAAG GAAG T AVal CGTG T	Proc CCCC TT T Vals GTCT C T T	AG G G G G G G G G G G G G G G G G G G	TTA T A hrif CCA G G G G G G G G G G G G G G G G G G	A T A IGTV AC AGCC A	A erPr CCCC T T roll CCAT A	CA CA COLy CGAA CG AG LETY CCTA T	A A G SAla GGCCC A T rIle CATC	T Phel TTCC C Val GTG	T T Leuk CTGC T(C Thr C ACCC A G G	A T C Slug SAGG SAGG	C A G IuAl AGGC heMe TCAT	A A A A A A A A A A A A A A A A A A A	A ValM GTCA A A CACG A A G	A G ITGA ITGA IYS GCA G	T C C YSL AGC A A C A A C T T	euL TGC AAT A euL TGC A	T T TGCO A A A EQUA: TGG: A	A (T (T (GGC/ A SpPI ATT G C C	isAs ACG/ T T ITC/ C C	G A C C SPLy ACA/ G G C G G	A G VSLe VSLe VSLE VSLE VSLE VSLE VSLE VSLE VSLE G G G	A euVa IGGJ T snPi ACC AGC SGG/	CA (CA (CA (AlGIn TCA TCA TCA TCA TCA TCA CAGAC	1383 2113 882 1 (210) 5 631 1473 972 1 (240) 5 721 1563 A 2293 A 1062
v-fgr; v-yes; c-src; v-fgr; v-yes; c-src; v-fgr; v-yes; c-src; v-fgr; v-yes; c-src;	GAG GAG LysTf AAGA A LeuT CTGTZ T	TT (GC (A C C VCALa ACGOC T T T	G G CG DLys GAAC C A AVal CGTC T	ProC CCCC T T Vals CCCCC T T	AG G G G G G G G G G G G G G G G G G G	TTA T A hrife CCA: A G sluG: AGG/	A T A PetSe IGTO ACC A	A PPPP CCCC IG T T T T T T T A	CA CA COLy CGAA CG AG LeTy CCTA T	A A G SAla GGCC A T rIle CATC	T Phel TTCC C Val GTG	T T Leu(CTGC TTGC (MCCCC A G	A T C Slug SAGG SAGG A C	C A G IuAl AGGC heMe TCAT A Xonl	aGln' GCAG T C C C STGT C ACA A C	A ValM GTCA A A Hiso CACG A A G	A G ITGA ITGA IYS GCA T G	T C C AGC A A C A C A C C T T	euL TGC AAT A euL TGC A	T T TGCC A A A EUAJ TGGC A	A (T (T (GGC/ A SpPI ATT G C C	isAs ACG/ T T NeLe ITC/ C	G A C C S PL J G G G G G	A G VSLe VSLe VSLE VSLE VSLE VSLE VSLE VSLE VSLE VSLE	A PeuVa TGGT T SnPi ACC SGG	CA (CA (AlGIn ROCA T C T C CAGA AG G G G G G G G G G G G G	1383 22113 882 1473 4203 972 1(240) 721 1563 42293 41062
v-fgr; v-yes; c-src; v-fgr; v-fgr; v-yes; c-src; v-fgr; v-yes; c-src; v-fgr; v-yes; c-src;	GAG GAG LysTi AAGA A LeuT CTGTZ T GlyG	TT (GC (A C C V(Ala ACOOC T T T	G G CG DLys GAAG C A CGTG T T DLeu	ProC ProC CCCC T T Vals CTCT T T ArgI	AG G G G G G G G G G G G G G G G G G G	TTA TA TA TA TA TA CCA: A G (luG: AGG/ 2roG.	A T A PetSe IGTO A A C C A A A C C A	A PerPriccoo IG T roll CCAN A	CA CA COLY CGAA CG AG CG AC T T	A A G SAla GGCC A T TILe CATC A PMet	T Phel TTICC C Val GTG C C	T T Leuk CTGC TTGC MCCCC A G Alak	A T C SluG SAGG SLUP SAGT A E SLUP	C A G IuAl AGGC heMe TCAT A xonl alAl	A aGln GCAG T C C C C C C C C C C C C C C C C C C	A ValM GTCA A A HisG CACG A A A GlyM	A G ITGA ITGA IYS GCAW T G	T C C YysL AGC A A C A A C T T T IaT	euL TGC AAT A euL TGC A	T T T T T T T T C C A A euA: T C G A A etG	A (T (rgH: GGC A SpPI GC C C	isAs ACG/ T T C C C	G A C SPL ACA G G C C C C C C C C C C C C C C C C	A G VSLe VGC1 VSAs G G G G SnTy	A euVa IGGI T SnPi ACCO J AGC J AGC J Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z	CA (CA (CA (AlGIn ROCA T C CAGA AG AG AG AG AC AL EHI:	1383 22113 882 1 (210) 5 631 1473 4 2203 972 1 (240) 5 721 1563 4 2293 4 1062 5 (270)
v-fg;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;	GAG GAG LysTh AAGAG A LeuT CTGTZ T GlyG	TT (GC (nrLev CGCTC A C C V(Ala ACGOC T T T T S LINASE	G G GCG DLys GAAG C A N N CGTG C T C CTTG C TTG	ProC ProC CCCC T T Vals CTC T T ArgI	AG G G G G G G G G G G G G G G G G G G	TTA TA TA TA TA TA TA TA TA TA TA TA TA	A T A etSe IGTO A GCO A A A A A A A T	A erPr CCCC IG T roll CCAT A euVa	CA CA COLY CGAA CG AG CG AG T T	A A G sAla GGCC A T rIle CATC A pMet CATC	T Phel TTCC C Val? GTG C C Alai	T T Leuk CTGC TTGC TTGC TTGC G ACCCC A G G	A T C SluG SAGG SluP SAGT A C SlnV CAGG	C A G Iual AGGC heffe TCAT A Xonl alal	aGln GCAG T C C ACA A C aGlu TCAG	A ValM GTCA A A HisG CACG A A A G GQCA	A G ITGA ITGA ITSS ITSS ICCA T G IELA	T C C YSLL AGC A A C C T T T IaT	euL TGC AAT A euL TGC A	T T T T T T T T C C A A euA: T G G C C A A E t G G	A (T (rgH: GGC/ A SpPI G C C LUA:	isAs ACG/ T T C C C SCAT	G A C C SpLy G G C A C A C A C A C A C A C A C A C	A ysLe ysA G G G SnTy ACT/	A euVa IGGI T snPi ACC AGC GGG/ yrII ACA	CA (CA (alGhr TGCA T Ci TGCA AG G i ICAGACi IeHi:	1383 22113 882 1 (210) 5 (31) 1473 4 2203 972 1 (240) 5 721 1563 4 2293 4 1062 5 (270) 5 811
v-fg; v-yes; c-src; v-fg; v-yes; c-src; v-fg; v-yes; c-src; v-fg; v-yes; c-src; v-fg; v-yes; c-src; v-fg; v-yes; c-src; v-fg; v-yes; c-src; v-fg; v-yes; c-src; v-fg; v-yes; c-src; v-fg; v-yes; c-src; v-fg; v-yes; c-src; v-fg; v-yes; c-src; v-fg; v-yes; c-src; v-fg; v-yes; c-src; v-fg; v-	GAG GAG Lystf AAGA A Leut CTGT T T GlyG GGCC	TT (GC (nrLev CGCTC A C C V(Ala ACGCC T T T T S LINASE	G G GCG DLys GAAG C A AVal CGTG T T DLeu TTTG	ProC ProC CCGC T T T Vals CGTGT C T T T ACGC CA	AG G G G G G G G G G G G G G G G G G G	hrife CCA: A G HuG: AGG/ ProG.	A T A PetSe IGTO AGCO A AGCO A AGCO A G	A erPr CCCC IG T roll CCAT A euVa	CA CA CA CG CGAA CG AG CG AG T T	A A A G SAla GGCC A T rIle CATC PMet CATC	T Phel TTCC C Val? GTG C C Alai GCA	T T Leux CTGC TTC (Ithr(ACCCC A G G Alac G	A T C Slug Slup SAGG A C Slup C AGT A C C AGT	C A G Iual AGGC heffe TCAT A Xonl alal TAGC	aGln GCAG T C C C C C C C C C C C C C C C C C C	A A GTCA A A A CACG CACG A A A G GCA	A G ITGA ITGA ITSA ITSA G G IELA	T C C YSL AGC A A C C T T T IaT CCT	euLi TGC AAT A euLi TGC A yrM	T T T T T T T C C A A E U A T C G A E t G	A (T (T (GGCi A SpPI A ATT G C C C IUA:	isAs ACG/ T T T ITCI C C C	G A C C SpLy G G CA G G S CA G G CA G G G G G G G G G G G G G G G	A 3 ysLe AGC1 ysAs G G G G SnT) ACT/	A SuVa T T SuPi ACC J ACC J ACC J T	CA (CA (alGh TGCA T C CAGA G G IeHi: TTCA C	1383 2113 882 1 (210) 5 631 1473 972 1 (240) 5 721 1563 A 2293 972 1 (240) 5 721 1 563 A 2293 5 (270) 5 811 1653
v-tai	GAG GAG LysTf AAGAG A LeuT CTGTZ T GlyG GGCC GA	TT (GC (GC (A) C C V(Ala ACGOO T T T C InAsp ACGOO T	G G CG DLys GAAC T A AVal CGTC T DLeu TTTC	ProC ProC CCCC T T Vals CCCC T T T ACCC CA	AG G G G G G G G G G G G G G G G G G G	TTAI TAI TAI TAI TAI CCA: A G G G G G G G G G G G G G G G G G G	A T A PetSe IGTO A C C C C C C	A PerPr CCCC IG T roll CCAI A A PeuVa ICGI	CA CA CA CG CGAA CG AG ICTA T T	A A A G SAla GGCC A T rIle CATC PMet CATC	T Phel TTCC C Val' GTG C C Alai GCAC	T T Leuk CTGC TC CTGC TC C TGC TC C TGC C T C T	A T C Slug SAGG SAGG C SluP SAGT A C SluV C AGG	C A G G IUAL AGGC	A aGln GCAG T C C C C C C C C C C C C C C C C C C	A A GTCA A A A CACG A A A G G G C A G G G C A	A G ITGA ITGA ITSS IGCAU T G IetA	T C C YysLJ AGC A A A erL GCT T IaT CCT. T	euLi TGC AAT A euLi TGC A yrM	T T T T T T T C C A A A euA: T T C G T T C T T	A (T (T (GGCi A SpPI G C C C LUAI AAC(G A	isAs ACG/ T T heLc C C C rgMs CAI	G A C C SpLy G G C C C C C C C C C C C C C C C C C	A 3 VSLe VGCT VSAS G G G G SnT)	A euVa rGGJ T snPi VCCC AGC GGGJ VTI T	CA (CA (AlGIn ROCA T C/ T C/ T C/ AGAC I I I I I I I I I I I I I I I I I I I	1383 22113 882 1(210) 631 1473 4203 972 4(240) 721 1563 42293 41062 5(270) 811 1653 2383
