The Genexpress IMAGE Knowledge Base of the Human Muscle Transcriptome: A Resource of Structural, Functional, and Positional Candidate Genes for Muscle Physiology and Pathologies

Geneviève Piétu,[7] Eric Eveno, Béatrice Soury-Segurens, Nicole-Adeline Fayein, Régine Mariage-Samson, Christiane Matingou, Elisabeth Leroy,[1,6] Claude Dechesne,[2] Sabine Krieger,[3] Wilhelm Ansorge,[3] Isabelle Reguigne-Arnould,[4] David Cox,[5] Anindya Dehejia,[1] Mihael H. Polymeropoulos,[1] Marie-Dominique Devignes, and Charles Auffray

Genexpress, Centre National de la Recherche Scientifique (CNRS) ERS 1984, 94801 Villejuif, France; [1]National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892 USA; [2]CNRS UMR 6548, 06108 Nice, France; [3]European Molecular Biology Laboratory, Heidelberg, Germany; [5]Department of Genetics, Stanford University, Stanford, California, 96305 USA

Sequence, gene mapping, and expression data corresponding to 910 genes transcribed in human skeletal muscle have been integrated to form the muscle module of the Genexpress IMAGE Knowledge Base. Based on cDNA array hybridization, a set of 14 transcripts preferentially or specifically expressed in muscle have been selected and characterized in more detail: Their pattern of expression was confirmed by Northern blot analysis; their structure was further characterized by full-insert cDNA sequencing and cDNA extension; the map location of the corresponding genes was refined by radiation hybrid mapping. Five of the 14 selected genes appear as interesting positional and functional candidate genes to study in relation with muscle physiology and/or specific orphan muscular pathologies. One example is discussed in more detail. The expression profiling data and the associated Genexpress Index2 entries for the 910 genes and the detailed characterization of the 14 selected transcripts are available from a dedicated Web server at http://idefix.upr420.vjf.cnrs.fr/IMAGE/Page_unique/welcome_muscles.html. The database has been organized to provide the users with a working space where they can find curated, annotated, integrated data for their genes of interest. Different navigation routes to exploit the resource are discussed.

[Tables A and B are available as supplementary information at www.genome.org and also at http://idefix.upr420.vjf.cnrs.fr/IMAGE/Page_unique/welcome_muscles.html.]

The transcript repertoire of human skeletal muscle has been characterized through sequencing of cDNA clones, resulting in a preliminary description of some 4000 distinct transcripts, representing most of the genes expressed at moderate or high level but only 20%–25% of the skeletal muscle transcriptome. Of those, <5% appear to be expressed preferentially or specifically in muscle based on their frequency of occurrence in different tissues in sequence databases or as measured by cDNA array hybridization (Auffray et al. 1995; Houlgatte et al. 1995; Lanfranchi et al. 1996; Piétu et al. 1996; Murano et al. 1997; Bortoluzzi et al. 1998).

More than 1000 genes corresponding to muscle transcripts were initially mapped to specific chromosomes using panels of human–rodent somatic cell hybrids (Auffray et al. 1995; Houlgatte et al. 1995; Murano et al. 1997) and to more precise chromosomal bands through radiation hybrid (RH) mapping (Gyapay et al. 1996; Schuler et al. 1996; Pallavicini et al. 1997; Deloukas et al. 1998; Bortoluzzi et al. 1998).

Many of the genes involved in inherited neuromuscular diseases have been identified through positional cloning and later confirmed by functional candidate approaches. For example, the gene responsible for a specific form of limb–girdle muscular dystrophy (LGMD) was mapped through linkage mapping to chromosome 15q15 (Fougerousse et al. 1994). Among the genes registered in the initial version of the Genexpress Index (Auffray et al. 1995; Houlgatte et al. 1995) and mapped to this region (Richard et al. 1994), one appeared to be expressed specifically in muscle,

Present addresses: [4]Rhône-Poulenc Rorer, 91000 Evry, France; [6]Novartis Pharmaceuticals Corporation, Gaithersburg, Maryland 20878 USA
[7]Corresponding author.
E-MAIL pietu@infobiogen.fr; FAX (33-1) 49583509.

9:1313–1320 ©1999 by Cold Spring Harbor Laboratory Press ISSN 1054-9803/99 $5.00; www.genome.org

encoding the calpain 3 subunit, and appeared therefore as a candidate gene for the disease (Chiannilkulchai et al. 1995). Subsequently, a specific form of this gene was demonstrated to be associated with LGMD2A (Richard et al. 1995).

This illustrates the value of integrating sequence, map, and expression information to facilitate the elucidation of the role of specific genes in human muscle physiology and the identification of the genes involved in >40 orphan muscular pathologies that have been associated with a specific chromosomal region of the human genome but for which no specific gene has been identified. To this end, we have developed the Genexpress IMAGE Knowledge Base of the human muscle transcriptome, which is based on the sequence and gene-mapping data registered in Genexpress Index2 (R. Mariage-Samson et al., in prep.), an upgraded and updated version of the Genexpress Index (Houlgatte et al. 1995), and on expression profiling data collected by cDNA array hybridization (Piétu et al. 1996), following a scheme developed for a prototype integrated resource for functional and computational genomics of the human brain transcriptome (Piétu et al. 1999).

