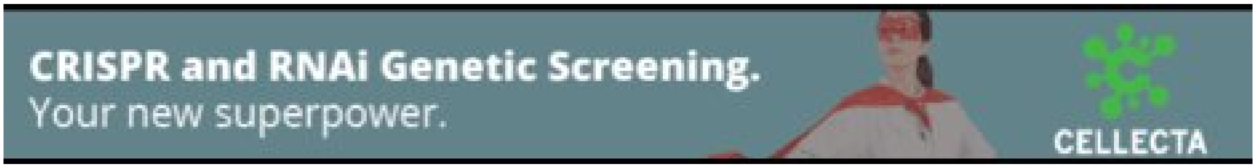
The Genexpress Index: a resource for gene discovery and the genic map of the human genome.

R Houlgatte, R Mariage-Samson, S Duprat, et al.

Genome Res. 1995 5: 272-304
Access the most recent version at doi:10.1101/gr.5.3.272



The Genexpress Index: A Resource for Gene Discovery and the Genic Map of the Human Genome

Rémi Houlgatte,[1,2,3,] Régine Mariage-Samson,[1,2,3] Simone Duprat,[2] Anne Tessier,[2] Simone Bentolila,[1,2] Bernard Lamy,[2] and Charles Auffray[1,2,4]

[1]Genexpress, Centre National de la Recherche Scientifique (CNRS) UPR420, 94801 Villejuif CEDEX, France; [2]Genexpress, Généthon, 91002 Evry CEDEX, France

Detailed analysis of a set of 18,698 sequences derived from both ends of 10,979 human skeletal muscle and brain cDNA clones defined 6676 functional families, characterized by their sequence signatures over 5750 distinct human gene transcripts. About half of these genes have been assigned to specific chromosomes utilizing 2733 eSTS markers, the polymerase chain reaction, and DNA from human–rodent somatic cell hybrids. Sequence and clone clustering and a functional classification together with comprehensive data base searches and annotations made it possible to develop extensive sequence and map cross-indexes, define electronic expression profiles, identify a new set of overlapping genes, and provide numerous new candidate genes for human pathologies.

During the last 20 years, since the first descriptions of eucaryotic cDNA cloning (Rougeon et al. 1975; Efstratiadis et al. 1976), cDNA studies have played a central role in molecular genetics. Early attempts to systematically analyze gene transcript repertoires were limited by available technologies (Milner and Sutcliffe 1983; Putney et al. 1983; Palazzolo et al. 1987; Sutcliffe 1988; Kato 1990), a situation that changed in the mid eighties with the development of fluorescent-based DNA sequencing (Smith et al. 1985, 1986). The initial proposal (Brenner 1990) and demonstrations of the potential of systematic sequencing and mapping of cDNA clones (Adams et al. 1991; Höög 1991; Hyde et al. 1991; Okubo et al. 1991; Wilcox et al. 1991) were soon followed by numerous medium- to large-scale cDNA sequencing studies in a variety of tissues and species in both animals and plants (Adams et al. 1992, 1993a,b; Gieser and Swaroop 1992; Khan et al. 1992; McCombie et al. 1992; Okubo et al. 1992, 1994; Uchimiya et al. 1992, 1994; Waterston et al. 1992; Höfte et al. 1993; Matsubara and Okubo 1993; Park et al. 1993; Takeda et al. 1993; Affara et al. 1994; Davies et al. 1994; Kerr et al. 1994; Konishi et al. 1994; Kurata et al. 1994; Liew et al. 1994; Murakawa et al. 1994; Newman et al. 1994; Nishiguchi et al. 1994; Nomura et al. 1994; Sasaki et al. 1994; Soares et al. 1994; Sudo et al. 1994; Auffray et al. 1995; Berry et al. 1995; Franco et al. 1995; Frigerio et al. 1995; Pawlak et al. 1995). This developing field has been recognized as an important component of the Human Genome Project for gene discovery (Collins and Galas 1993) and has been the subject of several reviews (Kato 1992; Southern 1992; Grausz and Auffray 1993; Matsubara and Okubo 1993; Sikela and Auffray 1993). Mapping of the corresponding human genes has not developed as rapidly, although it is a limiting step in the identification of disease genes (for review, see Hochgeschwender 1992; Parrish and Nelson 1993; Chen et al. 1994; Collins 1995). Thus, until recently, in addition to the 3700 genes registered in the Genome Data Base, a relatively limited number of gene transcripts characterized by cDNA sequencing have been assigned to a specific chromosome (Wilcox et al. 1991; Durkin et al. 1992; Gieser and Swaroop 1992; Khan et al. 1992; Polymeropoulos et al. 1992, 1993; Fukushima et al. 1994; Murakawa et al. 1994; Berry et al. 1995).

---

[3]The first two authors contributed equally to this work.
[4]Corresponing author.
E-MAIL remi@pauline.vjf.inserm.fr; FAX (33)(1)49 58 35 09.
The complete data contained in Appendices 1 and 2 can be found in electronic form at http://www.cshl.org/.

5:272–304 ©1995 by Cold Spring Harbor Laboratory Press ISSN 1054-9803/95 $5.00

We have recently reported the basic methodology and primary results of our Integrated Molecular Analysis of the human Genome and its Expression (IMAGE), a strategy to collect sequence signatures from collections of cDNA clones, quantitative hybridization signatures on high-density filters, and chromosomal assignment of the corresponding genes using human–rodent somatic cell hybrids (Auffray et al. 1995). We collected 26,938 sequence signatures from both ends of cDNA clones from a human skeletal muscle and a normalized infant brain (Soares et al. 1994) libraries and developed 2792 eSTS markers (Auffray et al. 1995) that were assigned to specific chromosomes. In this paper we report a detailed description of the various steps of secondary analyses of two-thirds of these sequences, and most of the eSTS markers to index >5750 distinct human gene transcripts by sequence clustering and data base comparison. We define electronic expression profiles of the transcripts in the various tissues in which they have been observed and the chromosomal assignments of about half of the corresponding genes. We describe sequencing and mapping cross-indexes to facilitate the use of this resource for gene discovery, the construction of the genic map of the human genome, and the identification of disease genes.

SEQUENCE CLUSTERING

Each sequence was compared with all the others obtained from the same library using the FASTA (Pearson 1990) program. When a sequence had >90% similarity over its entire length with a previously registered sequence, it was considered redundant and its analysis was stopped. Other sequences were considered unique and were searched for more limited overlaps by a second round of comparisons. Both processes involved automatic prefiltration and interactive validation steps, which provided links between sequences. Contigs were defined as a set of sequences grouped together by common links rather than multialignments. This made it possible to group sequences that cannot be aligned such as alternatively spliced sequences or sequences with low-quality segments at their ends, when a third sequence was overlapping with both of them (Fig. 1). This was also useful in the case of the brain library, as some cDNA clones appeared to be derived from unspliced mRNA.

A further advantage of this method is that erroneous links based on repeats, composition bias, or highly similar isogenes can be removed when detected at any stage, by inactivating it in the data base without any new calculation. Moreover, as in most cases we have obtained sequences from both ends of the cDNA clones, it was possible to use clone links to group contigs

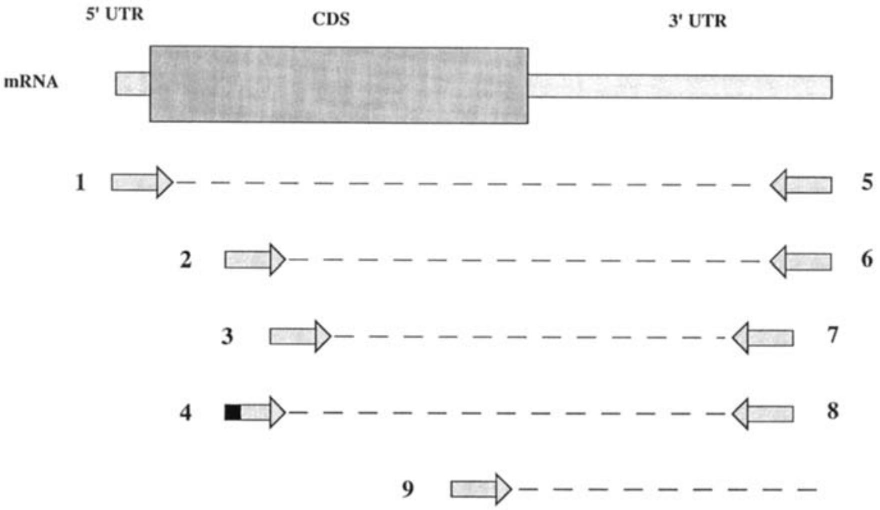


Figure 1 Sequence clustering strategy. Partial sequences (1–9) are represented by arrows, below the mRNA from which they were derived. Broken lines indicate individual clones. Sequence analysis allowed us to find redundant sequences (>90% similar over their entire lengths) such as sequence 6 (identical to sequence 5) or 8 (identical to sequence 7). These redundant sequences were not subject to further analysis. All nonredundant sequences were compared with each other to find overlapping sequences (sequences detected by FASTA, with a Opt parameter >120, >90% identities, and validated by users) such as sequences 5 and 7, 2 and 3, 3 and 4. This allowed us to cluster sequences into contigs defined as sets of sequences linked either by redundancy or overlaps (sequences 1, 2–4, 5–8, 9) and to cluster together some sequences that could not be aligned (sequences 2 and 4) because of a low-quality segment or alternative splicing (sequence 4 ■), if a third sequence (sequence 3) overlapped with them. Sequences derived from the same clone were further clustered into Families (sequences 1–8, 9). After data base comparisons, families defined as Identical or Homolog to the same gene transcript were clustered together (sequences 1–9).

together into families. When applied to the entire set, the 18,698 sequences derived from 10,979 cDNA clones were clustered into 12,713 contigs grouped in 6886 families.

DATA BASE COMPARISONS

All individual unique sequences were compared with the GenBank nucleic acid sequences (Benson et al. 1994), except the partial cDNA subsection and SWISS-PROT protein sequences (Bairoch and Boeckmann 1994), using the BLAST (Altschul et al. 1990; Gish and States 1993) family of programs to define their level of similarity with known sequences. To avoid spurious similarities resulting from composition bias or repeats, we used a masked version of SWISS-PROT using the XNU (Claverie and States 1993) program. For nucleic acid comparisons, we first compared the sequences with GenBank and a local repeat data base. Repeats detected with a high level of similarity were masked with the XBLAST (Claverie and States 1993) program, and all masked sequences were compared once again with GenBank, to detect similarities that could have been underscored because of the presence of a repeat. All sequences with no significant match ($P < 10^{-5}$) in any data base were classified as Unknown. To extend the initial matches detected by BLAST, all the other sequences were unmasked and subjected to LFASTA alignment with their nucleic best matches. This was necessary because insertions or deletions that are often present in partial cDNA sequences are not taken into account by BLAST that provides only local uninterrupted alignments, whereas FASTA integrates the presence of gaps. A direct FASTA search would have overcome this problem, but an initial BLAST search directly provides a measurement of the statistical significance of the matches (Altschul et al. 1994).

Sequences were classified as Known if identical to a human sequence, Homolog if highly similar to a vertebrate sequence, Related if having a partial similarity to a sequence of any species, Chimeric if having incoherent matches, or Unknown if no match appeared significant. Sequences from chimeric clones were not studied further. As correct classification depended on the available informations of the hit sequences, such as position of the match or existence of multiple isogenes, we chose to favor human interpretation of all matches with basic principles and no strict rules. For example, a complete but partial similarity with a nonhuman sequence can be more significant than a limited but identical match with a partial human sequence. This could be the case if the human gene has been partially sequenced or if this sequence has been split in different files, whereas the vertebrate homolog is only defined as an mRNA stored in a unique file. Therefore, if only the hit with the best score is taken into account, the identity with a known human sequence will not be detected. Similarly, the detection of 85% similarity with a vertebrate sequence could lead to classification of the sequence as homologous, but this is erroneous if the human homolog is already described and did not produce a more significant match than with the vertebrate sequence. In that case, the query sequence is probably derived from a related gene, such as an isogene. For all these reasons, a summary of all informations collected including a summary of all the data base matches, FASTA alignment with their nucleic best matches, text extraction of best-match files, six-frame translation, coding frame prediction, nucleic acid, or protein motif predictions was provided to the users for interpretation.

A functional assignment was obtained in most cases, but sometimes it was necessary to perform additional searches, such as multialignments of sequences, data base searches with a limited part of the query sequence, or text searches on data bases, in order to reach the final conclusion. The same functional assignment was derived for clones and families by gathering functional assignment of their constitutive sequences. As different sequences could have similarities with different files of the nucleic acid data base derived from a single gene, it was necessary to create cross-indexes between these files. These cross-indexes were mainly derived from SWISS-PROT cross-indexes, but many additionnal links were derived by comparison of each nucleic acid sequence with the entire GenBank. These cross-indexes link together different nucleic acid sequences derived from a unique gene, but also the encoded protein, as well as the Genome Data Base (GDB) (Fasman et al. 1994) locus or On-line Mendelian Inheritance in Man (OMIM) (Pearson et al. 1994) entries. This also allowed us to establish a special category of link between sequences derived from the same gene but in different organisms. The link of the nucleic acid sequences to SWISS-PROT files made it possible to access annotations of much better quality than those available in nucleic acid data bases which are much more limited and sometimes erroneous.

On the other hand, as there are nucleic acid sequences in GenBank that are not yet represented as proteins in SWISS-PROT, this strategy probably failed to detect some matches, because of differences in codon usage in different organisms. It will be possible to take them into account in the future using SWISS-PROT updates or by searching the GenPept nonredundant data base that connects translation of nucleic acid data bases together with protein data base entries.

Partial cDNA sequences were not searched for sequence similarity in the initial process but were searched for overlapping segments as in the clustering process. Tissue origin of the library was extracted to define electronic expression profiles.

TRANSCRIPT ASSIGNMENT

The cross-indexes allowed us to group different families classified as identical or homologous to the same gene that had not been clustered together by their sequence similarities (Fig. 1). The pool of 1276 such families was reduced by 210 (16.5%) during this process, leading to a total of 6676 functional families. This indicates that probably the number of families is an overestimation by ~16.5% of the number of gene transcripts characterized by partial sequencing. If this coefficient is applied to all families, we can estimate that the 6886 families are probably representative of 5750 different human gene transcripts. The situation appeared to be clearly different in the muscle compared with the brain library. In the muscle library, 24% of the known or homolog families were lost during the functional clustering step, whereas only 7% were lost in the brain library. This discrepancy is mainly attributable to the fact that, on average, 1.3 sequences are available for each clone in the muscle library (frequently, the 3′ end sequence was not obtained because of the difficulties of sequencing long poly(A) tails), whereas, on average, 1.8 sequences were available for each clone in the brain library that has relatively short poly(A) tracts by cloning design (Soares et al. 1994). This shows clearly that the gene transcript count can be largely overestimated when 3′ end sequences are not available, as is the case if only 5′ end sequencing is performed or if the cDNA library is random-primed. Failure to group together different families derived from the same gene transcript in the brain library was attributable to the absence of the 3′ end sequence (40% of the cases), alternative splicing (15%), alternative poly(A) site (33%), or internal priming of the mRNA (12%). Conversely, if 3′ sequence is always obtained, gene transcript overestimation should have occured only in 4.3% of the cases.

Of the 6676 functional families, it can be seen in Table 1 that 12.7% were already known in man, 3.2% are human homologs of a gene already known in another vertebrate species, 7.6% have a partial functional similarity with a known gene, and 76.4% displayed no significant similarity. Overall, 2515 have been characterized independently at least by partial sequencing. This also indicates that of the 5823 new gene transcripts described in this study, >4000 are described for the first time.

COMPARISON BETWEEN SKELETAL MUSCLE AND BRAIN LIBRARIES

Some differences between the muscle and brain libraries can be observed when comparing the most frequently represented gene transcripts (Table 2). In the muscle library, the 10 most redundant transcripts are derived from already known genes and represent ~20% of the clones of this library. The most abundant previously uncharacterized transcript in man appears to be derived from a gene mapped to chromosome 2 and related to the *Drosophila* Kelch protein (rank 13). In the brain library, the situation is completely different, as the most redundant transcript is the human homolog of rat neuronal olfactomedin-related protein localized in the endoplasmic reticulum and mapped to chromosome 9 and the nine most abundant transcripts are derived from four known, one homolog, and four unknown genes, which together represent only 1.4% of the library. These differences are in part attributable to variation in gene transcription in these tissues but mostly to the normalization of the brain library. This is also evident from the comparison between the functional classification of the different clones or gene transcripts in the two libraries (Table 3). In the muscle library, the most abundant transcripts encode cytoskeletal elements, metabolic proteins, or transport and storage proteins (functional redundancy of 6.31, 3.40, and 3.48, respectively). In the brain library, the most abundant transcripts encode proteins involved in signaling, but global functional redundancy is nearly identical for all the functional classes (ranging from 1.49 to 2.19), and, there-

Table 1. Summary of the Similarity Classification (*A*) and Chromosomal Assignment (*B*) of the Gene Transcripts and Assessment of Mapping Accuracy (*C*)

A	cDNA partial sequences		
Similiarity	registered	new	total
Identical	438	415	853
Homologous	104	112	216
Related	165	342	507
Unknown	1393	3707	5100
Total	2100	4576	6676

B	Gene transcript mapping			
Similarity	GDB	eSTS	not tested	total
Identical	481	101	271	853
Homologous	1	72	143	216
Related	3	180	324	507
Unknown	24	1919	3157	5100
Total	509	2272	3895	6676

	eSTS markers		
C	identical	partial	different
GDB	124 (80%)	10 (6%)	22 (14%)
Multiple	325 (86%)	20 (5%)	33 (9%)

(*A*) Comparison of individual sequences to nucleic acid (GenBank release 81.0 for muscle library, and release 84.0 for brain library) and protein (SWISS-PROT) release 28) data bases allowed to classify each gene transcript as Identical, Homolog, Related or Unknown. Further comparison with the partial cDNA sequence data base subsection allowed detection of entries overlapping with any sequence of a registered gene transcript. Each family having at least one entry overlapping one of its constitutive sequences was considered as Registered, or New if not.

(*B*) Each gene transcript having one eSTS derived from one of its partial sequences was considered as Assigned, or Known if identical to a human gene already mapped in GDB or GenBank.

(*C*) Some known genes (156) already mapped in GDB were also assigned in this study (GDB), and 584 redundant eSTS allowing 323 pairwise comparisons (Multiple) were examined to assess the external and internal accuracy of the method. The different eSTS markers were defined as Identical if they produced the same result, Partial if they had one chromosome assignment in common, or Different.

fore, clone redundancy is no more a reflection of gene transcription. Another consequence of the normalization is that even though a large number of sequences have already been obtained from brain libraries, additional sequencing of this normalized library will continue to yield a large proportion of new gene transcripts (as shown in Table 1A; discussed in Berry et al. 1995).

CHROMOSOMAL ASSIGNMENT

Expressed sequence tagged site (eSTS) markers were derived from unique sequences, mainly those obtained from the 3′ end of the transcripts and assigned to specific chromosomes using panels of human–rodent somatic cell hybrids. Of the 2792 markers initially developed, most provided an unambiguous assignment, whereas 114 (3%) detected several chromosomes (Auffray et al. 1995). To assess the coherence of the assignments, some eSTS markers were derived from known genes that had been localized previously, and in some cases, multiple eSTS markers were derived from different parts of the same transcript.

The sequence-clustering process described above allowed us to map at the chromosomal

Table 2. Most redundant transcripts found in the muscle or the brain libraries

A: Muscle library

Rank	GENX	% Muscle	% Brain	Seq	Tissues	eSTS	Map	Similarity	Species	Gene product
1	2965	5.10		1	FB		1q42.1-q42.3	Identical	Human	adult skeletal muscle alpha-actin
2	3195	3.38		1	FB	19	19q13.3	Identical	Human	creatine kinase M
3	2	2.20		3	HE	14	14q11.2-q13	Identical	Human	cardiac muscle beta-myosin heavy chain
3	3338	2.20		5	HE,HA	2	2q31-q32	Identical	Human	titin
5	3614	1.85				22	22q12	Identical	Human	myoglobin
6	3237	1.19					11q13.1	Identical	Human	muscle glycogen phosphorylase
7	2732	1.01		1	PI	19	19	Identical	Human	calpain small subunit
8	3430	0.92	0.05	19	BR,HI,FB,PI		12p13	Identical	Human	glyceraldehyde 3-phosphate dehydrogenase
9	3514	0.66	0.08	9	HI,FB,BR	12	12q23-q24.1	Identical	Human	endoplasmic reticulum calcium ATPase class 2
9	3560	0.66				2	2q35	Identical	Human	desmin
9	3889	0.66					19q13.4	Identical	Human	slow skeletal muscle troponin T
12	3377	0.62	0.01	6	HE,PI		6q25	Identical	Human	manganese superoxide dismutase
13	3470	0.57		6	CM,FB,HI,BR	3+10+16	16p22.1	Identical	Human	fructose 1,6 diphosphate aldolase A
13	4507	0.57				2		Related	Drosophilia	ring canal protein KELCH
15	3410	0.53	0.02			12	12	Identical	Human	skeletal muscle slow myosin-binding protein C
15	3628	0.53		15	HE,FLS,PI		13q14	Identical	Human	translationally controlled tumor protein
15	3727	0.53					12q23-q24.3	Identical	Human	cardiac/ventricular myosin light chain 2
15	352	0.53	0.01	2	BR,PI	1		Homolog	Rat	UNR protein with unknown function
19	3420	0.44					17pter-p12	Identical	Human	beta-enolase
20	3427	0.40		3	BR,PI,FB	16		Identical	Human	DNA-binding protein A
20	3122	0.40	0.02	1	BR	X+3	1q24-q32	Related	Human	cystein-rich protein

B: Brain library

Rank	GENX	% Muscle	% Brain	Seq	Tissues	eSTS	Map	Similarity	Species	Gene product
1	3111		0.29	15	BR,PI	9		Homolog	Rat	neuronal olfactomedin-related ER localized protein
2	3695		0.16	4	BR,HA		Xq21.33-q22	Identical	Human	myelin proteolipid protein
2	3013		0.16	2	FB,BR	5	5q21-q22	Identical	Human	adenomatous polyposis coli protein
2	4128		0.16	2	BR	8		Unknown		
5	3831		0.14	41	HL,BR,HI,FB,PI		2q	Identical	Human	tubulin alpha 1
6	1104		0.13	9	FB,BR		6p21.3	Identical	Human	tubulin beta-1
7	3739		0.11	9	HL,FB,BR			Unknown		
7	3978		0.11	2	BR	1+7		Unknown		
7	4027		0.11	2	CC			Unknown		
10	4156		0.10	3	HI,FB,BR		14q21-q22	Identical	Human	neuroendocrine-specific protein
10	2031		0.10	5	PI,BR		1q42-q43	Identical	Human	angiotensinogen
10	551		0.10	5	HI,FB,RE	5		Homolog	Rat	postsynaptic density protein 95
10	3789		0.10	3	HL,FB,BR	1		Unknown		
10	4127		0.10	2	BR,FB	7		Unknown		
15	4105		0.09	3	HL,FB,BR		1p36.1-p35	Identical	Human	stathmin
15	3060	0.04	0.09	12	HE,HL,FB,HI,PI,BR,FI	Y+3+4+6+14+22		Identical	Human	ribosomal protein L10
15	398		0.09			10		Unknown		
15	1503		0.09			7		Unknown		
15	2946		0.09			15		Unknown		
15	4054		0.09	1	HL	22		Unknown		
15	4255		0.09	2	BR	14		Unknown		

Data base extraction of the most redundant gene transcripts found in the muscle (*A*) or the brain (*B*) libraries. (GENX) Genexpress Index number. The complete description of the different columns is provided in Appendix I.

