Human Molecular Genetics, 1997, Vol. 6, No. 9 **1445–1450**

# The preliminary transcript map of a human skeletal muscle

**Alberto Pallavicini[1], Rosanna Zimbello[1], Natascia Tiso[2], Teresa Muraro[1], Luca Rampoldi[2], Stefania Bortoluzzi[2], Giorgio Valle[1,2], Gerolamo Lanfranchi[1,2] and Gian A. Danieli[2,*]**

[1]CRIBI Biotechnology Centre and [2]Department of Biology, University of Padua, via G. Colombo, 3, 35121 Padova, Italy

Received March 3, 1997; Revised and Accepted June 9, 1997

**By sequencing 11 405 individual expressed sequence tags (ESTs) from a cDNA library of a human skeletal muscle, we identified 1945 individual transcripts, 725 of which showed no correspondence with known human genes. We report here the chromosomal localization of 267 of these, obtained by radiation hybrid (RH) mapping. The map position of additional 242 ESTs from the same library, corresponding to known human genes, is also reported. The resulting information provides a preliminary genomic transcriptional profile of a human muscle. Several genes occur in clusters on different chromosomes. Moreover, chromosomes 17, 19, 21 and X appear to be significantly rich in muscle ESTs. By analysing several collections of ESTs from different tissues, we observed significant deviations in the distribution of ESTs by chromosome in fetal heart, adult brain and adult retina, supporting the hypothesis that a non-random localization of genes expressed in specific tissues might not be uncommon. The selective concentration of expressed genes in some chromosomes and in specific chromosomal subregions in a given tissue might reflect the existence of batteries of genes under the same control mechanisms, regulating tissue-specific gene expression.**

## INTRODUCTION

Large-scale sequencing of cDNA clones from human tissues is the most efficient approach for identifying new genes and for a systematic analysis of transcripts. The possibility of a comparative analysis of clones derived from different tissues (1), or from the same tissue at different times, discloses a new perspective in understanding the mechanisms of development and differentiation. The expression profiles of several tissue- or organ-specific libraries have been reported, including cardiac muscle (2,3), whereas little is known in this regard about human skeletal muscle (4,5). For this reason, we started a new project, supported by Telethon-Italy, with the specific aim of identifying expressed sequence tags (ESTs) from a 3′-end-specific cDNA library of human skeletal muscle. A new method was devised for this purpose, in order to circumvent the problem of the different clonability of mRNAs of different length, along with an ancillary method for the rapid identification of clones derived from the most abundant transcripts. To avoid the preferential cloning of short transcripts, random fragmentation of double-stranded cDNA was obtained by sonication, rather than by restriction digestion. Moreover, an effective selection was introduced for the 3′-end-specific cDNA tags to be cloned into vectors (6,7). By these methods, it was possible to obtain a library in which the relative frequency of each cDNA clone is likely to represent the actual abundance of the corresponding transcript in the tissue.

The chromosomal assignment of 115 ESTs from this catalogue, corresponding to new genes, was reported recently (8), showing that some chromosomes might be richer than others in genes expressed in skeletal muscle. Following this preliminary indication, we decided to clarify this issue by radiation hybrid (RH) mapping of a large sample of ESTs from our database, with the aim of building a preliminary genomic map of genes expressed in human skeletal muscle.

## RESULTS

### The skeletal muscle ESTs library

The cDNA was synthesized from muscle mRNA purified from a specimen of human *pectoralis maior*. The library consists of 450–550 bp long cDNA fragments, corresponding to the 3′-ends of skeletal muscle mRNAs. Up to now, 11 405 clones have been isolated by our group, 60% of them by direct sequencing and 40% by methods specially designed in our laboratory for the detection of the most abundant muscle transcripts. A brief description of the methods used for the construction and large-scale sequencing of the library is reported in Materials and Methods. These 11 405 ESTs were compared with each other by BLAST and, as a consequence, 1945 non-overlapping segments were obtained, 75% of which were single ESTs. The BLAST search against GenBank databases revealed that 608 transcripts corresponded to known human genes, whereas 1337 were apparently new genes. Among these 1337, only 27% resulted in a match with ESTs produced by other laboratories, ~18% were similar to known DNA sequences, whereas 725 ESTs showed no sequence similarity to any of the sequences available in current databases. The catalogue of the skeletal muscle ESTs produced in our laboratory is available on Internet (http://eos.bio.unipd.it) and is updated regularly.