v-fgi; v-yes; c-src; v-fgi; v-ge; v-fgi; v-fgi; v-fgi; c-src; v-fgi; v-fgi; v-fgi; c-src; v-fgi; v-ge; c-src; v-fgi; v-ge; c-src; v-fgi; v-ge; c-src; v-fgi; v-ge; c-src; v-fgi; v-ge; c-src; v-fgi; v-ge; c-src; v-fgi; v-ge; c-src; v-fgi; v-ge; c-src; v-fgi; v-ge; c-src; v-fgi; v-ge; c-src; v-fgi; v-ge; c-src; v-fgi; v-ge; c-src; v-fgi; v-ge; c-src; v-fgi; v-ge; c-src; v-fgi; v-ge; c-src; v-fgi; v-ge; c-src; v-fgi; v-ge; v-fgi; v-ge; v-fgi; v-ge; v	GAG GAG LysTf AAGAG A LeuT CTGTZ T GlyG GGCCZ GA A	TT (GC (GC (A C C C V(AL) A C C T T T T T T T T T T T T T T	G G CG CG CAAG CAAG CGTG T CGTG CGTG C C C	CC Prot CCCCC T T Vals GCC T T T ACCC CA AAA C	AG G G G G G G G G G G G G G G G G G G	TTA/ TA/ TA/ TA/ TA/ TA/ TA/ TA/ TA/ TA/	A T A IGTO A IGTO A G G G C G C	A erPr CCCC IG T roll CCAN A	CA CA COLy CGAA CG AG CG AG T T	A A A G SAla GGCC A T rile CATC PMet CATC	T Phel TTCC C Val' GTG C C Alai GCAC T T	T T Levic CTGC TTG C TGC T T C C C C C C C C C C	A T C Slug AGG AGG AGT A SluP C AGG AGT A A A A A	C A G G heMe TCAT A Xon1 TAGC T C T	A aGln GCAG T C C C C C C C C C C C C C C C C C C	A A GTCA A A A HisG CACG CACG A A A G G G G CA CACA	A G ITGA TGA G G HetA	T C C YSL AGC A A C A A C C T T C C T C T C T C C C C	euLi TGC AAT A ■ euLi TGC A yrM ACA	T T T T T T T T T T T T T T	A (T (T (GGCi A SpPl A TT G C C C IUA: A ACC G A G	isAs ACG T T heLa C C C C C C C C C C C C C C C C C C C	G A C SPL) G G G C C A C A C A C A C A C A C A C A	A J VSLe VSLe VSLE	A suVa T T snPi VCC J AGC J Z T T G	CA (CA (AlGIn ROCAU T C/ T C/ T C/ TCAU AGACI LeHi: TTCAU C C C C	1383 22113 882 1(210) 5 631 1473 4 2203 972 1(240) 5 721 1563 4 2293 4 1062 5 (270) 5 811 1653 2383 1152
v-fgi; v-gs; c-fgi; v-f	GAG GAG LysTf AAGAG A LeuT CTGT T T GlyG GGCC GA A	TT (GC (A C C C C C C C T T T T T T T T T T T T	G G CG CG CG CG CG CG CG CG CG CG CG CG CG	ProC CCCC T T Vals CCCCC C T T ACCC CA AA C	AG G G G G G G G G G G G G G G G G G G	TTA/ T A/ T A/ hr/% CCA: A G G G G G G G G G G C CCC/ A T	A T A EtSe IGTO A C A C C C C	A erPr CCCC IG T roll CCA3 A euVa IGGI	CA CA COLY CGAA CG AG CG AG T T	A A A G SAla GGCC A I rile CATC A PMet CATC	T Phel TTCC C Val? GTG C C Alai GCAC T T	T T Levic CTOC TTC C Thr C C Thr C C C T T T T	A T C GluG GAGG GAGG GAGT A C GluP GAGG A A A A A A	C A G I UAL AGGC	aGln GCAG T C C C C C C C C C C C C C C C C C C	A A GTCA A A HisG CACG A A A G G G CACG G G CA CACA A G G G CA	A G ITGA ITGA ITSS GCA T G HetA	T C C YSLI AGC A AGC A A C C T C C T C C T C C T C C T C C C C	euL TOC AAT A euL TGC A yrM ACA	T T T T T T C C A A A E UA: T C G A T C G T T T	A (T (T (GGC A SpPI A A C C C IUA: A A C C C IUA: G G G	isAs ACG/ T T T NeLa C C C C C C C C C C C C C C C C C C C	G A C SpL) G G C C C G G C C C G G C C C C C C C	A S YSLe AGC1 YSAs AGC4 G G G SnT) ACT7	A SuVa T T SnPi CO AGC AGC T T G	CA (CA (AlGh TCCA) T Ci CAGAI AG 3 A AGACI C C C C C	1383 22113 882 1 (210) 3 631 1473 4 2203 972 1 (240) 5 721 1563 4 2293 4 1062 5 (270) 5 811 1653 2383 1152
v-to: v-to:	GAG GAG LysTf AAGA A LeuT CTGT T T GlyG GGCC GA A	TT (GC (GC (CGCT(A) C C C C C C T T T T T T T T T C TT(T C T T C TT(C C C C	G G G G G G G G G G G A G	ProC CCCC T T Vals CCCCC T T T ACCC CA AAA C	AG G G G G G G G G G G G G G G G G G G	TTAA TTAA TTAA hrife CCA: A G G G G G G G G G C CCC A T T Sp I	A T A etSe IGTO AC AGC A C G G G C	A PrPr CCCC IG T roll CCA A PuVa PuVa PuVa	CA CA CA COLY CGAA CG AG CG AG T T ALAS T CTA	A A A G SAla GGCC A I rile CATC A PMet CATC	T Phel TTCC C Val? GTG GTG GCA T T	T T T LeuC CTGC TGC TGC TGC TGC T T T T	A T C SluG SAGG SluP SAGT A SINV CAGG A A A	C A G I UAL AGGC	aGln GCAG T C C C C C C C C C C C C C C C C C C	A ValM GTCA A A HisG CACG A A G GGCA	A G etL: TGA Exo IYS GCA T G HetA	T C C C YySL AQC A A C C T C C T C C T C C T C C T C C C C	euLA TGC AAT A euLL TGC A YrM ACA G	T T T T T T T C C A A A euA: T C G C A T T C G T T T E UA	A (T (T (GOC A SpPI A A C C C IUA A A C C C IUA I A A C C C IUA	isAs ACG/ T T T C C C C C C C C C C C C C C C C	G A C S S S S S S S S S S S S S S S S S S	A S YSLe AGC1 YSA2 G G G G SnT) ACT7 LeLA	A euVa IGGI T AGC GGGI Yr II G G YSAA	CA (CA (AlGh TCCA T Ci CAGAA AG 3 A AGACi C C C C C C C	1383 22113 882 1 (210) 5 (31) 1473 4 (240) 5 (220) 972 1 (240) 5 (270) 5 (270) 5 (270) 5 (270) 5 (270) 5 (210) 5
v-tai v v v v v v	GAG GAG LysTi AAGA A A LeuT CTGTZ T GlyG GGCCZ GA ArgAs	TT (GC () A C C VCA1a ACGCC T T T T T T C SpLev	G G	CC ProC CCCC T T Vals C C Argli ACC C Alad A C C	AG G G G G G G G G G G G G G G G G G G	TTAA TTAA TTAA hrife CCA: A G G G G G G G G G G G G G G G G G G	A T A Pet Se IGTO AC A A C A C C C C C C C C C C C C C C	A PrPr CCCC IG T roll CCA A PuVa IGG7 PuVa PuVa PuVa	CA CA CA CA CG AG CG AG LETY CTA T T ALAS	A A A G SAla GGCC A I rile CATC A PMet CATC YGlu GGAG	T Phel TTCC C Val GTG GTG C C Alai GCAC T T	T T LeuC CTGC TG CTGC C TG G G G G G G G G C T T T T	A T C SluG SluP SAGG A SluP SAGG A A A A A A A A A A A A A A	C A G IUAI AGGC	A aGln GCAG T C C C C C C C C C C C C C C C C C C	A A GTCA A A Hiso CACG A A GGCA GGCA	A G ITGA exos ITS G G letA T G letA	T C C C YySLI AQC A A C C T C C T C C T C T C T C T C C T C	euLa TGC AAT A euLCA A ACA G GCT	T T T T T T T T C C A A A euAa T G G T T C G T C G T T T T C C C C A A A A C C C C C C C C	A (T (GGC/ GGC/ A SpPI GGC/ G C C LUA: G C C LUA: G G G G IUA: C C	isAs ACG/ T T T neLa TTC C C C rgMa CAI A G rgLa GTC	G A C C S S S S S S S S S S S S S S S S S	A G YSLe YSLE Y	A SuVa T T ACC ACC ACC T T G YSAS	CA (CA (CA (CA (CA (CACA) T C/ T C/ T C/ T C/ T C/ CACA/ C C C C C C C C SpAS; ACCA/	1383 22113 882 1 (210) 5 (311 1473 4 2203 972 1 (240) 5 721 1563 4 2293 4 1062 5 (270) 5 811 1653 2383 1152 5 (300) 5 901
