Cytogenet Cell Genet 76:144–152 (1997)

Cytogenetics and Cell Genetics

Chromosome assignment of 115 expressed sequence tags (ESTs) from human skeletal muscle

T. Muraro,[1] D. Stephan,[2] N. Tiso,[3] R. Zimbello,[1] G.A. Danieli,[3] E.H. Hoffman,[2] G. Valle,[1] and G. Lanfranchi[1]

[1] CRIBI Biotechnology Centre, Università degli Studi di Padova, Padova (Italy);
[2] Department of Molecular Genetics and Biochemistry, University of Pittsburgh School of Medicine, Pittsburgh, PA (USA); and
[3] Dipartimento di Biologia, Università degli Studi di Padova, Padova (Italy)

Abstract. The chromosome assignment of 115 expressed sequence tags (ESTs) from human skeletal muscle, 101 of which identify unknown human genes, is reported. The ESTs were selected among over 4,000 obtained from systematic sequencing of a skeletal muscle cDNA library containing 3′ portions of the mRNAs. Chromosome assignments were obtained by PCR amplification of two panels of human × rodent somatic cell hybrids. Analysis of these preliminary data suggests a nonrandom distribution of muscle ESTs in the human chromosome complement. The unexpected occurrence of multiple chromosome localizations for some ESTs is discussed.

The generation of high number of expressed sequence tags (ESTs) by systematic random sequencing of cDNA libraries is a powerful approach for the rapid identification of new genes (Sikela and Auffray, 1993). Human ESTs have been extensively studied in order to map the corresponding genes on the chromosomes (Khan et al., 1992; Polymeropoulos et al., 1992, 1993; Durkin et al., 1994; Pappas et al., 1995; Takeda et al., 1995); the resulting sequence tagged sites (STSs) are now regarded as effective tools for the construction of a high-resolution physical map of the human genome (Adams et al., 1995; Hudson et al., 1995).

Because of its speed and simplicity, PCR amplification using EST-derived primers of human DNA contained in monochromosomal human × rodent hybrids (Abbott et al., 1990; Gardiner and Patterson, 1992), radiation hybrids (Cox et al., 1990; Barrett, 1992; Ceccherini et al., 1992), and YAC libraries (Green and Olson, 1990) is the technique most commonly chosen for STS generation. There are certain distinct advantages in using ESTs derived from the 3′ untranslated region of mRNA for STS mapping (Wilcox et al., 1991). First, this part of the gene is virtually free of introns (Hawkins, 1988); therefore, primers designed from such a region will amplify fragments in human genomic DNA whose length can be predicted from the cDNA sequence. Second, since the 3′ untranslated regions are poorly conserved, there is a good probability when somatic hybrids are used that the EST primers will selectively amplify a distinct human fragment even within a rodent background. Third, genes evolved from a common ancestor (such as globins, actins, and myosins) can be very similar in their coding region but normally are considerably different in the less-conserved 3′ untranslated part. Therefore, mapping by means of 3′ ESTs could avoid the bias of multicopy genes.

We have constructed a cDNA library containing the 3′ portions of human skeletal muscle mRNAs, and have so far obtained over 4,000 ESTs by systematic sequencing (Lanfranchi et al, 1996) (Table 1). In this paper, we report the chromosome assignment of 115 of our human muscle ESTs, 101 of which presumably identify novel human genes because they do

Supported by Telethon Italy (grant B.30).

Received 5 September 1996; revision accepted 8 November 1996.

Request reprints from Dr. Gerolamo Lanfranchi, CRIBI Biotechnology Centre, Università degli Studi di Padova, Via Trieste 75, 35121 Padova (Italy); telephone: 39-49-8276281; fax: 39-49-8276280; e-mail: lanfra@eos.bio.unipd.it.

KARGER

E-mail karger@karger.ch
Fax + 41 61 306 12 34
http://www.karger.ch

0301–0171/97/0764–0144$12.00/0
This article is also accessible online at:
http://BioMedNet.com/karger

not show any significant similarity to the existing databases or to the entire collection of ESTs so far produced from different human tissues. The mapping was done by PCR amplification of human × rodent somatic cell hybrids, followed by high-resolution agarose gel or SSCP analysis.

Materials and methods

Generation of the cDNA library

The first-strand cDNA was synthesized from pectoralis major muscle mRNA of an adult woman by AMV reverse transcriptase and an oligo-dT-*Not*I primer in which the 5′ terminal nucleotide was biotinylated. After completion of the second strand, the cDNA was sonicated, and the fragments were repaired with T4 DNA polymerase, ligated to *Bst*X1 nonpalindromic adaptors, and size-fractionated on agarose gel with a low melting point. Sonicated cDNA in the range of 450–550 bp was selected, and the 3′-specific cDNA fragments were then purified by binding to avidin-coated paramagnetic beads and released by *Not*I digestion. The cDNA was directionally cloned into a *Bst*X1- and *Not*I-digested pcDNAII plasmid vector (Invitrogen), electroporated into the *E. coli* TOP10F′ strain, and subjected to large-scale random sequencing (Lanfranchi et al., 1996).

Primer design

For each EST, specific forward and reverse primers were designed from Applied Biosystem electropherograms, using OLIGO ver 4.0 (National Biosciences). For ESTs represented in our catalog more than once, the consensus sequence was used. The primer pairs generate a series of PCR products varying in size between 60 and 270 bp. All of the oligonucleotides used in this work were synthesized by Genset.

Somatic cell hybrid panels and genomic DNA amplification

Two hybrid DNA panels were used for this work. The first was obtained from the Coriell Institute for Medical Research, and the second was the NIGMS human × rodent somatic hybrid panel 1 (Drwinga et al., 1993). Fifty nanograms of genomic DNA was used for amplification in 50 µl of PCR buffer (20 mM Tris-HCl [pH 8.3], 50 mM KCl, 2 mM $MgCl_2$, and 0.1% *[v/v]* Tween 20) containing 400 ng of the specific forward and reverse primers, 2.5 U of *Taq* polymerase (Perkin-Elmer) and 0.126 mM of each of the four dNTPs; 0.5 µCi of ^{32}P-dATP (specific activity, 3,000 Ci/mmol) was added for radioactive SSCP analysis. Cycling conditions were 3 min at 94°C, followed by 35 cycles of 20 s at 94°C, 33 s at 55°C, and 1 min at 72°C, and a final extension step for 5 min at 72°C.

Agarose gel electrophoresis

Twenty microliters of the PCR reactions were mixed with 5 µl of loading buffer (50 mM EDTA, 0.5% SDS, 25% glycerol, and 0.05% bromophenol blue) and separated on 2% Separide gel (Bethesda Research Laboratories) in TAE buffer (40 mM Tris acetate and 1 mM EDTA) containing ethidium bromide.