Based on a preliminary documentation of the expression profiles of 910 human gene transcripts by semiquantitative hybridization of an array of 1091 cDNA clones from a muscle library with complex probes derived from various mRNA sources (Piétu et al. 1996), we selected a set of 14 transcripts preferentially or specifically expressed in muscle and confirmed their pattern of expression by Northern blot analysis. The 14 transcripts were further characterized by full-insert cDNA sequencing and cDNA extension towards the 5′ end, and the map location of the corresponding genes was refined by RH mapping.

The entire set of expression profiling data for the 910 genes represented in the DNA array with the associated Genexpress Index2 entries and the detailed characterization of the 14 selected transcripts are available from a dedicated web site (http://idefix.upr420.vjf.cnrs.fr/IMAGE/Page_unique/welcome_muscles.html). We discuss in more detail through one specific example the difficulties encountered and the solutions adopted in the data integration process and its value for further characterization of the genes involved in human muscle physiology and pathology.

RESULTS

A dedicated web site has been constructed at http://idefix.upr420.vjf.cnrs.fr/IMAGE/Page_unique/welcome_muscles.html to provide access to the Genexpress IMAGE Knowledge Base, which integrates annotated and curated sequence, map, and expression data. The content of this web site is presented in a schematic form in Figure 1 and further described below.

Hybridization of 1091 Clones on High-Density Filters

The results of the expression-profiling experiments form the basis of the expression module of the Genexpress IMAGE Knowledge Base of the human muscle transcriptome. These results are based on previous work in which we reported the characterization of expression profiles of human gene transcripts in muscle by analyzing hybridization signatures on high-density filters carrying 1091 selected cDNA clones from a skeletal muscle library (Piétu et al. 1996). Each filter was hybridized in duplicate with cDNA probes derived from fetal heart, heart, brain, liver, testis, placenta, uterus, thymus mRNA. The 21,820 hybridization intensity values of these first-pass hybridization experiments (based on 1091 clones × 10 probes in duplicate) are presented in Table A of the web site.

A total of 629 clones (42%) are associated with hybridization values higher than the 1.96 threshold (95% confidence to differ from the population of weak signals) (Piétu et al. 1996) and can be ascribed to the categories of moderate to high abundance, whereas the remainder of the clones have intensity values that cannot be distinguished with confidence from background. Twenty-two percent of the clones have intensity values >1.96 with the muscle complex probe in duplicate experiments. The presence of repetitive sequences was demonstrated not to interfere with the hybridization signal intensity (Piétu et al. 1996).

Link to the Genexpress Index2

All of the clones of the human skeletal muscle cDNA library used in this study have been clustered and integrated into the Genexpress Index2 (R. Mariage-Samson et al., in prep.), an upgraded version of the Genexpress Index (Houlgatte et al. 1995). They correspond to 910 clusters or GENX. Clicking on the GENX identifier in Table A leads to the display of the corresponding cluster (called CLNVIEW in the Genexpress Index2), together with details on the clones, sequences, contigs, structural and coding properties, full contig sequence alignments, and relevant links to the corresponding UniGene and The Institute for Genomic Research (TIGR) entries.

Hybridization of 14 Gene Transcripts Preferentially Expressed in Muscle

We identified gene transcripts preferentially expressed in muscle by comparison of the hybridization signal intensity obtained with a probe derived from muscle mRNA to that obtained with probes derived from eight other tissues (fetal heart, heart, brain, liver, testis, placenta, uterus, thymus). Based on comparison of hybridization signal intensities, we have selected for detailed characterization 14 clones corresponding to transcripts preferentially expressed in muscle. These

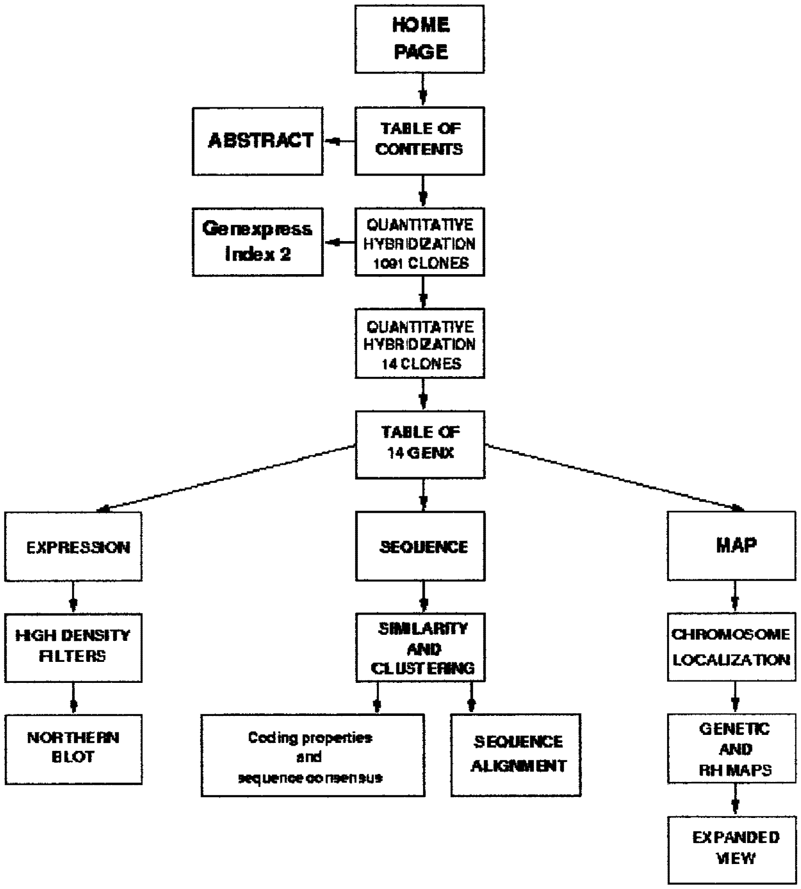


Figure 1 Schematic representation of the web-site content. The web site is available at http://idefix.upr420.vjf.cnrs.fr/IMAGE/Page_unique/welcome_muscles.html.

clones display hybridization values >3.29 (99% confidence to differ from the population of weak signals) for the muscle and/or the heart probes and <3.29 with all the other probes (Table 1; Table B on the web site).