Table 3. Functional classification and global redundancy of all gene transcripts found in the muscle or brain libraries

A: Muscle Library

Clones	Cytoskeleton	Nucleic-Acid Managing	Protein Managing	Metabolism	Recognition / Adhesion	Signalling	Transport / Storage	Unclassified	*Total*
Identical	343	34	128	281	13	54	93	46	992
Homologous	11	14	18	25	1	13	9	26	117
Related	56	29	23	24	2	19	6	35	194
Unknown								978	978
Total	410	77	169	330	16	86	108	1085	2281

Transcripts	Cytoskeleton	Nucleic-Acid Managing	Protein Managing	Metabolism	Recognition / Adhesion	Signalling	Transport / Storage	Unclassified	*Total*
Identical	33	22	60	60	11	35	20	25	266
Homologous	5	8	13	21	1	9	5	12	74
Related	27	19	21	16	2	13	6	19	123
Unknown								828	828
Total	65	49	94	97	14	57	31	884	1291
Redundancy	6.31	1.57	1.80	3.40	1.14	1.51	3.48	1.23	1.77

B: Brain Library

Clones	Cytoskeleton	Nucleic-Acid Managing	Protein Managing	Metabolism	Recognition / Adhesion	Signalling	Transport / Storage	Unclassified	*Total*
Identical	100	153	166	180	112	289	86	237	1323
Homologous	17	22	35	33	6	76	38	99	326
Related	36	112	82	69	32	108	30	161	630
Unknown								6419	6419
Total	153	287	283	282	150	473	154	6916	8698

Transcripts	Cytoskeleton	Nucleic-Acid Managing	Protein Managing	Metabolism	Recognition / Adhesion	Signalling	Transport / Storage	Unclassified	*Total*
Identical	41	100	81	96	50	150	41	133	692
Homologous	8	16	16	22	4	36	23	38	163
Related	21	74	53	45	22	72	17	101	405
Unknown								4384	4384
Total	70	190	150	163	76	258	81	4656	5644
Redundancy	2.19	1.51	1.89	1.73	1.97	1.83	1.90	1.49	1.54

Each individual gene transcript found in this study was classified using an arborescent functional classification based on the information of their best matches extracted from data bases. Main functional classes presented here are Cytoskeleton, Nucleic-Acid Managing (proteins involved in structure, repairing, replication or transcription of nucleic acids), Protein Managing (proteins involved in translation, maturation, targeting or degradation of proteins), Metabolism, Recognition and Adhesion (protein involved in cell or tissue adhesion, or recognition such as MHC), Signaling (proteins involved in intra- or extracellular signaling such as hormones or growth factors and their receptors, or proteins involved in second messenger pathways), Transport and Storage (proteins involved in chelation of inorganic compounds, transport across cell membranes, or polypeptide binding proteins), or Unclassified (protein with no known function, or new gene transcripts with no significant similarity). Functional classes are presented in terms of number of clones (Clones) or gene transcripts (Transcripts), and global functional redundancy (Redundancy) is defined as the number of clones divided by the number of gene transcripts for each functional class. Note that the sum of the gene transcripts in the two libraries is not equal to the total number of gene transcripts presented in this study, as some of them are found in both libraries.

level at least 2272 new genes using 2733 eSTS markers (Table 1B). In most of the cases (86%) when the corresponding gene had already been mapped, these eSTS markers provided a localization identical to that registered in GenBank or GDB (Table 1C). A better correlation (91%) is observed for multiple assigments of the same transcript. The major causes of discrepancy appeared to be the presence of an intron in the amplified sequences leading to the mapping of a related gene or pseudogene. This is clearly the case for the human fructose biphosphate aldolase A gene

Table 4. Human genes related to hypothetical genes found by systematic genome sequencing of model organisms

GENX	% Muscle	% Brain	Seq	Tissues	eSTS	Species	Gene product
3282		0.07	4	BR,CC,FB		C. elegans	hypothetical 80.7 kd protein ZC84.3
1860		0.03	1	BR	17	C. elegans	hypothetical 52.7kd protein T23g5.2 in chromosome III
127		0.03			7	C. elegans	hypothetical 88.1 kd protein
2796		0.03	2	BR		C. elegans	hypothetical 25.9kd protein C05B5.7 in chromosome III
4136		0.03				C. elegans	hypothetical 29.0kd protein ZK632.12 in chromosome III
3022		0.03	2	BR		C. elegans	hypothetical 63.5kd protein ZK353.1 in chromosome III
1827		0.02			22	C. elegans	hypothetical 52.7kd protein T23g5.2 in chromosome III
3896		0.02				C. elegans	hypothetical 46.4kd protein T16H12.5 in chromosome III
1412		0.02	2	FB,BR		C. elegans	hypothetical 54.9 kd protein CO2F5.7
3879		0.02			10	C. elegans	hypothetical 68.7kd protein ZK757.1 in chromosome III
1779		0.01				C. elegans	putative ATP-dependent RNA helicase K03H1.2 in chromosome III
2172		0.01	1	BR		C. elegans	putative ATP-dependent RNA helicase T26G10.1 in chromosome III
6824		0.01	1	HA	12	C. elegans	putative ATP-dependent RNA helicase T26G10.1 in chromosome III
930		0.01				C. elegans	putative acetylcholine regulator unc-18
321		0.01			4	C. elegans	hypothetical 17.0 kd ZK370.2
1162		0.01				C. elegans	hypothetical 51.6kd protein F59B2.5 in chromosome III
41		0.01	2	BR		C. elegans	hypothetical 152.4kd protein ZK370.4 in chromosome III
2642		0.01	2	BR		C. elegans	hypothetical 152.4kd protein ZK370.4 in chromosome III
6123		0.01			X	C. elegans	hypothetical 22.6kd protein F42H10.3 in chromosome III
532		0.01				C. elegans	hypothetical 23.3kd protein ZK688.3 in chromosome III
6669		0.01				C. elegans	hypothetical 29.0kd protein ZK632.12 in chromosome III
640		0.01				C. elegans	hypothetical 48.1 kd protein CO2F5.6
2514		0.01				C. elegans	hypothetical 50.4kd protein ZK637.1 in chromosome III
5133		0.01			22	C. elegans	hypothetical 58.3kd protein F42H10.7 in chromosome III
6779		0.01				C. elegans	hypothetical 64.5 kda protein ZK652.9
1978		0.01				C. elegans	hypothetical 67.6kd protein ZK637.3 in chromosome III
5722		0.01				C. elegans	hypothetical protein F44E2.6 in chromosome III
4445		0.02	1	BR	15	Yeast	hypothetical 21.9kD protein in MRPL6 5' region
1743		0.01				Yeast	hypothetical 34.9 kd protein in URA1
5816		0.01			17	Yeast	hypothetical 34.9 kd protein in URA1
2504		0.01	1	HL		Yeast	hypothetical 195.2 kd protein in GCN3-DAL80
2513		0.01				Yeast	hypothetical 45.6kd protein in TPD3 3' region
1547		0.01	2	BR		Yeast	hypothetical 75.5kd protein in SDH1-CIM5/YTA3 intergenic
6478	0.04				20	Yeast	hypothetical 44.1 kDa protein

(ALDOA; GENX-3470), for which one eSTS marker out of four correctly localized this gene on chromosome 16, whereas one mapped to chromosome 10, and the remaining two to chromosome 3 and 10, where the two aldolase A pseudogenes (ALDOAP1, ALDOAP2) are localized. Conversely, the four eSTS markers derived from the slow myosin binding protein C transcript (GENX-3410) and the six derived from the titin transcript (GENX-3338) always provided the same correct result (chromosome 12 and chromosome 2, respectively).

GENE TRANSCRIPT OVERVIEW

An overview of the characteristics of the 6676 gene transcripts described in this study is provided in Appendix 1 (functional similarities) and Appendix 2 (accession numbers and eSTS markers). It should be clearly pointed out that positional or functional informations have been obtained in many instances from the analysis of different sequences derived from the same gene transcript. As this set of sequences do not form in most cases a single contig from which a consensus can be derived, all information is gathered through links at this stage.

Some of the gene transcripts display significant functional similarity with new putative genes characterized by the systematic sequencing of different genomes (Table 4). This provides independent evidence that they are transcribed, at least in man, and underlines the complementarity between genomic and partial cDNA sequencing approaches. This also indicates the importance of the integration of data from different organisms, as many functional features can be derived from the complete sequence of a gene in another species.

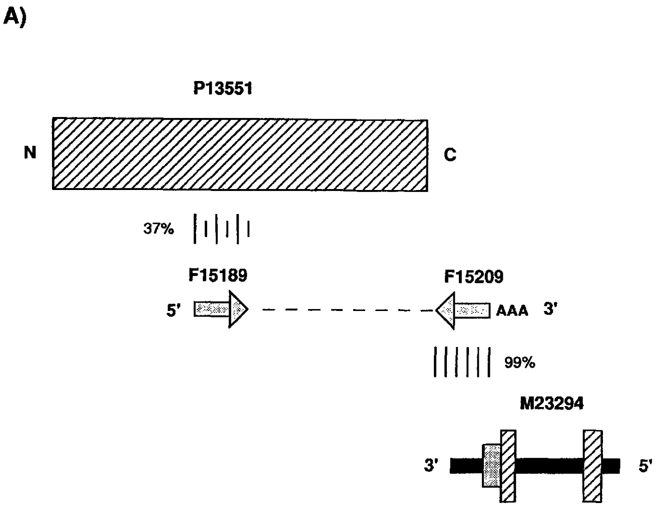
Another 27 gene transcripts are of particular interest because they appear to be derived from overlapping complementary strands of common segments of the genome. Figure 2 shows the sequence alignment of a gene transcript (GENX-5024) overlapping with the human β-hexosaminidase gene in their 3′ regions. This gene transcript exhibits significant similarity with elongation factor G from *Thermus aquaticus* (Fig. 2B) as well as with that of yeast mitochondrial elongation factor G (not available in data bases at the time of the initial search). The better similarity with this last sequence suggests that the GENX-5024 transcript may encode the human mitochondrial form of this factor. The other putative overlapping genes are listed in Appendix 1F, including two of them that had already been described (Ashworth 1993; Tsai et al. 1994). Some show a complex overlap and must be confirmed by direct genome sequencing and further detailed transcript analyses.

The relatively important percentage (0.4%) of putative overlapping gene transcripts found in this study suggests that the interaction between genes using common sequences for transcription or between mRNAs having complementary sequences is an important mechanism for the regulation of gene transcription. This may explain the conservation of some 3′-end untranslated regions during evolution. This also suggests that a higher percentage of genes may share a common overlapping segment with another gene and that the percentage observed here by partial sequencing could have been underestimated.

The main utility of the Genexpress Index is to provide new candidates for genes involved in human pathology with special reference to neuromuscular pathologies. Table 5A summarizes a set of 56 genes (~6.5% of the known genes) that have been characterized in this study and that were already known to be defective in relation with human pathologies. In addition, 17 other transcripts display similarities with known defective genes (Table 5B) and could therefore represent candidates for orphan pathologies. If a percentage of 6.5% of defective genes is applied to the whole set of genes described in this study, we probably characterized more than 400 genes involved in human pathology. The utility of partial cDNA sequencing for the identification of genes involved in neuromuscular diseases is underlined by the fact that the recently cloned Survival Motor Neuron gene (Lefebvre et al. 1995) involved in spinal atrophy was registered early in this study (GENX-4903).

CONCLUSIONS AND PERSPECTIVES

Partial sequencing of cDNA clones and chromosomal assignment of the corresponding genes are the preliminary steps in the identification of disease genes and need to be integrated in a more general way, including complete sequencing, precise mapping, and detailed expression, and functional studies. This is the purpose of the IMAGE consortium in which laboratories of various size and expertise collaborate worldwide to char-



B)

```
F15189   GGATGACTTATGTGCATTGGCATTTAAAGTTCTCCATGACAAGCAGCGAGG-CCACTGGT
          D  D  L  C  A  L  A  F  K  V  L  H  D  K  Q  R  G  P  L  V
P13551    G  P  L  A  A  L  A  F  K  I  M  A  D  P  Y  V  G  R  L  T

F15189   TTT-ATGCGCATTTACTCAGGCACTATAAAACCCCAGTTGGCCATTCATAATATTAATGG
          X  M  R  I  Y  S  G  T  I  K  P  Q  L  A  I  H  N  I  N  G
P13551    F  I  R  V  Y  S  G  T  L  T  S  G  S  Y  V  Y  N  T  T  K

F15189   AAACTGCACGGAGAGAATAAGTCGTCTGCTTTTGCCGTTTGCTGACCAACATGTAGAAAT
          N  C  T  E  R  I  S  R  L  L  L  P  F  A  D  Q  H  V  E  I
P13551    G  R  K  E  R  V  A  R  L  L  R  M  H  A  N  H  R  E  E  V

F15189   CCCTTCATTGACTGCTGGTAACATTGCTTTGACTGTTGGG-TTAAACATACTGCCACTGG
          P  S  L  T  A  G  N  I  A  L  T  V  G  X  K  H  T  A  T  G
P13551    E  E  L  K  A  G  D  L  G  A  V  V  G  L  K  E  T  I  T  G

F15189   AGACACCATTGTCTCATCCAAGTCCAGTGCATTAGCTGCA
          D  T  I  V  S  S  K  S  S  A  L  A  A
P13551    D  T  L  V  G  E  D  A  P  R  V  I  L
```

C)

```
F15209   GAATAAAATATTTTTATTGATTGA-CCTTTGACCCTCTATCTTATTACATTT
M23294   GAATAAAATATTTTTATTGATTGAACCTTTGACCCTCTATCTTATTACATTT
M13519   GAATAAAATATTTTTATTGATTGAA

F15209   AGAGGCCTGGCATCTTTCTTGTGAGACAAGCTTAAGGACACAAAAGAAACAC
M23294   AGAGGCCTGGCATCTTTCTTGTGAGACAAGCTTAAGGACACAAAAGAAACAC

F15209   AACTTTGTGATACCACTCT-CTGAAGGCTACTTATAGTAATAACTTCATAGA
M23294   AACTTTGTGATACCACTCTTCTGAAGGCTACTTATAGTAATAACTTCATAGA

F15209   TGA-CAGCATGATTCAGGTACAGTGGCTTTAAACATCAAAGCACATTTCTCA
M23294   TGAACAGCATGATTCAGGTACAGTGGCTTTAAACATCAAAGCACATTTCTCA

F15209   TTATATAAATTAAAACGGGTGGCTCCAGTGCCACTATCAAAGCTTA
M23294   TTATATAAATTAAA
```

Figure 2 Overlapping genes. (*A*) Schematic representation of a gene transcript (GENX-5024) related to *T. aquaticus* elongation factor G (P13551) and overlapping the human gene encoding β-hexosaminidase (M23294). Partial cDNA sequences are shown as arrows, coding regions, or proteins as hatched boxes, 3′-untranslated regions as dotted boxes, and introns or genomic DNA as solid boxes. Broken lines indicate unsequenced parts of the cDNA clones. Accession numbers are above their representation. Note that this transcript overlaps by 25 bp with the β-hexosaminidase transcript (GenBank M13519, not shown) and is identical to the untranscribed genomic DNA downstream of this gene. (*B*) Sequence alignments of the 5′ end sequence of GENX-5024 (lane *1*) and its putative translation (lane *2*) with *T. aquaticus* elongation factor G. (*C*) Sequence alignments of the 3′ end sequence of GENX-5024 (F15209) with the genomic (M23294) and the mRNA (M13519) sequences of the human β-hexosaminidase gene. Poly(A) signals are shown in boldface.

acterize common ordered collections of cDNA clones by a variety of complementary approaches. This includes positioning of all the genes characterized in this study in the integrated genetic, cytogenetic, and physical maps of the human genome using DNA fragments cloned in yeast artificial chromosomes (YACs) and other vectors and rearrangement, breakpoint, and radiation reduced somatic cell hybrids (Cox et al. 1990; Weissenbach et al. 1992; Cohen et al. 1993; Gyapay et al. 1994; Walter et al. 1994; Chumakov et al. 1995). The availability of the 2792 eSTS markers that we have developed (Auffray et al. 1995), together with >1200 others developed by Dr. Polymeropoulos (Polymeropoulos et al. 1992, 1993; M. Polymeropoulos, pers. comm.), provide the opportunity to double the number of mapped genes and therefore enhance the probability of identifying those involved in human pathologies.