*To whom correspondence should be addressed. Tel: +39 49 827 6215; Fax: +39 49 827 6209; Email: danieli@civ.bio.unipd.it

**Table 1.** Observed and expected distribution by chromosome of ESTs from different human tissues

| | HTM | | Skeletal muscle | | | Fetal heart | | | Brain | | | Retina | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Chr. | Genes | % | Obs | Exp | $\chi^2$ | Obs | Exp | $\chi^2$ | Obs | Exp | $\chi^2$ | Obs | Exp | $\chi^2$ |
| 1 | 2162 | 10.79 | 54 | 54.92 | 0.02 | 354 | 341.20 | 0.48 | 145 | 134.88 | 0.76 | 147 | 141.68 | 0.20 |
| 2 | 1600 | 7.99 | 34 | 40.65 | 1.09 | 224 | 252.51 | 3.22 | 84 | 99.82 | 2.51 | 31 | 104.85 | **52.02 |
| 3 | 1388 | 6.93 | 20 | 35.26 | 6.61 | 127 | 219.05 | **38.68 | 81 | 86.59 | 0.36 | 22 | 90.96 | **52.28 |
| 4 | 1013 | 5.06 | 12 | 25.73 | 7.33 | 129 | 159.87 | 5.96 | 35 | 63.20 | **12.58 | 82 | 66.38 | 3.67 |
| 5 | 974 | 4.86 | 19 | 24.74 | 1.33 | 140 | 153.71 | 1.22 | 53 | 60.77 | 0.99 | 67 | 63.83 | 0.16 |
| 6 | 1034 | 5.16 | 23 | 26.27 | 0.41 | 154 | 163.18 | 0.52 | 60 | 64.51 | 0.32 | 67 | 67.76 | 0.00 |
| 7 | 1111 | 5.55 | 16 | 28.22 | 5.29 | 174 | 175.33 | 0.01 | 64 | 69.31 | 0.41 | 73 | 72.81 | 0.00 |
| 8 | 865 | 4.32 | 17 | 21.97 | 1.13 | 105 | 136.51 | 7.27 | 43 | 53.97 | 2.23 | 67 | 56.69 | 1.88 |
| 9 | 852 | 4.25 | 28 | 21.64 | 1.87 | 135 | 134.46 | 0.00 | 44 | 53.15 | 1.58 | 58 | 55.83 | 0.08 |
| 10 | 887 | 4.43 | 20 | 22.53 | 0.28 | 137 | 139.98 | 0.06 | 45 | 55.34 | 1.93 | 74 | 58.13 | 4.33 |
| 11 | 1344 | 6.71 | 31 | 34.14 | 0.29 | 208 | 212.10 | 0.08 | 99 | 83.85 | 2.74 | 109 | 88.08 | 4.97 |
| 12 | 958 | 4.78 | 29 | 24.34 | 0.89 | 177 | 151.19 | 4.41 | 62 | 59.77 | 0.08 | 89 | 62.78 | *10.95 |
| 13 | 377 | 1.88 | 9 | 9.58 | 0.03 | 45 | 59.50 | 3.53 | 10 | 23.52 | 7.77 | 23 | 24.71 | 0.12 |
| 14 | 713 | 3.56 | 13 | 18.11 | 1.44 | 121 | 112.52 | 0.64 | 32 | 44.48 | 3.50 | 58 | 46.72 | 2.72 |
| 15 | 626 | 3.12 | 13 | 15.90 | 0.53 | 94 | 98.79 | 0.23 | 38 | 39.05 | 0.03 | 47 | 41.02 | 0.87 |
| 16 | 573 | 2.86 | 25 | 14.56 | 7.49 | 120 | 90.43 | 9.67 | 59 | 35.75 | **15.12 | 42 | 37.55 | 0.53 |
| 17 | 790 | 3.94 | 41 | 20.07 | **21.83 | 188 | 124.67 | **32.16 | 81 | 49.29 | **20.41 | 80 | 51.77 | **15.39 |
| 18 | 415 | 2.07 | 2 | 10.54 | 6.92 | 62 | 65.49 | 0.19 | 22 | 25.89 | 0.58 | 35 | 27.20 | 2.24 |
| 19 | 644 | 3.21 | 44 | 16.36 | **46.70 | 141 | 101.63 | **15.25 | 69 | 40.18 | **20.68 | 27 | 42.20 | 5.48 |
| 20 | 567 | 2.83 | 8 | 14.40 | 2.85 | 105 | 89.48 | 2.69 | 40 | 35.37 | 0.61 | 40 | 37.16 | 0.22 |
| 21 | 163 | 0.81 | 11 | 4.14 | *11.36 | 45 | 25.72 | **14.44 | 11 | 10.17 | 0.07 | 9 | 10.68 | 0.26 |
| 22 | 432 | 2.16 | 12 | 10.97 | 0.10 | 75 | 68.18 | 0.68 | 40 | 26.95 | 6.32 | 30 | 28.31 | 0.10 |
| X | 548 | 2.74 | 28 | 13.92 | **14.24 | 102 | 86.48 | 2.78 | 33 | 34.19 | 0.04 | 36 | 35.91 | 0.00 |
| Total | 20 036 | 100.00 | 509 | 509.00 | **140.02 | 3162 | 3162.00 | **144.19 | 1250 | 1250.00 | **101.61 | 1313 | 1313.00 | **158.48 |

Statistically significant deviations are indicated by *($P$<0.001) or by **($P$<0.0005).