Detailed Characterization of 14 Gene Transcripts Preferentially Expressed in Muscle

The three types of data, expression, sequence, and mapping, for each of the 14 GENX can be accessed on the web site through a table containing the appropriate links. Expression data are presented in 2 panels—one corresponds to the hybridization signal intensities (from Table B); the other presents the results of Northern blot analysis performed on a panel of RNA from eight human tissues, using as a probe a cDNA clone corresponding to each cluster. The muscle-restricted expression profile was confirmed for each of the 14 genes and allowed the determination of the size and number of transcripts.

Sequence data are the result of our cumulative cDNA clone and sequence clustering approach, registered in Genexpress Index2, and provide access to the structural features of the 14 selected muscle-specific transcripts. From each GENX cluster, 1–3 cDNA clones were selected as the most representative, through examination of the arrangement of clones with the CLNVIEW tool developed in Genexpress Index2, and were completely sequenced on both strands. Full-insert sequencing merged several previously disconnected contigs in 12/14 cases.

We then started to produce and sequence elongated cDNA copies of the transcripts. Of the 12 gene transcripts studied, 6 have been extended at their 5′ end on distances that vary from 0.8 to 1.9 kb. Additional sequence is represented in the CLNVIEW display of each GENX cluster.

The consensus sequences obtained have been updated in terms of sequence similarity with GenBank release 110.0, EMBL release 56.0, SWISS-PROT release 36.0, and SP-TREMBL release 8.0. Results are presented above the CLNVIEW display and classified according to sequence similarity to genomic DNA, mRNA, or protein.

Mapping data have been collected in silicio from the various maps available following a scheme described previously (Piétu et al. 1999) and completed through RH mapping experiments performed with the G3 (6 clusters) or the GB4 (4 clusters) panels.

The two most relevant data (selected mapping data) are displayed as a table below the schematic representation of the chromosome on the web site. The cytogenetic localization of the genes have been deduced from the cytogenetic data available for the genetic markers found in their vicinity. Orphan pathologies displaying muscular phenotype described so far in each region were retrieved from the GenAtlas data base, thus enabling possible correlation between selected genes and genetic disorder affecting muscle.

Complete mapping information has been schematized for each gene on a figure on the web site and is further accessible by clicking on the link to Genetic and RH maps. In this figure we did not attempt to propose a unique precise assignment for each gene but rather to provide a visual representation of the current state of knowledge that allows a researcher to keep track of the origin of the data used and to review it again if needed.

An Example of Integrated Results: The GENX-3587 Transcript

Integration of the three types of data, expression pro-

Table 1. Hybridization Signal Intensity of 14 Gene Transcripts Selected for Detailed Characterization

Genexpress	GenBank/EMBL accession no.		GENX		Fetal									
clone name	5′ end	3′ end	index 1	index 2	heart	Heart	Brain	Liver	Muscle 1	Muscle 2	Placenta	Thymus	Testis	Uterus
b-85g08	Z25032	Z28734	592	592	0.13	4.5	−0.42	−0.91	4.5	6.47	−0.73	−1.98	−0.07	−0.17
b-23c08	Z19393	F00425	702	702	0.22	3.93	1.73	0.82	4.32	5.3	2.8	−0.85	−1.9	−0.92
b-33h10	Z19491		6268	2543	0	3.16	−0.21	−0.72	4.33	4.91	1.33	−1.29	−1.25	−0.69
b-18e04	Z19354	Z28472	3446	3446	1.14	0.57	−0.69	−0.4	4.21	5.87	−1.39	0.13	−0.57	−1.19
b-28b09	Z19436	F00431	3471	3471	−0.11	1.37	1.3	−0.34	3.58	3.96	0.84	0.19	−0.02	−1.02
b-66b10	Z24882	Z28618	3587	3587	1.26	4.38	0.6	−1.2	4.48	4.74	2.38	0.69	−1.6	−1.44
b-21h02	F00106	Z19383	3733	3733	−0.75	7.02	−0.59	0.42	7.12	11.11	−0.22	−2	0.52	−1.66
b-94d06	Z28781	Z25106	4705	4705	0.07	0.3	−2.1	−0.28	4.62	6.95	−1.68	−0.85	−2.31	−0.45
b-84b12	Z25020	F00426	6163	6163	−0.43	1.08	−0.66	−1.19	4.44	7.11	1.6	−1.54	−1.41	0.03
b-20d07	Z19369	F00429	6166	6166	−1.04	5.14	1.37	1.32	4.76	6.18	1.31	−1.27	0.87	0.52
b-17a03		Z19342	6206	6206	−0.25	0.17	−0.75	−0.84	4.5	8.13	−0.92	−0.41	0.55	−1.17
b-36b02	Z24799		6278	6278	−0.5	4.94	2.09	−0.42	5.65	7.36	0.13	1.74	0.14	0.48
b-18a04	Z28469	Z28468	4549	115511	0.69	6.22	−1.65	−0.3	6.37	8.82	−0.78	−1.89	−1.15	−1.32
b-b1c12	Z25269		3432	115621	−0.37	1.7	−0.54	0.25	4.66	9.25	−2.34	−1.15	−0.51	−1.01

Clones were selected based on hybridization intensity Ri > 3.29 for the muscle probe and Ri < 3.29 for the other probes, except heart and fetal heart probes. Hybridization intensity values used for the selection are shadowed.

files, sequence analysis, and mapping, is illustrated in Figure 2 for the case of the GENX-3587 transcript.