Polyacrylamide gel electrophoresis and silver staining

Two microliters of the amplification products was mixed with 4 µl of loading buffer (95% formamide, 0.1% xylene cyanol, 0.1% bromophenol blue, and 10 mM EDTA). The samples were heated at 95°C for 10 min, chilled on ice, and loaded on a 9% polyacrylamide slab gel (20 × 20 × 0.1 cm) composed of 5.5 M urea (BDH), 9% of a 29:1 solution of acrylamide and bisacrylamide (Bio Rad), 0.08% of ammonium persulfate, and 0.08% of TEMED in TBE buffer (90 mM Tris-HCl [pH 8], 40 mM boric acid, and 2 mM EDTA). The gels were run, flanked by a 45°C water jacket, in the same TBE buffer at 500 V for 2–4 h, according to the size of the amplification fragments. The gels were silver stained as described by Santos et al. (1993).

SSCP analysis and autoradiography

One microliter of ^{32}P-labeled PCR samples was mixed with 1 µl of the above-mentioned loading buffer, denatured for 5 min at 95°C, and loaded on a 5% polyacrylamide slab gel using 0.5 × TBE as buffer. Electrophoresis was performed at a constant power of 25 W for 2 h at 4°C. After drying, the slabs were exposed to X-ray film at −80°C, using an intensifying screen.

Table 1. Human skeletal muscle EST catalog[a]

	Number	Percent
Total number of ESTs sequenced	4,370	
Total number of transcripts identified	934	
Transcripts identified by a single EST	719	77.0
Transcripts identified by a group of ESTs (> 1)	215	23.0
Already known genes	284	30.4
Already sequenced ESTs	133	14.2
New ESTs	517	55.4
New ESTs		
Similar to other genes	138	26.7
Similar to other ESTs	72	14.0
Containing *Alu* repeats	86	16.6
Without any similarity	221	42.7

[a] This table summarizes the principal features of the EST catalog obtained from systematic sequencing of the human skeletal muscle cDNA library containing the 3′ regions of mRNA.

Results

The library that we have constructed from human skeletal muscle mRNA contains cDNA fragments of 450–550 bp in length that correspond specifically to the 3′ portions of the mRNAs (Lanfranchi et al., 1996). Using a random sequencing approach, we have so far produced some 4,000 3′ ESTs. Information about the library is summarized in Table 1.

A sample of 115 human muscle ESTs from this library was selected for mapping. One hundred and one ESTs showed no match with sequences in the current databases, seven were similar to human ESTs produced by other research groups, and seven were similar to already known genes whose chromosome assignment has not yet been established.

Three different systems were used for checking the amplification products obtained from the human × rodent somatic hybrids: agarose gel electrophoresis, polyacrylamide gel electrophoresis, and single-strand conformation analysis (Fig. 1). The last technique was used whenever the other methods produced dubious results.

ESTs corresponding to the slow twitch myosin alkali 1 chain isoform (Cohen-Haguenauer et al., 1989), skeletal β-tropomyosin (Hunt et al., 1995), and α-actinin (Youssouffian et al., 1990) were used as positive controls and were assigned to their correct chromosomes (3, 9, and 14 respectively). Since oligo design was based on sequence from the 3′ termini of human cDNAs, in most cases primer pairs failed to amplify rodent DNA. However, in 5% of the amplifications, discrete bands of mouse or hamster parental genomic DNA were produced. In all these instances, the resulting human and rodent bands were clearly separable by agarose or polyacrylamide gel electrophoresis.

Table 2 summarizes the chromosome location of the 115 human skeletal muscle ESTs. ESTs 832, 917, 1448, 1481, 1676, and 3104 were mapped on two distinct chromosomes, and this result was confirmed using both the hybrid panels. Interestingly, five out of six ESTs showing multiple chromosomal assignment involve sexual chromosomes, and two of them are located on both the X and Y chromosomes.