The value of this resource to complement the positional paradigm (Collins 1995) is exemplified by the process that led to the identification of the gene responsible for the recessive form of limb-girdle muscular dystrophy (Richard et al. 1994, 1995; Chiannikulchai et al. 1995). In monitoring the progress of the genic map, toward the goal of 100-kb resolution (Cox et al. 1994), intervals of 50–250 Mb (chromosome), 10–50 Mb (cytogenetic interval), 1–10 Mb (integrated genetic map), and <1 Mb (cloned and mapped DNA fragment) are important milestones. As it is estimated that the human genome contains some 60,000 to 80,000 genes (Antequera and Bird 1993, 1994; Fields et al. 1994), the current genic map is only ~10% complete at the first level of chromosomal resolution. To contribute to the progress toward a dense and comprehensive genic map, we are currently extending the set of eSTS markers to the remaining indexed genes in collaboration with IMAGE consortium collaborators. This has already allowed regional localiza-

Table 5. Gene transcripts identical (A) or showing a partial similarity (B) to known genes involved in human pathology

A

GENX	% Muscle	% Brain	Seq	Tissues	eSTS	Map	Gene product	Pathology
5549		0.01	5	HL,BR,FB,FI		9q34	Set gene	acute undifferentiated leukemia
2412		0.01	9	CM,BR		5q21	polyposis locus protein 1	adenomatous polyposis coli
2327		0.01			9	9q33	plasma gelsolin	amyloidosis finnish type
1589		0.01	2	PI,BR		18q12.1	transthyretin	amyloidosis type I
5042		0.01	1	HL	17	17q21-q22	DNA topoisomerase II, alpha isozyme	ataxia-telangiectasia
676		0.01	2	HE,HI			alpha-catenin	cancer invasion and metastasis
2674		0.01				17p12-p11.2	peripheral myelin protein 22	Charcot-Marie-Tooth disease type Ia, Dejerine-Sottas syndrome
4270		0.02				9q34.1	c-ABL	chronic/acute myeloid leukemia
5570		0.01				8p22	lipoprotein lipase	chylomicronemia syndrome
5224		0.01				10q25.3-q26	fibroblast growth factor receptor 2	craniofacial dysostosis I
655		0.06	2	BR		20pter-p12	major prion protein	Creutzfeld-Jakob disease
3632	0.04	0.02			X	Xp21.3-p21.1	dystrophin	Duchenne and Becker's muscular dystrophies
5449		0.01	2	HL,BR		1p36.3-p36.2	procollagen-lysine 2-oxoglutarate 5-dioxygenase	Ehlers-Danlos syndrome type VI
5161		0.01			7	7q21.3-q22.1	collagen alpha 2 (I)	Ehlers-Danlos syndrome type VII-A2
3013		0.16	2	FB,BR	5	5q21-q22	adenomatous polyposis coli protein	familial adenomatous polyposis
3423	0.13	0.01				15q23-q25	electron transfer flavoprotein alpha-subunit	glutaric aciduria II
6544	0.04		1	HL		19q13.3	electron transfer flavoprotein beta-subunit	glutaric aciduria type IIb
2527	0.31	0.02	1	BR	12	1cen-q32	muscle 6-phosphofructokinase	glycogen storage disease VII
3467	0.09		1	PI	8	9q34.1	adenylate kinase isoenzyme 1	hemolytic anemia
6313	0.04				19	19q13.1	glucose-6-phosphate isomerase	hemolytic anemia
2719	0.04				2+10	20cen-q13.1	S adenosyl-homocysteine hydrolase	hypermethioninemia
776		0.03				15q14-q15	isovaleryl-CoA dehydrogenase	isovaleric acidemia
1175		0.01				Xq25-q26.1	Lowe's oculocerebronal syndrome protein	Lowe's oculocerebrorenal syndrome
846		0.01	1	BR		1p35-p34	tissue alpha-L-fucosidase	lysosomal storage disease fucosidosis
1341	0.13	0.01	13	HL,HI,FB,PI	20	20q13.2-q13.3	guanine nucleotide-binding G(s) alpha subunit	Maccune-Albright syndrome
6195	0.18		2	BR	2+19	19q13.1	skeletal muscle ryanodine receptor	malignant hyperthermia
5029		0.01			8	8q13-q22.1	carbonic anhydrase 2	Marble brain disease
408		0.02	1	CM		6p21.2-p21.1	methylmalonyl-CoA mutase	methylmalonic acidemia
6218	0.04				6	12	mevalonate kinase	mevalonic aciduria
53		0.02	4	HE,FB,BR		17p13.3	PAF acetyl hydrolase	Miller Dieker lissencephaly
3869		0.05	1	BC		Xq27.3-q28	iduronate 2-sulfatase	mucopolysaccharidosis type II
1142		0.01				10q11.2	ret proto-oncogene	multiple endocrine neoplasia type 2A/2B, Hirschprung's disease
4819		0.01				22q12.2	merlin	neurofibromatosis 2
5601		0.01				11p15.4	sphingomyeline phosphodiesterase	Niemann-Pick type A/B
2416		0.01				7p	hexokinase D	non-insulin-dependent diabetes mellitus type II
2456		0.01				16q22.1	phophatidylcholine-sterol acyltransferase	Norum disease, fish eye disease
391		0.03				10q26	ornithine aminotransferase	ornithinemia with gyrate atrophy of choroid and retina
146		0.01				3q13	UMP synthase	orotic aciduria type 1
3695		0.16	4	BR,HA		Xq21.33-q22	myelin proteolipid protein	Pelizaeus-Merzbacher disease
5974		0.01	1	BR		16p13.3	polycystic kidney disease 1 protein	polycystic kidney disease I
4906		0.01				3p22-p23	peroxisomal 3-ketoacyl-CoA thiolase	pseudo-Zellweger syndrome
1619		0.01	1	BR		22q13-qter	alpha-N-acetylgalactosaminidase	Schindler disease, Kanzaki disease
1501		0.01				Xq28	glucose 6-phosphate 1-dehydrogenase	severe chronic hemolytic anemia
3108	0.09	0.02	5	FB,BC,BO	9+11	11p15.4	hemoglobin beta	sickle cell anemia
2915		0.01	1	HI		17q21	prohibitin	sporadic breast cancer
3190		0.03	1	BR		17q11.2	neurofibromin	type I neurofibromatosis
1900		0.05					glycine dehydrogenase	type I nonketotic hyperglycinemia
1956		0.01				18q21.33	Bcl-2	type II chronic lymphatic leukemia
2373		0.01				4q11-q21	osteopontin	urinary stone
4821		0.01				12p13.3-p13.2	von willebrand factor	Von Willebrand disease
5774		0.01	2	HL,FB		3p14.3	transketolase	Wernicke-Korsakoff syndrome
4863		0.01	1	FB		Xp22.32	steryl sulfatase	X-linked ichthyosis
643		0.01				Xp11.2	delta-aminolevulinate synthase	X-linked sideroblastic anemia
3367		0.02	4	FB,PI		3p25	XP-C repair complementing protein (p58/HHR23B)	xeroderma pigmentosum
5687		0.01				13q32	DNA repair protein complementing XP-G cells	xeroderma pigmentosum group G
4921		0.01	1	BR		8q21.1	peroxysome assembly factor 1	Zellweger syndrome I

Table 5. *(continued)*

B

GENX	% Muscle	% Brain	Seq	Tissues	eSTS	Map	Gene product	Pathology
75		0.02			6		transforming protein RFP	oncogenic product
236		0.01			12	Xq26.1	hypoxanthine phosphoribosyltransferase	Lesch-Nyhan syndrome, gout
323		0.01	1	BR		1p36.1-p35	hydroxymethylglutaryl-CoA lyase	hydroxymethylglutaricaciduria
691		0.01				Xq28	glucose 6-phosphate 1-dehydrogenase	severe chronic hemolytic anemia
796		0.01				1q23-q25.1	antithrombin 3	serious spontaneous thrombosis
1075		0.01	1	HL		19q13.1-q13.2	B-cell lymphoma 3-encoded protein	lymphoma 3
1184		0.01					fibrillin 1	Marfan syndrome
1908		0.02			X	Xq21.33-q22	tyrosine-protein kinase atk	X-linked agammaglobulinemia type 1
1996		0.02				1p13.1	3-beta hydroxy-5-ene steroid dehydrogenase type II	severe depletion of steroid formation
2477		0.01	1	CM	14	6q22.3-q23.1	arginase	argininemia
3995	0.04	0.01			19	1p36.3-p36.2	procollagen-lysine 2-oxoglutarate 5-dioxygenase	Ehlers-Danlos syndrome type VI
4806		0.01				Xq26.1	hypoxanthine phosphoribosyltransferase	Lesch-Nyhan syndrome, gout
5095		0.01					glycine cleavage system H protein	type III nonketotic hyperglycinemia
5519		0.01			16	16p13.3	polycystic kidney disease 1 protein	polycystic kidney disease I
5831		0.01			1	9q34.1	adenylate kinase isoenzyme 1	hemolytic anemia
5870		0.01			19	11q23	zinc finger protein HRX	translocation in acute leukemias
6809		0.01				21q22.3	cystathionine beta-synthase	homocystinuria

tion of eSTS markers assigned to several chromosomes at one to a few megabase resolution (Callen et al. 1995; Rosier et al. 1995; C. Auffray et al., unpubl.).

As a fraction of ~15% of the eSTS markers also amplify the corresponding genes in the mouse, it will be possible to use them directly to integrate the human and mouse maps. These markers are also versatile tools that can be used in RT–PCR experiments to monitor the expression profiles of the transcripts in a large number of tissues, cell lines, and species in various physiological and pathological situations. Cross-referencing with the results collected in the study of the genomes of other model organisms will be instrumental in deciphering the functions of the large number of new genes indexed in this study and other similar programs.

MATERIAL AND DATA AVAILABILITY

The cDNA sequences have been registered to the EMBL Data Library, and the eSTS markers in GDB. The Genexpress index is available on the Internet by anonymous FTP at the address ftp.infobiogen.fr in the /pub/db/Genexpress directory with the associated help file READ_ME. In addition, they can be downloaded or consulted on the Genome Research World Wide Web server at address http://www.cshl.org. Information on the IMAGE Consortium is available from Dr. Greg Lennon, Department of Energy Genome Center, Lawrence Livermore: http://www.bio.llnl.gov/bbrp/genome/genome.htlm; cDNA clones can be obtained from Dr. Hans Lerach, Reference Library Data Base, Max Planck Institute für Molekulare Genetik, Berlin, Germany: http://gea.lif.icnet.uk/ and from Dr. Keith Gibson, Human Genome Program Resource Center, Hinxton, UK: http://www.hgmp.mrc.uk/. Primer pairs will be available from Research Genetics. Other requests concerning additional information and reagents should be directed to the Email address: Remi@pauline.vjf.inserm.fr.

NOTE ADDED IN PROOF

Adams et al. (1995. *Nature* **377:** S3–S174) have described an "Initial assessment of human gene diversity and expression patterns based on 83 million nucleotides of cDNA sequence" at the time this paper was accepted for publication.

ACKNOWLEDGMENTS

We thank all members of the Genexpress Program team and the staff at Généthon and CNRS for their dedication and support. The IMAGE consortium is coordinated by C.A. in collaboration with Dr. Greg Lennon at the Department of Energy Genome Center in Livermore, Dr. Mihaelis Polymeropoulos at the National Center for Human Genome Research, National Institutes of Health, Bethesda, and Dr. Bento Soares at Columbia University, New York. The Genexpress Program is supported by Association Française contre les Myopathies (AFM), CNRS, and grants from the Ministère de la Recherche et de l'Espace (MRE), Groupement de Recherches et d'Etudes de Génomes (GREG), and the European Union Biomed 1 Program to C.A.

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

Adams, M.D., J.M. Kelley, J.D. Gocayne, M. Dubnick, M.H. Polymeropoulos, H. Xiao, C.R. Merril, A. Wu, B. Olde, R.F. Moreno, A.R. Kerlavage, W.R. McCombie, and J.C. Venter. 1991. Complementary DNA sequencing: Expressed sequence tags and Human Genome project. *Science* **252:** 1651–1656.

Adams, M.D., M. Dubnick, A.R. Kerlavage, R. Moreno, J.M. Kelley, T.R. Utterback, J.W. Nagle, C. Fields, and J.C. Venter. 1992. Sequence identification of 2,375 human brain genes. *Nature* **355:** 632–635.

Adams, M.D., A.R. Kerlavage, C. Fields, and J.C. Venter. 1993a. 3,400 new expressed sequence tags identify diversity of transcripts in human brain. *Nature Genet.* **4:** 256–267.

Adams, M.D., M.B. Soares, A.R. Kerlavage, C. Fields, and J.C. Venter. 1993b. Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. *Nature Genet.* **4:** 373–380.

Affara, N.A., E. Bentley, P. Davey, A. Pelmear, and M.H. Jones. 1994. The identification of novel gene sequences of the human adult testis. *Genomics* **22:** 205–210.

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Altschul, S.F., M.S. Boguski, W. Gish, and J.C. Wooton. 1994. Issues in searching molecular sequence databases. *Nature Genet.* **6:** 119–129.

Antequera, F. and A. Bird. 1993. Number of CpG islands and genes in human and mouse. *Proc. Natl. Acad. Sci.* **90:** 11995–11999.

———. 1994. Predicting the total number of human genes. *Nature Genet.* **8:** 114.

Ashworth, A. 1993. Two acetyl-CoA acetyltransferase genes located in the t-complex region of the mouse chromosome 17 partially overlap the Tcp-1 and Tcp-1x genes. *Genomics* **18:** 195–198.

Auffray, C., B. Béhar, F. Bois, C. Bouchier, C. Da Silva, M.-D. Devignes, S. Duprat, R. Houlgatte, M.-N. Jumeau, B. Lamy, F. Lorenzo, H. Mitchell, R. Mariage-Samson, G. Piétu, Y. Pouliot, C. Sébastiani-Kabaktchis, and A. Tessier. 1995. IMAGE : Integrated Molecular Analysis of the human Genome and its Expression. *C.R. Acad. Sci. Paris (Life sciences)* **318:** 263–272.

Bairoch, A. and B. Boeckmann. 1994. The Swiss-Prot protein sequence data bank: Current status. *Nucleic Acids Res.* **22:** 3578–3580.

Benson, D.A., M. Boguski, D.J. Lipman, and J. Ostell. 1994. GenBank. *Nucleic Acids Res.* **22:** 3441–3444.

Berry, R., T.J. Stevens, N.A.R. Walter, A.S. Wilcox, T. Rubano, J.A. Hopkins, J. Weber, R. Goold, M.B. Soares, and J.M. Sikela. 1995. Gene-based sequence-tagged-sites (STSs) as the basis for a human gene map. *Nature Genet.* **10:** 415–423.

Brenner, S. 1990. The human genome: The nature of the enterprise. *CIBA Found. Symp.* **149:** 6–17.

Callen, D., S.A. Lane, H. Kozman, G. Kremmidiotis, S.A. Whitmore, M. Lowenstein, N.A. Doggett, N. Kenmochi, D.C. Page, D.R. Maglott, W.C. Nierman, K. Murakawa, R. Berry, J.M. Sikela, R. Houlgatte, C. Auffray, and G.R. Sutherland. 1995. Integration of transcript and genetic maps of chromosome 16 at near 1 Mb resolution: Demonstration of a hot spot for recombination at 16p12. *Genomics* **29:** 503–511.

Chen, E., M. d'Urso, and D. Schlessinger. 1994. Functional mapping of the human genome by cDNA localization *versus* sequencing. *BioEssays* **16:** 693–698.

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

Chumakov, I.M., P. Rigault, I. LeGall, C. Bellané-Chantelot, A. Billault, S. Guillou, P. Soularue, G. Guasconi, E. Poullier, I. Gross, et al. 1995. A YAC contig map of the human genome. *Nature* **377:** S175–S297.

Claverie, J.-M. and D. States. 1993. Information enhancement methods for large scale sequence analysis. *Comput. Chem.* **17:** 191–201.

Cohen, D., I. Chumakov, and J. Weissenbach. 1993. Première génération de la carte physique du génome humain. *C.R. Acad. Sci. Paris (Life Sciences)* **316:** 1484–1488.

Collins, F.S. 1995. Positional cloning moves from perditional to traditional. *Nature Genet.* **9:** 347–350.

Collins, F. and D. Galas. 1993. A new five-year plan for the US human genome project. *Science* **262:** 43–46.

Cox, D.R., M. Burmeister, E.R. Price, S. Kim, and R.M. Myers. 1990. Radiation hybrid mapping: A somatic cell genetic method for constructing high resolution maps of mammalian chromosomes. *Science* **250:** 245–250.

Cox, D.R., E.D. Green, E.S. Lander, D. Cohen, and R.M. Myers. 1994. Assessing mapping progress in the human genome project. *Science* **265:** 2031–2032.

Davies, R.W., A.B. Roberts, A.J. Morris, G.W. Griffith, J. Jerecic, S. Ghandi, K. Kaiser, and A. Savioz. 1994. Enhanced access to rare brain cDNAs by prescreening libraries: 207 new mouse brain ESTs. *Genomics* **24:** 456–463.

Durkin, A.S., D.R. Maglott, and W.C. Nierman. 1992. Chromosomal assignment of 38 human brain expressed sequence tags (ESTs) by analyzing fluorescently labeled PCR products from hybrid cell panels. *Genomics* **14:** 808–810.

Efstratiadis, A., F.C. Kafatos, A.M. Maxam, and T. Maniatis. 1976. Enzymatic *in vitro* synthesis of globin genes. *Cell* **7:** 279–288.

Fasman, K.H., A.J. Cuticchia, and D.T. Kingsbury. 1994. The GDB human genome database anno 1994. *Nucleic Acids Res.* **22:** 3462–3469.

Fields, C., M.D. Adams, O. White, and J.C. Venter. 1994. How many genes in the human genome? *Nature Genet.* **7:** 345–346.

Franco, G.R., M.D. Adams, M.B. Soares, A.J. Simpson, J.C. Venter, and S.D. Pena. 1995. Identification of new Schistosoma mansoni genes by the EST strategy using a directional cDNA library. *Gene* **152:** 141–147.

Frigerio, J.M., P. Berthezene, P. Garrido, E. Ortiz, S. Barthellemy, S. Vasseur, B. Sastre, I. Seleznieff, J.C. Dagorn, and J.L. Iovanna. 1995. Analysis of 2166 clones from a human colorectal cancer cDNA library by partial sequencing. *Hum. Mol. Genet.* **4:** 37–43.

Fukushima, A., K. Okubo, H. Sugino, N. Hori, R. Matoba, T. Niiyama, K. Murakawa, J. Yoshii, M. Yokoyama, and K. Matsubara. 1994. Chromosomal assignment of HepG2 3′-directed partial cDNA sequences by Southern blot hybridization using monochromosomal hybrid cell panels. *Genomics* **22:** 127–136.

Gieser, L. and A. Swaroop. 1992. Expressed sequence tags and chromosomal localization of cDNA clones from a subtracted retinal pigment epithelium library. *Genomics* **13:** 873–876.

Gish, W. and D.J. States. 1993. Identification of protein coding regions by database similarity search. *Nature Genet.* **3:** 266–272.

Grausz, D. and C. Auffray. 1993. Strategies in cDNA programs. *Genomics* **17:** 530–532.

Gyapay, G., J. Morissette, A. Vignal, C. Dib, C. Fizames, P. Millasseau, S. Marc, G. Bernardi, M. Lathrop, and J. Weissenbach. 1994. The 1993-94 Généthon human genetic linkage map. *Nature Genet.* **7:** 246–339.

Hochgeschwender, V. 1992. Toward a transcriptional map of the human genome. *Trends Genet.* **8:** 41–44.

Höfte, H., T. Desprez, J. Amselem, H. Chiapello, M. Caboche, A. Moisan, M.F. Jourjon, J.L. Charpenteau, P. Berthomieu, D. Guerrier, J. Giraudat, F. Quigley, F. Thomas, D.Y. Yu, R. Mache, M. Raynal, R. Cooke, F. Grellet, M. Delseny, Y. Parmentier, G. Marcillac, C. Gigot, J. Fleck, G. Philipps, M. Axelos, C. Bardet, D. Tremousaygue, and B. Lescure. 1993. An inventory of 1152 expressed sequence tags obtained by partial sequencing of cDNAs from Arabidopsis thaliana. *Plant J.* **4:** 1051–1061.

Höög, C. 1991. Isolation of a large number of novel mammalian genes by a differential cDNA library searching strategy. *Nucleic Acids Res.* **19:** 6123–6127.

Hyde, D.R., K.L. Mecklenburg, J.A. Pollock, T.S. Vihtelic, and S. Benzer. 1991. Twenty Drosophila visual system cDNA clones: One is homolog of human arrestin. *Proc. Natl. Acad. Sci.* **87:** 1008–1012.

Kato, K. 1990. A collection of cDNA clones with specific expression patterns in mouse brain. *Eur. J. Neurosci.* **2:** 704–711.

———. 1992. Finding new genes in the nervous system by cDNA analysis. *Trends Neurosci.* **15:** 319–323.

Kerr, S.M., S. Vambrie, S.J. McKay, and H.J. Cooke. 1994. Analysis of cDNA sequences from mouse testis. *Mammalian Genome* **5:** 557–565.

Khan, A.S., A.S. Wilcox, M.H. Polymeropoulos, J.A. Hopkins, T.J. Stevens, M. Robinson, A.K. Orpana, and J.M. Sikela. 1992. Single pass sequencing and physical and genetic mapping of human brain cDNAs. *Nature Genet.* **2:** 180–185.

Konishi, K., Y. Morishima, E. Ueda, Y. Kibe, K. Nonomura, K. Yamanishi, and H. Yasuno. 1994. Cataloguing of genes expressed in human keratinocytes: Analysis of 607 randomly isolated cDNA sequences. *Biochem. Biophys. Res. Commun.* **202:** 976–983.

Kurata, N., Y. Nagamura, K. Yamamoto, Y. Harushima, N. Sue, J. Wu, B.A. Antonio, A. Shomura, T. Shimizu, S.Y. Lin, T. Inoue, A. Fukuda, T. Shimano, Y. Kuboki, T. Toyama, Y. Miyamato, T. Kirihara, K. Hayasaka, A. Miyao, L. Monna, H.S. Zhong, Y. Tamura, Z.X. Wang, T. Momma, Y. Umehara, M. Yano, T. Sasaki, and Y. Minobe. 1994. A 300 kilobase interval genetic map of rice including 883 expressed sequences. *Nature Genet.* **8:** 365–372.

Lefebvre, S., L. Bürglen, S. Reboullet, O. Clermont, P. Burlet, P. Viollet, B. Bénichou, C. Cruaud, P. Millasseau, M. Zeviani, D. Le Paslier, J. Frézal, D. Cohen, J. Weissenbach, A. Munnich, and J. Melki. 1995. Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* **80:** 155–165.

Liew, C.C., D.M. Hwang, Y.W. Fung, C. Laurenssens, E. Cukerman, S. Tsui, and C.Y. Lee. 1994. A catalogue of genes in the cardiovascular system as identified by expressed sequence tags. *Proc. Natl. Acad. Sci.* **91:** 10645–10649.

Matsubara, K. and K. Okubo. 1993. Identification of new genes by systematic analysis of cDNAs and database construction. *Curr. Opin. Biotech.* **4:** 672–677.

McCombie, W.R., M.D. Adams, J.M. Kelley, M.G. FitzGerarld, T.R. Utterback, M. Khan, M. Dubnick, A.R. Kerlavage, J.C. Venter, and C. Fields. 1992. *Caenorhabditis elegans* expressed sequence tags identify gene families and potential disease gene homologues. *Nature Genet.* **1:** 124–131.

Milner, R.J. and J.G. Sutcliffe. 1983. Gene expression in rat brain. *Nucleic Acids Res.* **11:** 5497–5517.

Murakawa, K., K. Matsubara, A. Fukushima, J. Yoshii, and K. Okubo. 1994. Chromosomal assignments of 3′-directed partial cDNA sequences representing novel genes expressed in granulocytoid cells. *Genomics* **23:** 379–389.

Newman, T., F.J. de Bruijn, P. Green, K. Keegstra, H. Kende, L. McIntosh, J. Ohlrogge, N. Raikhel, S. Somerville, M. Thomashow, E. Retzel, and C. Somerville. 1994. Genes galore: A summary of methods for accessing results from the large-scale partial sequencing of anonymous Arabidopsis cDNA clones. *Plant Physiol.* **106:** 1241–1255.

Nishiguchi, S., T. Joh, K. Horie, Z. Zou, T. Yasunaga, and K. Shimada. 1994. A survey of genes expressed in undifferentiated mouse embryonal carcinoma F9 cells: Characterization of low-abundance mRNAs. *J. Biochem.* **116:** 128–139.