Taking advantage of the BODY MAP database, the above catalogue of muscle ESTs was compared with the expression profiles of human fibroblasts, granulocytes, adipose tissue and aortic endothelium. The comparison involved the first 100 most frequent ESTs from skeletal muscle and all the ESTs reported in each catalogue. Overlaps were found, involving only genes for ribosomal proteins (all tissues), glyceraldehyde-3-phosphate dehydrogenase (fibroblasts only) and translationally controlled tumour protein (aortic endothelium only). Moreover, some ESTs corresponding to α- and β-globin genes were found in the skeletal muscle catalogue, due to the presence of blood in the muscle specimen. These findings suggest that the cDNA library from which the ESTs were derived is sufficiently representative of genes expressed in skeletal muscle.

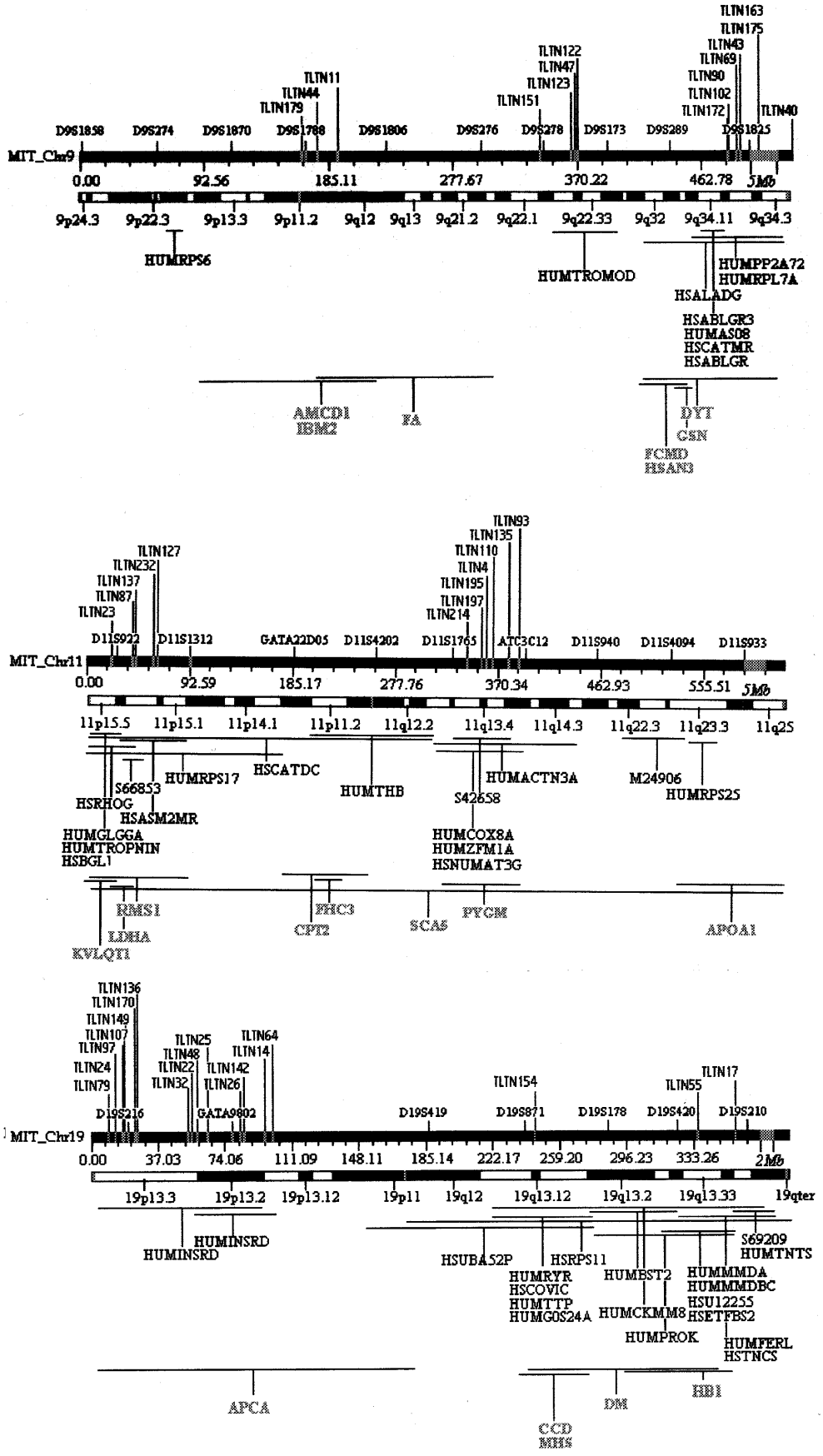
## RH mapping of muscle ESTs

Specific primers were designed to amplify 267 skeletal muscle ESTs from our library, corresponding to new genes. After PCR amplification of the GeneBridge4 panel (9), the retention profiles were submitted to the Whitehead Institute (MIT, USA), where they were analysed by the RHMAPPER program (10). The amplification of 37 ESTs produced exceedingly high or low retention profiles. In these cases, only the chromosomal assignment was possible, by monochromosomal cell hybrids (11). For the remaining 230 ESTs, a precise map position was obtained. The 267 skeletal muscle ESTs corresponding to new genes, along with their PCR primers and their map position are listed in supplementary material to this paper, published on HMG OnLine. These ESTs, potentially corresponding to new muscle genes, were called TLTN-ESTs (TLTN stands for TELETHON, the financing Agency).