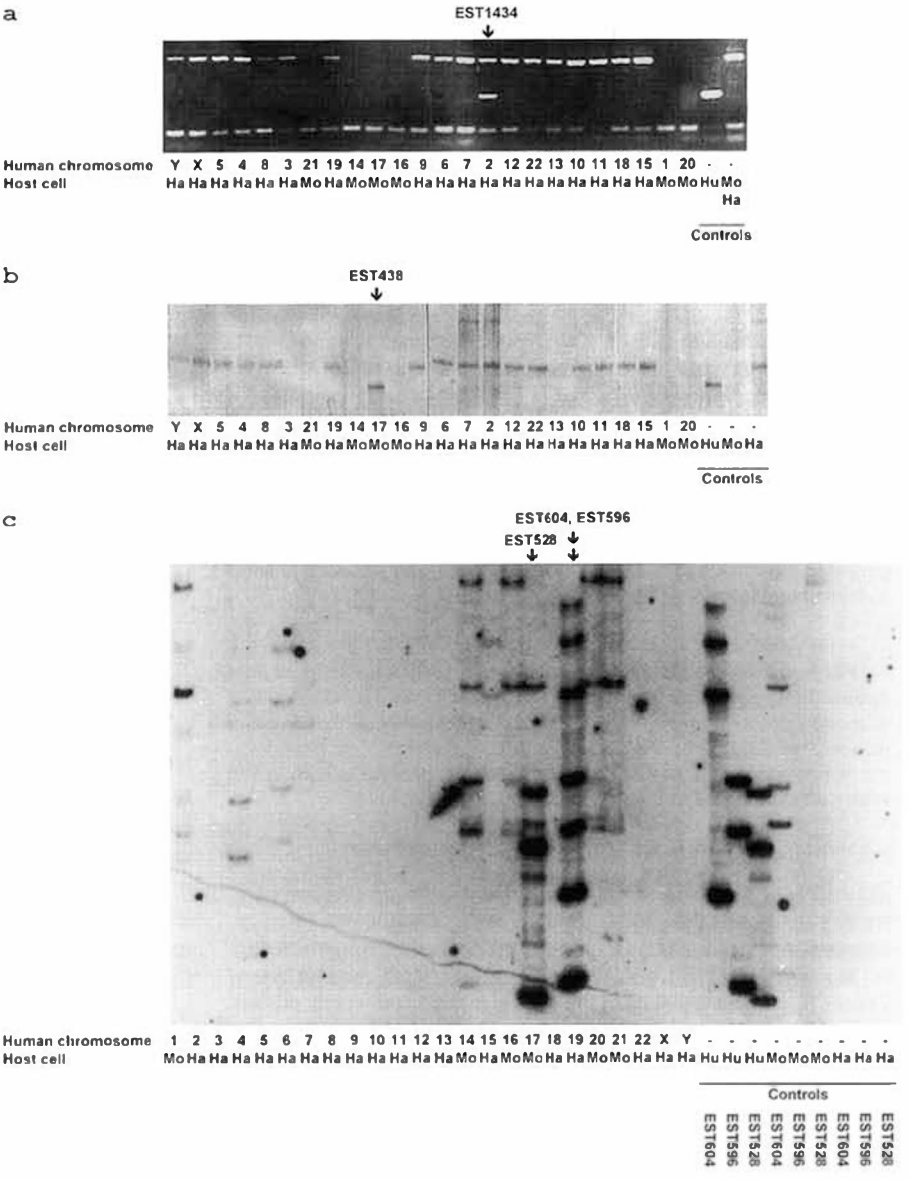

Fig. 1. Examples of the mapping results obtained with the three different techniques adopted in this work. Each lane contains PCR products from amplifications of monochoromosomal cell lines with EST primers. In the rows at the bottom of each panel are indicated the human chromosome and the parental rodent genome contained in each hybrid cell line. Polymerase chain reactions from the rodent and human parental cell lines were run as controls in adjacent lanes (Ha = hamster, Mo = mouse, Hu = human). Arrows show the hybrid cell line that resulted in positive amplifications with EST-specific primers: (**a**) EST 1434 assigned to chromosome 2 by Separide agarose gel electrophoresis; (**b**) EST 438 assigned to chromosome 17 by polyacrylamide denaturing gel and silver staining; (**c**) mapping of EST 528 to chromosome 17, as well as ESTs 596 and 604 to chromosome 19, by multiplexed SSCP. Cross-species amplification products are easily differentiated from the human product by the SSCP technique.

Table 3 and Fig. 2 show the distribution of 111 skeletal muscle ESTs (excluding four ESTs that showed multiple chromosome localizations on the autosomes) among human chromosomes (excluding the Y chromosome), compared with the distribution of 1,975 human genes for which the map position was already established (Location Data Base, August 1996 [Morton et al., 1992]). Forty-seven ESTs (42% of the total sample) were assigned to four chromosomes only (viz., 1, 9, 16, and 19), whereas no ESTs were assigned to chromosomes 18 and 21.

A nonrandom distribution of skeletal muscle ESTs is suggested by the significant paucity of ESTs on chromosome 11 and by an excess of ESTs on chromosomes 9 and 19.

Since the size of the sample of skeletal muscle ESTs is very small in comparison to the real number of genes localized on chromosomes, the hypothesis of a Poisson distribution was tested. A significant deviation from expectation was found ($\chi^2 = 65.333$, df = 22, $P < 0.005$).

Table 2. Chromosome assignment of 115 3′ ESTs derived from human skeletal muscle[a]

Skeletal muscle EST code	BLASTN probability	Occurrences out of 4,370 ESTs	PCR fragment size (bp)	Type of analysis	Chromosome assignment	Primers (5′→3′)
00374	$> e^{-30}$	2	201	DPGE	1	*Forward:* GAAGGGCTAGCCCACCGCAGCAGAC *Reverse:* GGCCCTAGGCCAGACAGGCACTGAC
00738	$> e^{-30}$	1	149	DPGE	1	*Forward:* GGTGCTTCACGATGCTGCTCTGGTGATTCT *Reverse:* AGCATTCTCTCCATCCCACCCCCTCACTGA
01265[b]	$4.8\ e^{-30}$	1	265	AGE	1	*Forward:* ATGTTAAGCTGCTGGAAACCTCATG *Reverse:* GTTCAGCTTGAGGAGTATAACTAAA
01401	$> e^{-30}$	1	242	AGE	1	*Forward:* GCCCAGCCGAGGGCTTAGGAGAGAG *Reverse:* GAAAGGATCTGAAAAGCGTAATAAG
01423	$> e^{-30}$	4	59	SSCP	1	*Forward:* TCCATTTTGNCCATAGATGGGC *Reverse:* CAGAATCTGGAAGGGGGATGAG
01438	$> e^{-30}$	1	171	AGE	1	*Forward:* CTGAAAATGGGCACATCTTCAACTG *Reverse:* AGGTGCCCATGATTGTCACCATCAC
02251	$> e^{-30}$	1	60	SSCP	1	*Forward:* TTTACCAACTAGCGCTGGGAG *Reverse:* ACAGGAAAGGAGGACAGATGAGG