Nomura, N., N. Miyajima, T. Sazuka, A. Tanaka, Y. Kawarahayasi, S. Sato, T. Nagase, N. Seki, K. Ishikawa, and S. Tabata. 1994. Prediction of the coding sequences of unidentified human genes. 1. The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by analysis of randomly sampled cDNA clones from human immature myeloid cell line KG-1. *DNA Res.* **1:** 27–35.

Okubo, K., H. Hori, R. Matuba, T. Niiyama, and K. Matsubara. 1991. A novel system for large-scale sequencing of cDNA by PCR amplification. *DNA Sequence* **2:** 137–144.

Okubo, K., H. Hori, R. Matuba, T. Niiyama, A. Fukushima, Y. Kojima, and K. Matsubara. 1992. Large scale cDNA sequencing analysis of quantitative and qualitative aspects of gene expression. *Nature Genet.* **2:** 173–179.

Okubo, K., J. Yoshii, H. Yokouchi, M. Kameyama, and K. Matsubara. 1994. An expression profile of active genes in human colonic mucosa. *DNA Res.* **1:** 37–45.

Palazzolo, M.J., D.R. Hyde, K. VijayRaghavan, K. Mecklenburg, S. Benzer, and E. Meyerowitz. 1987. Use of a new strategy to isolate and characterize 436 drosophila cDNA clones corresponding to RNAs detected in adult heads but not in early embryos. *Neuron* **3:** 527–539.

Park, Y.S., J.M. Kwak, O. Kwon, Y.S. Kim, D.S. Lee, M.J. Cho, H.H. Lee, and H.G. Nam. 1993. Generation of expressed sequence tags of random root cDNA clones of *Brassica napus* by single-run partial sequencing. *Plant Physiol.* **103:** 359–370.

Parrish, J.E. and D.L. Nelson. 1993. Methods for finding genes a major rate-limiting step in positional cloning. *Genet. Anal. Techniques Applic.* **10:** 29–41.

Pawlak, A., C. Toussaint, I. Lévy, F. Bulle, M. Poyard, R. Barouki, and G. Guellaën. 1995. Characterization of a large population of mRNAs from human testis. *Genomics* **26:** 151–158.

Pearson, W.R. 1990. Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol.* **183:** 63–98.

Pearson, P., C. Francomano, P. Foster, C. Bocchini, P. Li, and V. McKusick. 1994. The status of On line Mendelian Inheritance in Man (OMIM) medio 1994. *Nucleic Acids Res.* **22:** 3470–3473.

Polymeropoulos, M.H., H. Xiao, J.M. Sikela, M. Adams, J.C. Venter, and C.R. Merril. 1992. Chromosomal assignment of 46 brain cDNAs. *Genomics* **12:** 492–496.

———. 1993. Chromosomal distribution of 320 genes from a brain cDNA library. *Nature Genet.* **4:** 381–386.

Putney, S.D., W.C. Herlihy, and P. Schimmel. 1983. A new troponin T and cDNA clones for 13 different muscle proteins, found by shotgun sequencing. *Nature* **302:** 718–721.

Richard, I., O. Broux, N. Chiannilkulchai, F. Fougerousse, V. Allamand, N. Bourg, L. Brenguier, C. Devaud, P. Pasturaud, C. Roudaut, F. Lorenzo, C. Sébastiani-Kabaktchis, R.A. Schultz, M.H. Polymeropoulos, G. Gyapay, C. Auffray, and J.S. Beckmann. 1994. Regional localization of human chromosome 15 loci. *Genomics* **23:** 1001–1009.

Richard, I., O. Broux, V. Allamand, F. Fougerousse, N. Chiannilkulchai, N. Bourg, L. Brenguier, 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.

Rosier, M.-F., I. Reguigne-Arnould, P. Couillin, M.-D. Devignes, and C. Auffray. 1995. Regional assignment of 68 new human gene transcript on chromosome 11. *Genome Res.* **5:** 60–70.

Rougeon, F., P. Kourilsky, and B. Mach. 1975. Insertion of a rabbit β-globin gene sequence into an *E. coli* plasmid. *Nucleic Acids Res.* **2:** 2365–2378.

Sasaki, T., J. Song, Y. Koga-Ban, E. Matsui, F. Fang, H. Higo, H. Nagasaki, M. Hori, M. Miya, E. Murayama-Kayano, T. Takiguchi, A. Takasuga, T. Niki, K. Ishimaru, H. Ikeda, Y. Yamamoto, Y. Mukai, I. Ohta, N. Miyadera, I. Havukkala, and Y. Minobe. 1994. Toward cataloguing all rice genes: Large-scale sequencing of randomly chosen rice cDNAs from a callus cDNA library. *Plant J.* **6:** 615–624.

Sikela, J. and C. Auffray. 1993. Finding new genes faster than ever. *Nature Genet.* **3:** 189–191.

Smith, L.M., W. Hunkapiller, T.J. Hunkapiller, and L.E. Hood. 1985. The synthesis of oligonucleotides containing an aliphatic amino group at the 5′ terminus: Synthesis of fluorescent DNA primers for use in DNA sequence analysis. *Nucleic Acids Res.* **13:** 2399–2412.

Smith, L.M., J.Z. Sanders, R.J. Kaiser, P. Hughes, C. Dodd, C.R. Connell, C. Heiner, S.B.H. Kent, and L.E. Hood. 1986. Fluorescence detection in automated DNA sequence analysis. *Nature* **321:** 674–679.

Soares, M.B., M.D.F. Bonaldo, P. Jelenc, L. Su, L. Lawton, and A. Efstratiadis. 1994. Construction and characterization of a normalized cDNA library. *Proc. Natl. Acad. Sci.* **91:** 9228–9232.

Southern, E.M. 1992. Genome mapping: cDNA approaches. *Curr. Opin. Genet. Dev.* **2:** 412–416.

Sudo, K., K. Chinen, and Y. Nakamura. 1994. 2058 expressed sequence tags (ESTs) from a human fetal lung cDNA library. *Genomics* **24:** 276–279.

Sutcliffe, J.G. 1988. mRNA in the mammalian central nervous system. *Annu. Rev. Neurosci.* **11:** 157–198.

Takeda, J., H. Yano, S. Eng, Y. Zeng, and G.I. Bell. 1993. A molecular inventory of human pancreatic islets: Sequence analysis of 1000 cDNA clones. *Hum. Mol. Genet.* **2:** 1793–1798.

Tsai, J.-Y., M.L. Namin-Gonzales, and L.M. Silver. 1994. False association of human ESTs. *Nature Genet.* **8:** 321–322.

Uchimiya, H., S. Kidou, T. Shimazaki, S. Takamatsu, H. Hashimoto, R. Nishi, S. Aotsuka, Y. Matsubayashi, N. Kidou, M. Umeda, and A. Kato. 1992. Random sequencing of cDNA libraries reveals a variety of expressed genes in cultured cells of rice (Oryza sativa L.). *Plant J.* **2:** 1005–1009.

Walter, M.A., D.J. Spillett, P. Thomas, J. Weissenbach, and P. Goodfellow. 1994. A method for constructing radiation hybrid maps of whole genomes. *Nature Genet.* **7:** 22–28.

Waterston, R., C. Martin, M. Craxton, C. Huynh, A. Coulson, L. Hillier, R. Durbin, P. Green, R. Shownkenne, N. Halloran, M. Metzstein, T. Hawkins, R. Wilson, M. Berks, Z. Du, K. Thomas, J. Thierry-Mieg, and J. Sulston. 1992. A survey of expressed genes in *Caenorhabditis elegans. Nature Genet.* **1:** 114–123.

Weissenbach, J., G. Gyapay, C. Dib, A. Vignal, J. Morissette, P. Millasseau, G. Vaysseix, and M. Lathrop. 1992. A second-generation linkage map of the human genome. *Nature* **359:** 794–801.

Wilcox, A.S., A.S. Khan, J.A. Hopkins, and J.M. Sikela. 1991. Use of 3′ untranslated sequences of human cDNAs for rapid chromosome assignment and conversion to STSs: Implications for an expression map of the genome. *Nucleic Acids Res.* **19:** 1837–1843.

Received July 13, 1995; accepted in revised form September 28, 1995.

APPENDICES: THE GENEXPRESS INDEX

Only a small fraction of Appendices 1D and 2A–B are shown in print as examples. The remaining parts are available as an electronic supplement together with the full text of the other appendices at http://www.cshl.org/ in the Genome Research section. Please follow the links from the home page.

APPENDIX I: FUNCTIONAL SIMILARITIES

Classification of each gene transcript has been designated as Known (A), Homolog (B), Related (C), or Unknown (D) on the basis of data base similarities. A special classification was necessary for Overlapping transcripts (E), or New transcripts in known (already sequenced) regions (F).

Information available in Appendix I are the Genexpress Index number (GENX), the clone frequency in muscle (% muscle) or brain (% brain) libraries, the number of partial cDNA sequences registered in public data bases linked to this gene transcript (Seq), the chromosomal localization obtained with Genexpress eSTS markers (eSTS), or already available in GDB or GenBank (Map), the species (Species), and gene product name (Gene Product). In new (E) or overlapping (F) subsections, "DNA region" or "overlapped gene" refer to the gene product name having an identical or overlapping sequence, whereas "similarity" and "gene product" refer to the true functional similarity of the transcript.

"Tissues" indicates the origin of the cDNA partial sequences, abbreviated as follows: (AE) aortic endothelium; (BC) adult brain cortex; (BO) adult bone; (BR) brain; (BS) brain striatum; (CC) colorectal cancer; (CM) colon mucosa; (EK) epidermis keratinocyte; (FA) fetal adrenals; (FB) fetal brain; (FI) fibroblast; (FK) fetal kidney; (FLS) fetal liver and spleen; (HA) heart atrium; (HE) heart; (HG) HepG2 liver cell line; (HI) hippocampus; (HL) HL60 cell line; (LF) lung fibroblast; (LI) liver; (PI) pancreatic islets; (PL) placenta; (RE) retinal pigment epithelium; (RH) adult rhabdomyosarcoma cell line; (TC) temporal cortex; (TE) adult testis; (TH) THP-1 cell line; (UN) unclassified. A few gene transcripts were found identical to short genomic sequences such as STSs and were classified in the "known gene transcripts" subsection, to show their similarity, even though they should be considered as new gene transcripts.

APPENDIX II: DATA BASE CROSS-INDEXES

Indicated are accession numbers (A) or eSTS markers (B) linked to the different gene transcripts. The first column (in bold) indicates the Genexpress Index number (GENX). Accession numbers correspond either to Genexpress sequences (normal typing) or sequences from other groups (accession numbers preceded by #) present in GenBank. The different columns for eSTS markers indicate the Genexpress Index number (col. 1), GDB D-number (col. 2), sequences of forward (col. 3) or reverse (col. 4) primers, and expected (col. 5) and observed (col. 6) sizes for the genomic PCR product.

Appendix 1A: Known gene transcripts

GENX	% Muscle	% Brain	Seq	Tissues	eSTS	Map	Gene product
2	2.20		3	HE	14	14q11.2-q13	cardiac muscle beta-myosin heavy chain
3	0.09	0.06	2	HI,HE		3p21	RhoA
6		0.01				8	carbonic anhydrase-related protein
10		0.01	1	HL			pre-B cell enhancing factor
12		0.01	3	BR		1	profilin 2
33		0.01	1	FB		4q31.3-q33	soluble guanylate cyclase beta-1
42		0.01	5	TH,BR,PI			histone H3.3
50		0.05				5q15-q21	neuroendocrine convertase 1
53		0.02	4	HG,FB,BR		17p13.3	PAF acetyl hydrolase
55		0.01					mitochondrial matrix protein P1
85		0.01				15q26	aggrecan 1
92		0.05	2	PI,BR			initiation factor 4A-I
106		0.01				1q25-q31	nicein B2
114		0.06	1	FLS	9		protein-tyrosine phosphatase delta
121		0.07	6	HL,HI,FB,BR	17	17q25	metalloproteinase inhibitor
124		0.02	2	HI,BR		19q13.4	protein kinase C gamma
128		0.02	3	HL,BR,FLS			randomly sampled cDNA D31767
133		0.01	2	HL,BR			related lysyl-tRNA synthetase
146		0.01				3q13	UMP synthase
176		0.01					NOT
191		0.01			1	1p21	collagen alpha 1 (XI)
192		0.02					gravin
197		0.01	2	HL,BR		21q22.3	collagen alpha 1 (XVIII)
199	0.04	0.01	5	CM,HI,BR	12	12p13.3-p12.3	alpha 2 macroglobulin
205		0.01			1		53PB2 p53-binding protein
226		0.06	1	HI			calnexin
236		0.01				Xq28	bone/cartilage proteoglycan I precursor
262		0.01					TGF-beta receptor type 3
277		0.01				13q13	caseine kinase 1
278		0.01				6p24.2-p23	endothelin-1
279		0.01				15q24-q25	cathepsin H
303	0.13	0.08	15	HL,FB,HI,PI,BR	X+3+4+7		elongation factor 1 gamma
314		0.03					methylmalonate-semialdehyde dehydrogenase
317	0.04	0.01	1	BR		3p21.1-p12	non-specific delta-amino levulinate synthase 1
334		0.01			17		STS UT228
341		0.02	2	CM,BR			B12 protein
344		0.01	1	HL		8q24.12-q24.13	c-myc proto-oncogene
350		0.01					G25k GTP-binding protein, brain isoform
357		0.01	9	HI,BR		8p21	clusterin
361		0.01				11q13	phospholipase C-Beta-3
365		0.01					60S ribosomal protein L7
374		0.01	1	HG		16p13.1	GST1-HS GTP-binding protein
380		0.01	1	BR	17	17	replication protein A 70kD DNA-binding subunit
387		0.02				16q22.1	calretinin
391		0.03				10q26	ornithine aminotransferase
405		0.01	1	HL	3		coatomer complex
408		0.02	1	CM		6p21.2-p21.1	methylmalonyl-CoA mutase
410		0.08	1	BR	11		metabotropic glutamate receptor 5a
424		0.02	3	BR			oriP binding protein (OBP-2)
441		0.02	2	HI,FI		6p21.3	antigen peptide transporter 1
443		0.01			1		DNA M82819
452		0.02	4	HI,BR			high mobility group protein HMG1
454		0.01	4	HG,BR			protein phosphatase-1 gamma 1
459		0.01	1	FB		17	psoriasis-associated fatty acid binding protein homolog
469		0.02	4	HI,BR		8p21	neurofilament subunit NF-L
474		0.02	1	HG			mitochondrial 60S ribosomal protein L3
483	0.04	0.01			1		STS UT5790
492		0.01	1	BR	5	5q31	dinucleotide repeat polymorphism at locus D5S178
507		0.01				21q22	glutamate receptor (GLUR5)
515		0.02				6q14-q21	5'nucleotidase precursor
521		0.01					beta-arrestin 1
549		0.01	1	HI		7q22	cAMP-dependent protein kinase 2 beta regulatory
554		0.03	3	BR		18	MHC binding protein 2
561	0.13	0.01	1	HA	1+3		randomly sampled cDNA D14662
569		0.01				17p13.3	mRNA for D17S379-E gene
586		0.01	3	HG,TE,HA			Rab-1A
587		0.01					replicative polymerase accessory protein related
593		0.01					beta catenin
604		0.01	2	FB,BR			kinesin light chain
617		0.02					GC-box binding protein
625	0.04	0.01					cyclin G1
643		0.01				Xp11.2	delta-aminolevulinate synthase
644		0.01	2	BR		3p25	c-raf-1
647		0.01	3	BR		9	transducin-like enhancer protein 1
648		0.01					anion exchange protein 3
653		0.01					cAMP-responsive element modulator alpha isoform
655		0.06	2	BR		20pter-p12	major prion protein
662		0.02				17	sodium-potassium ATPase beta-2 subunit
667		0.01					protein phosphatase PP-X
669		0.01	1	BR		6q21	proto-oncogene tyrosine-protein kinase FYN
672		0.01				4q31.3-q32	neuropeptide Y receptor type 1
676		0.01	2	HG,HI			alpha-catenin
728		0.01				Xq28	STS SWXD139
733		0.01				17q11.2-q12	corticotropin releasing factor receptor
743		0.01	1	BR			transmembrane protein p63
744		0.02	2	HL,PI		8p22	cathepsin B
753		0.02	1	HL		12q12-q13	CD63
763		0.01				2p22-p21	interferon-induced double stranded RNA-activated protein kinase
765	0.04	0.02	2	BR	12		periodic tryptophan protein related 56kd protein
776		0.03				15q14-q15	isovaleryl-CoA dehydrogenase
779		0.03	1	HG		5q23-q31	protein phosphatase 2A alpha catalytic subunit
781		0.01	1	FB		3q28	somatostatin I
791	0.04	0.03	1	FI		12q21.3-q23	bone proteoglycan II
800		0.01					myeloid cell differenciation protein MCL1
804	0.04	0.01					serine/threonine protein kinase receptor R1
809		0.01					pre-T/NK cell associated protein 3C1
810		0.02	1	BR		7q21-q22	guanine nucleotide-binding protein G(i) alpha 1
814		0.01	1	BR		17q21.1-q21.2	B-box protein within the BCRA1 region
815		0.01	1	FB			transcription factor ISGF3
826		0.03	1	HI			dematin
828		0.01					leukotriene A-4 hydrolase
830		0.01	1	BR		X	ubiquitin-like protein GdX
844	0.04	0.05	1	HI	3		microtubule associated protein 4
846		0.01	1	BR		1p35-p34	tissue alpha-L-fucosidase
849		0.01	1	BR			randomly sampled cDNA D31885
862	0.22	0.08	12	HL,FB,BR,PI,FI,TE	8+17		Sui1 translation initiation factor isolog
868	0.04	0.08	7	BR,FB,FI	17	17q21	microtubule associated Tau protein
872		0.02					elongation factor TF2S

Appendix 1A: Known gene transcripts (*continued*)

874	0.04	0.03	1	BR			calmodulin-dependent calcineurin A
877		0.01					randomly sampled cDNA D28475
901		0.01					carboxyl methyltransferase
903		0.02	1	HL		21q22.3	c-ETS-2
914		0.02	2	BR		11q22.3-q23	Thy-1
915		0.02	2	CM,FA		8p23.1-p22	farnesyl-diphosphate farnesyltransferase
929		0.02				21q22.1	interferon gamma receptor beta chain
950		0.02	1	HL		14q23	Max
972	0.09	0.05	10	HI,BR	1		glutamine-synthetase
977		0.01					protein phosphatase PP2A 55kd regulatory subunit neuronal isoform
980		0.01				21q	STS in D21S337 DNA marker
985		0.01	5	PI,LI			alpha-1-antichymotrypsine
998		0.01					B94 protein
1017		0.01					tyrosine kinase receptor
1032		0.01				1p36.1-p35	sodium-proton exchanger 1
1033		0.01				9q34-qter	endoglin
1045	0.04	0.01	4	HL,FB,FLS,BC			DNA-binding protein
1062		0.05	1	BR		4q11-q12	PDGF alpha receptor
1080		0.01	2	BR		17	granulin
1098		0.01				1p31	medium-chain acyl-CoA dehydrogenase
1104		0.13	9	FB,BR		6p21.3	tubulin beta-1
1108		0.02	3	FB,BR			MADS/MEF2-family transcription factor MEF2C
1115		0.01				6	S-adenosylmethionine decarboxylase 1
1118		0.01				5q34-q35	GABA-A receptor alpha 1 subunit
1119		0.01	1	HG		11q12	high mobility group box SSRP1
1121	0.18	0.05	2	HI	1	1q21-q23	sodium-potassium ATPase alpha2
1129		0.02	4	HG,HL,FB			major nuclear matrix protein matrin-3
1141		0.03	5	HI,PI,BR			14-3-3 protein theta
1142		0.01				10q11.2	ret proto-oncogene
1146		0.01					secretogranin 2
1151		0.01				19p13.3	guanine nucleotide-binding protein G(y) alpha
1159		0.01	1	BR			triplet-repeat containing cDNA CTG-B37
1161		0.01				19q13.2-q13.3	DNA ligase 1
1175		0.01				Xq25-q26.1	Lowe's oculocerebronal syndrome protein
1196		0.01	3	FB		2	prothymosin alpha
1225		0.01				8q22-q23	syndecan-2
1234		0.02	5	HG,FB,BR			motor protein
1240		0.01				19p13.2	Rab-3A
1249		0.01					transcription factor HTF4A
1250		0.03	3	BR,PI			TPA-inducible c54 mRNA
1256		0.01					RNA helicase RCK
1280		0.01					kruppel zinc finger LDR152
1285		0.02				10q24-q25	Mxi1
1298		0.01	1	HL			NuMA gene (clone T33)
1319		0.01					aspartyl beta-hydroxylase
1325		0.02	1	BR			Bcl-xL
1341	0.13	0.01	13	HL,HI,FB,PI	20	20q13.2-q13.3	guanine nucleotide-binding G(s) alpha subunit
1342	0.04	0.01	1	BR		11q13	protein phosphatase PP-1 alpha
1350		0.01	1	PI			RSU-1
1351		0.01				7	sorcin SR1
1380	0.04	0.01	9	HI,BR,PI			ADP-ribosylation factor 1
1382		0.01				19p13.3	lamin B1
1407		0.02	1	BR		16p12	protein kinase C beta
1411		0.01	1	FB			cadherin 13
1423		0.01				3p26-p25	plasma membrane calcium ATPase brain isoform 2
1434		0.01					cAMP-response element binding protein CREB
1450		0.03	4	BR			clone 1NIB210 from infant brain
1458		0.02	1	HI		16q13	guanine nucleotide-binding protein G(0) alpha
1464		0.06	2	BR		12cen-q21	synaptotagmin 1
1472		0.02					CD83
1481		0.03				15q11-q13	gamma-aminobutyric-acid receptor alpha-5 subunit
1492		0.02	1	BR		5q33	glutamate receptor subunit GluH1
1493		0.02				21q22.1	phosphoribosylamine-glycine ligase
1496	0.04	0.02	2	HI,BR		17q11.2-q12	thyroid hormone receptor alpha
1498		0.01				6p21	HMG-I(Y)
1501		0.01				Xq28	glucose 6-phosphate 1-dehydrogenase
1510		0.01	1	BR		14q22-q24	HSP70 protein 2
1522		0.01				17q25.1	peroxisomal fatty acyl-coA oxidase
1528		0.01				20q12-q13.1	DNA topoisomerase I
1529		0.01	2	HL,HI		20q12-q13.1	ribophorin II
1530		0.01	1	BR			randomly sampled cDNA D14695
1542		0.02	1	BR		19p13.3-p13.2	complement C3
1548		0.01					STS UT5929
1549		0.02	3	BR			CLA-1
1560		0.03	2	FB,BR		22q11.2	guanine nucleotide-binding protein G(Z) alpha
1572		0.01					activin receptor type II
1574		0.02					sterol regulatory element binding protein 2
1581		0.01					growth-arrest-specific protein
1588	0.04	0.01			11	11q24.1	docking protein alpha
1589		0.01	2	PI,BR		18q12.1	transthyretin
1596		0.01	2	HL,BR		11q23-q25	amyloid-like protein 2
1602		0.01	1	BR			randomly sampled cDNA D21064
1606	0.09	0.02	4	FB,BR	2+10		actin-related protein
1619		0.01	1	BR		22q13-qter	alpha-N-acetylgalactosaminidase
1630		0.01	1	FB		8p12-p11	DNA polymerase beta
1639		0.06			X	X	membrane transport protein XK
1643		0.01	1	BR			rho GDP-dissociation inhibitor 1
1644		0.02					T-complex protein 1 gamma
1650		0.02					histone H3.1
1658		0.02				15q22-qter	mannose 6-phosphate isomerase
1663		0.01					autoantigen PM-SCL
1682		0.02	1	HG		1p35-p31.3	glucose transporter 1 erythrocyte/brain
1693		0.01	1	AE			caldesmon
1705		0.02	1	BR			tight junction protein ZO-1
1713		0.01				13q34	liver protein similiar to pig hepatic protein
1719		0.02	3	BR			phosphodiesterase gene
1720		0.01				12q13	early response protein NAK1
1723		0.01	2	BR		17	pyrroline-5-carboxylate reductase
1727		0.01	1	HL			clone 76 mRNA
1753		0.01	2	FB		14q32.3	creatine kinase B
1759		0.01	1	HL			randomly sampled cDNA D25538
1780		0.01	2	FB		21	sodium channel protein brain I alpha subunit
1783	0.04	0.01			6	6q25-q27	cation-independent mannose-6-phosphate receptor
1785		0.01				3p22-p21.33	beta galactosidase
1789		0.01				9p21-p12	atrial natriuretic peptide receptor B
1801		0.01	2	FB,BR			tubulin beta-5
1823		0.01	1	BR			finger protein HF.12
1828		0.01	1	BR		11p13	paired box protein PAX-6
1835		0.02					aryl sulfotransferase
1850		0.01				2	erythrocyte beta adducin
1852		0.02	1	FB			hnRNP complex K
1866		0.01	2	BR			transmembrane receptor Ror1