Northern blot analysis (Fig. 2A) confirms the preferential muscle expression of this transcript. A strong signal corresponding to a 4-kb mRNA was detected in muscle and to a lesser extent in heart, whereas faint signals were also visible in pancreas and placenta.

The results of our cumulative clone and sequence clustering, full-insert sequencing, and elongated cDNA approaches for GENX-3587 are displayed in Figure 2B. All of the sequences from the GENX-3587 cluster were present in the corresponding Unigene cluster Hs.10632 (Build#72), which contains another 62 sequences, and they were distributed in three TIGR clusters (HGI Release 3.3) (11 in THC197898 containing 3 more sequences, 7 in THC176652 containing 11 additional sequences, and one singleton). Full-insert sequencing of one clone (yb84b08, IMAGE clone 39953) was performed leading to merge the two contigs initially present in Genexpress Index2 (Fig. 2B).

Starting from the 5′ region of the consensus sequence (represented by GenBank accession no. Z42230) and using various RACE (5′ rapid amplication of cDNA elongation) techniques and DNA library screening by PCR reactions, the cluster was elongated by about 1 kb leading to a 2852-bp consensus sequence. A 270-amino-acid reading frame was detected from nucleotide 1 to 621 of the consensus.

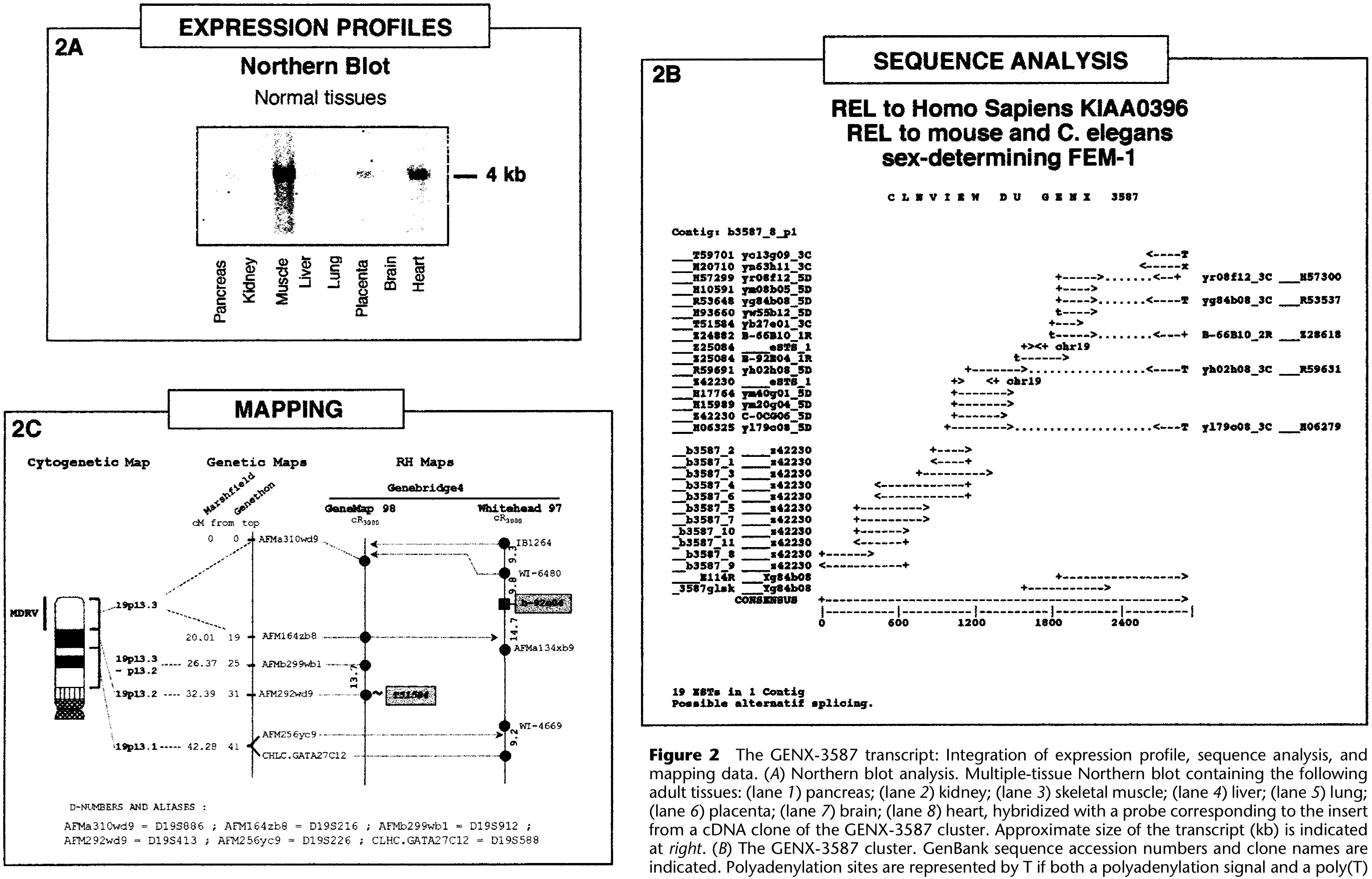
Sequence similarity search in databases revealed that the GENX-3587 consensus sequence is related to a protein encoded by the human KIAA0396 sequence (TREMBL accession no. O43146). It is also related to the mouse (Ventura-Holman et al. 1998) and *Caenorhabditis elegans* sex-determining protein FEM-1 (Spence et al. 1990). Furthermore, it also contains a human erythrocyte ankyrin motif (P16157, Lambert et al. 1990; Lux et al. 1990) and is identical to a human genomic sequence on chromosome 19. Nevertheless, the precise function of the GENX-3587 gene remains to be elucidated.