02455	$> e^{-30}$	1	125	SSCP	1	*Forward:* GTCGTTGGAAGAGTACAAGCTGATC *Reverse:* ATGATCCTCCTCAGGACCCTTGGGG
02638	$> e^{-30}$	1	113	SSCP	1	*Forward:* CTCCCAAAGAATAACATCTTCCAG *Reverse:* TTTGCCAAAAGGAACACAGAGG
02817	$> e^{-30}$	1	154	SSCP	1	*Forward:* GGTGAAGGAGATGGTGAAGGAGTTC *Reverse:* TTAGCAGTGAGGGTCCAGGTTTAC
03036	$> e^{-30}$	1	123	SSCP	1	*Forward:* GGGAAGGAAGCCAGGTGGATAC *Reverse:* TCAGTTACCACCCCGTATGTCTGG
03052[c]	$7.7\ e^{-93}$	2	225	AGE	1	*Forward:* ACTCTGTTGAAATTCTTGCATAGAG *Reverse:* ACCAAAAAGGCTTTCATTTTCTGTC
03072	$> e^{-30}$	1	58	SSCP	1	*Forward:* TGATGGCTGAAGGAGACGC *Reverse:* GACTCCACAACACACAACTTCCC
03123	$> e^{-30}$	1	120	SSCP	1	*Forward:* CCAAAAACTTTATGGAGCTAGTACTGGTC *Reverse:* GGAATTGCATGAGGTCAGAAAGAC
00116	$> e^{-30}$	1	182	DPGE	2	*Forward:* TCTGGCCATTCACTTACCCACTCTC *Reverse:* TTCCCTCAATTAGGCCAGAAACTCA
00838[d]	$9.2\ e^{-107}$	2	138	AGE	2	*Forward:* CATGACAAATACGGTAATGCTGTAT *Reverse:* ATTCCTCCATTCCTTTGGGGTAACT
00892	$> e^{-30}$	1	216	SSCP	2	*Forward:* ATGCTAGCGACAGTCTTTCTTACTG *Reverse:* AGCAGAATCCCATCGTAACAGTTCT
01281	$> e^{-30}$	1	224	DPGE	2	*Forward:* CAGGCCCAGAACAAGGCTGTTGTTACTATG *Reverse:* TGATTGGTATAAAAAGGTGTGTTTGTCACT
01434	$> e^{-30}$	1	253	AGE	2	*Forward:* GGAGGGTCCTAAGAGGAAGGTGACT *Reverse:* ATAGTTGTATCCCTTTGTGTGATAT
01055	$> e^{-30}$	1	98	DPGE	3	*Forward:* AAGAAGAAGGCATCTGTTTACACAC *Reverse:* ACTGTTCTGGAGGCTGGCGGTCTAA
01091	$> e^{-30}$	1	121	SSCP	3	*Forward:* GCGATTGGTCCTCTACACTATCGTG *Reverse:* CGAGCCTGAGTCTATGGGTTTTTG
01307	$> e^{-30}$	1	140	DPGE	3	*Forward:* GCGGGCACCCTCAGCCGGGCTCTTT *Reverse:* CCTTAATGTGTGTGTTTATTTACAA
02082	$> e^{-30}$	1	145	SSCP	3	*Forward:* GGAGAGGAGCTAGAGATGGGTTCAC *Reverse:* GCTGAAAGTCCAAACATGGGG
02135	$> e^{-30}$	1	112	SSCP	3	*Forward:* TCTCTTGGGGCAACCCATTC *Reverse:* GATTTCCGAAAACTCCACACCC
00011	$> e^{-30}$	1	225	DPGE	4	*Forward:* TGTTCAGTTTGCTATGGCTTGTATG *Reverse:* CAAGGAGCTGGCACTTGTTGACTAG
00149	$> e^{-30}$	3	169	SSCP	4	*Forward:* AAAGAAGCTGAGCTGAACACATTAC *Reverse:* TTAAAATAGGTGTCCTGATAACTTC
03260	$> e^{-30}$	1	77	SSCP	4	*Forward:* ATGCTGAGTTGGGAGGATC *Reverse:* GTAGAGAGGGGTCTTGTTATG
01479	$> e^{-30}$	1	109	AGE	5	*Forward:* TGTCCTGAATTAGCAACCCTGACAC *Reverse:* TTAAGTGGTCAGCTTGGCAGCTCTG
03043	$> e^{-30}$	2	121	SSCP	5	*Forward:* CAGTTCAGACAGCACCACCACTAG *Reverse:* GAATACCCTTCGCACAGAACCAG
03272	$> e^{-30}$	1	122	SSCP	5	*Forward:* TTCTCTGTACTCTGGGCGTGCTG *Reverse:* CTAGGTGGAGACGTTTGGTAAC
03332	$> e^{-30}$	1	162	SSCP	5	*Forward:* GCAGGGGAAGGAGATTGTGTATG *Reverse:* TTGATGCGAGTCTGGAGTCAGG
00042	$> e^{-30}$	1	131	SSCP	6	*Forward:* CAAGTGTGACCTGGGATTTTATTTA *Reverse:* ATACATGATGTGTTTAGGGTTACAT
01406	$> e^{-30}$	1	203	AGE	6	*Forward:* TGGTACCTGTGTCCTGCTCAGAACT *Reverse:* TATTCAGGCTCTCTCCCAACTACTG
01550	$> e^{-30}$	1	115	AGE	6	*Forward:* CTGTGCTGGTGGAGAGGTCCTAGAG *Reverse:* GAAAAAGCCAATGACAATCAATTAG
01984	$> e^{-30}$	1	164	AGE	6	*Forward:* TARGGAATTTGGTGACACAATATCT *Reverse:* CAAATATAGGAAATAGAAGCTATCT
02045	$> e^{-30}$	1	225	AGE	7	*Forward:* TGGTTCAGAGATAAATTGGCTAATT *Reverse:* TCACAAAACTGTAATTCAGGTATAA