Appendix 1A: Known gene transcripts (*continued*)

1868		0.02	2	BR		2p25	red cell acid phosphatase 1
1872		0.01					transcription factor LSF
1881		0.01				5q23.3-q31.2	lysyl oxidase
1896		0.02	4	HG,RE,BR	3	3p25	(chromosome 3p25) membrane protein
1900		0.05					glycine dehydrogenase
1910		0.01	2	HI,BR		22q11	TUP1-like enhancer of split gene 1
1934		0.01				14q24	C-1-tetrahydrofolate synthase cytoplasmic
1951	0.09	0.01	1	HI			adenylcyclase associated protein CAP2
1952		0.02			2		DNAJ protein homolog HSJ1
1953		0.01				4	glutamine PRPP amidotransferase
1955		0.05	1	BR	4		smg GDS
1956		0.01				18q21.33	Bcl-2
1969		0.02	2	FB,BR		2q34	fibronectin
1970		0.01	1	TE		1q31	laminin gamma-1
1980		0.01	3	FB,BR		5q13	microtubule associated protein 1B
1998	0.09	0.01	1	BR			initiation factor 4 gamma
2001		0.01				3p21.2-p21.1	glycine cleavage system T-protein
2002		0.01					myelin transcription factor 1
2003		0.01					integrin beta 5
2004		0.01					B70 antigen
2009		0.01	2	HI		17q21	glial fibrillary acidic protein
2012		0.01				4	STS4-666
2016		0.01				1p13.3	muscle glutathione S-transferase
2030	0.04	0.02			3	3p22-p23	mismatch repair protein homolog
2031		0.10	5	PI,BR		1q42-q43	angiotensinogen
2043		0.01					steroid hormone receptor ERR1
2051		0.01	1	PI		17q25	protein disulfide isomerase
2060		0.02	3	HG,BR			randomly sampled cDNA D21260
2065		0.01	1	BR		22	14-3-3 protein eta
2077	0.04	0.03	1	HG	6		HSP70
2088	0.04	0.01	1	HL	6		HLA-B-associated transcript 3
2096		0.01	1	CM		11p15	tumor suppressor HTS1
2103		0.01				9q33-q34	tenascin
2110		0.01	1	BR		Xp21.1	RP3
2118		0.02	2	FB,BR			Lupus Ku autoantigen protein p86
2124		0.01	1	BR			lymphocyte clathrin light chain B
2125	0.13	0.01	6	HG,FB,BR,RH			acidic ribosomal phosphoprotein P0
2133	0.04	0.01	2	BR	7	7p13-p11.2	2-oxoglutarate dehydrogenase E1
2135		0.01					activator 1 37kd subunit
2137		0.01				19p13.3	ubiquitin conjugating enzyme
2143	0.04	0.01	1	CM		1	calpain 2 large subunit
2146		0.01	1	BR		7q35	aldose reductase
2147	0.09	0.02	1	PI	1+4	1q22-q25	sodium-potassium ATPase beta-1
2154		0.01					protein disulfide isomerase-related protein ERP72
2186		0.01				5q22-q32	monocyte differenciation antigen CD14
2202		0.01					myoblast cell surface antigen 24.1D5
2204		0.01	1	BR		19q13.1	receptor tyrosine kinase UFO
2206		0.02					IL2 receptor alpha-chain kappa B binding protein
2211		0.02	2	HG,BR		17q22	hepatic leukemia factor
2222		0.02				8q21.3-q22.1	calbindin 1
2223		0.05				5q13.3-q14	3-hydroxy-3-methylglutaryl-CoA reductase
2227		0.01					FKBP-rapamycin associated protein FRAP
2230		0.01	1	HI		4q26-q28	annexin 5
2232		0.02				6p21.3	HLA-B-associated transcript 2
2242		0.01				3q21-q25	ribophorin 1
2249		0.01	1	HA			randomly sampled cDNA D14664
2258		0.01	2	BR			DAP-kinase
2259	0.04	0.03	5	PI,BR,FA	5	5q	peptidyl-glycin alpha-amidating monooxygenase
2270	0.09	0.03	1	BR	10		ret fused gene RFG (RET/PTC3)
2280		0.02	1	BR			growth factor receptor-bound protein 2
2289		0.01					L-serine dehydratase
2290		0.05					cadherin 8
2295		0.01	1	HL		19p13.3-p13.2	DNA (cytosine-5)-methyl-transferase
2296		0.02	4	BR,TE			leucine-rich 130kD protein
2300	0.09	0.03	3	FB,BR	16	16p12	mitochondrial ubiquinol-cytochrome C reductase core protein 2
2319		0.01				1q21.1	high affinity immunoglobulin gamma FC receptor 1
2327		0.01			9	9q33	plasma gelsolin
2330		0.01					protein phosphatase PP-2A 130kd regulatory subunit
2344	0.13	0.05	2	CM,PI	13		small ribonucleoprotein U6
2363		0.01				14q24	TGF beta-3
2369		0.01				16p	DNA-3-methyladenine glycosidase
2373		0.01				4q11-q21	osteopontin
2376		0.01				Xq24-q25	STS sWXD595
2382	0.04	0.01			3		cellular nucleic acid binding protein (CNBP)
2391		0.01	1	BR			prolylcarboxypeptidase
2399		0.01	1	BR			paraneoplastic cerebellar degeneration-associated antigen
2412		0.01	9	CM,BR		5q21	polyposis locus protein 1
2416		0.01				7p	hexokinase D
2439		0.01	1	FB		16q13-q21	DNA-directed RNA polymerase II 33kd polypeptide
2440	0.04	0.01			13	13q34	collagen IV alpha 1
2445		0.02				2	integrin alpha 6
2456		0.01				16q22.1	phophatidylcholine-sterol acyltransferase
2460		0.01					Kruppel-like zinc finger HKR-T1
2476		0.07	4	AE,HI,BR		5q31-q33	osteonectin
2487		0.01	2	FB,FLS		3p21.3-p21.2	inosine-5'-monophosphate dehydrogenase 2
2499		0.01				10q21.1-q21.2	annexin VII
2506		0.05	6	CM,FB,HI,BR		21q21.2	Alzheimer's disease amyloid A4 protein
2527	0.31	0.02	1	BR	12	1cen-q32	muscle 6-phosphofructokinase
2536		0.02	1	BR			arginine-rich nuclear protein
2540		0.01				1q22-q24	Ski oncogene
2543	0.04	0.02			21		papillomavirus E6 oncogenic protein associated protein
2554		0.06	6	BR	17	17p13.3	guanine nucleotide regulatory protein ABR
2586		0.01				7q35-q36	anion exchange protein 2
2588		0.01					male enhanced antigen (Mea)
2593		0.01	1	FB			glypican
2602		0.01	1	BR		7q11.2	NADPH-cytochrome P450 reductase
2603		0.02				1	replication protein A 32kd subunit (RP-A)
2611		0.01					basic transcription factor P62
2616	0.13	0.05	1	BR	4		CDC2-related protein kinase (PISSLRE)
2619		0.02	3	HI,BR		19p13.3	basigin
2625		0.01	3	CM,PI,BR		1p34	LAR protein
2632		0.01	1	BR			T-cell proteine-tyrosine phosphatase
2647		0.02	2	BS,BR		5	STS D5S478 containing CA repeat
2661		0.01					putative receptor protein
2667	0.09	0.02			X+7+12+14		initiation factor 4B
2674		0.01				17p12-p11.2	peripheral myelin protein 22
2676		0.01				7q36	homeobox protein EN-2
2690		0.01	1	HG			bone morphogenetic protein 2B
2691		0.01					H326 mRNA
2692	0.04	0.02	3	HL,PI,HA	17	17q23-q24	cAMP-dependent protein-kinase 1 alpha regulatory
2697		0.02					U2 small ribonucleoprotein B"
2698		0.01	1	BR		6p21.3	G9a

Appendix 1A. Known gene transcripts (*continued*)

2709		0.01	4	BR,PI	10		43 kd inositol polyphosphate 5-phosphatase
2719	0.04				2+10	20cen-q13.1	S adenosyl-homocysteine hydrolase
2726		0.02	2	HG,BR			plasma glutathione peroxidase
2728	0.13				6	6q22.1	cardiac phospholamban
2730	0.04					13q34	collagen IV alpha 2
2731	0.09				1	1p13	AMP deaminase 1
2732	1.01		1	PI	19	19	calpain small subunit
2738	0.04				4		T-cell specific tyrosine kinase
2741		0.02				12q11-q12	contactin 1
2747		0.05			10		randomly sampled cDNA D31883
2755		0.01					tyrosine kinase receptor HEK2
2769		0.02				Xq24	lysosomal-associated membrane glycoprotein 2
2781		0.01				Xq21.1-q21.3	Znf6
2798		0.02	4	HL,BR,CC		1p31-p32	epidermal growth factor receptor substrate
2839		0.01	1	HA			skeletal muscle 165kd protein
2861		0.03	1	BR		D26350	inositol 1,4,5-triphosphate type 2 receptor UNK
2862		0.01					rolipram-sensitive 3'-5' phosphodiesterase
2872		0.01				19p13.2-p13.3	N-acetylglucosamine receptor
2888	0.04	0.03	2	FB,BR			cytochrome C
2891		0.01				17q24-q25	adrenoxine oxydoreductase (NADPH)
2902		0.01			4	4	STS 4-1312
2915		0.01	1	HI		17q21	prohibitin
2933		0.01				7q21	multidrug resistance protein 1
2937		0.01					adrenomedullin
2945		0.01	1	HL			transformation-sensitive protein IEF SSP 3521
2950	0.04	0.02	3	BR	4	4p16.3	erythrocyte adducin alpha subunit
2956		0.05	1	BR	17		clone KDB1.2 with (CAC)n/(GTG)n repeat
2957		0.01				4q25-q34.3	glutamate receptor 2
2964		0.06				4p16.3	neuron-specific protein
2965	5.10		1	FB		1q42.1-q42.3	adult skeletal muscle alpha-actin
2967		0.02			18	18p11.2	receptor-like protein-tyrosine phosphatase MU
2972	0.22						fast skeletal muscle troponin C
2973		0.05	1	BR	6		malate oxidoreductase
2979		0.01					Kruppel-type zinc finger
2984	0.13	0.02	1	CC	19		natural killer cell enhancing factor B
2993	0.09	0.03	5	HG,HI,BR	4+6+8	6p21	HSP-90 beta
3000		0.02					hMCM2
3006	0.13	0.07	7	RE,BR,PI,FI			voltage-dependent anion channel isoform 2
3013		0.16	2	FB,BR	5	5q21-q22	adenomatous polyposis coli protein
3015		0.06	3	BR,PI	4		synuclein
3018	0.04	0.08	1	PI	4+7	4	randomly sampled cDNA D25274
3025		0.02					ndf19 neu differenciation factor
3028	0.04	0.06	4	HL,CM,HI,PI	6	6p21.3	HLA-B
3030		0.02	3	PI,BR			ADP-ribosylation factor 4
3051	0.18	0.02			4	4q35	skeletal muscle ADP,ATP carrier protein
3060	0.04	0.09	12	HG,HL,FB,HI,PI,BR,FI	+3+4+6+14+2		ribosomal protein L10
3061		0.03	2	HI,RH		Xq28	CDM
3067	0.09	0.05	3	HL,HI	2+17		ubiquitin (3 repeats)
3071		0.05					randomly sampled cDNA D29643
3078		0.02					randomly sampled cDNA D29958
3099		0.02	2	BR		10q24-q26	alpha-2A adrenergic receptor
3108	0.09	0.02	5	FB,BC,BO	9+11	11p15.4	hemoglobin beta
3125		0.02				Xq21.3-q22	BTK region clone ftp-3
3145	0.04	0.03	2	HL,FB	2	2q12-qter	nucleolin
3147		0.05			7		beta-2 chimaerin
3154		0.03			19		RSRFR2
3160		0.02				12q13	transmembrane 4 superfamily protein SAS
3161		0.02	2	BR			randomly sampled cDNA D21262
3164		0.03	11	HI,BR		17	fructose biphosphate aldolase C
3167		0.02	1	FB			S-adenosylmethionine synthetase gamma isoform
3170		0.06	2	BR		6p21.2-p21.1	lactoylgluthatione lyase
3174		0.02				2cen-q13	Ras-related protein RAL-B
3176		0.02				14q32.33	Ig Mu chain C
3183		0.03	2	HL,BR	1	6p21.3	HLA-E
3187		0.02	1	HL		12p13.3	glucose transporter type 3
3190		0.03	1	BR		17q11.2	neurofibromin
3194	0.04	0.01	1	BR	9+17		NRF1 protein
3195	3.38		1	FB	19	19q13.3	creatine kinase M
3198		0.02					extracellular signal regulated kinase 3 (ERK3)
3207		0.03					focal adhesion kinase
3212		0.03	4	BR			glutamate transporter D26443
3215		0.02				5p11-p12	zinc finger ZNF131
3219	0.35				2	2q33-q34	skeletal muscle myosin light chain 1 and 3
3230		0.05	6	FB,PI		4p14	ubiquitin carboxyl-terminal hydrolase isozyme L1
3237	1.19					11q13.1	muscle glycogen phosphorylase
3243		0.08	2	BR		12	tubulin alpha 4
3248		0.03	2	HG,BR		3p21.2-p21.1	inter-alpha-trypsin inhibitor complex component II
3252		0.01				19q12	RNA polymerase II 14.5kda subunit
3259		0.02				12q13	cell division protein kinase 4
3272	0.04					19	muscle cytochrome C oxidase 7A
3274	0.04	0.03	6	HG,PI,BR			HSP-90 alpha
3275		0.02	1	BR	2	2q33-q34	IGF binding protein 5
3284		0.03	3	HI,BR	4	4	carboxypeptidase H
3306		0.03				18p11.1-q11.2	alpha-2-plasmin inhibitor
3316		0.02				1q41-q43	1D-myo-inositol-triphosphate 3-kinase B
3321		0.02	5	BR,FB,HI		6	T-complex-associated testes expressed 1 homolog
3329	0.09		10	HG,FLS			ribosomal protein L23A
3330		0.03	1	PI	1	1q21-q23	histone H2B.1
3338	2.20		5	HE,HA	2	2q31-q32	titin
3351	0.04		1	HL	1		phospholipase C alpha
3352	0.09		2	FB,BO		16p13.3	hemoglobin alpha
3353		0.03	1	HI			neuron specific calcium-binding protein hippocalcin
3358		0.02	1	BR	6	6q21-qter	neuromedin-B receptor
3359	0.04		1	BR			leukemia virus receptor 2
3360	0.09	0.02	2	FB,BR	7		randomly sampled cDNA D26068
3364		0.02	3	FB,PI,BR			randomly sampled cDNA D13641
3367		0.02	4	FB,PI		3p25	XP-C repair complementing protein (p58/HHR23B)
3368	0.13				3	3p21.3-p21.2	slow-twich muscle B/ ventricular myosin light chain 1
3372		0.03	4	BR,FB			protocadherin 43
3374		0.01	2	BR		10p15	dihydrodiol dehydrogenase
3377	0.62	0.01	6	HG,PI		6q25	manganese superoxide dismutase
3383		0.05	3	HL,FLS,BR		15q21-q22.2	beta 2 microglobulin
3392	0.04					14q32.33	Rac protein-kinase alpha
3393		0.03	1	BR		Xq13	X-linked PEST-containing transporter
3402	0.13	0.01	2	HL,BR	X		TNF receptor
3405	0.09	0.02	5	HG,FB,PI,CC			calmodulin
3406		0.02	1	BR		8q13	corticoliberin
3410	0.53	0.02			12	12	skeletal muscle slow myosin-binding protein C
3411	0.35					15q2	alpha-tropomyosin
3420	0.44					17pter-p12	beta-enolase
3423	0.13	0.01				15q23-q25	electron transfer flavoprotein alpha-subunit
3426	0.09	0.01	8	HL,FB,BR,PI	X+7+11		HSC70