Mapping data, which were entirely absent in the GeneMap'98, were all produced in this study either through PCR-typing using the GB4 panel (marker b-92e04) followed by integration in the Whitehead framework using the Whitehead RH server or through score submission (marker T51584, RH80896) at the Sanger RH server to establish a link with a GeneMap'98 framework marker. Unfortunately no G3-based mapping data was available for any marker associated with this gene and our typing assays on the G3 panel were made unsuccessful as no linkage was found. The two sets of GB4-based data are schematized in Figure 2C. Integration with the genetic map was immediate for the GeneMap'98 data but rather difficult with the Whitehead data. From the two genetic markers mapped on this framework in the vicinity of the b-92e04 marker, one (AFMa134xb9) had no coordinate in any genetic map, and the other (CHLC-GATA27C12) was mapped in the Marshfield genetic map but had no cytogenetic assignment in GenAtlas. An additional genetic marker possessing these two features was therefore searched for in the intervals of the Whitehead map and was finally found at tier 2 (lod score < 1) as the AFM256yc9 (D19S226) marker (see Fig. 1C). These difficulties encountered in integrating the data lead to a considerable enlargement of the cytogenetic interval (19p13.1–p13.3) associated with the GENX-3587 gene. However, and although this should be taken with great caution, calculations involving the conversion of the genetic cM and Whitehead cR_{3000} scales into Mb strongly suggest that the interval containg the b-92e04 marker is entirely included in the 19p13.3 cytogenetic band. Interestingly enough, the 19p13.3 band corresponds to the cytogenetic localization of an orphan genetic disorder with a muscle phenotype: a muscular dystrophy (MDRV, OMIM 601846) described in GenAtlas as autosomal dominant with rimmed vacuoles and typical inclusion bodies. Further studies are required to determine whether or not the GENX-3587 gene could constitue a bona fide candidate gene for this pathology. It remains that the conjunction of localization at 19p13.3 and muscle-specific expression gives weight to this assumption. Availability of additional data concerning the encoded protein could also help in the future to formulate hypotheses concerning the function of this gene in muscle physiology and/or pathology. In summary, this example demonstrates how the integrated data available in the muscle module of the Genexpress IMAGE Knowledge Base could be used to identify novel candidate genes for orphan genetic disorders affecting human muscles.

DISCUSSION

The approach described here, based on the collection and integration of sequence, mapping, and expression annotated data, constitutes a further development of our IMAGE Knowledge Base of human transcriptomes. Entry into and navigation through the Genexpress IMAGE Knowledge Base of the muscle transcriptome can be envisioned in a variety of different ways.

Five genes illustrate a possible navigation route taking advantage of sequence information to identify structural candidate genes based on the relatedness of the sequence of the gene and its products (transcripts and proteins) to structures of known function. A BLAST analysis using as a query the sex-determination protein FEM-1 would yield as hits partial sequence data corresponding to the GENX-3587 transcript. The Genexpress IMAGE Knowledge Base would then provide access to complete sequence information, together with expression and mapping data. The same is true for four other GENX transcripts characterized in this



```
C L E V I E W   D U   G E N X   3587

Contig: b3587_8_p1

___T59701 yc13g09_3C                                                <----T
___H20710 yn63h11_3C                                               <-----x
___H57299 yr08f12_5D                                  +----->.......<--+   yr08f12_3C ___H57300
___H10591 ym08b05_5D                                  +---->
___R53648 yg84b08_5D                                  +---->.......<----T  yg84b08_3C ___R53537
___H93660 yw55b12_5D                                  t---->
___T51584 yb27e01_3C                                 +--->
___Z24882 B-66B10_1R                                 t----->.......<---+  B-66B10_2R ___Z28618
___Z25084 ____eSTS_1                             +><+ chr19
___Z25084 B-92E04_1R                            t------>
___R59691 yh02h08_5D                      +------->.................<----T  yh02h08_3C ___R59631
___Z42230 ____eSTS_1                    +>   <+ chr19
___H17764 ym40g01_5D                    +------->
___H15989 ym20g04_5D                    +------->
___Z42230 C-0CG06_5D                    +------>
___H06325 yl79c08_5D                   +-------->...................<---T  yl79c08_3C ___H06279

__b3587_2 ____z42230                  +---->
__b3587_1 ____z42230                  <----+
__b3587_3 ____z42230                +--------->
__b3587_4 ____z42230          <------------+
__b3587_6 ____z42230          <------------+
__b3587_5 ____z42230       +---------->
__b3587_7 ____z42230       +---------->
_b3587_10 ____z42230       +------->
_b3587_11 ____z42230       <-------+
__b3587_8 ____z42230  +------->
__b3587_9 ____z42230  <-----------+
_____E114R ___Yg84b08                                   +----------------->
_3587glsk ___Yg84b08                               +----------->
          CONSENSUS   +-------------------------------------------------->
                      |----------|----------|----------|---------|---------|
                      0         600       1200       1800      2400

19 ESTs in 1 Contig
Possible alternatif splicing.
```