## A preliminary genomic transcriptional profile of a skeletal muscle

Among the ESTs identified in the course of the project, 608 were found to correspond to known human genes. For 242 of these (40%) the map position was already available in the OMIM, GenBank or LDB databases. By pooling these data with the information obtained from the RH mapping of the 267 TLTN-ESTs, a preliminary

**Figure 1.** Comprehensive map of human skeletal muscle ESTs assigned to chromosomes 9, 11 and 19. For each chromosome, both the RH map with reference markers and the cytogenetic map are shown. At the extreme right of each RH map, a grey segment indicates (in Mb) the scale of the map. The localization of ESTs corresponding to new genes (TLTN) are reported on the RH map, whereas the localization of ESTs corresponding to known genes are on the cytogenetic map. The localization of genes involved in neuromuscular disorders (symbols in grey) is also reported. The comprehensive map of 509 ESTs is available as supplementary material on HMG OnLine.


TLTN163
TLTN175
TLTN43
TLTN69
TLTN122
TLTN47
TLTN123
TLTN90
TLTN102
TLTN11
TLTN44
TLTN179
TLTN151
TLTN172
TLTN40
D9S1858
D9S274
D9S1870
D9S1788
D9S1806
D9S276
D9S278
D9S173
D9S289
D9S1825
MIT_Chr9
0.00
92.56
185.11
277.67
370.22
462.78
5Mb
9p24.3
9p22.3
9p13.3
9p11.2
9q12
9q13
9q21.2
9q22.1
9q22.33
9q32
9q34.11
9q34.3
HUMRPS6
HUMTROMOD
HUMPP2A72
HUMRPL7A
HSALADG
HSABLGR3
HUMAS08
HSCATMR
HSABLGR
AMCD1
IBM2
FA
DYT
CSN
FCMD
HSAND
TLTN93
TLTN135
TLTN110
TLTN127
TLTN232
TLTN4
TLTN137
TLTN195
TLTN87
TLTN197
TLTN23
TLTN214
D11S922
D11S1312
GATA22D05
D11S4202
D11S1765
ATC3C12
D11S940
D11S4094
D11S933
MIT_Chr11
0.00
92.59
185.17
277.76
370.34
462.93
555.51
5Mb
11p15.5
11p15.1
11p14.1
11p11.2
11q12.2
11q13.4
11q14.3
11q22.3
11q23.3
11q25
HUMRPS17
HSCATDC
S66853
HSRHOG
HSASM2MR
HUMTHB
HUMACTN3A
S42658
M24906
HUMRPS25
HUMGLGGA
HUMTROPNIN
HSBGL1
HUMCOX8A
HUMZFM1A
HSNUMAT3G
HMS1
LDHA
KVLQT1
FHC3
CPT2
SCA5
PYGM
APOA1
TLTN136
TLTN170
TLTN149
TLTN107
TLTN97
TLTN24
TLTN79
TLTN25
TLTN48
TLTN22
TLTN32
TLTN142
TLTN26
TLTN64
TLTN14
TLTN154
TLTN55
TLTN17
D19S216
GATA9B02
D19S419
D19S871
D19S178
D19S420
D19S210
MIT_Chr19
0.00
37.03
74.06
111.09
148.11
185.14
222.17
259.20
296.23
333.26
2Mb
19p13.3
19p13.2
19p13.12
19p11
19q12
19q13.12
19q13.2
19q13.33
19qter
HUMINSRD
HUMINSRD
HSUBA52P
HSRPS11
HUMRYR
HSCOVIC
HUMTTP
HUMG0S24A
HUMBST2
HUMCKMM8
HUMPROK
S69209
HUMTNTS
HUMMMDA
HUMMMDBC
HSU12255
HSETFBS2
HUMFERL
HSTNCS
APCA
DM
HB1
CCD
MHS

distribution of muscle genes among human chromosomes was obtained.

Figure 1 shows the chromosomal distribution of the muscle ESTs on chromosomes 9, 11 and 19, considered in this study. The comprehensive map of 509 ESTs is published as supplementary material in HMG OnLine. In the same map, the position of disease loci involved in neuromuscular disorders is reported. The general impression is that the distribution of muscle genes along human chromosomes is not random. Twenty seven TLTN-ESTs corresponding to new genes (~10% of the total) appear to map in subchromosomal regions spanning <1 Mb each (1p36.32, 1p32.3p21, 9q34.11, 19p13.3). In particular, five ESTs map on 1p36.32, and seven map very close to each other on 19p13.3. Moreover, several ESTs expressed in muscle and corresponding to known genes are concentrated in specific chromosomal regions (1q21, 2p23, 6p21.3, 9q34.1, 11p15.5, 12p13, 17q21, 19q13.1, 19q13.3, 21q22.3 and Xq28). It appears that some chromosomal subregions (6p21.3, 9q34.2, 11p15.4, 11q13.4, 17p13.2) are particularly rich both in ESTs corresponding to new genes and in ESTs of known genes. Furthermore, several known disease loci for neuromuscular disorders map in correspondence with the above clusters.