v-fgi;;;;; c-si;;;; i;;;;;; v-fgi;;;;; c-si;;;;; i;;;;;;; c-fgi;;;;;; c-fgi;;;;;; i;;;;;;; v-fgi;;;;; c-fgi;;;;;; i;;;;;; v-fgi;;;;; c-fgi;;;;;; i;;;;;; v-fgi;;;;; c-fgi;;;;;; i;;;;;; v-fgi;;;;; c-fgi;;;;;; i;;;;;;;	GAG GAG LysTf AAGAG A LeuT CTGT2 T GlyG GGCC2 GA A ArgAs CGCG	TT (GC () A C C C V(Ala C C T T T T T T C T T T C T T C T T C T T C T T C	G G GCG ULys GAAG GAAG GAAG T DLeu TTTG C C C C C C C C	CC ProC CCCC T T Vals CC C ArgI AcC C A A A C C Ala/ C C	AG G G G G G G G G G G G G G G G G G G	TTA T A CT A CCA: A G G G G G G G G G G G G G G G G G G	A T A IGTO A A INLA G G C G C C ILELA	A PrPr CCCC IG T roll CCA3 A PuVa PuVa PuVa IGGT T	CA CA CA CG AG CG AG CG AG T T AlAs TGGA	A A A G SAla GGCC A I rIle CATC A PMet CATC YGlu GGAG	T Phell TTTCC C Val ¹ GTG C C Allai GCA T T Argli	T T T Leux CTOC T T C T T T T T Leux	A T C SluG SAGG SAGG SAGT A C SINV CAGG A A A LlaC T T	C A G IUAI AGGC AGGC AGGC AGGC AGGC AGGC AGGC AG	A aGln C C T C C C C C C C C C C C C C C C C	A A GICA A A A A A A A A A A A A A A A A A A	A G ITGA exo ITGA G G IetA T G G IetA	T C C VysL AGC A A A A A C C T C T C T C T C T C C T C C C C	euLa TGC AAT A euL TGC A YTM ACA G GCT GC	T T T T T T T T T T T T T T T T T T T	A (T (GGC) A SpPI A ATT G C C IUA: A ACC G G A A C C IUA: C C C	isAs ACG/ T T T heLa C C C rgMc C C C rgMc C A G C S T C	G A C C G G G G G G C A C A C A C A C A	A G YSLe YSLE Y	A SuVa T T SnPi ACC A ACC A C C T T G ySAs	CA (CA (CA (CA (CA (CA (CA (CA (1383 22113 882 1383 2213 882 1473 4203 972 1473 4203 972 1473 4203 972 1473 4203 972 1473 4203 972 1563 42293 41062 5(270) 5(1) 5(3) 1563 42293 41062 5(270) 5(1) 5(3) 1563 2383 1152 5(300) 7(30
v-terility v-vession v-terility v-vession v-terility v-vession v-terility v-vession v-terility v-vession v-terility v-vession v-terility v-vession v-terility v-vession v-terility v-vession v-terility v-vession v-terility v-vession v-terility v-vession v-terility v-vession v-terility v-vession ve	GAG GAG Lystf AAGAG A Leut CTGTZ T GlyG GGCCZ GA A ArgAs CGCGG A G	TT (GC (C C C C C C C C C C C C C C C C C C	G G GCG ULYS GAAG GAAG GAAG T PLEU TTTG C C A C C C C C C C C	Argi Argi Argi Argi Argi Argi Argi C Argi C Argi C Argi C Argi C C C C C C C C C C C C C C C C C C C	AG G G G G G G G C C G C C C C C C C C C	TTA/ T A/ CCA: A G (luG: AGG/ AGG/ A T T ssnII) ACA:	A T A PetSe IGTO A C C C C C C C C C C T	A PrPr CCCC IG T roll CCA3 A PuVa IGG7 T T T	CA CA CA CGAA CG AG ICTY T T AlAs ICTA A IGGA	A A A G SAla GGCC A I rile CATC A CATC YGlu GGAG A C	TTCC C Val ¹ GTG C C Alai GCA T T Argi C C AAT	T T T Leuk CTOC T T C T T T T T T T T T	A T C SluG SAGG SAGG SAGG SAGG A A A LlaC T T	A IGH A G IUAI AGGC heffe TCAT A Xon1 alAl TAGC T C YSLy GCAA	A aGln GCAG T C C C C C C C C C C C C C C C C C C	A ValM A A A A CACG A A GGCA GGCA AlaA GCAC T	A G ITGA exco ITS G G HetA T G G SoPPI ACT	T C C VysL AGC A A A A A C C T C C T C C T C C T C C T C C C C	euL TCC AAT A euL TCC A YTM ACA G G CT C CC TC	T T T T T T T T T T T T T T C C	A (T (T (GGCi A SpPl AATT G C C LUA: AACO G A G G A A C C C C A A	isAs ACG/ T T T heLc C C C rgMs G C A G G T	G A C C G G G G G G C A C A C A C A C A	A ysLee AGCT ysAs G G G G G C SnT) ACT/ IeL) AG	A SuVa T SnPi ACC ACC ACC ACC T C G VSAs ACC YSAs	CA C CA C CA C CA C C C C C C C C C C C	1383 22113 882 (210) 631 1473 4203 972 1(240) 5721 1563 42293 41062 5(270) 811 1653 4293 41062 5(270) 581 152 5(300) 5931 152 5(300) 5931 1743 2383 1152 5(300)
v-tai iiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiii	GAG GAG LysTf AAGAG A LeuT CTGT T T GlyG GGCC GA A A GGCG A G G	TT (GC (C C C C C C C C C C C C C C C C C C	G G GCG DLys GAAG C A AVal CGTG T C DLeu TTTG C C C C C C C C C C C C C C C	CC Prot CCCCC T T Vals CCCCC T T ACCCCA C ALAA C C ALAA C	AG G G G G G G G C C C C C C C C C C C C	TTA TTA TTA CCA: A G SluG: A CCC/ A T T snII ACA:	A T A PetSe IGTO A AGCO A A G G C G C G C T	A PrPr CCCC IG T F T ProII CCAT A PuVa IGGT T T T T	CA CA CA CA CG AG CG AG CG AG T T AlAs TGGA	A A A G SAla GGCC A I rIle CATC A PMet CATC YGlu GGAG A A A	T Phel TTCC C Val? GTG GTG GCAC T T ArgI CCGC	T T T T LeuX CTOC T T T T T T T T T	A T C Slup SAGG SAGG SINV CAGG A A A A CGT T T T	heme G heme TCAT A xonl TAGC T C ysLy GCAA	A aGln GCAG T C C ACA A C C C C C C C C C C C C C C C	A A A A A A A A A A A A A A A A C A C A	A G ITGA exco ITSA T G G IetA T G S S P P I S S P P	I C C YSLL C YSLL A C C A A A P erL C C T I a T C C T G C T C C C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C C C A C C A C C C A C C C A C C C A C C C A C C C A C C C A C C C A C C C C A C C C C A C C C A C	euL TCC AAT A euL TCC A YTM ACA G G C TC G C TC G C	T T T T T T T T T T T C C	A (T (T (GGCi A SpPl AATT G C C C IUA: AACO G A G G IUA: AACO C C A A T	isAs ACG/ T T T heLc C C C rgMs G C A G C G T G T G	G A C C C S S S S S S S S S S S S S S S S	A S VSLee VSLEE VSL	A SuVa T T SuPi ACC A CCCC A CCC A CCC A CCCCCC	CA C CA C CA C CA C T C C C C C C C C C C C C C C C C C C	1383 22113 882 1383 2213 882 1373 882 1473 4203 972 1473 4203 972 1473 4203 972 1473 42203 972 151 152 152 152 152 152 152 15
	GAG GAG LysTf AAGAA A LeuT CTGTZ T GlyG GGCCZ GA A CGCCG A G G G	TT (GC (CGCT(A C C C C C C C C C C C C C C C C C C C	G G GCG DLys GAAC T AVal CGTC T DLeu TTTC C C C C C C C C C C C C C C	ProC ProC CCCC T T Vals GTC C T T T Vals C C A A A C	AC G G G G G G G G G G G G G G G G G G G	TTAA T A CCA: A G (luG) ACA: ACA: SnII ACA:	A T A etSe IGTO A C C C C C C C C T	A PrPr CCCC IG T F T PuVa IGGT T T T T	CA CA CA CG AG CG AG T T AlAs AIGIA MTCG A G	A A A G SAla GGCC A I rile CATC A CATC YGlu GGAC A C A	T Phel TTCC C GTG GTG GTG C C C C Alai GCAC T T AArgl CGG	T T T T Leux CTGC T T CTGC T T T T T T T	A T C Slug SAGG SAGG A A A A A A A A A A A A T T T T	heme G heme TCAT A xonl alal TAGC T C ysLy GCAA	aGln GCAG T C C C C C C C C C C C C C C C C C C	A A GTCA A A A A A A GCACG A A A GCAC T C	A G TGA TGA IYS CCA T G LetA TGG SpP	T C C YSL AACC A A A A A A C C T T C C T T G C T T G C C T T G C C C C	euLl TGC AAT A euLl TGC A A A A A A A C I YIM G C TC G C C C	T T T T T T T T T T T T T T T C	A (T (T (GGCi A SpPI A A T G C C C IuAi G C C IuAi G G C C A A T	isAs ACG T T T C C C C C C C C C C C C C C C C	G A C C C S S S S S S S S S S S S S S S S	A ysLee ysAs G G G G G C SnT) LeL Y TG AG TG	A SuVa T T SuPi ACC A C C C T T G C A C A	CA (CA (CA (CA (A) T C) T C) T C) T C) T C) T C) T C) T C	1383 22113 882 1 (210) 5 (311 1473 4 (240) 5 (270) 5 (270) 5 (270) 5 (270) 5 (270) 5 (270) 5 (270) 5 (270) 5 (300) 7 901 1743 2473 1242