➤

Table 2 (continued)

Skeletal muscle EST code	BLASTN probability	Occurrences out of 4,370 ESTs	PCR fragment size (bp)	Type of analysis	Chromosome assignment	Primers (5'→3')
02477	$> e^{-30}$	1	64	SSCP	7	*Forward:* AGCCGGCTGAAATGCTCCAAC *Reverse:* AGATTCCCATCAGTCCAAACTGG
02618	$> e^{-30}$	1	143	SSCP	7	*Forward:* GCAAGATCAACACTGAGTATCCCTG *Reverse:* GTAGGATGGAACACAGGAAATTGG
01839	$> e^{-30}$	1	158	DPGE	8	*Forward:* TCCTCTTCCCAGGGGCTGCCAAGTC *Reverse:* AGGGGATGTGGGGCTGGGGGAGAAC
02659	$> e^{-30}$	1	90	SSCP	8	*Forward:* CTTGTCCTGACCTGTTCTCGG *Reverse:* CGGCTCCTGTTTTATTGCCTTCGG
00295	$> e^{-30}$	1	187	SSCP	9	*Forward:* GTCAACGTGAGCAAGCGTGATTATGATGAG *Reverse:* ATAAAACAGAAGCCCCGCCAGGTATGATAG
00547[e]	$4.3\ e^{-43}$	3	154	AGE	9	*Forward:* CCTCCCCACGTCTCCCCCTTGTGTC *Reverse:* GGGGCTGGCCCTCACAGGTTGTTGA
01210	$> e^{-30}$	1	174	DPGE	9	*Forward:* CTCTGAGATTTTGAGTTCTTCTGCAGAGAT *Reverse:* TTTTAAATGGCGCCCACTGGTCTCATACAG
02044	$> e^{-30}$	1	120	AGE	9	*Forward:* CCCTACCACCTTTTGGTGCCACACT *Reverse:* GTTCCTGATACGGCCCCGGCAGTCA
02053	$> e^{-30}$	1	149	AGE	9	*Forward:* AGTGCTCCTCTGGGCTCTTGAGTTG *Reverse:* AGAAAAGTCATCTGTGAGTGACTAG
02246	$> e^{-30}$	1	68	SSCP	9	*Forward:* CACAAGCTGCATCAGGCTCTCT *Reverse:* ACAGTGCTGACCCCACAGAGTCTT
02585	$> e^{-30}$	1	130	SSCP	9	*Forward:* CCTCCACAACTGAATATAGTGGCTG *Reverse:* AAAGCACTGTGCCTACCAGGTG
02713	$> e^{-30}$	1	103	SSCP	9	*Forward:* GCCTCTGCCAAGTTCTGCATTG *Reverse:* TTCCTAAATCCCTCCCTGGACC
02855	$> e^{-30}$	1	115	SSCP	9	*Forward:* ATCAACGATGGGAAGCTGGG *Reverse:* GGTGTCCTTTAATCAGTTGGCAGG
03094	$> e^{-30}$	1	106	SSCP	9	*Forward:* TCTTGCCCACACTCAGTTCCTG *Reverse:* CACATGCAGATCACAAAGCGG
00355[f]	$1.5\ e^{-56}$	8	197	AGE	10	*Forward:* TGGAGAGGGAATGCTGAGCAGATAG *Reverse:* TGGTGGCAGAGCAAATCCCATAAAC
00686[g]	$1.6\ e^{-93}$	5	190	AGE	10	*Forward:* TCAGGGTCTGGGGAAGAGGCTGATC *Reverse:* GAGCCAGCACACACGGTACAGAAAG
00788	$> e^{-30}$	1	227	SSCP	10	*Forward:* TTTTGGAAGGAGAGGTCTTTAATAG *Reverse:* AATGTCTACTTTGTGCTAGGTACTC
01040[h]	$6.7\ e^{-86}$	3	200	DPGE	10	*Forward:* CTAGGCCTCTGGCTTCCTGGTAGAG *Reverse:* ACAAAAATCTCGAAGTCCATTTAAT
01825	$> e^{-30}$	1	145	AGE	10	*Forward:* ATCTGAAACGTCCTAAGCAGAGTTA *Reverse:* CACATAAGATTATTCAAAGCCATAG
02860	$> e^{-30}$	1	157	SSCP	10	*Forward:* GCCTAATGATAGTGGGCGAGAGA *Reverse:* TCATGCAGAAAAAAGGCTGGG
00545	$> e^{-30}$	3	179	DPGE	11	*Forward:* CTGATCGAGAAGCTGCTCAATTATG *Reverse:* TTCCAGAGTGGGGTATGGGATACAC
01705	$> e^{-30}$	5	60	SSCP	11	*Forward:* AGTGCTAGGATTGATTATGTG *Reverse:* AAACACTTGGCAGCCCTTGG
01379	$> e^{-30}$	1	118	AGE	12	*Forward:* GGAGCGATGAGGCAAGGATTGTCAG *Reverse:* ACACTGAGATCCATCGAAGCTGTGT
01396	$> e^{-30}$	1	253	AGE	12	*Forward:* ATCGTTGGCTTCACAGTATGACTGA *Reverse:* AATGGTAAACAGTAAAAGAGTAATT
01485	$> e^{-30}$	1	148	AGE	12	*Forward:* AAAATCATCCTCCAAGTGCCAGACA *Reverse:* TGCTACTATGGGATTTCGTTTTCTG
02777	$> e^{-30}$	1	53	SSCP	12	*Forward:* CCCCCCATTGCCACCTTCACTCCTG *Reverse:* AAGAGTCTGAGGCTAATCAGC
03256	$> e^{-30}$	1	60	SSCP	12	*Forward:* TGTGCATCACTGGGCTTGG *Reverse:* ATTAGGTTCCACTGCAGGGCTGG
01287	$> e^{-30}$	1	192	DPGE	13	*Forward:* CTCAAGGTCGTCAACCTCGGTATTC *Reverse:* TACACAGAGAAAATCCACTGATATT
01370	$> e^{-30}$	1	148	AGE	13	*Forward:* CCTTATCTATGCCTCTGCTCTTAGA *Reverse:* GTAGAGACAGACAATAAACAAATAC
01224	$> e^{-30}$	1	109	DPGE	14	*Forward:* TCAGCCACAGACAGCCGTAGCCAGGATCTC *Reverse:* AGAACTGTCTAGAGGCCAATTTAGAGCCAG
01912	$> e^{-30}$	1	181	AGE	14	*Forward:* TTTCCACAGGAAGCAGATGGAGATC *Reverse:* TACGTGACAACAAGCTGGCCTAGAG
03174	$> e^{-30}$	1	153	SSCP	14	*Forward:* GAATGGGACAGACCCCACAAC *Reverse:* GGCAATCTCACAGGCAGCAC
00761	$> e^{-30}$	1	222	SSCP	15	*Forward:* GAGTGGTGGCCTCTCTGAGATTTCCTAGAG *Reverse:* CGTGCCAATCAGTAGCATGGGACAAAGTAA
00964	$> e^{-30}$	2	154	SSCP	15	*Forward:* TTTGTCAACAAAATCTGCTTGCTGCAGTAG *Reverse:* CTCAATACTGTTGAACAACAAGATAACACA
00125[i]	$2.7\ e^{-57}$	2	150	DPGE	16	*Forward:* GGGCTTAGCCTCCAAGTGGGACTAC *Reverse:* AGTGGTACAACTGTTTGGCATAACA
00129	$> e^{-30}$	1	205	SSCP	16	*Forward:* AGCCTTCTGCTGCCCTTGCTTAACA *Reverse:* GCTGCTTTAAATGACAACTTCTAAC

Table 2 (continued)