3427	0.40		3	BR,PI,FB	16		DNA-binding protein A
3430	0.92	0.05	19	BR,HI,FB,PI		12p13	glyceraldehyde 3-phosphate dehydrogenase
3436	0.04				8+13		histone H2A.Z
3438	0.22		1	HL	8+18		non-sarcomeric myosin regulatory light chain 2
3445	0.09				3		1,4 alpha-glucan branching enzyme
3450	0.09		1	HG		12p13	triose phosphate isomerase
3462	0.22	0.06	63	HG,BR,FB,PI,FK,FA			elongation factor 1 alpha
3464	0.31	0.05	4	FB,BR	8		muscle-brain cAMP dependent protein kinase inhibitor (PKI-alpha)
3465	0.09						Kruppel-like zinc finger
3467	0.09		1	PI	8	9q34.1	adenylate kinase isoenzyme 1
3470	0.57		6	CM,FB,HI,BR	3+10+16	16p22.1	fructose 1,6 diphosphate aldolase A
3474	0.04						cytochrome C odidase 5A
3483	0.26		2	BR,HE	11	11q22.3-q23.1	alpha B-crystallin
3485	0.31				1	1q32	slow skeletal muscle troponin I
3512	0.13					Xq12-q13	muscle phosphorylase kinase alpha chain
3514	0.66	0.08	9	HI,FB,BR	12	12q23-q24.1	endoplasmic reticulum calcium ATPase class 2
3523	0.13		4	HL,HI,FB	5		guanine nucleotide-binding protein beta subunit-like protein 12.3
3524	0.09					17	phenylethanolamine N-methyl transferase
3543	0.09		6	HG,FB,TE,FI		5	ribosomal protein S20
3546	0.04				5	5p14-p12	growth hormone receptor
3552	0.22		2	HI,BR	3	3p21	mitochondrial ubiquinol cytochrome C reductase core protein 1
3553	0.04	0.06	5	HI,BR	9		TL7 cDNA from LNCaP cell-line
3559	0.04	0.01			1	1q41-q42	NAD+ ADP ribosyltransferase
3560	0.66				2	2q35	desmin
3565	0.09		1	BR	18		skeletal muscle 190kD titin-associated protein
3566	0.04						muscle acylphosphatase
3567	0.09				19		tristetraproline
3574	0.04		2	FB,BR	1		polyadenylate binding protein 2
3583	0.13		1	FB	3+9+15		ribosomal protein L4
3593	0.09	0.01	3	HL,FB,BR			laminin receptor
3603	0.18		5	HL,FB,HI,PI	X		OM protein
3607	0.26		2	FB,HA	1+4	1q42-q43	skeletal muscle alpha2-actinin
3614	1.85				22	22q12	myoglobin
3617	0.18				12	12q24.3	restin
3625	0.04	0.01	1	BR			mitochondrial peptidylproline cis-trans isomerase
3628	0.53		15	HG,FLS,PI		13q14	translationally controlled tumor protein
3632	0.04	0.02			X	Xp21.3-p21.1	dystrophin
3645	0.09						ribosomal protein L6
3648	0.04				9		thymosin beta 4
3650	0.09		1	PL			non-muscle/smooth muscle myosin alkali light chain
3660	0.09		2	PI		1q21	pyruvate kinase
3665		0.07	1	BR		1p36-p31.2	guanine nucleotide-binding protein G(i)/G(s)/G(t) beta subunit 1
3671		0.02	4	HL,FB,BR			protein phosphatase PP2A, 65kd regulatory subunit, alpha isoform
3672		0.06	1	BR	9		antigen OX40 homologue
3685	0.04						heat shock protein HSP72 homolog
3686	0.09	0.05	2	BR,HA	2		UDP-glucose pyrophosphorylase
3688		0.02	1	HL		12	cation-dependent mannose 6-phosphate receptor
3689		0.03					IEF 7442
3690		0.03					transaldolase
3692	0.09		4	HG,HL,TH,FI			ribosomal protein S13
3695		0.16	4	BR,HA		Xq21.33-q22	myelin proteolipid protein
3698	0.18		1	LF			ribosomal protein L41
3700		0.02	3	HG,HI		11p15.1-p14.1	lactate dehydrogenase M chain
3709		0.03				12q13	Erb-B3 receptor protein-tyrosine kinase
3724		0.02	2	BR			milk fat globule protein HMFG
3727	0.53					12q23-q24.3	cardiac/ventricular myosin light chain 2
3736		0.02	2	FB,BR		5q12-q13	phosphatidylinositol 3-kinase regulatory alpha subunit
3741		0.08	6	FB,HI,BR			synaptosomal associated protein 25 isoform 2
3742	0.31		1	HL	12	12p13-qter	mitochondrial ATP synthase beta chain
3749		0.02					moesin
3752		0.02	2	CM,BR			TALLA-1
3768		0.03			10		homeobox protein PHOX1
3779		0.02	4	HI,BR			triplet-repeat containing cDNA CCA12
3788		0.03	2	BR		14q24.3	c-Fos
3791		0.02	1	BR		15q13-q14	neuroendocrine protein 7B2
3796		0.02				17q	tre oncogene (deubiquitinase)
3803		0.05	1	BR	X+1	Xq28	55kd erythrocyte membrane protein
3807		0.02	1	HA	7		ELP-1
3813		0.02					cellular proto-oncogene (c-mer)
3826		0.06	2	BR,FI	8		Rab-2
3831		0.14	41	HL,BR,HI,FB,PI		2q	tubulin alpha 1
3846	0.26				8	8q13-q22	carbonic anhydrase 3
3856		0.02				6q22-q27	ezrin
3859	0.26		3	HG	X	17p12-q11	ribosomal protein L19
3868		0.02	1	HL	6		serine-kinase
3869		0.05	1	BC		Xq27.3-q28	iduronate 2-sulfatase
3886		0.02	8	PI,BR,FLS,BC,FB		14q24-q31	calmodulin 1
3889	0.66					19q13.4	slow skeletal muscle troponin T
3890		0.03	1	PL	X		tubulin beta-2
3894		0.02	1	FB			glutamate decarboxylase GAD67
3895	0.04	0.01	2	FB	5	5q31-q33	ribosomal protein S14
3904		0.02					E2F-related transcription factor
3905		0.03	1	BR			FGF
3906	0.31				9		beta-tropomyosin
3907		0.02					randomly sampled cDNA D26445
3919		0.05				5	CDK-activating kinase
3920		0.02				2p22-p21	MutS homolog MSH2
3928		0.02	1	BR		1q22-23	MHC-related antigen CD1 R3
3929		0.03				4q21-q25	alcohol dehydrogenase class III chi chain
3948		0.02					randomly sampled cDNA D14658
3949		0.02					extracellular proteinase inhibitor homologue, HE4gene
3969		0.06	4	CM,FB,BR	9+13	Xp22.32/Yp11.3	ADP-ATP translocase 3
3971		0.02				4	randomly sampled cDNA D25248
3973	0.09		2	HI,FB		19	AES1
3975		0.07	5	BR,PI			hnRNP C
3984		0.02				2	thymidylate kinase
3986		0.02	2	CM,HE	20	20p11.2	brain glycogen phosphorylase
3987		0.05	1	BR			smooth muscle myosin heavy chain isoform SMemb
4001		0.03					interferon-induced guanylate binding protein 1
4004		0.03	1	BR	4+6		GTP-binding protein RAN
4006		0.03	1	FB	3	3q11-q13	MRC ox-2 antigene
4010		0.08	8	HI,BR,FLS	10		cytochrome B561
4018	0.04	0.05	2	BR	8		phosphatidylserine synthase
4032	0.04	0.01				3q27-qter	apolipoprotein D
4037		0.03	1	HG		9q21.1	cytosolic aldehyde dehydrogenase
4041		0.03	1	HG		11p13	CD59 glycoprotein
4048		0.02				19q13.2	DNA repair protein XRCC1
4051	0.09	0.03			2+12		mitochondrial long-chain enoyl-CoA hydratase
4057		0.07	2	HL,FI		22pter-q11.2	vacuolar ATP synthase subunit E
4064		0.02	3	FB,BR		7pter-p14	glutamine cyclotransferase
4066	0.04	0.03	1	BR	4	4q25-q27	brain ankyrin
4069		0.02	1	BR			paraneoplastic encephalomyelitis antigen HUD

4073		0.02					putative serine/threonine-protein kinase p78
4077	0.09		4	HG			yeast ribosomal protein S28 homologue
4081		0.02			17		randomly sampled cDNA D30756
4086		0.03	2	BR		1q32.1	MAP kinase phosphatase 1
4089	0.09	0.01	4	HL,EK,HI	X	Xq13.3	phosphoglycerate kinase 1
4104	0.04	0.01	2	HI,BR	14		nucleosome assembly protein
4105		0.09	3	HL,FB,BR		1p36.1-p35	stathmin
4106		0.05	1	BR	4		activator 1 140kd subunit
4119	0.04	0.05	2	BR	10	10q24.1-q25.1	cytosolic aspartate aminotransferase
4134		0.03	3	HG,HI,PL			peptidylproline cis-trans isomerase A
4137		0.03	1				MAC30
4138	0.09	0.03			9		isoleucyl-tRNA synthetase
4149	0.04	0.01	3	CM,PI,CC			clone tec10 mRNA
4156		0.10	3	HI,FB,BR		14q21-q22	neuroendocrine-specific protein
4157		0.03					TAFII70
4162	0.09	0.03	3	HL,HI	2	2q37	HDL binding protein
4164		0.02			20		translational initiation factor 2 beta subunit (e1F-2-beta)
4167		0.02	1	FB			extracellular signal regulated kinase 2 (ERK2)
4206		0.03	1	BR		19q13.2	poliovirus receptor
4218		0.01				17q24-qter	nuclear p68 protein
4220	0.13		1	CM		1p22.1	phosphoglucomutase 1
4229	0.04	0.02	2	HL,BC		1p36-p21	non-specific lipid-transfer protein
4238		0.02	6	PL,LI			placental thrombin inhibitor
4252		0.05	1	HL	3	3q26.2-qter	transferrin receptor
4256	0.13		4	HL,FB		6p21.3	ribosomal protein S18
4262		0.03			17		endometrial antigen
4265		0.02	1	FB			gal beta1,3(4)GlcNAc alpha2,3-sialyltransferase
4270		0.02				9q34.1	c-ABL
4275	0.04	0.02	1	HG			ribosomal protein S5-like
4290		0.03				6q21-q22.2	myristoylated alanine-rich C-kinase substrate
4297		0.02	1	BR			14-3-3 protein zeta
4299	0.04	0.01				9q31	protein tyrosine phosphatase PTP-H1
4304		0.03	1	HI		Xq13-q26	fibroblast ADP,ATP carrier protein
4313		0.02	1	BR		4p16.3	alpha-2-macroglobulin receptor-associated protein
4314		0.03	2	BR		1p21	alpha-amylase
4315		0.02	1	BR		16	extracellular signal regulated kinase 1 (ERK1)
4320	0.04	0.01	1	HI			mitochondrial ATP synthase B chain
4323		0.02	2	BR			leukophysin
4328		0.03				5	COUP transcription factor
4331	0.04	0.01	2	CM,BR	22	12p	mitochondrial NADH-ubiquinone oxidoreductase 39kD subunit
4353	0.04	0.02	2	FB,BR	2	2p21	non-erythrocytic beta-spectrin 1
4372		0.03				1q32	complement decay-accelerating factor
4376		0.03	1	FB			fibulin
4385		0.05					transcription initiation factor IIB
4403		0.03	5	HI,BR			beta adaptin
4404		0.02				4	lymphocyte differentiation antigen CD38
4408	0.04	0.01	1	BR	1		glycogenin
4466		0.02			10	10q25.2-q26.3	uroporphyrinogen III synthase
4472	0.04	0.01				Xq28	creatine transporter
4475	0.04		1	HL		16q24.1	cytochrome C oxidase 4
4479	0.04						glia maturation factor beta
4481	0.04		1	BR	17	17q21	2',3' cyclic nucleotide 3' phosphodiesterase
4489	0.04		2	BR	5		peptide binding protein 74
4497	0.04					4	SAICAR synthetase
4519	0.04				22	22q13.1	adenylosuccinate lyase
4527	0.04						quinone oxidoreductase
4531	0.04	0.01	1	HG	12	12p12.1	transforming protein P21/K-Ras-2
4532	0.04				2	2q31-q32	nebulin
4547	0.04	0.01	1	BR	8	6	microsatellite AFM135xh2
4598	0.04				20	20q12-q13.1	phospholipase C gamma 1
4612	0.04				2	2q14-q21	Rab-6
4693	0.04					7	microsatellite AFM036xg5
4694	0.04		2	FB,PI	10	10p13	BMI-1
4803	0.13	0.01	6	HG,PI,BR	6		ubiquitin
4814		0.01	3	HL,BR			mitochondrial 2-oxoglutarate/malate carrier protein
4816		0.01			2	2p22-p21	guanine nucleotide exchange factor SOS-1
4819		0.01				22q12.2	merlin
4821		0.01				12p13.3-p13.2	von willebrand factor
4837		0.01	1	HL		12	uracil-DNA glycosylase 1
4838		0.01				1p34.1-p32	ribosomal protein S8
4855		0.01				6p21.3	HLA-DR alpha
4863		0.01	1	FB		Xp22.32	steryl sulfatase
4869		0.02			1	1p13	sodium-potassium ATPase alpha-1
4890		0.01					low-affinity IgG Fc receptor
4906		0.01				3p22-p23	peroxisomal 3-ketoacyl-CoA thiolase
4919		0.01					HCF1 gene related (VP16 protein)
4921		0.01	1	BR		8q21.1	peroxysome assembly factor 1
4936		0.01	16	FB,BR			tubulin beta-4
4968		0.01	1	FLS		1q32	complement factor H
4985	0.04	0.01					ras-related GTP-binding protein
4990		0.01				Xp22.3-p22.2	ribose-phosphate pyrophosphokinase 2
4995		0.01	1	BR			mitochondrial transcription factor 1
4997		0.01				21q	MAP-kinase p49
4998		0.01			1		probable G protein-coupled receptor EDG-1
5013		0.01			17		pigment epihelium-differenciation factor
5018		0.01					nicotinamide N-methyltransferase
5029		0.01			8	8q13-q22.1	carbonic anhydrase 2
5035		0.01			3	3p25	biotinidase
5042		0.01	1	HL	17	17q21-q22	DNA topoisomerase II, alpha isozyme
5044		0.01	1	HL	19	19q13	urokinase plasminogen activator surface receptor CD87
5048		0.01	9	FB,BR	4	4p	randomly sampled cDNA D25269
5053		0.01	2	HG			proteasome C9
5057		0.01			2	2q12-q21	acyl-CoA-binding protein
5090		0.02	1	HI	X	10q23.3	glutamate dehydrogenase
5096		0.01	2	BR			mitochondrial ATP synthase gamma chain
5099		0.01				11q13	rod outer segment membrane protein 1
5109		0.01				1	regulator of chromosome condensation RCC1
5112		0.01	1	BR		7	dipeptidyl aminopeptidase like protein
5136		0.01				17q12-q21.1	insulin-like growth factor binding protein 4
5149		0.01	2	HI,PI	8		randomly sampled cDNA D31887
5158		0.02	1	BR	3	3p21.1	chromosome 3p21.1 gene sequence
5161		0.01			7	7q21.3-q22.1	collagen alpha 2 (I)
5163		0.01	2	FB	12		CCAAT displacement protein CASP
5196		0.01	1	BR			Sec4-like protein Rar
5197		0.01					gluthation S-transferase theta 1
5215		0.01					E4TF1-53
5219		0.01				16q22.1	proteasome-like subunit MECL-1
5224		0.01				10q25.3-q26	fibroblast growth factor receptor 2
5236		0.01				9	homeobox protein PXB3
5258		0.01				21q22.3	interferon-regulated resistance GTP-binding protein MXA
5267		0.01	2	HL,BR		20	small nuclear ribonucleoprotein binding protein B
5274		0.01				6p21.3	MHC class III HLA-RP1

Appendix 1A: Known gene transcripts (*continued*)

5276		0.01			1		randomly sampled cDNA D31891
5294		0.01				15q11.2-q12	gamma-aminobutyric-acid receptor beta-3 subunit
5319		0.01					helix-loop-helix basic phosphoprotein
5322		0.01			10	10q24	nuclear factor NF-KAPPA-B P49
5364		0.01	3	HI,BR			integral membrane protein E16
5368		0.01				4p	randomly sampled cDNA D25250
5369		0.01	2	HG,FB		22	X-box binding protein
5370		0.01					2-oxoisovalerate dehydrogenase
5386		0.01				11q11-q13.1	pyruvate carboxylase
5412		0.01					myo-inositol-1 (or 4) monophosphatase
5414		0.01					diacylglycerol kinase
5426		0.02	1	BR	9		phosphotyrosyl phophatase activator
5439		0.01	3	PI,BR			T-plastin
5442		0.01	2	HG			transcription factor BTF3b
5449		0.01	2	HL,BR		1p36.3-p36.2	procollagen-lysine 2-oxoglutarate 5-dioxygenase
5451		0.01	3	HG,BR			TL35 mRNA from LNCaP cell line
5465		0.01	2	HG,BR			proliferation-associated protein PAG
5493		0.01				Xp11.4-p11.3	monoamine oxidase A
5498		0.01			1		clone A9A2BRB7 containing (CAC)n/(GTG)n repeat
5506		0.01					zinc finger protein IA-1
5515		0.01	1	HL		4	AF-4 mRNA
5534		0.01	2	BR			death associated protein DAP-1
5549		0.01	5	HL,BR,FB,FI		9q34	Set gene
5570		0.01				8p22	lipoprotein lipase
5580		0.01	5	CM,BR,FLS		12q24.2-qter	endoplasmin
5597		0.01	2	CM,BR		1p22.1-qter	succinate dehydrogenase (ubiquinone) flavoprotein subunit
5601		0.01				11p15.4	sphingomyeline phosphodiesterase
5602		0.01					transforming growth factor induced protein
5610		0.01	1	BR		Xp22.3	GS1 protein of unknown function
5623		0.01	1	BR	10		CpG clone A5
5629		0.01	2	FB,BR		14q	protein farnesyltransferase beta subunit
5652		0.01	3	HL,TE			protein-tyrosine phosphatase G1
5668		0.01	1	CM		20pter-p12	proliferating cell nuclear antigen
5670		0.01	3	HG,FB,BR			transformation-related protein
5674		0.01				15q25-q26	furin
5682		0.01	1	FB		3q13.1-q13.2	neuromodulin
5687		0.01				13q32	DNA repair protein complementing XP-G cells
5694		0.01	1	HL			9G8 splicing factor
5696		0.01					randomly sampled cDNA D13634
5697		0.01				11q13.5-q14.1	adrenodoxin
5701		0.01					glycosyl-phosphatidylinositol anchor
5718		0.01				6q23-q24	interferon-gamma receptor alpha chain
5727		0.01					tyrosine phosphatase (IA-2/PTP)
5742		0.01	3	BR			GP36b glycoprotein
5743		0.01	2	BR			B-cell activation gene BL34
5751		0.01	1	TE		1p22-p21	tissue factor precursor
5770		0.01			15		randomly sampled cDNA D29956
5774		0.01	2	HL,FB		3p14.3	transketolase
5778		0.01					integrin alpha 3 chain
5780		0.01				3p21.3-p21.2	acylamino-acid-releasing enzyme
5815		0.01			5		randomly sampled cDNA D31766
5846		0.01			8		Gem GTPase
5890		0.01	6	HL,BR		12	testis enhanced gene transcript
5905		0.02	1	FB	12	12q21-q23	plasma membrane calcium-transporting ATPase isoform 1B
5909	0.04					12	Myl-6
5924		0.01					transcription factor E2F-like protein
5926		0.01					ERC-55
5936	0.04	0.01	4	BR		9q34.1	non-erythroid alpha-spectrin
5938		0.02				7q31-q33	protein tyrosine phosphatase zeta
5946		0.01					TL25 mRNA from LNCaP cell line
5963		0.03	1	BR			giantin
5964		0.01					randomly sampled cDNA D14660
5972	0.04	0.01	2	FB,BR	X		serine/threonine-protein kinase PCTAIRE-1
5974		0.01	1	BR		16p13.3	polycystic kidney disease 1 protein
5989		0.02	2	BR,CC	12	12q14	Ras-related protein Rap-1B
5995		0.02	16	HI,BR		18q23	myelin basic protein
5998		0.01	3	HG,BR		12	mitochondrial phosphate carrier protein
6007		0.01				18q21.1-q21.31	ferrochelatase
6012		0.01	1	BR		15q14-q21	inositol 1,4,5-triphosphate 3-kinase
6013		0.01				19q13.1	myelin associated glycoprotein
6014	0.04	0.03	1	HL		22q11.21-q11.23	catechol-O-methyltransferase
6055		0.01					phospholipase C beta 2
6057		0.02					ETO
6060		0.02			3	3p24-p22	Rab-5A
6065		0.01					heat-shock factor 2
6103		0.02				3q21-q22	propionyl CoA carboxylase beta subunit
6111		0.01					vacuolar ATPase VA68
6151	0.04		1	HI		15	calpain P94 large subunit
6182	0.09						elongation factor 1 alpha 2
6186	0.04				11	11q22.3-q23.1	mitochondrial acetoacetyl-CoA thiolase
6195	0.18		2	BR	2+19	19q13.1	skeletal muscle ryanodine receptor
6197	0.09		6	HG,HL,PI,FB	1	11q13	ferritin heavy chain
6202	0.04		1	PI	19+20	19q13.3-q13.4	ferritin light chain
6210	0.04						slow skeletal muscle troponin C
6215	0.04	0.01	1	FB		11p15.5	cathepsin D
6217	0.18		1	BR	19	19pter-q12	elongation factor 2
6218	0.04				6	12	mevalonate kinase
6220	0.04						adipocyte fatty-acid binding protein
6227	0.04					9p21	ribosomal protein S6
6236	0.13		4	HL,BR	10	10q21-22	cerebroside sulfate activator protein
6242	0.09		3	HG,BR	16		mitochondrial ATP synthase alpha
6251	0.04						PM5
6254	0.04		2	HG,FB	11		acidic ribosomal phosphoprotein P2
6256	0.04				14		protein phosphatase 2C alpha
6261	0.04						LIF receptor
6262	0.04		1	FB		2q32	CREB2
6276	0.04						yeast Sec7 homologue
6286	0.04					11q23.1	dihydrolipoamide S-acetyl transferase
6287	0.04				2	2q14-q21	glycophorin C
6288	0.04				16		cellular adhesion regulatory molecule
6295	0.04		3	BR,FB	X+12	Xq13.1/Yp11.3	40 S ribosomal protein S4 X isoform
6301	0.09		1	HI	X+1+12+16	16q12-q22	mitochondrial aspartate aminotransferase
6309	0.04					2cen-q13	beta B inhibin
6311	0.04						ribosomal protein L37
6313	0.04				19	19q13.1	glucose-6-phosphate isomerase
6323	0.04		1	BR	10		sphingolipid activator protein 1 et 2
6326	0.04					2cen-q13	brain vacuolar ATP synthase subunit B
6330	0.04				3	3p21	dystroglycan DAG1
6333	0.04						cAMP-dependent protein-kinase 2-alpha regulatory chain
6334	0.04				4		lysosomal sialoglycoprotein
6336	0.04		2	FB,BR	13	12p12.2-p12.1	lactate dehydrogenase H chain
6337	0.04						ribosomal protein S7

Appendix 1A: Known gene transcripts (*continued*)