	GAG GAG LysTr AAGAC A LeuT CTGTZ T GlyG GGCCZ GA A ArgAa CGCGG A G	TT (GC (urLee COCTA A C C VCA1a ACGOO T T C T C T C T C T C T C T C T C T C	G G G GCG GAAG GAAG GAAG GCGTC I I CGTC CC I Arg GAGG CC CC I CC I	Argi Argi Argi Argi Argi Argi Argi Argi	AC G G G G G G G G G G G G G G G G G G G	TTAA T A CCA: A G (luG) ACA: ACA: SnII ACA:	A T A etSe IGTO A C C C C C C C C C C C C C C C C C C	A Pr Pri CCCC IG T roll CCAY A PuVz PuVz PuVz T T T	CA CA CA CA CG AG LETY CTA T T ALASA A G	A A A G SAla GGCC A I rile CATC A CATC CATC VGlu GGAG A C A	T Phel TTCC C Val C GTG C C C C C C C C C C C C C C C C C	T T T Leuk CTGC T C CTGC A Leuk C T T T Leuk T T	A T C Slug SAGG SLUP A A C SLUP A A C SLUP A A A A A A A A C T T T T	I ual G G Iual AGGC TCAT A Xonl TAGC T C T C C YSLy GCAA	A aGln GCAG T C C C C C C C C C C C C C C C C C C	A A GTCA A A A A A A GCACG A A A A C A C A C A	A G TGA TGA T G G LetA TGG SpP	T C C YSL AACC A A A A A C C T T C C T T G C T T G C C T T G C C C C	euLi TGC AAT A E ULI TGC A A A A A A A C I YLM G C TC G C	T T T T T T T T T T T T T T T C	A (T (T (GGCi A SpPI A A T G C C C IuAi G C C IuAi G C C C A A T	isAs ACG T T T ITC C C C G G G G G T G G T	G A C C S S S S S S S S S S S S S S S S S	A ysLee ysAs G G G G G C SnT) IcLy TG AG	A SuVa T T SuVa ACC Z ACC Z ACC Z C T G G Z SGCJ A C A C A	CA (CA (CA (CA (CA (AlG)) T C/ T C/ T C/ T C/ T C/ T C/ C C C C C C C C C C C C C C C C C C	1383 22113 882 1 (210) 5 (311 1473 4 2203 972 1 (240) 5 721 1563 2383 1062 5 (270) 5 811 1653 2383 1152 5 (300) 7 901 1743 2473 1242
	GAG GAG LysTf AAGAC A LeuT CTGT2 T GlyG GGCC2 GA A CGCGG A G GluT GluT	TT (GC (urLee CGCTV A C C VrAla ACGCC T T C C UnAsp Lee ACGCTV T C C T C C C C C C C C C C C C C C	G G G GCG GAAG GAAAG C A C GTC T T C C C C C C C C C C C C C C C C	Proc CCCC TT T Vals GTCT C T T Argli ACC C Ala/A C C Cys0 C	AG G G G G G G G G G G G G G G G G G G	TTAI TTAI TTAI CCA: A G G G G G G G G G G G CCC/ A T SnIJ ACA: (30)	A T A INTA A INTA A C C C C C C C C C C C C C C C C C C	A Pr Pr CCCC IG T roll CCAY A PuVa IGG7 T T T	CA CA CA CA CG AC CG AC CG AC T T ALASS A CG A G	A A A G SAla GOCC A T PILE CATC A CATC CATC CATC CATC CATC CATC C	T Phel TTCC C Val GTG GTG C C C Alai GCA T T ArgI CCCC	T T T Leuk CTGC T C CTGC A C CTGC T T T Leuk T T	A T C Slug SAGG A SluP SaGT A A C SlnV A C SlnV A C T T T T T	I ual G G heme TCAT A Xon1 alal TAGC T C ysLy GCAA T	aGln aGCAG T C C C C C C C C C C C C C C C C C C	A ValM GTCA GTCA A A A CACG GCA GCA GCA T C	A G TGA TGA T G G LYS G A T G SpPl ACT	I C C YSLL AQC A A A A A A A C T I a T C T G C T C C T C C C C AQC A A C C C AQC A A C C C C C	euLa TGC AAT A euL TGC A G G G C TC G C	T T T T T T T T C C	A (T (T (GGC) A SpPI ATT G C C C IUAI G C C IUAI G C C IUAI G C C A A T	isAs ACG T T T ITC C C C G G G G G G G G	G A C C S S S S S S S S S S S S S S S S S	A JSLee VSLEE VSLEE	A SuVa T T SuVa ACC Z Z ACC Z Z Z Z Z Z Z Z Z Z Z Z Z Z	CA (CA (CA (CA (CA (CA)) T C/ T C/ T C/ T C/ T C/ T C/ C C C C C C C C C C C C C C C C C C	1383 22113 882 (210) 631 1473 4203 972 (240) 721 1563 42293 41062 5(270) 811 1653 2383 1152 (300) 700 1743 2473 1242
	GAG GAG LysTf AAGAG A LeuT CTGTZ T GlyG GGCCZ GGCCZ GA A GlyG GACG G GluT GACTZ	TT (GC (A C C C C C C C C C C C C C C C C C C C	G G G G G G G G G G G G G G G G G G G	Argi Argi Argi Argi Argi Alad C Alad C Cyst C Cyst	AG G G G G G G G G G G G G G G G G G G	TTAM TTAM TTAM TTAM CCA: A G G G G G G G G G G G G G G G G G G	A T A DELSE DELC ACC A ACC A ACC A ACC A ACC A ACC A A C	A erPr CCCC IG T roll CCA A euVa IGGI T T T	CA CA CA CA CG AG CG AG T T AlAS TGGA A G G	A A A G SAla GGCC A T rIle CATC A CATC CATC CATC CATC CATC A A CA	T Phel TTCC C Val GTG C C C Alai GCA T T T ArgJ C GGG	T T T Leuk CTOC T T C C T T T T T T	A T C Slug AGG AGG AGG A SluP A C SluP A C SluP A C C C C C C C C C C C C C C C C C C	heme G Iual AGGC TCAT A Xonl TAGC T C YSLy GCAA	aGln GCAG T C C ACA A C C C C C C C C C C C C C C	A ValM GTCA A A A CACG GCCA GCCA GCCA T C	A G ITGA ITGA T G G LetA T G G SpPP	T C C YSL AQC A A C C C T C C T C C T C C T C C C C	euL TGC AAT A euL TGC A G G G C T C G C	T T T T T T T T C C	A (T (GGCi A SpPl A A T G C C C IUA: G G C C IUA: G C C A A G G C C A A T	isAs ACG/ T T T heLe C C C C C C C G G G T G G T G	G A C C SPL S C A C A C C C S S C C C C S S C C C C	A J J J J J J J J J J J J J	A SuVa GGT T SuPi VCC A A G G G CA T G G VSAs A A	CA (CA (CA (CA (IGCA(T C) T C) T C) T C) T C) T C) T C) T C)	1383 22113 882 (210) 631 1473 4203 972 (240) 721 1563 42293 41062 (270) 811 1653 2383 1152 (300) 7 901 1743 2473 1242
	GAG GAG LysTf AAGAG A LeuT CTGTZ T GlyG GGCZ GA A A GGCG A G G GIUT GAGTZ	TT (GC (ur Lev CGCTK A C C VCALL ACGCC T T T T C SpLev ACGAX T T C SpLev ACCTK T C C	G G G G G G G G G G G G G G G G G G G	CYSC CYSC CYSC CYSC CYSC CYSC CYSC CYSC	AG G G G G G G G G G G G G G G G G G G	TTAA TTAA TTAA hriffe CCA: A G (luG: A G CCCA A T T SnII A CA (30 ; 9: 170 (24)	A T A T A T T A C C C C C C C C C C C C C	A Profil CCCA A Profil CCA A Profil CCA A Profil CCA T T T T	CA CA CA CG AG CG AG T T AlAs TGGA A G	A A A G sAla GGCCC A T rIle CATC PMet CATC YGlu GGAC A C A	T Phel TTCC C C Val? GTG C C C Alai GCAC T T Argli C C C Alai AAT	T T T Leux CTOX T T C T T T T T T T	A T C Slug SAGG SAGG SAGG SAGG A A A A A A A A A A	I ual G Iual AGGC heffe TCAT A Xonl TAGC T C ysLy GCAA T	A Galan Galan Galan Galan ACA A C C C C C C C C C C C C C C C C C	A ValM GTCA A A A A A A A GCCA GCCA A A A A GCCA T C	A G ITGA ITGA ITGA T G IELA T G IELA T G	T C C YSLIAOC AAOC AAOC AAOC AAOC T T G C T T G C T G C T C C C C AOC C C C AOC C C C C C C C C C	euL TGC AAT A euL TGC A A A A A A A C A G C TC G C	T T T T T T T T T T T T T T T C	A (T (GOCI A SpPl ATT G C C C LUAI G C C C LUAI G G C C LUAI C C C A A T	I SAS ACG/ T T T T C C C C C C C C C C C C C C C	G A C C S S C C S S C C S S C C S S C C S S C C S S S C C S S S C C S S S C C S S S C C S S S C S S S C S	A 3 ysLec ysAs G G G G G C S S S T C T G C T C T C T C T C T C	A SuVa T T SuPi SCO A A C C T T G X SCO T T G X SCO T A	CA (CA (CA (AlG) T C/ T C/ T C/ T C/ T C/ AGAC/ C C C C SpAS; ACGAC T A A A A A	1383 22113 882 1383 2213 882 1373 882 1473 4203 972 1473 4203 972 1473 4203 972 1563 42293 41062 5(270) 5811 1563 2383 1152 5(300) 5901 1743 2422 1242