Skeletal muscle EST code	BLASTN probability	Occurrences out of 4,370 ESTs	PCR fragment size (bp)	Type of analysis	Chromosome assignment	Primers (5′→3′)
00385	$> e^{-30}$	3	60	SSCP	16	*Forward:* AGAGCCTGGGCTCTGCTGGAC *Reverse:* TGAGGAATAGGAGAACAC
01785	$> e^{-30}$	1	174	AGE	16	*Forward:* GGCAGCTCCGTGGAGGGGGCTTTAA *Reverse:* GACTGATGAAATCAGAACCAAAAGA
01974	$> e^{-30}$	1	142	AGE	16	*Forward:* CCCTCCCCCTGGAGGGCATGGTGTC *Reverse:* GACAGCCCTGGACTGGGGCTATGAA
02073	$> e^{-30}$	1	133	AGE	16	*Forward:* TCAGTCCCATAAGGGCAGCCTTGTG *Reverse:* GCAGTGATGCTTTAATCCCCCTGTT
02094	$> e^{-30}$	1	76	SSCP	16	*Forward:* GGATGCTGCTGGAGCTGAATC *Reverse:* TCCGGAGTTTATTCACTTCCAGC
02792	$> e^{-30}$	1	142	SSCP	16	*Forward:* GCAGTCAGCTTTGGAATACTGTTG *Reverse:* CAGGGGACCTTGAAAATCACTAAG
03341	$> e^{-30}$	1	176	SSCP	16	*Forward:* CGGAAAAACTAAGCCCCTCCTTC *Reverse:* GCAGCAGGACATGCTTTATTCTG
00015 [j]	$6.5\ e^{-103}$	33	170	DPGE	17	*Forward:* GGTGGCAGAGGCAGTGGGCACTAAG *Reverse:* CTTCTGCCTAGAGCGTTCCAGATTC
00245	$> e^{-30}$	4	219	SSCP	17	*Forward:* TGACCCCTGGCAGCAAGAGCTCTAC *Reverse:* GCAGGCACAGTCCTGGGAACACTAG
00438	$> e^{-30}$	1	143	DPGE	17	*Forward:* TGAACCGAATCCCACTTGGCAAGTT *Reverse:* GGAGGGAGCTCAGCAGGCCCAGAAG
00604	$> e^{-30}$	1	143	SSCP	17	*Forward:* TGTCCCAGTCCCATGCCCATTCCCATCACT *Reverse:* TGAGTAGATCTTATTTGTATTCATCATGTC
01100	$> e^{-30}$	1	111	SSCP	17	*Forward:* CCTGCATACATACCTGCAAACAAC *Reverse:* TACGGTCACAAGGCTACAGAGGTC
02633	$> e^{-30}$	1	112	SSCP	17	*Forward:* ATTCCGTCTGTACCAACCCCAGTG *Reverse:* CTTTATGTCTCTGAGCTGTGCCTTC
00046	$> e^{-30}$	1	187	DPGE	19	*Forward:* AACCCAGGGCTGTGGACCCAGACTG *Reverse:* GAGTTCACAAGAAACTGATCTTTAC
00528	$> e^{-30}$	1	216	SSCP	19	*Forward:* TGTCCTGCAGGTCCCTGCGGAGATC *Reverse:* GTGGAGAGATAGCCCATGATTTAAG
00557 [k]	$3.3\ e^{-78}$	2	111	DPGE	19	*Forward:* TCTACACCCGCGTGGCGAGCTATGC *Reverse:* GACTTCATTGCTCGGGACTTTGTTG
00594 [l]	$8.6\ e^{-97}$	4	191	DPGE	19	*Forward:* TGAGGGGAAGCTGGAGCCCCAACTT *Reverse:* CCTTTATCAGAGACACGGTTCTGTT
00596	$> e^{-30}$	3	179	SSCP	19	*Forward:* CGTGCTGAGTTCTGTGTTTCTCTGAATAGA *Reverse:* CCCAACAGCATCTGTAAACATCGGGTACAA
00648	$> e^{-30}$	1	106	DPGE	19	*Forward:* AGTCTGTTAACTTCTTTGATTGTGCTAGTT *Reverse:* ATCACTGACAAATGAGGTACTGTGTGTATT
00836	$> e^{-30}$	2	115	DPGE	19	*Forward:* TATCTCTTTGCATACCCCATGTCTT *Reverse:* AACCCACCCATCCTGATGCCTCTCT
01240 [m]	$3.9\ e^{-103}$	2	129	DPGE	19	*Forward:* AGCTGCTCCAGGACCTAGAGAAGAA *Reverse:* GTGCCCCAGAGGCTGCTGGGTCTTG
02028	$> e^{-30}$	1	261	AGE	19	*Forward:* TGCTCAGTGAGAGACGCCAGACACA *Reverse:* CGTGCGTATGTGAGATATAATGTAC
02198	$> e^{-30}$	1	66	SSCP	19	*Forward:* TTTGCCCATCCTGCTCTCCT *Reverse:* ATCCAGCAGTGAGAAGGAGCCA
02587	$> e^{-30}$	1	123	SSCP	19	*Forward:* CGTGCAGATACTTCATAGCCGTG *Reverse:* CCTGTGTTTCAAAGTCCAGTGATG
02670	$> e^{-30}$	1	115	SSCP	19	*Forward:* CGTACCACTGTTGATGTGAAGCAC *Reverse:* AATGAAGGGGTCCAGGTGACAC
03031	$> e^{-30}$	1	128	SSCP	19	*Forward:* GGAGCTGTTTGATCGCTCTGTG *Reverse:* ATAGCCACTGCCCCAAAAGG
03209	$> e^{-30}$	1	222	SSCP	19	*Forward:* GGAGATAACAGAGGATTTGCACAGG *Reverse:* GAGTTTATCAGGCTATCTCAGCC
00387	$> e^{-30}$	3	159	DPGE	20	*Forward:* AGGATCGGAATGCGGGGTCGAGAGCTGATG *Reverse:* CATGGGCTGGGATTGATGGGAAGGGACAGA
00451 [n]	$2.0\ e^{-71}$	16	247	DPGE	20	*Forward:* CGACAGGAATGCAGACGGCTACATC *Reverse:* TGGGGACCCGGCAGGGCGGAGTCTC
01301	$> e^{-30}$	1	127	DPGE	20	*Forward:* CCCTCATCGCCTGGGGCCGGGAGTC *Reverse:* CCCCAGGATGGTGGGGAGGGATGAA
01498 [o]	$1.4\ e^{-53}$	1	154	AGE	20	*Forward:* CACCCCCAGTAGGTCCCATCTGAAG *Reverse:* CCAAGAGTAAGGAGGAGAGAGACAG
01385	$> e^{-30}$	1	212	AGE	22	*Forward:* TTGGCGTGTGACAGCAGTTCTACAG *Reverse:* AGTAGTAGTCCAAACAGGAAACTAG
02652	$> e^{-30}$	1	148	SSCP	22	*Forward:* TGGGGAGGATTTACACACAGACC *Reverse:* TCAGGAATTAGCCTTGCTCCAC
00705	$> e^{-30}$	1	182	SSCP	X	*Forward:* AGGAGAAGATTAAACTTTCATATT *Reverse:* GAATTTTACAACGTAGCTCTAGAT
01347	$> e^{-30}$	1	128	DPGE	X	*Forward:* ATATGATTTTATGGAGAATGATATG *Reverse:* CATAATTTAAAAGGTGAAGAACTAA
02942	$> e^{-30}$	1	107	SSCP	X	*Forward:* ACTGTAATCTTGCAGTCTCCCCAG *Reverse:* AGCAGCTCCAGGTAAGTGGATGAG

Table 2 (continued)