6364	0.04				1	1p11-qter	microsomal epoxide hydrolase
6368	0.04		2	HL,FI			hnRNP homolog
6373	0.04				3+9		ribosomal protein L30 (Yeast RPL30 homolog)
6377	0.04	0.02	2	HL,FB		3q22-q25	polyadenylate binding protein
6378	0.04		1	TE	5+6	6q25-q27	T-complex protein 1 alpha subunit
6379	0.04		1	FB	5	5q23-q31	ubiquitin conjugating enzyme E2-17kD
6384	0.04					6	GMP reductase
6396	0.09	0.03			6	6p21.3	HLA-A
6399	0.04				17	17pter-p11	skeletal muscle beta-myosin heavy chain
6403	0.04		3	BR,FLS			ribosomal protein S3A
6411	0.04					15	peptidylproline cis-trans isomerase B
6412	0.04		2	FB,PI			Y-box binding protein
6419	0.04					7q22	acetylcholinesterase
6437	0.04					6p21.3	HLA-C
6448	0.04					6p21.3	HLA-DR beta
6457	0.09						glycogen debranching enzyme
6458	0.04				X		ovarian granulosa cell 13.0kD protein HGR74
6463	0.04				20	1p36-p22	spermidine synthase
6480	0.04				4	4	long chain fatty acid CoA ligase
6481	0.04					14q11.2-q13	cardiac alpha-myosin heavy chain
6502	0.04				7	7p12	EGF receptor
6503	0.04		6	BR,PI,BO	11+14		putative ribosomal protein L13
6504	0.04				1		mitochondrial ubiquinol-cytochrome C reductase 11kD protein
6510	0.04						ribosomal protein L23
6513	0.04					Xp22.3	putative adhesion molecule
6514	0.04				22	22q12.1-q13.2	metalloproteinase 3 inhibitor
6517	0.04				20		retinoic acid-inducible transglutaminase homologue
6523	0.04		2	HL,HA	5	11q23.3	ribosomal protein S25
6524	0.09				5	5	sarcomeric mitochondrial creatine kinase
6526	0.04					19q13.1	cytochrome C oxidase 6B
6533	0.04					3q25.1-q25.2	neprilysin
6536	0.04				1	1p31	glucose transporter type 5
6538	0.04					11p15.5	transforming protein P21/H-Ras-1
6539	0.04				17	17p13.1	phosphoprotein p53
6541	0.04						randomly sampled cDNA D13642
6544	0.04		1	HL		19q13.3	electron transfer flavoprotein beta-subunit
6546	0.04		1	HI	X+15	15q24-q25	pyruvate kinase M2
6547	0.04				1+12		phosphatidyl-ethanolamine binding protein
6567	0.04		1	HI		11p13	catalase
6593	0.04						ORF M68864
6598	0.04				17		proteasome delta chain
6599	0.04					5q31-q33	HLA-DR antigens associated invariant chain
6606	0.04						calpain inhibitor
6611	0.04				4	4	STS 4-248
6620	0.04				20	20	major centromere autoantigen B
6621	0.04		6	HG,FB,FLS			ribosomal protein S3 homolog
6637	0.04		1	BR		12q13	hnRNP A1
6638	0.04						hnRNP A2
6653		0.01					nuclear respiratory factor-1
6682		0.01					randomly sampled cDNA D13643
6683		0.01					Id-2
6691		0.01					P-selectin ligand
6694		0.01	2	BR			randomly sampled cDNA D25217
6709		0.01			7	7p13-p12	IGF binding protein 3
6711		0.01			9+10	9q22.1-q22.2	cathepsin L
6718		0.01				6p21.3	complement C4
6733		0.01	1	BR	4	17q25	splicing factor SC35
6747		0.01					smooth muscle protein 22-alpha
6749		0.01			12+17		protein tyrosine kinase
6752		0.01					serine/threonine-protein kinase receptor R5
6759		0.01					coatomer beta' subunit
6781		0.01					deoxycytidylate deaminase
6787		0.01				2q37.3	collagen VI alpha 3
6823		0.01	1	HI			DNAJ protein homolog 1
6841		0.01				10q24	urokinase-type plasminogen activator
6845		0.01				6p21.3	retinoic acid receptor RXR-beta
6881		0.01			9		glutamate transporter U08989
6888		0.01					vacuolar ATP synthase subunit C
6901	0.04		3	HL,PI		17p11-qter	cytoskeletal gamma-actin
6903	0.04	0.03	18	HI,FB,BR,PI,PL		7p15-p12	cytoplasmic actin 1 (beta)
6916		0.05					randomly sampled cDNA D17793

Appendix 1B: Homolog gene transcripts

GENX	% Muscle	% Brain	Seq	Tissues	eSTS	Species	Gene product
1	0.09					Bovine	mitochondrial NADH-ubiquinone oxidoreductase 49kD subunit
214		0.08	7	FB,BR	8	Mouse	SCG10 gene
263		0.01	1	FLS	8	Pig	17 beta-estradiol dehydrogenase
297		0.03	2	BR		Mouse	selenium binding protein
311		0.01	1	BR	1	Hamster	monocarboxylate transporter 1
351		0.02				Pig	thimet oligopeptidase
352	0.53	0.01	2	BR,PI	1	Rat	UNR protein with unknown function
358		0.01				Rat	beta-alanine synthase
447		0.01				Rat	adenylate cyclase, olfactive type 3
473		0.05	14	FB,BR		Rat	beta tubulin T beta 15
541		0.01	7	BR,TE		Sheep	6-phosphogluconate dehydrogenase, decarboxylating
551		0.10	5	HI,FB,RE	5	Rat	postsynaptic density protein 95
562		0.01	1	HI		Rat	protein-tyrosine phosphatase striatum-enriched
566		0.03	1	BR		Bovine	vacuolar H-ATPase subunit D
621		0.02				Rat	hem-2
697		0.03	3	BR		Rat	150kd dynein-associated polypeptide
759	0.04	0.01			8	Bovine	mitochondrial NADH-ubiquinone oxidoreductase 13kD B subunit
790	0.09	0.03	5	HI,BR,HE	5+22	Pig	aconitate hydratase (aconitase)
827		0.01				Rat	hemin-sensitive initiation factor 2a kinase
907	0.13	0.05	1	BR	16	Rabbit	phosphorylase kinase beta chain
1078		0.07	2	FB,BR		Mouse	membrane glycoprotein M6B
1166		0.01				Mouse	KYBP
1169		0.01				Bovine	neurophilin
1180		0.01				Mouse	mdm-1
1222		0.01				Mouse	uridine kinase
1241		0.01	2	HL,BR		Hamster	arginyl tRNA synthetase
1275		0.01				Rat	N-heparan sulfate sulfotransferase
1290		0.01				Mouse	centrocyclin
1296		0.05	7	BR		Mouse	serum inducible kinase SNK
1367		0.01				Mouse	brain protein I47 (fragment)
1455		0.01				Rat	glutamate receptor subunit GluR6
1527		0.01				Chicken	probable G protein-coupled receptor 6H1
1577		0.03	1	BR		Rat	cytoplasmic dynein 74kd intermediate chain
1582		0.03	1	BR		Rabbit	cardiac ryanodine receptor
1610		0.02				Rat	long-chain-fatty-acid-coA ligase brain isozyme
1805		0.01	1	BR		Duck	malate oxidoreductase
1817		0.01				Rat	chloride channel protein 2
1855		0.01	3	BR		Bovine	gamma-COP
1873		0.01	2	BR		Rat	syntaxin A
1874		0.01	1	BR		Rat	L-arginine:glycineamidinotransferase
1878		0.01	1	FB		Rat	insulin-induced growth response protein
1899		0.01				Rat	brain sodium dependent glutamate/aspartate transporter
1914		0.01	2	BR,CC		Rabbit	lambda-crystallin
1915		0.01	1	HL		Mouse	H beta 58
1928		0.02	1	BR		Rat	NMDA receptor glutamate-binding subunit
1943		0.01	1	BR		Rat	protein phosphatase inhibitor 1
2024	0.04	0.01	1	BR		Rat	histone H2A.1
2028		0.01				Dog	Rab10
2041		0.01	1	BR		Bovine	inorganic pyrophosphatase
2049		0.02	6	HL,BR,FB		Rat	cytoplasmic dynein heavy chain
2095		0.01				Rat	CaBP1 calcium binding protein
2129		0.01				Bovine	phosphatidylinositol 3-kinase regulatory beta subunit
2148		0.01	1	FB		Mouse	T-cell antigen receptor alpha chain
2177		0.03				Mouse	NDPP-1
2302		0.01			1	Mouse	putative GTP-binding protein Mov10
2336		0.01				Rat	glycogen synthase kinase-3 beta
2346		0.03	4	HL,BR,HA	22	Mouse	brain protein E46
2386		0.01				Mouse	novel beta 1-4 galactosyltransferase
2396		0.01	1	FB	16	Mouse	glutamate (NMDA) receptor subunit epsilon 1
2401		0.02	1	BR	3	Rat	gamma-aminobutyric acid transporter
2434		0.01	2	BR		Rat	brain specific Na+-dependent inorganic phosphatecotransporter
2462		0.01	3	FB,BR		Mouse	amyloid-like protein 1
2473		0.01	1	BR		Rat	f-spondin
2480		0.01				Mouse	serotonin receptor 5HT7
2546	0.04	0.01	1	BR	2	Bovine	cAMP-regulated phosphoprotein (ARPP-19)
2558		0.06	7	BR,FB,BC		Bovine	p87 transporter-like protein
2684		0.01				Mouse	serine/threonine kinase (sak-a)
2708		0.01				Bovine	chlorine channel P64
2711		0.01				Bovine	NAD+-dependent isocitrate dehydrogenase
2740		0.01				Rat	major vault protein
2749		0.01	2	FLS,BR		Mouse	NGF-inducible protein TIS21
2784		0.01				Rat	synaptotagmin 3
2813		0.02	1	BR		Dog	mucin
2822		0.01				Mouse	mesoderm-specific unknown protein
2844		0.01				Bovine	actin-like protein
2899		0.01				Mouse	ARP-1 (COUP family of nuclear orphan receptors)
2917	0.04	0.02	9	FB,HI,BR	17	Rat	14-3-3 protein epsilon
2920		0.02	3	BR		Mouse	RNA/DNA binding protein RNPS1
2990	0.04	0.01	1	HA	5	Bovine	mitochondrial nicotinamide nucleotide transhydrogenase
2992		0.03			1	Rat	voltage-gated potassium channel
3097		0.02				Rat	sodium channel protein I
3111		0.29	15	BR,PI	9	Rat	neuronal olfactomedin-related ER localized protein
3131		0.02				Rat	potassium channel KShIIIA
3140		0.03	1	BR		Dog	Rab22
3152	0.13	0.02	5	FB,HI	7	Rat	14-3-3 protein gamma
3155		0.03				Rat	100kDa protein
3173		0.02	1	HL	3	Chicken	dynein light chain A
3193		0.03				Mouse	invasion inducing protein Tiam-1
3220		0.05	4	HL,FB,PI,BR		Rat	translation initiation factor 5
3221		0.03				Rat	vesicle associated membrane protein VAMP-2
3279	0.04	0.02	1	FB		Rat	nuclear factor 1
3287		0.02	1	HL		Mouse	glucosaminyl N-deacetylase / N-sulfotransferase
3357	0.09	0.01	2	HL,BR	1	Rat	proteasome RN3 subunit
3400		0.07	1	FB	19	Rat	glycogen syntase kinase-3 alpha
3403		0.07	1	BR	19	Mouse	synapse specific phosphoprotein F1.20
3412	0.04	0.01			3	Bovine	mitochondrial NADH-ubiquinone oxidoreductase SGDH subunit
3422	0.26		2	HL,PI	2	Mouse	hnRNP X
3429	0.18				6	Rabbit	junctional sarcoplasmic reticulum glycoprotein (triadin)
3455	0.04	0.01			1+10	Dog	signal recognition particle 9 protein (SRP9)
3460	0.04				14	Bovine	mitochondrial ATP synthase coupling factor B
3477	0.18					Rabbit	skeletal muscle myosin regulatory light chain 2 type 2
3487	0.09		2	HA		Bovine	mitochondrial NADH-ubiquinone oxidoreductase MLRQ subunit
3490	0.04	0.01	2	BR	11	Bovine	mitochondrial NADH-ubiquinone oxidoreductase 51kD subunit
3501	0.04		2	HG		Xenopus	ribosomal protein S27 homologue
3528	0.13		2	HL,TH	8	Rat	ribosomal protein L30
3549	0.13				1	Rabbit	skeletal muscle calsequestrin
3562	0.04	0.01	1	HG	2	Mouse	differentially expressed homolog of HepG2

Appendix 1B: Homolog gene transcripts (*continued*)

3570	0.09					Rat	fast skeletal muscle troponin T
3585	0.09	0.01	2	TE	2	Xenopus	double stranded RNA binding protein A
3620	0.04				5+7	Bovine	mitochondrial NADH-ubiquinone oxidoreductase B22 subunit
3647	0.09					Rat	neonatal calcium ATPase
3669		0.02	1	BR		Mouse	morphogenic protein Fif
3715		0.02			20	Bovine	1-phosphatidylinositol 4,5-biphosphate phosphodiesterase beta 1
3720		0.08	9	HI,FB,BR	9	Bovine	UNC18 homolog
3723		0.03	8	BR		Rat	integrase-like FE65
3732		0.02			1+3	Rat	TRAP-complex gamma subunit
3737		0.02				Mouse	potassium channel subunit (m-eag)
3751		0.02				Bovine	neurexin III beta
3804		0.02			12	Rat	initiation factor Eif-2
3876		0.02			4	Rat	calmodulin-dependent protein kinase 2 delta
3900		0.03	1	FB		Mouse	membrane glycoprotein M6-A
3942	0.09	0.06	4	HI,BR	7	Pig	mitochondrial malate dehydrogenase
3954		0.02				Rat	opioid binding protein / cell adhesion molecule
3956		0.02			10	Mouse	SDF-1-beta
3960		0.06	2	HG,BR	2	Rat	visinin-like protein 1
4025		0.05				Rat	beta-coat protein
4116	0.04	0.03	1	BR	2	Pig	cytoplasmic malate dehydrogenase
4183		0.02			1	Rat	Rab geranylgeranyl transferase beta subunit
4198		0.02				Rat	interferon related putative protein
4217		0.02	1	PI		Mouse	Rab-18
4230		0.02				Bovine	multiubiquitinating enzyme
4236		0.02	4	HG,HL,BR		Rat	DNA binding protein URE-B1
4254		0.03				Mouse	protein-tyrosine phosphatase kappa
4274		0.02				Mouse	enhancer-trap-locus 1
4286	0.04	0.01	2	HE		Rat	protein phosphatase 1 catalytic subunit
4295		0.02	3	BR		Mouse	inositol 1,4,5-triphosphate-binding protein type 1 receptor
4303	0.04					Rat	ribosomal protein L3
4340		0.05	1	BR		Bovine	epsilon-COP
4368		0.02	3	HI,BR		Rat	syntaxin B
4371	0.04	0.02				Rat	gephyrin
4394		0.02	1	BR		Rat	mitochondrial glycerol-3-phosphate dehydrogenase
4398	0.04	0.02	7	HG		Mouse	TNF receptor related protein
4450	0.04					Mouse	syntrophin-2
4461		0.02				Rat	potassium channel
4473	0.09					Rabbit	sarcoplasmic reticulum calcium ATPase fast twitch skeletal muscle
4502	0.04				7	Rabbit	glycogen-associated protein phosphatase regulatory subunit
4521	0.04					Rat	gamma-glutamylcysteine synthetase light subunit
4538	0.04					Rat	basic transcription element binding protein
4600	0.13		4	HG,HA	14	Chicken	cofilin
4648	0.04				8	Rat	extracellular signal regulated kinase 3 (ERK3)
4670	0.04				2	Mouse	putative calcium binding protein MO25
4747	0.04					Rat	putative calcium pump
4776	0.04				10	Chicken	integrin alpha 8
4782	0.04		1	BR	20	Mouse	RNA-binding protein
4800		0.01				Rat	neurexin
4859		0.01	2	HI,RE		Rat	c kinase substrate calmodulin-binding protein RC3
4928		0.01				Rat	metabotropic glutamate receptor 3
4973		0.01	3	BR		Bovine	leukemia virus cell receptor
5008		0.01				Bovine	poly(A) polymerase
5033		0.02	2	FB,BR		Chicken	NR-CAM cell adhesion molecule
5087		0.01				Rabbit	endopeptidase
5157		0.01				Rat	transcription factor RZR-beta
5171		0.01	1	HI	17	Rat	clathrin-coated vesicle/synaptic vesicle proton pump 116kd subunit
5186		0.01				Mouse	protein overexpressed in testicular tumors
5188		0.01				Dog	sec61 homologue
5194		0.01			17	Mouse	B6D2F1 clone 2A-1
5227		0.01			14	Rat	Rab geranylgeranyl transferase alpha subunit
5231		0.01	1	BR		Rat	metabotropic glutamate receptor 2
5250		0.01			3	Rat	sodium and chloride-dependent GABA transporter 3
5304		0.01				Rabbit	bleomycin hydrolase
5326		0.01	4	HG,CM,BR	9	Mouse	surfeit 4 protein
5476		0.01	2	BR	11	Mouse	flap endonuclease-1
5478		0.01	1	BR		Bovine	pyruvate dehydrogenase (lipoamide) phosphatase
5546		0.01	1	BR		Bovine	neurocalcin
5560		0.01			6	Mouse	ribosomal protein S6 kinase II alpha 1
5574		0.01			11	Rat	LL5
5578		0.01				Rat	3-methyl 2-oxobutanoate dehydrogenase (lipoamide) kinase
5775		0.01			3	Rat	integral membrane glycoprotein GP210
5827		0.02	3	HI,BR		Bovine	GDP dissociation inhibitor for SMG P25A
5960		0.01				Pig	protein P97
6075	0.04				15	Mouse	ZFP29
6077	0.04		2	HI,TE	1	Hamster	mevalonate transporter
6170	0.04		2	HL,FB		Bovine	mitochondrial NADH-ubiquinone oxidoreductase 30kD subunit
6173	0.04		2	HG,HL		Bovine	mitochondrial NADH-ubiquinone oxidoreductase PDSW subunit
6174	0.04		1	HG	4+5	Bovine	ubiquinol-cytochrome C reductase subunit
6193	0.09		1	BR	3	Rat	clathrin coat assembly protein AP50
6257	0.09				22	Rat	clathrin heavy chain
6280	0.04					Rabbit	cytoplasmic glycerol-3-phosphate dehydrogenase
6293	0.04					Rat	V-1 protein
6310	0.04					Bovine	mitochondrial adenylate kinase 2A
6317	0.04		6	HG,FB	5+18+20	Mouse	ribosomal protein L27A
6319	0.04		1	TH		Rat	ribosomal protein L11
6321	0.04					Mouse	single strand DNA binding protein P9
6332	0.04					Rat	clathrin associated protein P17 (AP17)
6339	0.04		1	CM	12	Bovine	mitochondrial ATP synthase D chain
6349	0.04					Pig	prolyl endopeptidase
6354	0.04		2	PI		Rat	Rab14
6361	0.04				5+7	Bovine	protein kinase C inhibitor 1 (PKCI-1)
6405	0.04					Dog	chloride channel
6424	0.04		3	FB,PI,FA		Mouse	interleukin 10 (IL10)
6479	0.04		1	HL		Bovine	mitochondrial NADH-ubiquinone oxidoreductase 14.5kD A subunit
6531	0.04		1	HG	17	Bovine	mitochondrial NADH-ubiquinone oxidoreductase 15kD subunit
6555	0.04					Mouse	nedd-1 protein
6562	0.04				15	Rat	guanidinoacetate N-methyltransferase
6587	0.04		1	FB	3	Rat	ribosomal protein L8
6604	0.04				17	Rat	myosin 1 heavy chain
6608	0.04				5	Rat	adenylate cyclase type 2
6629	0.04					Mouse	TIS7
6649	0.04					Mouse	Evi-1
6659		0.01				Rat	protein kinase C regulated chloride channel
6669		0.01	3	BR		Rat	cysteine-rich protein 2
6693		0.01				Mouse	kinesin-like protein KIF2
6713		0.01				Rat	calcium/calmodulin-dependent protein kinase type 2 alpha
6773		0.01	1	BR	3	Rat	plasma membrane calcium ATPase brain isoform 2
6780		0.01				Mouse	protein p97