FIG. 2. Nucleotide sequence of the exons of the c-fgr gene. The sequences were compared with those of the v-fgr (26), v-yes (18), and human c-src (1) genes. Only nucleotides in the v-fgr, v-yes, and human c-src genes that differ from that of the human c-fgr gene are shown. Numbers of nucleotide and amino acids (in parentheses) are given at the end of each line. The c-fgr nucleotide sequence is numbered from the first nucleotide of the predicted exon 4. The numbers of other three oncogenes are from each references. The nucleotide sequence of human c-src gene was reported about the exons 6 through 12.

	v- 1	fgr ike p	act	tin- tion														
c-fgr; v-fgr;	EGDWWEA	ARSL (MTQ	SS(IM	GKTO FE FI	7 CIPS N	NYVA	PVDS	SIQ	AEEWY	FGKI	g rki	DAERO)LLS	PG	NPQC AR	AFI	LI V	(60) (311)
v-yes;		I	АT		Y		Α			м		I	N		QR	I	v	(458)
c-fgr;	RESETTI	KGAY	SL	SIRD	NDQT	RGDH	VKHY	κı	RKLDM	GGYY:	TT	RVQF1	ISVQ	EL	VQHJ	MEN	٨N	(120)
v-fgr;					EA				Т	•		A				v		(371)
v-yes;			c		EV	N			Ň			AE	E L	K	K	RH	IA.	(518)
c-sic;			C	v5 1	C INPA	K LAN			2		5	1	Г	Q	AI	SN	14	(234)
c-fgr;	DGLCNLI	IAP	CT	IMKP	OLIC	LAKD	AWEI	SR	SSITL	ERRLO	G TO	CFGD\	MLG	TW	NGST	KV	٩V	(180)
v-fgr;	н	ТΑ		Г	М					Q				Μ				(431)
v-yes;	HK	TTV	P	ΓV	Q			P	ELR	VK	Q	E	М		Т		I	(578)
c-src;	HR	TTV	P	rs	Q			Ρ	ELR	VK	Q	Е	М			R	I	(294)
c-forr;	KTLKPG	MSP	KAJ	FLEE	AQVM	KLLR	HDKL	vo	LYAW	SEEP	I YI	VTEFN	1CHG	SLI	LDFI	KNI	PΕ	(240)
v-fgr;			:	5	I					Р					Е	D(2	(491)
v-yes;	L	м	Е	Q	I	к		Ρ					TΚ			EX	3	(638)
c-src;			Е			к	Е					Y	SK			GI	т	(354)
c-far:	GODLRLI	POLV	DM	AAOV	AEGM	AYME	RMNY	ин	RDLRA		/ GE	RLACI		FG	LARI	лĸ	œ	(300)
v-fgr;	Т						D					v				Е	N	(551)
v-yes;	KF K			I	D	I					D	NV				Е	N	(698)
c-src;	КY			I	s	v		v			1	NV	v			E	N	(414)
c-for:	EYNPCO	(30)	5)															
v-fgr;	R	(55	7)															
v-yes;	TAR	(704	4)															
c-src;	TAR	(42)))															