Skeletal muscle EST code	BLASTN probability	Occurrences out of 4,370 ESTs	PCR fragment size (bp)	Type of analysis	Chromosome assignment	Primers (5′→3′)
03049	$> e^{-30}$	1	74	SSCP	X	*Forward:* TGTGCAAAACTCACTTCTGAG *Reverse:* TGAAAGTCAGCTATCGGG
01448	$> e^{-30}$	3	80	SSCP	1 and 16	*Forward:* TGGTGTAGCGGGCAGCTGC *Reverse:* ATGCTTAATCTGCCGACTCAGG
03104	$> e^{-30}$	1	164	SSCP	5 and X	*Forward:* AACAAGAAGGTTCCTTTACCCCC *Reverse:* AGTTCCTTTCCATGCAGCACAC
01676	$> e^{-30}$	1	145	AGE	17 and Y	*Forward:* GAGTGGGACATGGGCACTCTTATTT *Reverse:* TAAAGCTAGTTCCTATGGCAGACAC
1481	$> e^{-30}$	1	181	AGE	19 and Y	*Forward:* AAACTGCCTCAGACTCCCCCTAGAG *Reverse:* CAGTACCCCAAATGTCTAAGATATT
00832	$> e^{-30}$	1	148	DPGE	X and Y	*Forward:* CAGCGTTCTCTTCACGTCTCTAACA *Reverse:* GTCCAACTTTACAGCATTAAATAAG
00917	$> e^{-30}$	1	163	DPGE	X and Y	*Forward:* GTCTCGCTCCCAGCCATTTGCTGGGATGAC *Reverse:* GGAAACATTCCGAGGGAAAGCAGTTCACAG

[a] Using specific primers for PCR amplification of DNA from human × rodent somatic cell hybrid panels, 56 ESTs were mapped by radiolabeled SSCP (SSCP), 28 by silver-stained denaturing polyacrylamide gel (DPGE), and 30 by agarose gel (AGE) analysis. A similarity search against current databases was performed with the BLASTN computer program (Altschul et al., 1990), and the resulting probability values are reported in the second column. A BLASTN probability value greater than e^{-30} indicates ESTs identifying unknown genes, whereas values between e^{-30} and e^{-60} were considered as an indication of similarity, and values smaller than e^{-60} as an indication of identity.

[b] EST 1265, similar to EST 68E07 from the Stratagene human skeletal muscle cDNA library (Genexpress).

[c] EST 3052 = EST 22901 from Soares infant brain cDNA library (Genexpress).

[d] EST 838 = human cytochrome C oxidase (coxVIIb).

[e] EST 547, similar to human epithelial tropomyosin (TM1).

[f] EST 355, similar to EST 76H08 from the Stratagene human skeletal muscle cDNA library (Genexpress).

[g] EST 686 = EST 96B06 from the Stratagene human skeletal muscle cDNA library (Genexpress).

[h] EST 1040 = human sphingolipid activator protein.

[i] EST 125, similar to human cytochrome C oxidase subunit IV (cox1).

[j] EST 15 = EST 98B10 from the Stratagene human skeletal muscle cDNA library (Genexpress).

[k] EST 557 = human adipsin/complement factor D.

[l] EST 594 = human mRNA for HHR23A protein from a nucleotide excision-repair complex.

[m] EST 1240 = EST 153822 from the Soares human breast cDNA library (WashU-Merck EST project).

[n] EST 451 = human fast skeletal troponin C.

[o] EST 1498, similar to EST a1C-0oB05 from the Soares human fetal brain cDNA library (Genexpress).

Table 3. Comparison between the chromosome distribution of known mapped genes and muscles ESTs

Chromosome	Number of mapped genes	Gene density (%)	Expected number of ESTs (exp)	Observed number of ESTs (obs)	Chi-square value $[(\text{exp} - \text{obs})^2/\text{exp}]$
1	183	9.3	10.323	14	1.310
2	119	6.0	6.660	5	0.414
3	80	4.1	4.551	5	0.044
4	88	4.5	4.995	3	0.797
5	66	3.3	3.663	4	0.031
6	94	4.7	5.217	4	0.284
7	73	3.7	4.107	3	0.298
8	43	2.2	2.442	2	0.080
9	94	4.8	5.328	10	**4.097**
10	71	3.6	3.996	6	1.005
11	138	7.0	7.770	2	**4.285**
12	83	4.2	4.662	5	0.025
13	35	1.8	1.998	2	0.000
14	78	3.9	4.329	3	0.408
15	47	2.4	2.664	2	0.166
16	126	6.4	7.104	9	0.506
17	91	4.6	5.106	6	0.157
18	22	1.1	1.221	0	1.221
19	161	8.1	8.991	14	**2.791**
20	39	2.0	2.220	4	1.427
21	28	1.4	1.554	0	1.554
22	46	2.3	2.553	2	0.120
X	170	8.6	9.546	6	1.317
Total	1,975	100.0	111	111	22.34[a]

[a] The χ^2 value, referred to the entire distribution (degrees of freedom = 22), is not significant; but a nonrandom deviation from the expectation is observed for chromosomes 9, 11, and 19.

Discussion

ESTs derived from the 3′ ends of mRNA have worked very well as probes for precise and unequivocal localization of the corresponding genes to human chromosomes, owing to the great range of sequence variability in this region. The amplification of human DNA with specific primers for our 3′ muscle ESTs always proved effective and specific, and more than 80% of the ESTs could have been successfully assigned to their respective chromosome directly after Separide agarose gel electrophoresis. Therefore, this method may be recommended for mass mapping of ESTs.

The muscle-specific tags considered in this study undoubtedly constitute a sample of minimal size, compared to the total number of genes which could be active in a given tissue; however, they provide preliminary information on the distribution of skeletal muscle genes among human chromosomes.

The apparent nonrandom distribution of these ESTs in the human chromosome complement might derive from an intrinsic heterogeneity of gene content per chromosome, but evidence recently obtained from the chromosome assignment of a large number of ESTs points to a correlation between gene density per chromosome and chromosome length, with the exception of a significant excess found in chromosomes 1, 11, 17, 19, and 22 and a deficiency in the X chromosome (Hudson et al., 1995). An excess of assignment of skeletal muscle ESTs to chromosomes 1 and 19 was also observed in the present study and could be interpreted as a consequence of the relative abundance of genes on these chromosomes.

However, the observed genomic distribution of muscle ESTs showed statistically significant deviation for chromosomes 9, 11, and 19, even taking into account the observed gene density per chromosome. Therefore, the idea that muscle-specific genes might be clustered on certain chromosomes is worth consideration.

A second interesting point is the multiple chromosome localization of six ESTs derived from our library. Since ESTs showing this peculiarity constitute about 5% of the total ESTs considered in this study, the occurrence of genes having identical copies on different chromosomes may be more common than previously thought, at least for skeletal muscle genes. In a previous study on 320 genes obtained from a brain cDNA library, only five cases (1.5%) of multiple assignment were reported (Polymeropoulos et al., 1993). Considering the specificity of the 3′ untranslated region of mRNAs, this result could mean that these multiple copies of genes on different chromosomes may have a recent evolutionary origin. The possible existence of expressed pseudogenes should also be considered. The determination of the full-length sequence of these ESTs will be of great help in elucidating this point.