The expected distribution of ESTs by chromosome was calculated according to the chromosomal gene density, as reported by the Human Transcript Map (HTM) database (12), the largest collection of mapped human ESTs available so far. The observed distribution deviates significantly from the expectations ($\chi^2$ = 140.02, 22 df; $P$ <0.001), although for 19 chromosomes (involving in total 89% of the genes reported so far) there is no deviation between observed and expected data (Table 1). The deviation from the expectation is highly significant for chromosomes 17 ($\chi^2$ = 21.83, 1 df; $P$ <0.0005), 19 ($\chi^2$ = 46.70, 1 df; $P$ <0.0005), 21 ($\chi^2$ = 11.36 , 1 df; $P$ <0.001) and X ($\chi^2$ = 14.24, 1 df; $P$ <0.0005), which appear enriched in ESTs expressed in muscle.

### EST distribution by chromosome in different tissues

The distribution of ESTs by chromosome in different tissues was investigated by taking advantage of the option 'query ESTs by chromosome and tissue' of the LENS database. This database links and resolve the names and identifiers of clones and ESTs generated by the I.M.A.G.E. consortium (13). For each selected tissue, the ESTs mapped in each chromosome, corresponding to distinct UNIGENE (14) (Unique Human Gene Sequence Collection) clusters, were identified and counted. The genomic distribution of such ESTs was compared with the expectation calculated according to the HTM database (12).

Fetal heart is the only muscle tissue available for the search in LENS. Its genomic distribution of ESTs significantly deviates from the expectation ($\chi^2$ = 144.19, 22 df; $P$ <0.0005). In particular, chromosomes 17, 19 and 21, as in skeletal muscle, seem richer in expressed genes than other chromosomes (Table 1).

Five additional LENS ESTs collections were analysed, all from human adult tissues (brain, retina, liver, lung and ovary). The genomic distribution of ESTs from liver, ovary and lung showed no deviation from the expectation (data not shown), whereas significant deviations were observed in brain ($\chi^2$ = 101.61, 22 df; $P$ <0.0005) and in retina ($\chi^2$ = 158.48, 22 df; $P$ <0.0005) (Table 1).

## DISCUSSION

An attempt to provide a preliminary genomic transcriptional profile of a given tissue may be jeopardized by two major biases: (i) the complexity (in terms of cell types) of the source tissue from which the library was obtained; and (ii) the poor correspondence between relative frequency of each EST in the library and the actual abundance of the corresponding mRNA in the tissue.

In the present study, the cDNA library may be considered representative of the skeletal muscle mRNAs, since an extremely low fraction of ESTs found correspondence in libraries from possibly contaminant tissues. This does not imply that all the ESTs obtained in our study are muscle-specific, but simply that they correspond to transcripts of the muscle cells.

The strategy adopted for the construction of the cDNA library was designed specifically to preserve the actual relative abundance of each mRNA (6). This was confirmed by Northern blot analyses (6) and by the screening of other cDNA libraries prepared from skeletal muscle mRNA (Valle *et al.*, submitted).

The mapping of 509 skeletal muscle ESTs to human chromosomes (Fig. 1) showed the presence of several clusters of ESTs. This was expected, since several genes coding for relevant muscle proteins are known to occur in functional clusters or to belong to multigene families, like actins, myosins, troponins and tropomyosins. However, it is interesting to note that, in general, ESTs mapping in a given cluster did not show sequence similarity among themselves.

Several genes involved in neuromuscular disorders map in chromosomal subregions rich in ESTs expressed in muscle. It is possible that some still unknown disease genes correspond to ESTs mapped by the present study. However, the difference in resolution power between the classical linkage map and the RH map hinders their direct identification.

Chromosomes 17, 19, 21 and X appear particularly rich in genes expressed in muscle. The deviation from the expected distribution ($P$ >0.0005) is statistically very significant even at the stringent level established by the Bonferroni's correction. A comparable deviation was observed also when considering the distribution by chromosome of a very large sample of ESTs from fetal heart. It is interesting to note that also in this muscle tissue, chromosomes 17, 19 and 21 appear enriched in expressed genes. These findings substantiate the suspicion of a non-random distribution of muscle ESTs, raised by a preliminary investigation on 115 sequence tags from skeletal muscle (8). On the other hand, the analysis of the genomic distribution of ESTs from other human tissues suggests that this phenomenon might be not uncommon.