FIG. 3. Comparison of the amino acid sequences of the c-fgr, v-fgr, (26), v-yes (18), and human c-src (1) genes. Only amino acid residues of the compared oncogene products that differ from that of the human c-fgr gene product are shown. The amino acid sequence of exons 4 and 5 of human c-src has not been reported yet. Numbers of amino acids are given in parentheses at the end of each line. The standard one-letter abbreviation system is used.

by 11 introns (37). Very recently, Anderson et al. (1) reported that the human c-src gene also consists of 12 coding exons. The lengths of the exons of chicken and human c-src genes are identical, utilizing common splicing positions (1, 37). These results indicated that introns were acquired before the divergence of Aves and Mammals. Interestingly, the predicted splicing site positions around the kinase domain of the human c-fgr gene were also identical with those of the chicken and human c-src genes (Fig. 2). The predicted exons 4 to 10 of the c-fgr were numbered according to these of the corresponding exons of chicken c-src. This coincidence seemed to support the predicted splicing positions.

Allocation of the human c-fgr gene to chromosome 1. We determined the chromosomal location of the c-fgr gene by karyotype analysis of human-mouse hybrid cells and Southern blot analysis of DNAs of the hybrid cells. With a c-fgr-specific probe (probe A in Fig. 1), HindIII-digested DNAs of 12 independent human-mouse hybrid cells were analyzed by Southern blotting (Fig. 4A). From the results of a hybridization experiment and the human chromosome content determination by the differential staining method (39), the human c-fgr gene was assigned to chromosome 1 (Table 3). For confirmation of this result, the Southern blot filter was rehybridized with human N-ras-specific probes after washing out the c-fgr-specific probe (Fig. 4B), since the human N-ras gene has been located on chromosome 1 by several groups (5, 12, 32). The DNAs of the hybrid cells that retained the human c-fgr sequence were also found to contain the N-ras-specific sequence. Furthermore, the relative intensities of the bands detected with the c-fgr probe and N-ras probes seemed to correlate well. Thus, it was confirmed that the human c-fgr gene is located on chromosome 1.

Expression of human c-fgr gene. To examine the expression of the c-fgr gene, we carried out RNA blot hybridization. We used a 220-bp SacI-SacI fragment containing a 154-bp sequence of exon 10 as a c-fgr-specific probe to

detect the transcript (probe B in Fig. 1). Stringent hybridization conditions were adopted to avoid cross-hybridization with the transcripts of the yes and src genes. The result of RNA blot hybridization is shown in Fig. 5. Weak expression could be detected in placenta after 8 days of exposure. The major transcript was 2.6 kb long, and two additional longer species of 3.5 and 4.7 kb were also detected. The transcript of the c-yes gene was 4.8 kb long (33); its expression was found to be lower in placenta cells than in A431 cells (33), in which no expression of the c-fgr gene was detectable. The three species of c-fgr gene transcripts are probably not the result of cross-hybridization with c-yes mRNA.

Relatively higher expression was found in a human leukemia cell line, IM-9, which was considered to be derived from a lymphocyte cell established by Epstein-Barr virus infection (8). However, the expression level of the c-fgr gene in IM-9 cells was almost the same as that of the c-yes gene observed in various tissues and thus could not be regarded as overexpression. No amplification or rearrangement of the c-fgr gene was detected in IM-9 cells (data not shown). The

TABLE 2. Nucleotide and amino acid sequence homology^a

_	% Homology										
Oncogene	Nucleotide sequence	Amino acid sequence									
v-fgr	90.0	90.7									
v-yes	68.7	77.5									
v-src	71.9	73.0									
c-src	72.9	73.1									

^a Nucleotide and amino acid sequence homology between the c-fgr gene and three viral oncogenes, v-fgr (26), v-yes (18), and v-src (36), and human c-src gene (1). Since the 5' extremity of the sequenced region corresponds to the v-fgr gene actinlike sequence (26, 28), the 5' termini of 52 bases (17 amino acid residues) were omitted here in calculation of homologies. The homology between the two human cellular oncogenes c-fgr and c-src was calculated about the exons 6 through 10.



FIG. 4. Identification of human c-fgr (A) and N-ras (B) genes in DNA of human-mouse somatic hybrid cells. High-molecular-weight DNAs were prepared from human placenta, mouse B82 cells, and human-mouse hybrid cells by proteinase K digestion and phenol extraction. The DNAs were digested with HindIII restriction endonuclease and subjected to electrophoresis on 1% agarose gel and then Southern blot analysis with the ³²P-labeled human c-fgr locusspecific probe (A). The c-fgr-specific probe detected a fragment of about 11 kb. Subsequently, the filter was rehybridized to human N-ras-specific probes (34), which detected two HindIII fragments of 3.5 and 9 kb (B). The probe also detected an additional 8-kb fragment that was derived from the mouse genome in the DNAs of the hybrid cells and the mouse cells. Lanes; 1 and 14, DNAs from human placenta and mouse B82 cells, respectively; 2 to 13, DNAs from hybrid clones 1A, II-5, III-1, A1, II-6, Bm, 1B1, 3-3, 7-2, 7D4, 6-3, and 3D3, respectively.

significance of the high expression of the *c-fgr* gene in IM-9 cells remains to be investigated.