The building of tissue-specific EST databases and the use of PCR amplification on panels of human × rodent somatic cell hybrids, integrated by tests for tissue specificity, are opening the way for construction of "functional" maps of the human genome. For this purpose, we have set up a World Wide Web site (http://eos.bio.unipd.it) where routinely updated information on our EST collection (sequences, Applied Biosystem electropherograms, mapping data by radiation hybrids [work in progress], PCR primers, homology data, etc) is accessible to the scientific community.

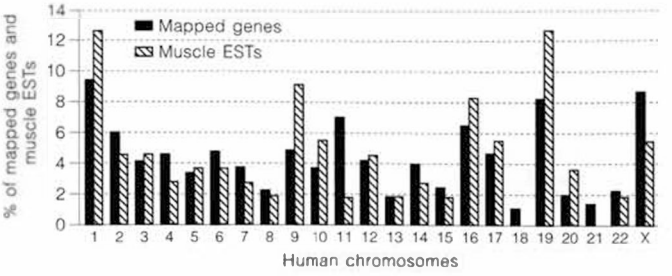

Fig. 2. Distribution of skeletal muscle ESTs on human chromosomes. The figure shows the distribution of 111 human skeletal muscle ESTs (dotted bars) among the human chromosome complement and of the known mapped genes, according to the Location Data Base (black bars). Data are expressed as percentages of the total muscle ESTs and genes considered for this figure.

References

Abbott C, Piaggio G, Ammendola R, Solomon E, Povey S, Gounari F, De Simone V, Cortese R: Mapping of the gene TCF2 coding for the transcription factor LFB3 to human chromosome 17 by polymerase chain reaction. Genomics 8:165–167 (1990).

Adams MD, Kezlavage AR, Fleischmann RD, Fuldnez RA, et al: Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequences, in The Genome Directory, supplement to Nature 377:3–174 (Macmillan, London 1995).

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. J molec Biol 215:403–410 (1990).

Barrett JH: Genetic mapping based on radiation hybrid data. Genomics 13:95–103 (1992).

Ceccherini I, Romeo G, Lawrence S, Breuning MH, Harris PC, Himmelbauer H, Frischauf AM, Sutherland GR, Germino GG, Reeders ST, Morton NE: Construction of a map of chromosome 16 by using radiation hybrids. Proc natl Acad Sci, USA 89:104–108 (1992).

Cohen-Haguenauer O, Burton PJ, Van-Cong N, Cohen A, Masset M, Buckingham M, Frezal J: Chromosomal assignment of two myosin alkali light-chain genes encoding the ventricular slow skeletal muscle isoform and the atrial fetal muscle isoform (MYL3, MYL4). Hum Genet 81:278–282 (1989).

Cox DR, Burmeister M, Price ER, Kim S, Myers RM: Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. Science 250:245–250 (1990).

Drwinga HL, Toji LH, Kim CH, Green AE, Mulivor RA: NIGMS human/rodent somatic cell hybrid panels 1 and 2. Genomics 16:311–314 (1993).

Durkin AS, Nierman WC, Zoghbi H, Jones C, Kozak CA, Maglott DR: Chromosome assignment of human brain expressed sequence tags (ESTs) by analyzing fluorescently labeled PCR products from hybrid cell panels. Cytogenet Cell Genet 65:86–91 (1994).

Gardiner K, Patterson D: The role of somatic cell hybrids in physical mapping. Cytogenet Cell Genet 59:82–85 (1992).

Green ED, Olson MV: Systematic screen of yeast artificial chromosome libraries by use of the polymerase chain reaction. Proc natl Acad Sci, USA 87:1213–1217 (1990).

Hawkins JD: A survey on intron and exon length. Nucl Acids Res 16:9893–9908 (1988).

Hudson TJ, Stein LD, Gezety SG, Ma J, et al: An STS-based map of the human genome. Science 270:1945–1954 (1995).

Hunt CCJ, Eyre HJ, Akkari PA, Meredith C, Dorosz SM, Wilton SD, Collin DF, Laing HG, Baker E: Assignment of the human beta tropomyosin gene (TPM2) to band 9p13 by fluorescence in situ hybridization. Cytogenet Cell Genet 71:94–95 (1995).

Khan AS, Wilcox AS, Polymeropoulos MH, Hopkins JA, Stevens TJ, Robinson M, Orpana AK, Sikela JM: Single pass sequencing and physical and genetic mapping of human brain cDNAs. Nature Genet 2:180–185 (1992).

Lanfranchi G, Muraro T, Caldara F, Pacchioni B, Pallavicini A, Pandolfo D, Toppo S, Trevisan S, Scarso S, Valle G: Identification of 4,370 expressed sequence tags (ESTs) from a 3'-end specific cDNA library of human skeletal muscle by DNA sequencing and filter hybridization. Genome Res 6:35–42 (1996).

Morton NE, Collins A, Lawrence S, Shields DC: Algorithms for a location database. Ann hum Genet 56:223–232 (1992).

Pappas GJ, Polymeropoulos MH, Boyle JM, Trent JM: Regional assignment by hybrid mapping of 36 expressed sequence tags (ESTs) on human chromosome 6. Genomics 25:124–129 (1995).

Polymeropoulos MH, Xiao H, Glodek A, Gorski M, Adams MD, Moreno RF, Fitzgerald MG, Venter JG, Merril CR: Chromosomal assignment of 46 brain cDNAs. Genomics 12:492–496 (1992).

Polymeropoulos MH, Xiao H, Sikela JM, Adams M, Venter JG, Merril CR: Chromosomal distribution of 320 genes from a brain cDNA library. Nature Genet 4:381–386 (1993).

Santos FR, Pena SDJ, Epplen JT: Genetic and population study of a Y-linked tetranucleotide repeat DNA polymorphism with a simple non-isotopic technique. Hum Genet 90:655–656 (1993).

Sikela JM, Auffray C: Finding new genes faster than ever. Nature Genet 3:189–191(1993).

Takeda J, Espinosa R III, Eng S, Le Beau MM, Bell GI: Chromosomal assignment and tissue distribution of novel expressed sequence tags from a human pancreatic islet cDNA library. Genomics 29:276–281 (1995).

Wilcox AS, Khan AS, Hopkins JA, Sikela JM: Use of 3' untranslated sequences of human cDNAs for rapid chromosome assignment and conversion to STSs: implications for an expression map of the genome. Nucl Acids Res 19:1837–1843 (1991).

Youssoufian H, McAfu M, Kwiatkowski DJ: Cloning and chromosomal localization of the human cytoskeletal alpha-actinin gene reveals linkage to the α-spectrin gene. Am J hum Genet 47:62–72 (1990).