The selective concentration of expressed genes in some chromosomes and in specific chromosomal subregions in a given tissue might reflect the existence of batteries of genes under the same control mechanisms, regulating tissue-specific gene expression.

## MATERIALS AND METHODS

### cDNA library, large-scale DNA sequencing and local EST database

The cDNA library was prepared from a surgical specimen of *pectoralis maior* muscle of an adult woman. RNA was reverse transcribed and primed by an oligo(dT)–*Not*I carrying a biotinylated 5′ terminus. After completion of the second strand,

the cDNA was sonicated, end-repaired, ligated to *Bst*XI non-palindromic adaptors and size-fractionated on low-melting agarose gel. Sonicated cDNA of 450–550 bp was selected and the 3′-specific cDNA fragments were purified by binding to avidin-coated paramagnetic beads and subsequently released by *Not*I digestion. The cDNA was cloned directionally into a *Bst*XI- and *Not*I-digested pcDNAII plasmid vector (Invitrogen) and electroporated into the *Escherichia coli* TOP10F′ strain (6).

For large-scale DNA sequencing, the cDNA library was diluted in liquid medium in 96-well plates at a density of 0.5 c.f.u. per well. After overnight incubation at 37°C grown wells were arrayed on fresh plates containing glycerol and stored at –80°C. Templates for DNA sequencing were then generated by PCR, using 5 µl of the bacterial culture and a pair of universal primers (7). Following the inspection of the agarose gel electrophoretic separation of the PCR products, 5 µl of amplification products were used for DNA sequencing by dye-primer chemistry. The DNA sequence analysis was performed on an ABI 377 automatic sequencer (Applied Biosystems). Handling of samples was performed with the aid of a Biomek 2000 robotic station (Beckman). A checking system was developed which allows the identification of cDNA clones belonging to the 10 most abundant muscle transcripts (adding up to 40% of the total mRNA) before sequencing. A set of 'interference primers', specific for the 10 most abundant mRNAs, are added to universal primers in the PCR reaction. In the presence of one of the abundant transcripts, the interference primers would compete for the template and, hence, a double band would result in the agarose gel electrophoretic separation. In this case, the clone would not be processed for sequencing (7).

A local database of skeletal muscle ESTs, identified by systematic sequencing in our laboratory, is available on Internet (http://eos.bio.unipd.it).

### Primer design

Primers, synthesized from Genset, were designed specifically based on the original DNA sequence of each TLTN-EST present in our database, using the software OLIGO (Version 4.0) or PRIMER3 (15). Primers were generally 18–22 bp long with a GC content between 40 and 60%. The $T_m$ range was 57–61°C. In the selection of primers, we excluded as target sequences regions very close (~20 bp) to the poly(A) tract, where polyadenylation signals could be present.

### Amplification of radiation hybrids

The RH mapping was performed by the GeneBridge4 whole-genome Radiation Hybrid Panel (Research Genetics) consisting of 93 genomic DNAs from the same number of human-on-hamster somatic cell lines, plus the two control DNAs (HFL donor and A23 recipient) (9).

Twenty ng of genomic DNA was used for amplification in 10 µl of PCR buffer (16.6 mM $(NH_4)_2SO_4$; 67 mM Tris–HCl pH 8.3; 0.01% Tween-20; 1.5 mM $MgCl_2$) containing 800 nM each of the forward and reverse primers, 0.2 U of DNA polymerase (RTB polymerase; Bioline, Italy) and 25 µM of each of the four dNTPs. The mix was overlaid with 5 µl of mineral oil.

After primer optimization, the reactions were carried out in 96-well plates (93 GeneBridge lines, plus three controls). The preparation of the reaction mixtures and the delivery of the DNA samples on to the plates were performed by aid of a Beckman Biomek 2000 robotic station. The PCR reactions were carried out in a PTC-225 Peltier Thermal Cycler (MJ Research).

Cycling conditions were: 1 min 15 s at 94°C, followed by 35 cycles of 15 s at 94°C, 25 s at the working annealing temperature and 30 s at 72°C, and a final extension step for 1 min 30 s at 72°C.

The PCR products were mixed with 5 µl of loading buffer (30% glycerol; 0.25% Orange G; 25 mM EDTA) and separated on a 2.5% horizontal agarose gel in TAE buffer (40 mM Tris-acetate and 1 mM EDTA) stained with ethidium bromide. The gel was electrophoresed in the same buffer at 100 V for 1 h.

### Analysis of the retention profiles

The retention profiles were deduced from the results of the electrophoretic separations and were submitted to the Whitehead Institute/MIT Center for Genome Research (USA), where they were processed by the program RHMAPPER (10) and placed in the Whitehead framework of markers.

One centiRay (cR) corresponds to a 1% frequency of breakage between the involved markers, at a given dose of radiation used for the chromosome fragmentation. The GeneBridge panel was obtained with a dosage of 3000 rad, and 1 cR corresponded to ~300 kb (16).