DISCUSSION

fgr and yes are distinct oncogenes. The fps gene of Fujinami avian sarcoma virus and the fes gene of Gardner feline sarcoma virus are known to correspond to a common cellular counterpart in different animals (11). A similar relation is also accepted for the raf gene of a mouse sarcoma virus and the mil gene of an avian sarcoma virus (16). In the src family, yes and fgr are closely related. The amino acid sequence of the kinase domain of the v-fgr gene shows more



FIG. 5. Expression of the c-fgr gene. RNAs were isolated from human placenta (lane 1), human embryo fibroblast cells (lane 2), A431 cells (lane 3), K562 cells (lane 4), and IM-9 cells (lanes 5 and 6). Samples of 2 to 3 μ g of polyadenylated RNAs were fractionated by electrophoresis through a 2.2 M formamide–1% agarose gel (20) and transferred directly to a nitrocellulose filter. The filter was hybridized to the ³²P-labeled c-fgr-specific probe (probe B in Fig. 1) and was exposed for 8 days (lanes 1 to 5) or 15 h (lane 6) after washing out the probe. The sizes of the transcripts were estimated with *Hind*III-digested, ³²P-labeled λ DNA as a size marker.

than 80% homology with that of the v-yes gene (26). The yes gene has been identified as an oncogene of an avian sarcoma virus (15). On the other hand, the fgr oncogene was found in a feline transforming virus (31). It remained to be determined whether these two genes were derived from distinct cellular loci or from a common proto-oncogene of chickens and cats. In the present study, we isolated a human c-fgr clone and obtained specific probes. Using these probes, we found that the c-fgr gene was on chromosome 1, and we examined its expression. In previous studies, we located the c-yes gene on chromosome 18 and another yes-related locus, which is probably a pseudogene, on chromosome 6 (33). Thus, the two oncogenes, the fgr gene and the yes gene, were proved to be distinct members of the src family. The results of RNA blotting also supported this conclusion; the c-yes mRNA is 4.8 kb long (33), whereas the major transcript of the c-fgr

TABLE 3. Segregation of c-fgr gene with human chromosomes in mouse-human cell hybrids^a

	Presence of chromosome:																		Hybridization						
Hybrid	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	x	Y	to c-fgr
1a	_	_	+	+	_	+	_	+	-	_	+	+	+	+	+	+	+	_	_	+	+	+	+	+	_
II-5	-	+	+	+	+	+	+	-	-	+	_	+	_	+	+	-	+	+	-	+	+	+	+	_	-
III-1	_	-	+	+	+	+	+	+	-	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	_
A1		_	+	+	+	+	_	+	_	_	+	-	+	+	+	-	+	+	+	+	+	+	+	+	-
II-6	+	+	_	+	+	+	-	+	-	-	+	_	+	+	+	-	+	+	-	+	+	+	_	-	+
Bm	+	+	+	+	+	+	+	+	-	+	+	+		+	+	+	+	+	+	+	+	+	+	+	+
1 B 1	+	_	+	—	-	-	+	_	-	-	-	_	+	+	_	_	+	-	+	+	_	-	-	-	+
3-3	+	+	_	+	+	_	-		_	-	_	_	_	-	+	-	+	-	+	_	+	-	-		+
7-2	_	_	+	_	+	-	-	-	_	-	-	+	+	+	-		+	-	-	+	+	-	_	_	-
7D4	-		-	+	+	_	+	+	_	_	+	+	+	+	+	+	+	+	+	+	+	-	+	-	_
6-3	_	-	-	+	+	+	+	+	-	-	+	+	-	+	+	-	+	-	+	+	+	+	-	_	-
3D3	-	-	-	-	-	-	-	+	-	_	+	-	-	+	+	-	-	_	-	+	+	-	+	-	-

^a The human chromosome content was determined by the method of differential staining of human and mouse chromosomes (39).

gene is 2.6 kb long. These findings clearly show that the yes and fgr genes are distinct oncogenes.

Recently, Le Beau et al. (19), using the method of in situ hybridization of a DNA probe of human c-src genomic sequence, detected two distinct loci, 1p34-p36 and 20q12-q13, as src-related loci. The c-src locus located on chromosome 20 was isolated and well characterized recently (1, 10). In the 3' two-thirds of the coding sequence, 98% of the amino acid residues are conserved in human and chicken c-src (1). Very recently, Parker et al. (30) reported isolation of a clone of the src-related gene, c-src-2, localized on chromosome 1. They also reported the nucleotide sequence of the two predicted exons of the c-src-2 gene, which correspond to exons 11 and 12 of the c-src gene. The predicted splicing positions were identical with those of the chicken and human c-src genes. The restriction maps of the c-src-2 and c-fgr were seemed to resemble each other, and both genes are located on chromosome 1. The exon-intron structures of the two are identical with that of the c-src gene, at least in their sequenced regions. In addition, the sequence of the c-src-2 gene shows extensive homology with that of v-fgr (94% nucleotide sequence homology and 95% amino acid sequence homology). Furthermore, our preliminary sequence analysis confirmed the presence of the sequence of exon 12 reported by Parker et al. (30). All of the results of our analyses support the idea that the c-src-2 gene is the c-fgr gene.

Generation of the c-fgr gene. The exon-intron structure of the c-fgr gene was found to be identical with those of the chicken (37) and human (1) c-src genes. We also found that the same exon-intron structure in the human c-yes gene. The lengths of exons 7, 8, and 9 of the c-yes genes are 156, 180, and 77 bp, respectively, which exactly coinicide with those of the c-src gene and the c-fgr gene (unpublished result). These results suggest that the three proto-oncogenes were generated by gene duplications of an ancestral gene harboring the exon-intron structure, as in the case of the generation of the globin gene family (7, 21). Three src-related genes were also identified by hybridization experiments in the genome of Drosophila melanogaster (14, 35). Partial nucleotide sequence analysis of the most closely src-related gene, names Dsrc, has been reported (13). The predicted amino acid sequence of the gene showed definite homology with the v-src product (54% in amino acid sequence), and the gene was found to contain one intron in the sequenced region. However, the position of the intron of the Dsrc is distinct from those of the c-src genes of chickens (37) and humans (1). In addition, recent studies showed that the splicing points of cellular counterparts of other members of the src gene family, mouse c-abl gene (38) and human c-raflmil gene (3), were different from those of the chicken c-src gene. Accordingly, it seems likely that the src, fgr, and yes genes were formed in the most recent split within the src gene family in evolution. Thus, it is likely that the unidentified amino-terminal structure and the normal function of the c-fgr gene product have some analogy to those of the c-src gene product. Analysis of the structure of the amino-terminal portion of the fgr gene product will be helpful in defining the function of the cellular oncogenes.

ACKNOWLEDGMENTS

We thank J. Sukegawa for valuable discussion.

This work was supported in part by a Grant-in-Aid for Special Project Research, Cancer-Bioscience, from the Ministry of Education, Science and Culture of Japan. M.N. acknowledges a fellowship from Japan Society for the Promotion of Science.