D-NUMBERS AND ALIASES :

AFMa310wd9 = D19S886 ; AFM164zb8 = D19S216 ; AFMb299wb1 = D19S912 ;
AFM292wd9 = D19S413 ; AFM256yc9 = D19S226 ; CLHC.GATA27C12 = D19S588

Figure 2 The GENX-3587 transcript: Integration of expression profile, sequence analysis, and mapping data. (*A*) Northern blot analysis. Multiple-tissue Northern blot containing the following adult tissues: (lane *1*) pancreas; (lane *2*) kidney; (lane *3*) skeletal muscle; (lane *4*) liver; (lane *5*) lung; (lane *6*) placenta; (lane *7*) brain; (lane *8*) heart, hybridized with a probe corresponding to the insert from a cDNA clone of the GENX-3587 cluster. Approximate size of the transcript (kb) is indicated at *right*. (*B*) The GENX-3587 cluster. GenBank sequence accession numbers and clone names are indicated. Polyadenylation sites are represented by T if both a polyadenylation signal and a poly(T) tail are present at the 5′ end of the sequence and by x if only the polyadenylation signal is present at the 5′ end of the sequence. (*C*) Localization of the GENX-3587 gene in the integrated genome maps. The cytogenetic map presented is based on GenAtlas data. The genetic map is displayed with a scale in cM from the top of the chromosome according to the Généthon and Marshfield maps (data from Genome Database). The distances between markers of the RH maps are indicated in cR_{3000}. The marker names are shadowed for results obtained in this study and italicized for results obtained in silicio from a RH server (T51854). The microsatellite and other framework markers are represented as circles, whereas the GENX-3587 marker b-92e04 is shown as a square and was assigned on the Whitehead map with a significant (>3) lod score. Results obtained from the RH server are presented as a link (~) to a framework marker in GeneMap'98.

study. The GENX-4705 transcript encodes a protein strongly related to the rat mitochondrial and liver cytosolic very-long-chain acyl-CoA thioesterase (Lindquist et al. 1998; Svensson et al. 1998) and to the rat acyl-CoA hydrolase (Yamada et al. 1998). Furthermore, it is identical to *Homo sapiens* clone zap128 mRNA, which encodes a protein of unknown function. The GENX-6206 transcript appears to be related to the mouse mRNA for a kinesin-like protein (Nomura et al. 1994), with a probable role in transport of mitochondria along microtubules. A region of the GENX-3446 transcript appears to be distantly related to a human transcript encoding a putative transcription factor XPRF (Quaderi et al. 1997; Van den Veyver et al. 1998). The GENX-6163 gene product is related to *Schizosaccharomyces pombe* phosphatidyl synthase.

Another possible navigation route takes advantage of map information: Where is the gene precisely located, and is there a human pathology associated with the corresponding genomic region? Starting from the cytogenetic localization associated with a human pathology, information concerning human gene transcripts mapped in that region could be retrieved from the Genexpress IMAGE Knowledge Base. Not only the precise mapping with physical linkage to the closest genetic markers, but also complete sequence information and expression data are thus immediately available.

In this study, five orphan pathologies can be considered as possible entry point in search of positional candidate genes: an autosomal dominant muscular dystrophy (MDRV, at 19p13.3, GENX-3587), a hypoplasia of a facial muscle (ACF, 22q11, GENX-6163), the Charcot-Marie Tooth neuropathy type 2A (CMT2A, at 1p36, GENX-6206) in which muscle weakness and amyotrophy have been observed with normal nerve conduction velocity, as well as two heart defects observed either alone (ARVD1, 14q23–q24, GENX-4705) or in combination with other features in the complex DiGeorge syndrome (DGCR, 22q11.2, GENX-6163). In the case of the ventricular septal defect (AVD, GENX-115261), two cytogenetic positions are indicated as the genetic disorder involved is a translocation between the two loci.

The muscle module of our IMAGE Knowledge Base with the 14 examples documented on the web site now provides an updated and integrated vision of current biological knowledge on a set of transcripts preferentially or specifically expressed in muscle and a first step toward a representation of the entire muscle transcriptome. This will require inclusion of the great deal of biological knowledge already registered in the literature and the collection of missing informations by a variety of existing and emerging techniques, with the active participation of the community of biologists who are generating and using the IMAGE Consortium resources.

METHODS

Expression Profiling by Semiquantitative cDNA Array Hybridization

The array of 1091 human muscle cDNA clones on high-density filters, the capture of hybridization signals, and the identification and quantitation of the spots were as described (Piétu et al. 1996).

Northern Blot Analyses

Northern blots containing 2 µg of poly(A)$^+$ mRNA from eight adult tissues were purchased from Clontech (MTN blots). Probe preparation and hybridization of the membranes were performed as previously described (Piétu et al. 1999). Actin and ubiquitin cDNAs were used as probes to check the presence of similar levels of RNA in each lane.

Clustering of cDNA Clones, Sequences, and Genic Markers

The 1091 clones of the human muscle cDNA array correspond to 910 clusters of clones, sequences, and eSTS markers assembled in the Genexpress Index, or Index1 (Houlgatte et al. 1995). To extend the annotation of these clusters, referred as GENX clusters, we relied on a second generation, updated and upgraded version of Index1, called Index2, which contains 63,000 GENX clusters (R. Mariage-Samson et al., in prep.). Information provided by Index2 was presented in detail in Piétu et al. (1999).