The localization of each TLTN-EST on the cytogenetic map was established by taking advantage of the position of the flanking markers and by comparison with the Location Data Base (LDB) (17).

The comprehensive map in which TLTN-ESTs corresponding to new genes are located with respect to known genes and to genes known to be involved in neuromuscular disorders was obtained by assembling data from OMIM, GenBank and Neuromuscular Disorders:Gene Location (18).

### Collection of data from remote databases

Data were collected from several databases:

BODYMAP (http://www.imcb.osaka-u.ac.jp/bodymap)
dbEST (http://www3.ncbi.nlm.nih.gov/dbEST)
GDB (http://gdbwww.gdb.org/gdb/)
GenBank (http://www3.ncbi.nlm.nih/Entrez)
Human Transcript Map (http://www.ncbi.nlm.nih.gov/SCIENCE96/)
LENS (http://agave.humgen.upenn.edu/lens/index.html)
LDB (http://cedar.genetics.soton.ac.uk/public_html)
OMIM (http://www3.ncbi.nlm.nih.gov/OMIM)
UNIGENE (http://www.ncbi.nlm.nih.gov/UniGene)

### Calculating the expected distribution and statistical tests

The number of genes per chromosome was deduced from the Human Transcript Map catalogue (12). The resulting distribution was used for calculating the expected distribution of ESTs by chromosomes. The Y chromosome was excluded from calculations, since the original cDNA library was obtained from the muscle of a female subject.

The statistical significance of the deviation of the observed distribution from the expectation was tested by a $\chi^2$ test, with 22 degrees of freedom, with the level of significance established at 0.002 (i.e. 0.05/23), according to the Bonferroni correction (19). When considering single chromosomes, the statistical significance of the deviation was tested both by a $\chi^2$ test (obs/exp) with 22/23 degrees of freedom and by a $\chi^2$ test in which the deviation for that

specific chromosome was compared with the deviation observed for the whole genome (excluding the chromosome involved), with one degree of freedom. Also in these tests, the level of statistical significance was established as specified above.

## SUPPLEMENTARY MATERIAL

Further supplementary material is available on HMG OnLine.

## ACKNOWLEDGEMENTS

Professor I. Barrai and Dr G. Bertorelle provided helpful advice on the statistical treatment of data. Dr Gibson, CBIL University of Pennsylvania facilitated the access to the LENS database. This work was financially supported entirely by TELETHON-Italy (grant no. B30 to G. Lanfranchi and G. Valle). A. Pallavicini, N. Tiso, L. Rampoldi and S. Bortoluzzi are PhD students of 'Dottorato in Scienze Genetiche', University of Ferrara, Italy.