LITERATURE CITED

- Anderson, S. K., C. P. Gibbs, A. Tanaka, H.-J. Kung, and D. J. Fujita. 1985. Human cellular src gene: nucleotide sequence and derived amino acid sequence of the region coding for the carboxy-terminal two-thirds of pp60^{c-src}. Mol. Cell. Biol. 5:1122–1129.
- Bishop, J. M. 1983. Cellular oncogenes and retroviruses. Annu. Rev. Biochem. 52:301–354.
- Bonner, T. I., S. B. Kerby, P. Sutrave, M. A. Gunnell, G. Mark, and U. R. Rapp. 1985. Structure and biological activity of human homologs of the *raf/mil* oncogene. Mol. Cell. Biol. 5:1400-1407.
- Chirgwin, J. M., A. E. Pyzybyla, R. J. McDonald, and W. J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18:5294-5299.
- 5. Davis, M., S. Malcolm, A. Hall, and M. J. Christopher. 1983. Location of the human N-ras oncogene to chromosome 1 cen-p21 by *in situ* hybridization. EMBO J. 2:2281-2283.
- Denhardt, D. T. 1966. A membrane filter technique for the detection of complementary DNA. Biochem. Biophys. Res. Commun. 23:641-646.
- 7. Efstratiadis, A., J. W. Posakony, T. Maniatis, R. M. Lawn, C. O'Connell, R. A. Spritz, J. K. DeRiel, B. G. Forget, S. M. Weissman, J. L. Slightom, A. E. Blechl, O. Smithies, F. E. Baralle, C. C. Shoulders, and N. J. Proudfoot. 1980. The structure and evolution of the human β -globin gene family. Cell 21:653–668.
- Fahey, J. L., D. N. Buell, and H. C. Sox. 1971. Proliferation and differentiation of lymphoid cells: studies with human lymphoid cell lines and immunoglobulin synthesis. Ann. N.Y. Acad. Sci. 190:221-234.
- 9. Fukuhara, S., J. D. Rowley, D. Variakojis, and D. L. Sweet. 1979. Chromosome abnormalities in poorly differentiated lymphocytic lymphoma. Cancer Res. 39:3119–3128.
- Gibbs, C. P., A. Tanaka, S. K. Anderson, J. Radul, J. Barr, A. Ridgway, H.-J. Kung, and D. J. Fujita. 1985. Isolation and structural mapping of a human c-src gene homologous to the transforming gene (v-src) of Rous sarcoma virus. J. Virol. 53:19-24.
- Groffen, J., N. Heisterkamp, M. Shibuya, H. Hanafusa, and J. R. Stephenson. 1983. Transforming genes of avian (v-fps) and mammalian (v-fes) retroviruses correspond to a common cellular locus. Virology 125:480–486.
- 12. Hall, A., C. J. Marshall, N. K. Spurr, and R. A. Weiss. 1983. Identification of transforming gene in two human sarcoma cell lines as a new member of the *ras* gene family located on chromosome 1. Nature (London) 303:396–400.
- Hoffman, F. M., L. D. Fresco, H. Hoffman-Falk, and B. Shilo. 1983. Nucleotide sequences of the *Drosophila src* and *abl* homologs: conservation and variability in the *src* family oncogenes. Cell 35:393–401.
- Hoffman-Falk, H., P. Einat, B. Shilo, and F. M. Hoffman. 1983. Drosophila melanogaster DNA clones homologous to vertebrate oncogenes: evidence for a common ancestor to the src and abl cellular genes. Cell 32:589–598.
- Itohara, S., K. Hirata, M. Inoue, M. Hatsuoka, and A. Sato. 1978. Isolation of a sarcoma virus from a spontaneous chicken tumor. Gann 69:825–830.
- Jansen, H. W., R. Lurz, K. Bister, T. I. Bonner, G. E. Mark, and U. R. Rapp. 1984. Homologous cell-derived oncogenes in avian sarcoma virus MH2 and murine sarcoma virus 3611. Nature (London) 307:281-284.
- 17. Kawai, S., M. Yoshida, K. Segawa, H. Sugiyama, R. Ishizaki, and K. Toyoshima. 1980. Characterization of Y73, an avian sarcoma virus: a unique transforming gene and its product, a phosphopolyprotein with protein kinase activity. Proc. Natl. Acad. Sci. USA 77:6199-6203.
- 18. Kitamura, N., A. Kitamura, K. Toyoshima, Y. Hirayama, and

M. Yoshida. 1982. Avian sarcoma virus Y73 genome sequence and structural similarity of its transforming gene product to that of Rous sarcoma virus. Nature (London) **297**:205–208.

- Le Beau, M. M., C. A. Westbrook, M. O. Diaz, and J. D. Rowley. 1984. Evidence for two distinct c-src loci on human chromosome 1 and 20. Nature (London) 312:70-71.
- Lehrach, H., D. Diamond, J. M. Wozney, and H. Boedtker. 1973. RNA molecular weight determination by gel electrophoresis under denaturing conditions, a critical reexamination. Biochemistry 16:4743-4751.
- Maniatis, T., E. F. Fritsch, J. Lauer, and R. M. Lawn. 1980. The molecular genetics of human hemoglobins. Annu. Rev. Genet. 14:145–178.
- Maniatis, T., R. C. Hardison, E. Lacy, J. Lauer, C. O'Connell, D. Quon, D. K. Sim, and A. Efstratiadis. 1978. The isolation of structural genes from libraries of eukaryotic DNA. Cell 15:687-701.
- Messing, J. 1983. New M13 vectors for cloning. Methods Enzymol. 101:20-78.
- Mount, S. M. 1982. A catalogue of splice junction sequences. Nucleic Acids Res. 10:459–472.
- Naharro, G., C. Y. Dunn, and K. C. Robbins. 1983. Analysis of the primary translation product and integrated DNA of a new feline sarcoma virus, GR-FeSV. Virology 125:502–507.
- Naharro, G., K. C. Robbins, and E. P. Reddy. 1984. Gene product of v-fgr onc: hybrid protein containing a portion of actin and a tyrosine-specific protein kinase. Science 223:63-66.
- Naharro, G., S. R. Tronick, S. Rasheed, M. B. Gardner, S. A. Aaronson, and K. C. Robbins. 1983. Molecular cloning of integrated Gardner-Rasheed feline sarcoma virus: genetic structure of its cell-derived sequence differs from that of other tyrosine kinase-coding *onc* genes. J. Virol. 47:611-619.
- Nishizawa, M., K. Semba, T. Yamamoto, and K. Toyoshima. 1985. Human c-fgr gene does not contain coding sequence for actin-like protein. Jpn. J. Cancer Res. (Gann) 76:155–159.
- Norrander, J., T. Kempe, and J. Messing. 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene 26:101–106.
- 30. Parker, R. C., G. Mardon, R. V. Lebo, H. E. Varmus, and J. M.

Bishop. 1985. Isolation of duplicated human c-*src* genes located on chromosome 1 and 20. Mol. Cell. Biol. **5:**831–838.

- 31. Rasheed, S., M. Barbacid, S. Aaronson, and M. B. Gardner. 1982. Origin and biological properties of a new feline sarcoma virus. Virology 117:238-244.
- Ryan, J., P. E. Barker, K. Shimizu, M. Wigler, and F. Ruddle. 1983. Chromosomal assignment of a family of human oncogenes. Proc. Natl. Acad. Sci. USA 80:4460–4463.
- 33. Semba, K., Y. Yamanashi, M. Nishizawa, J. Sukegawa, M. Yoshida, M. Sasaki, T. Yamamoto, and K. Toyoshima. 1985. Location of the c-yes gene on the human chromosome and its expression in various tissues. Science 227:1038–1040.
- 34. Shimizu, K., M. Goldfarb, M. Perucho, and M. Wigler. 1983. Isolation and preliminary characterization of the transforming gene of a human neuroblastoma cell line. Proc. Natl. Acad. Sci. USA 80:383–387.
- 35. Simon, M. A., T. B. Kornberg, and J. M. Bishop. 1983. Three loci related to the *src* oncogene and tyrosine-specific protein kinase activity in *Drosophila*. Nature (London) **302**:837–839.
- 36. Takeya, T., R. A. Feldman, and H. Hanafusa. 1982. DNA sequence of the viral and cellular *src* gene of chickens. I. Complete nucleotide sequence of an *Eco*RI fragment of recovered avian sarcoma virus which codes for gp37 and pp60^{src}. J. Virol. 44:1–11.
- Takeya, T., and H. Hanafusa. 1983. Structure and sequence of the cellular gene homologous to the RSV src gene and the mechanism for generating the transforming virus. Cell 32:881-890.
- Wang, J. Y. J., F. Ledley, S. Goff, R. Lee, Y. Groner, and D. Baltimore. 1984. The mouse c-abl locus: molecular cloning and characterization. Cell 36:349–356.
- Yoshida, M. C., T. Ikeuchi, and M. Sasaki. 1975. Differential straining of parental chromosomes in interspecific cell hybrids with a combined quinacrine and 33258 Hoechst technique. Proc. Jpn. Acad. 51:184–187.
- Yoshida, M. C., M. Sasaki, K. Mise, K. Semba, M. Nishizawa, T. Yamamoto, and K. Toyoshima. 1985. Regional mapping of the human proto-oncogene c-yes-1 to chromosome 18 at band q21.3. Jpn. J. Cancer Res. (Gann) 76:559–562.