Production of Elongated cDNA

For each gene, three antisense oligonucleotides (GSP for gene-specific primer) were designed at the 5′-most region of the previously determined sequence: one for gene-specific reverse transcription and two for nested PCR amplification. Human skeletal muscle mRNA (Clontech) was reverse transcribed using the Superscript II enzyme (Life Technologies) in the presence of either oligodT or GSP. The Marathon procedure (Clontech) was then used to produce a pool of cDNAs that can serve as substrate for PCR amplification of the 5′ regions of these cDNAs. Alternatively, the 5′ ends of cDNA clones were specifically amplified from a pooled total cDNA library (human skeletal muscle 5′-Stretch Plus, nonoriented cDNA library, or Matchmaker-oriented cDNA library, Clontech) using GSP as an antisense primer and, as sense primer, an oligonucleotide designed in the vector upstream the cloning site at the 5′ end of the cDNA insert. The PCR products were then cloned into the pCR2.1-TOPO vector according to the TOPO-TA cloning kit (Invitrogen) and the insert size was then checked by PCR with the M13 forward and reverse primers. At least three clones were prepared according to the Wizard DNA mini-preparation (Promega) and sequenced on both strands (Génome Express, Paris, France). Sequence alignment between three overlapping clones was used to eliminate mismatches generated by PCR misincorporation or sequencing errors. The Genetics Computer Group (GCG) package of programs was used for assembling and aligning partial cDNA sequences and for generating the consensus sequence.

Integrated Gene Mapping

Available mapping data concerning a given GENX cluster were retrieved from the appropriate web sites according to the procedure described previously (Piétu et al. 1999). De novo mapping was performed with G3 and Genebridge4 RH panels as described (Gyapay et al. 1996; Stewart et al. 1997).

In some cases RH scores found for a given marker in RHdb were submitted to the RH server available at Stanford for mapping with the G3 panel and at the Sanger Centre for mapping with the GB4 panel. Results are for the two-point RH analysis.

The conversion of the various scales to a common kb scale was performed as previously described (Piétu et al. 1999).

The web site was implemented as a dedicated server at CNRS (http://idefix.upr420.vjf.cnrs.fr/IMAGE/Page_unique/welcome_muscles.html) by Brainstorm, Paris.

ACKNOWLEDGMENTS

This work was supported by the CNRS and grants from the European Union to C.A. (GENE-CT-93-0089) and from the BIOMED2 programs (EURO-IMAGE Consortium, BMH4-CT-97-2284) to C.A. and W.A. and from Association Française contre les Myopathies (AFM) to C.A. E.E. and C.M. were supported by a grant from Rhône-Poulenc Rorer.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Auffray, C., G. Béhar, F. Bois, C. Bouchier, C. Da Silva, M.D. Devignes, S. Duprat, R. Houlgatte, M.N. Jumeau, B. Lamy et al. 1995. IMAGE: Integrated molecular analysis of the human genome and its expression. *C. R. Acad. Sci.* **318:** 263–272.

Bortoluzzi, S., L. Rampoldi, B. Simionati, R. Zimbello, A. Barbon, F. d'Alessi, N. Tiso, A. Pallavicini, S. Toppo, N. Cannata et al. 1998. A comprehensive, high-resolution genomic transcript map of human skeletal muscle. *Genome Res.* **8:** 817–825.

Chiannilkulchai, N., P. Pasturaud, I. Richard, C. Auffray, and J.S. Beckmann. 1995. A primary expression map of the chromosome 15q15 region containing the recessive form of limb-girdle muscular dystrophy (LGMD2A) gene. *Hum. Mol. Genet.* **4:** 717–725.

Deloukas, P., G.D. Schuler, G. Gyapay, E.A. Stewart, C. Soderlund, P. Rodriguez-Tome, L. Hui, T.C. Matise, J.S. Beckman, S. Bentolila et al. 1998. A physical map of 30,000 human genes. *Science* **282:** 744–746.

Fougerousse, F., O. Broux, I. Richard, V. Allamand, A.P. de Souza, N. Bourg, L. Brenguier, C. Devaud, P. Pasturaud, C. Roudaut et al. 1994. Mapping of a chromosome 15 region involved in limb-girdle muscular dystrophy. *Hum. Mol. Genet.* **2:** 285–293.

Gyapay, G., K. Schmitt, C. Fizames, H. Jones, N. Vega-Czarny, D. Spillett, D. Muselet, J.F. Prud'Homme, C. Dib, C. Auffray et al. 1996. A radiation hybrid map of the human genome. *Hum. Mol. Genet.* **5:** 339–346.

Houlgatte, R., R. Mariage-Samson, S. Duprat, A. Tessier, S. Bentolila, B. Lamy, and C. Auffray. 1995. The Genexpress Index: A resource for gene discovery and genic map of the human genome. *Genome Res.* **5:** 272–304.

Lambert, S., H. Yu, J.T. Prchal, J. Lawler, P. Ruff, D. Speicher, M.C. Cheung, Y.W. Kan, and J. Palek. 1990. cDNA sequence for human erythrocyte ankyrin. *Proc. Natl. Acad. Sci.* **87:** 1730–1734.

Lanfranchi, G., T. Murano, F. Caldara, A. Pacchioni, D. Pandolfo, S. Toppo, S. Trevisan, S. Scaro, and G. Valle. 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.