## REFERENCES

1. Okubo, K., Hori, N., Matoba, R., Niiyama, T., Fukushima, A., Kojima, Y. and Matsubara, K. (1992) Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. *Nature Genet.*, **2**, 173–179.
2. Liew, C.C., Hwang, D.M., Fung, Y.W., Laurenssen, C., Cukerman, E., Tsui, S. and Lee, C.Y. (1994) A catalogue of genes in the cardiovascular system as identified by expressed sequence tags. *Proc. Natl Acad. Sci, USA*, **91**, 10645–10649.
3. Tanaka, T., Ogiwara, A., Uchiyama, I., Takagi, T., Yazaki, Y. and Nakamura, Y. (1996) Construction of a normalized directionally cloned cDNA library from adult heart and analysis of 3040 clones by partial sequencing. *Genomics*, **35**, 231–235.
4. Houlgatte, R., Mariagesamson, R., Duprat, S., Tessier, A., Bentolila, S., Lamy, B. and Auffray, C. (1995) The Genexpress index: a resource for gene discovery and the genetic map of the human genome. *Genome Res.*, **5**, 272–304.
5. Pietu, G., Alibert, O., Guichard, V., Lamy, B., Bois, F., Leroy, E., Mariagesamson, R., Houlgatte, R., Soularue, P. and Auffray, C. (1996) Novel gene transcripts preferentially expressed in human muscles revealed by quantitative hybridization of a high density cDNA array. *Genome Res.*, **6**, 492–503.
6. Lanfranchi, G., Muraro, T., Caldara, F., Pacchioni, B., Pallavicini, A., Pandolfo, D., Toppo, S., Trevisan, S., Scarso, S. and Valle, G. (1996) Identification of 4370 expressed sequence tags from a 3′ end specific cDNA library of human skeletal muscle by DNA sequencing and filter hybridization. *Genome Res.*, **6**, 35–42.
7. Pacchioni, B., Trevisan, S., Gomirato, S., Toppo, S., Valle, G. and Lanfranchi, G. (1996) A semimultiplex PCR technique for the screening of abundant transcripts during systematic sequencing of cDNA libraries. *Biotechniques*, **21**, 644–649.
8. Muraro, T., Stephan, D., Tiso, N., Zimbello, R., Danieli, G.A., Hoffman, E.H., Valle, G. and Lanfranchi, G. (1997) Chromosomal assignment of 115 expressed sequence tags (ESTs) from human skeletal muscle. *Cytogenet. Cell Genet.*, **76**, 144–152.
9. Walter, M.A., Spillet, D.J., Thomas, P., Weissenbach, J. and Goodfellow, P.N. (1994) A method for constructing radiation hybrid maps of whole genomes. *Nature Genet.*, **7**, 22–28.
10. Slonim, D., Stein., L., Kruglyak, L. and Lander, E. RHMAPPER: an interactive computer package for constructing radiation hybrid maps. (http://www-genome.wi.mit.edu/ftp/pub).
11. Gardiner, K. and Patterson, D. (1992) The role of somatic cell hybrids in physical mapping. *Cytogenet. Cell Genet.*, **59**, 82–85.
12. Schuler, G.D., Boguski, M,S., Stewart, E.A., Stein, L.D., Gyapay, G., Rice, K., White, R.E., Rodrigueztome, P., Aggarwal, A., Bajorek, E., Bentolila, S., Birren, B.B., Butler, A., Castle, A.B., Chiannilkulchai, N., Chu, A., Clee, C., Cowles, S., Day, P.J.R., Dibling, T., Drouot, N., Dunham, I., Duprat, S., East, C., Edwards, C., Fan, J.B., Fang, N., Fizames, C., Garrett, C., Green, L., Hadley, D., Harris, M., Harrison, P., Brady S., Hicks, A., Holloway, E., Hui, L., Hussain, S., Louisditsully, C., Ma, J., Macgilvery, A., Mader, C., Maratukulam, A., Matise, T.C., McKusick, K.B., Morissette, J., Mungall, A., Muselet, D., Nusbaum, H.C., Page, D.C., Peck, A., Perkins, S., Piercy, M., Qin, F., Quackenbush, J., Ranby, S., Reif, T., Rozen, S., Sanders, C., She, X., Silva, J., Slonim, D.K., Soderlund, C., Sun, W.L., Tabar, P., Thangarajah, T., Vegaczamy, N., Vollrath, D., Voyticky, S., Wilmer, T., Wu, X., Adams M.D., Auffray, C., Walter, N.A.R., Brandon, R., Dehejia, A., Goodfellow, P.N., Houlgatte, R., Hudson, J.R., Ide, S.E., Iorio, K.R., Lee, W.Y., Seki, N., Nagase, T., Ishikawa, K., Nomura, N., Phillips, C., Polymeropoulos, M.H., Sandusky, M., Schmitt, K., Berry, R., Swanson, K., Torres, R., Venter, J.C., Sikela, J.M., Beckmann, J.S., Weissenbach, J., Myers, R.M., Cox, D.R., James, M.R., Bentley, D., Deloukas, P., Lander, E.S. and Hudson, T.J. (1996) A gene map of the human genome. *Science*, **274**, 540–546.
13. Lennon, G., Auffray C., Polymeropoulos M., and Soares M.B. (1996) The image consortium: an integrated molecular analysis of genomes and their expression. *Genomics*, **33**, 151–152.
14. Boguski, M.S. and Schuler, G.D. (1995) ESTablishing a human transcript map. *Nature Genet.*, **10**, 369–371.
15. Rozen, S. and Skaletsky, H.J. (1996) PRIMER: a computer program for automatically selecting PCR primers. (http://www-genome.wi.mit.edu/genome_software/other/primer3.html).
16. Hudson, T.J., Stein, L.D., Gerety, S.S., Ma, J.L., Castle, A.B., Silva, J., Slonim, D.K., Baptista, R., Kruglyak, L., Xu, S.H., Hu, X.T., Colbert, A.M.E., Rosenberg, C., Reevedaly, M.P., Rozen S., Hui, L., Wu, X.Y., Vestergaard, C., Wilson, K.M., Bae, J.S., Maitra, S., Ganiatsas, S., Evans, C.A., Deangelis, M.M., Ingalls, K.A., Nahf, R.W., Horton, L.T., Anderson, M.O., Collymore, A.J., Ye, W.J., Kouyoumjian, V., Zemsteva, I.S., Tam, J., Devine, R., Courtney, D.F., Renaud, M.T., Nguyen, H., Oconnor, T.J., Fizames, C., Faure, S., Gyapay, G., Dib, C., Morissette, J., Orlin, J.B., Birren, B.W., Goodman, N., Weissenbach, J., Hawkins, T.L., Foote, S., Page, D.C. and Lander, E.S. (1995) An STS-based map of the human genome. *Science*, **270**, 1945–1954.
17. Collins, A., Frezal, J., Teague, J. and Morton, N.E. (1996) A metric map of humans: 23,500 loci in 850 bands. *Proc. Natl Acad. Sci. USA*, **93**, 14771–14775.
18. Kaplan, J.C. and Fontaine B. (1996) Neuromuscular disorders: gene location. *Neuromusc. Disord.*, **6**, I–IV.
19. Simes, R.J. (1986) An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, **73**, 751–754.