Lindquist, P.J., L.T. Svensson, and S.E. Alexson. 1998. Molecular cloning of the peroxisome proliferator-induced 46-kDa cytosolic acyl-CoA thioesterase from mouse and rat liver-recombinant expression in Escherichia coli, tissue expression and nutritional regulation. *Eur. J. Biochem.* **215:** 631–640.

Lux, S.E., K.M. John, and V. Bennett. 1990. Analysis of cDNA for human erythrocyte ankyrin indicates a repeated structure with homology to tissue-differentiation and cell-cycle control proteins. *Nature* **344:** 36–42.

Murano, T., D. Stephan, A. Pallavicini, N. Tiso, R. Zimbello, G.A. Danieli, E.H. Hoffman, G. Valle, and G. Lanfranchi. 1997. Chromosomal assignment of 115 expressed sequence tags (ESTs) from human skeletal muscle. *Cytogenet. Cell Genet.* **76:** 144–152.

Nomura, N., T. Nagase, N. Miyajima, T. Sazuka, A. Tanaka, S. Sato, N. Seki, Y. Kawarabayasi, I. K., and S. Tabata. 1994. Prediction of the coding sequences of unidentified human genes. II. The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by analysis of cDNA clones from human cell line KG-1. *DNA Res.* (Suppl.) **1:** 251–262.

Pallavicini, A., R. Zimbello, N. Tiso, T. Murano, L. Rampoldi, S. Bortoluzzi, G. Valle, G. Lanfranchi, and G.A. Danieli. 1997. The preliminary transcript map of a human skeletal muscle. *Hum. Mol. Genet.* **6:** 1445–1450.

Piétu, G., O. Alibert, V. Guichard, B. Lamy, F. Bois, E. Leroy, R. Mariage-Samson, R. Houlgatte, P. Soularue, and C. Auffray. 1996. Novel gene transcripts preferentially expressed in human muscles revealed by quantitative hybridization of a high density cDNA array. *Genome Res.* **6:** 492–503.

Piétu, G., R. Mariage-Samson, N.A. Fayein, C. Matingou, E. Eveno, R. Houlgatte, C. Decraene, Y. Vandenbrouck, F. Tahi, M.D. Devignes et al. 1999. The Genexpress IMAGE knowledge base of the human brain transcriptome: A prototype integrated resource for functional and computational genomics. *Genome Res.* **9:** 195–209.

Quaderi, N.A., S. Schweiger, K. Gaudenz, B. Franco, E.I. Rugarli, W. Berger, G.J. Feldman, M. Volta, G. Andolfi, S. Gilgenkrantz et al. 1997. Opitz G/BBB syndrome, a defect of midline development, is due to mutations in a new RING finger gene on Xp22. *Nat. Genet.* **17:** 285–291.

Richard, I., O. Broux, N. Chiannilkulchai, F. Fougerousse, V. Allamand, N. Bourg, L. Brengier, C. Devaud, P. Pasturaud, C. Roudaut et al. 1994. Regional localization of human chromosome 15 loci. *Genomics* **3:** 619–627.

Richard, I., O. Broux, V. Allamand, F. Fougerousse, N. Chiannilkulchai, N. Bourg, L. Brengier, C. Devaud, P. Pasturaud, C. Roudaut et al. 1995. Mutations in the proteolytic enzyme calpain 3 cause limb-girdle muscular dystrophy type 2A. *Cell* **81:** 27–40.

Schuler, G.D., M.S. Boguski, E.A. Stewart, L.D. Stein, G. Gyapay, K. Rice, R.E. White, P. Rodriguez-Tomé, A. Aggarwal, E. Bajorek et al. 1996. A gene map of the human genome. *Science* **274:** 540–546.

Spence, A.M., A. Coulson, and J. Hodgkin. 1990. The product of fem-1, a nematode sex-determining gene, contains a motif found in cell cycle control proteins and receptors for cell-cell interactions. *Cell* **60:** 981–990.

Stewart, E.A., K.B. McKusick, A. Aggarwal, E. Bajorek, S. Brady, A. Chu, N. Fang, D. Hadley, M. Harris, S. Hussain et al. 1997. An STS-based radiation hybrid map of the human genome. *Genome Res.* **7:** 422–433.

Svensson, L.T., S.T. Engberg, T. Aoyama, N. Usuda, S.E. Alexson, and T. Hashimoto. 1998. Molecular cloning and characterization of mitochondrial peroxisome proliferator-induced acyl-CoA thioesterase from rat liver. *Biochem. J.* **329:** 601–608.

Van den Veyver, I.B., T.A. Cormier, V. Jurecic, A. Baldini, and H.Y. Zoghbi. 1998. Characterization and physical mapping in human and mouse of a novel RING finger gene in Xp22. *Genomics* **51:** 251–261.

Ventura-Holman, T., M.F. Seldin, W. Li, and J.F. Maher. 1998. The murine fem1 gene family: Homologs of the Caenorhabditis elegans sex-determination protein FEM-1. *Genomics* **54:** 221–230.

Yamada, J., K. Suga, T. Furihara, M. Kitahara, T. Watanabe, M. Hosokawa, T. Satoh, and T. Suga. 1998. cDNA cloning and genomic organization of peroxisome proliterator-inducible long chain acyl-CoA hydroxylase from liver cytosol. *Biochem. Biophys. Res. Commun.* **248:** 608–612.

Received April 14, 1999; accepted in revised form October 4, 1999.