Appendix 1C: Related gene transcripts

GENX	% Muscle	% Brain	Seq	Tissues	eSTS	Map	Species	Gene product
22		0.01			9		Yeast	GTP-binding protein GTR1
27	0.04	0.01	1	HL	1	22q13.31-qter	Human	NADH-cytochrome B5 reductase
41		0.01	2	BR			C. elegans	hypothetical 152.4kd protein ZK370.4 in chromosome III
64		0.01					E. coli	Ala-tRNA-1a
67		0.01			22	11q11	Human	oxysterol-binding protein
75		0.02			6		Human	transforming protein RFP
77		0.01				12	Human	uracil-DNA glycosylase 1
78		0.01				21q22.3	Human	cAMP-regulated mRNA
109		0.01			5		Mouse	modifier 2 protein
112		0.01			12		Mouse	natural resistance associated macrophage protein
122		0.01			5	4q28	Human	microsomal UDP-glucuronosyltransferase 2B7
126		0.02					Pig	isocitrate dehydrogenase (NADH), mitochondrial
127		0.03			7		C. elegans	hypothetical 88.1 kd protein
132		0.01	1	BR	2		Rat	glutaminase kidney isoform
134		0.01	2	BR			Human	zinc finger ZNF135
151		0.01					Human	retinoblastoma binding protein 1
165		0.01					Drosophilia	developmental protein eyes
179		0.01			9		Human	probable G protein-coupled receptor EDG-1
186		0.01				7q21	Human	multidrug resistance protein 1
212		0.01	2	BR			Drosophilia	dosage compensation regulator
215	0.04	0.01	2	PI,CC	2		Yeast	tRNA-processing protein SEN3
217		0.01					Chicken	transforming growth factor beta 2
221		0.02			3	Xp22.1	Human	zinc-finger X-chromosomal
236		0.01			12	Xq26.1	Human	hypoxanthine phosphoribosyltransferase
260		0.01	1	BR	19		Rat	thyrotropin releasing hormone receptor
269		0.01					Human	Sec4-like protein Rar
271		0.03			3		Drosophilia	dosage compensation regulator
273		0.01	2	BR			Mouse	alpha-mannosidase 2
300		0.01					Mouse	deubiquitinase
302		0.01	1	CC			Rat	5S RNA
304		0.01					Human	cadherin 12
306		0.01	4	BR,TE	20		Yeast	chromosome segregation protein CSE1
308		0.01				4q13.3-q21.1	Human	deoxycytidine kinase
321		0.01			4		C. elegans	hypothetical 17.0 kd ZK370.2
323		0.01	1	BR		1p36.1-p35	Human	hydroxymethylglutaryl-CoA lyase
331		0.01			8		E. coli	peptide methionine sulfoxide reductase (protein reparase)
337		0.01					E. coli	molybdopterin biosynthesis MOEB
343		0.03	3	HG,HA			Yeast	transcriptional regulatory protein RPD3
355		0.01	5	FB,BR			Ray	synaptic vesicle membrane protein VAT-1
367		0.06	1	BR			Chicken	zinc finger protein CTCF
373		0.01					Mouse	glycoprotein GP38
377		0.02	1	HA	6		Yeast	serine-rich RNA polymerase I suppressor
409		0.01	2	FB,BR	5	11p12	Human	inducible membrane protein R2
421		0.01			9		Human	p80-coilin pseudogene
429		0.01			4		Drosophilia	cadherin-related tumor suppressor protein FAT
439		0.01			18		Rat	zinc finger protein
442		0.01	3	BR	12	1p21-p13	Human	transforming protein RHOC
448		0.01					Rat	stress activated protein kinase alpha 1
449		0.01					Yeast	MSP1
467		0.01					Pig	sperm surface protein PH-20
489		0.01			2		Human	ribosomal protein S26
505		0.01			20		Drosophilia	crooked neck
510		0.02	4	HL,FB			Rat	liver nuclear protein p47 (eIF-4A-related)
532		0.01					C. elegans	hypothetical 23.3kd protein ZK688.3 in chromosome III
546		0.01					Human	transforming growth factor beta-1 binding protein
575		0.02	3	BR		1q21	Human	aryl hydrocarbon receptor nuclear translocator
580		0.01				12	Human	oligo-adenylate synthetase 1
591		0.01	1	BR			Yeast	nuclear protein SNF4
599		0.02	1	BR			Human	53PB2 p53-binding protein
616		0.01	3	CM,BR			Yeast	cell division control protein
634		0.02	1	PI	13		Human	rag cohort (rch1)
640		0.01					C. elegans	hypothetical 48.1 kd protein CO2F5.6
642		0.01	1	BR			Rat	calbindin D28
646		0.01					Chicken	100kd integral membrane glycoprotein SC-1
684		0.01					Human	calmodulin retropseudogene CaMII-psi2
686		0.02	1	BR			Drosophilia	Goliath
688		0.02					Rat	cytoplasmic dynein heavy chain
691		0.01				Xq28	Human	glucose 6-phosphate 1-dehydrogenase
706		0.02			17		Mouse	pericentrin
707		0.03	1	HL	7		Rat	insulin-induced growth-response protein (CL-6)
720		0.01					Human	cartilage glycoprotein 39
723		0.01	1	HI			Rat	guanine nucleotide dissociation stimulator RALGDSB
726		0.01					Human	LIMK (protein kinase with 2 LIM/double ZF motif)
731		0.01	1	PI			Mouse	brain protein DN38
741		0.01				9p21-p13	Human	N-acetyllactosamine synthase
755		0.01					Human	fibulin
768		0.03					E. coli	arginyl tRNA synthetase
775		0.01					Human	G-gamma and A-gamma globin
782		0.01	1	HI			Chicken	zyxin
793		0.01					Bovine	G protein gamma-5 subunit
796		0.01				1q23-q25.1	Human	antithrombin 3
822		0.01					Rabbit	sarco(endo)plasmic reticulum ca2+-ATPase (SERAC2)
835		0.01					Human	randomly sampled cDNA D26445
859		0.02	1	BR			Pseudomonas	tropinesterase
893		0.03					Mouse	U7 snRNA
930		0.01					C. elegans	putative acetylcholine regulator unc-18
1007		0.01					Human	heterogenous nuclear RNA W16W
1016		0.02					Rat	NADH-cytochrome B5 reductase
1043		0.01					Drosophilia	box-b binding factor-2
1049		0.02					Rat	ornithine decarboxylase antizyme
1051	0.04	0.01					Yeast	intracellular protein transport protein USO1
1056		0.01				17q24	Human	somatostatin receptor isoform 2
1075		0.01	1	HL		19q13.1-q13.2	Human	B-cell lymphoma 3-encoded protein
1077		0.01				Xp21.3-p21.1	Human	dystrophin
1088		0.01	2	BR,BC			Rat	phosphodiesterase I and pyrophosphatase
1139		0.01					Mouse	neural cell adhesion molecule NCAM-140
1162		0.01					C. elegans	hypothetical 51.6kd protein F59B2.5 inchromosome 3
1165		0.01	1	FB		11q23.1-q23.2	Human	zinc finger ZNF128
1183		0.01					E. coli	GTP-binding protein
1184		0.01					Human	fibrillin 1

Appendix 1C: Related gene transcripts (*continued*)

1190		0.01					Yeast	serine-rich RNA polymerase I suppressor
1227		0.03				1q32.1	Human	MAP kinase phosphatase 1
1233		0.01					Drosophilia	cadherin-related tumor suppressor protein FAT
1244		0.01	1	HI		Xp11.22-cen	Human	gastrin-releasing peptide receptor
1248		0.02					Plethodon	WNT-10B protein
1270		0.05	3	HL,BR,FB	11+19		Mouse	substrate of protein tyrosine kinase receptors and p60v-src
1276		0.01					Human	focal adhesion kinase
1305		0.01	2	HL,FB			Dog	glycoprotein 25L
1306		0.01	1	HA		21	Human	zinc finger domains ZF21.3
1320		0.01					Bovine	cartilage leucine rich protein
1324		0.03					Yeast	YME1
1326		0.01	1	FI			Chicken	sodium potassium transporting ATPase beta 2
1335		0.03					Drosophilia	claret segregational protein
1356		0.01	3	FB,PI,CC			Tobacco	thioredoxin
1364		0.01					Chicken	probable G protein-coupled receptor 6H1
1397		0.01	4	CM,BR,BC			Mouse	nicotine acetylcholine receptor beta subunit gene
1412		0.02	2	FB,BR			C. elegans	hypothetical 54.9 kd protein CO2F5.7
1418		0.03				1p13	Human	CD53
1439		0.02					Pig	inhibin beta a-subunit
1448		0.02					Drosophilia	white protein
1507		0.01					Dog	cytochrome P450 2B
1536		0.01	1	HA			Yeast	phenylalanyl-tRNA synthatase alpha chain cytplasmic
1547		0.01	2	BR			Yeast	hypothetical 75.5kd protein in SDH1-CIM5/YTA3 intergenic
1573		0.01					Human	zinc finger protein 7
1579		0.05			21		Amoeba	myosin heavy chain IB
1595	0.09	0.06	4	HL,FB,FA,HA	20		Rat	14-3-3 protein beta
1616		0.02	1	BR			Human	DNAJ protein homolog 1
1635		0.01					Yeast	tRNA isopentenyltransferase
1662		0.01				13q13	Human	caseine kinase 1
1672		0.01	1	BR			Rat	acyl-peptide hydrolase
1678		0.01					Rat	TR4 orphan receptor (steroid receptor)
1683		0.01					Drosophilia	cadherin-related tumor suppressor protein FAT
1689		0.01					Yeast	tRNA-splicing endonuclease positive effector
1698		0.01				8p21.1-p11.2	Human	erythrocyte ankyrin
1700		0.01	2	BR,FB			Human	splicing factor CC1.3
1707		0.01	2	BR			Xenopus	gastrula zinc finger protein XLCGF57.1
1721		0.03	1	BR			Drosophilia	casein kinase 2, alpha chain
1743		0.01					Yeast	hypothetical 34.9 kd protein in URA1
1767		0.01					Human	heat shock factor (HSF1)
1779		0.01					C. elegans	putative ATP-dependent RNA helicase K03H1.2 in chromosome III
1791		0.01					Dictyostelium	developmentally regulated protein kinase 1
1816	0.04	0.03			16		Mouse	teg292
1827		0.02			22		C. elegans	hypothetical 52.7kd protein T23g5.2 in chromosome III
1842		0.01	1	BR			Mouse	GABA transporter
1856		0.02					Drosophilia	serine/threonine kinase Fused
1860		0.03	1	BR	17		C. elegans	hypothetical 52.7kd protein T23g5.2 in chromosome III
1867		0.01	1	CM			Human	zinc finger ZNF139
1883		0.01					Rat	guanine nucleotide releasing protein P140 Ras-GRF
1908		0.02			X	Xq21.33-q22	Human	tyrosine-protein kinase atk
1912		0.01					Xenopus	ribosomal protein S6 kinase II beta
1929		0.01					Rat	skeletal muscle myosin light chain kinase
1947		0.02	1	BR			Yeast	BCS1 protein
1957		0.01	1	BR			Yeast	zinc finger protein GCS1
1965	0.04	0.01	5	HL,BR	1		Chicken	ribosomal protein S6 kinase II alpha
1974		0.01	3	FB,PI,HA			Rat	adducin-like
1978		0.01					C. elegans	hypothetical 67.6kd protein ZK637.3 in chromosome III
1996		0.02				1p13.1	Human	3-beta hydroxy-5-ene steroid dehydrogenase type II
2042		0.03	5	HI,BR	19		Vaccinia virus	protein K4
2047		0.02	2	HL,BR			Yeast	leucine-tRNA synthetase
2068		0.01					Vaccinia virus	protein K4
2076		0.01					Mouse	guanine nucleotide releasing protein P140 Ras-GRF
2090		0.01	4	HL,BR			B. subtilis	initiation factor IF2
2106		0.01	3	HG,BR			Bovine	CSSM015 microsatellite
2138		0.01					Rat	neuraxin
2140		0.01				3p21-p22	Human	transcription factor
2158		0.01	1	BR			Neurospora	sulfate permease II
2162		0.01					S. pompe	CDC5
2172		0.01	1	BR			C. elegans	putative ATP-dependent RNA helicase T26G10.1 in chromosome III
2174		0.01					Xenopus	oocyte zinc-finger protein XLCOF26
2182		0.01	1	BR			Thermus a.	succinyl-CoA synthetase beta chain
2188		0.01					Rat	guanine nucleotide releasing protein P140 Ras-GRF
2194		0.01					Yeast	SCO1
2209		0.01				17q21.3	Human	MDC protein
2225		0.01				1	Human	HKR3 protein
2228		0.01	3	CM,BR			H. marismortui	30S ribosomal protein HS6
2238		0.01					Klebsiella terr.	acetolactate synthase
2250		0.01				3q22-q25	Human	polyadenylate binding protein
2254		0.01	2	BR			Vaccinia virus	hypothetical 5.3kd protein
2263		0.02	1	BR			Cryptococcus n.	ADP ribosylation factor
2267		0.01					Mouse	ECA39
2287		0.01				12p13	Human	glyceraldehyde 3-phosphate dehydrogenase
2298		0.01					Human	myelin transcription factor 1
2341		0.01					Chicken	cartilage matrix protein
2345		0.02					Human	erythrocyte band 7 integral membrane
2362		0.05			5		Drosophilia	ubiquitin-conjugating enzyme E2-16 kD
2380		0.01			X		Mouse	brain protein DN38
2395		0.01					Chicken	epithelial-cadherin
2398		0.01			19		Yeast	32.3kd protein in APE1/LAP4-MBR1 intergenic region
2403		0.01					Human	PGB4G7 gene
2405		0.01	1	HE	X		Yeast	transcriptional regulatory protein RPD3
2413		0.01	3	BR,CC	7+11		Human	DNAJ protein homolog HSJ1
2423		0.01					Drosophilia	suppressor of forked protein
2453		0.01					Human	CMRF35
2477		0.01	1	CM	14	6q22.3-q23.1	Human	arginase
2478	0.18	0.05			1		Human	dual specificity phosphatase tyrosine/serine
2490		0.01					Chicken	lamin B
2504		0.01	1	HL			Yeast	hypothetical 195.2 kd protein in GCN3-DAL80
2513		0.01					Yeast	hypothetical 45.6kd protein in TPD3 3' region
2514		0.01					C. elegans	hypothetical 50.4kd protein ZK637.1 in chromosome III
2520		0.01				17q	Human	tre oncogene (deubiquitinase)
2550		0.01	2	BR			Dictyostelium	coronin
2556		0.03	3	FB,BR			Salmonella t.	protease DO
2562		0.02	1	BR			Amoeba	myosin Ic heavy chain
2565		0.03			12		Mouse	oocyte maturation OM-1

Appendix 1C: Related gene transcripts (*continued*)

2566		0.01					Arabidopsis th.	cell division control protein 2 homolog A
2571		0.01					Human	ribosomal protein S6 kinase
2578		0.03	2	BR	7		Drosophilia	Slit protein
2590		0.01	1	HA			Rat	galactose-specific lectin
2598		0.02	4	HI,BR	5		Bovine	stress activated protein kinase alpha I
2610		0.01					Mouse	ZFP28
2620		0.01					Yeast	general negative regulator of transcription subunit 2 (NOT2)
2627	0.04	0.01			1+9		Bovine	chlorine channel P64
2629		0.01	1	BR		Yp11.3	Human	zinc-finger Y-chromosomal protein
2631		0.01					Drosophilia	Slit protein
2638		0.02	7	TC,FB,BR		Xq26.3-q27.1	Human	proto-oncogene dbl
2642		0.01	2	BR			C. elegans	hypothetical 152.4kd protein ZK370.4 in chromosome III
2646		0.01					Bovine	mitochondrial nicotinamide nucleotide transhydrogenase
2657		0.02					Human	liver cytochrome C oxidase 7A
2662		0.01	3	HL,FB,BR			Bovine leukemia virus	gag polyprotein
2703		0.02				1p22.1-qter	Human	succinate dehydrogenase (ubiquinone) flavoprotein subunit
2724	0.04				2	2q31-q32	Human	nebulin
2767		0.01					Chlamydomonas r.	ADP-ATP translocase
2794		0.03	3	TH,HI,FB			Yeast	proteolipid protein PPA1
2796		0.03	2	BR			C. elegans	hypothetical 25.9kd protein C05B5.7 in chromosome III
2810		0.01				10q23-q24	Human	interferon induced protein 56kd
2863		0.01	1	BR			Yeast	transcriptional regulatory protein RPD3
2885		0.01					Human	telomere associated repeat sequence
2889		0.01	1	HI	1	17	Human	pyrroline-5-carboxylate reductase
2896		0.01			11		Mouse	perforin
2905		0.01				12cen-q21	Human	synaptotagmin 1
2910		0.01	1	BR			Trichoderma r.	elongation factor 1 alpha
2930		0.01					Yeast	pre-mRNA splicing factor RNA helicase PRP28
2939		0.01					Rat	peroxisomal enoyl hydratase-like
2947		0.01				16p13.1	Human	GST1-HS GTP-binding protein
2961		0.01					Yeast	clathrin associated protein AP17
2963		0.02	4	PI,BR,HE		10q23-q25	Human	insulin-degradating enzyme
2977		0.03					Mouse	synaptotagmin 4
2985		0.02	5	HL,BR,FB,BC	19	6p12	Human	hnRNP G
3003		0.03					Mouse	potassium channel mShal
3008		0.02	4	HI,BR			Mouse	FK506-binding protein
3022		0.03	2	BR			C. elegans	hypothetical 63.5kd protein ZK353.1 in chromosome 3
3027		0.06					Xenopus	ornithine decarboxylase
3036		0.03			3	6p21.3	Human	HLA class-2 66kb-region
3041	0.04				2		Chicken	smooth muscle myosin light chain kinase
3050		0.02			10		Human	peptidylproline cis-trans isomerase A
3058		0.03					Rat	putative zinc finger protein
3068		0.03	1	FB			Rat	calbindin D28
3103		0.02	2	CM,BR			C. elegans	ubiquitin-conjugating enzyme E2-17 kD
3116		0.02					Human	ribosomal protein homologous to yeast S24
3119	0.09	0.03	3	PI,BR	X+9+16		Rat	MAP-1A and 1B light chain 3
3121		0.02	1	BR			Mouse	gamma adaptin
3122	0.40	0.02	1	BR	X+3	1q24-q32	Human	cystein-rich protein
3159		0.02	1	BR		1p22.1-qter	Human	succinate dehydrogenase (ubiquinone) flavoprotein subunit
3169		0.02					E. coli	sigma cross-reacting protein 27A
3172		0.02	1	HI		19	Human	HKR1
3186		0.03					Duck	malate oxidoreductase
3189		0.03	2	BR		1p36-p31.2	Human	guanine nucleotide-binding protein G(i)/G(s)/G(t) beta subunit 1
3239		0.02				4q25-q27	Human	brain ankyrin
3249		0.02			6	1p34.2-p33	Human	erythroid protein 4.1
3282		0.07	4	BR,CC,FB			C. elegans	hypothetical 80.7 kd protein ZC84.3
3299	0.04	0.03	2	HL,HI			Anacystis nidulans	elongation factor Tu
3302	0.04					11q13-q14	Human	skeletal muscle alpha3-actinin
3303		0.03	1	BR			Human	putative centromeric DNA
3313		0.03					Yeast	elongation factor 3
3326		0.02					Rabbit	phosphoglucomutase isoform 2
3334		0.02	1	PI			Human	HSP47
3370		0.02	2	BR,CC	7	7q21-q22	Human	serum paraoxonase/arylesterase
3384	0.26					2q31-q32	Human	titin
3389		0.02	1	HL			C. elegans	meiotic spindle formation protein MEI-1
3404		0.02	2	TE			Human	heat shock 70kD protein HSP70RY
3442	0.13				17	17pter-p11	Human	skeletal muscle beta-myosin heavy chain
3443	0.04				X+2		Rat	hexokinase 2
3449	0.09				1	1q24-q32	Human	cystein-rich protein
3463	0.09		1	CC			Human	heat shock protein HSP27
3468	0.09					22q13	Human	interleukin-2 receptor beta chain
3486	0.09				17	17pter-p11	Human	skeletal muscle beta-myosin heavy chain
3504	0.09					17pter-p12	Human	perinatal cardiac myosin heavy chain
3517	0.09		2	HL,CC	1		Mouse	MSP23 mRNA
3531	0.09				4	13	Human	retinoblastoma suspectibility gene
3533	0.04	0.01			14		Dianthus c.	glutathione S-transferase 2
3536	0.26				10	2q35	Human	villin
3539	0.31						Rabbit	titin
3545	0.04						B. subtilis	glutamyl-tRNA synthetase
3584	0.13					Xq28	Human	non-muscle filamin
3589	0.09				2		Rabbit	titin
3604	0.09				8+9+15		Bovine	mitochondrial adenylate kinase 3
3616	0.04				19	Xp11.3-p11.23	Human	ZNF81
3644	0.09				2		Rabbit	titin
3653	0.09				3	6p21.3	Human	HLA class-2 66kb-region
3668		0.02	3	BR			Rat	protein phosphatase 2B catalytic subunit 1
3684		0.02	5	HL,HI,BR			Yeast	probable peroxisomal targeting signal receptor
3704		0.02	4	FB,BR			Xenopus	proto-oncogene tyrosine protein kinase Yes
3731		0.02	2	BR	17	16	Human	carbonic anhydrase 5
3756	0.04						Drosophilia	serendipity locus protein H-1
3767		0.03				19q13.1	Human	zinc finger ZFP-36
3781		0.02				19p13.1-p12	Human	zinc finger protein ZNF91
3784	0.09				9		Bovine	mitochondrial adenylate kinase 3
3785		0.02			3		Rat	MRC ox-2 antigene
3786		0.03			3		Yeast	DNA repair protein RAD5
3798	0.22		2	HE		2q31-q32	Human	nebulin
3800		0.06	2	BR	1		Bovine	auxilin
3843		0.05	3	HL,BR	1+3+14		Clostridium a.	DNAJ protein
3866	0.09				7		Bombyx mori	glycyl-tRNA synthetase
3879		0.02			10		C. elegans	hypothetical 68.7kd protein ZK757.1 in chromosome III
3884		0.02	2	FB	20		Mouse	ETO (putative transcription factor)
3896		0.02					C. elegans	hypothetical 46.4kd protein T16H12.5 in chromosome III
3902	0.04	0.03	6	CM,HI,BR	16		Yeast	nuclear transport protein NIP1
3910		0.02	1	FB	10		Drosophilia	lethal(1)discs large-1 tumor suppressor

Appendix 1C. Related gene transcripts (continued)

3917		0.02			12	2q35	Human	villin
3923		0.02	4	HL,BR,FLS	11		Mouse	radixin
3941		0.02	2	FB			Human	membrane metalloendopeptidase
3961		0.02	9	HI,BR,FB			Mouse	amyloid-like protein 1
3983		0.02			5		Mouse	alpha-mannosidase 2
3990		0.02					Rat	proline-rich protein
3995	0.04	0.01			19	1p36.3-p36.2	Human	procollagen-lysine 2-oxoglutarate 5-dioxygenase
4002		0.02	7	BR	7		Mouse	Ca++/calmodulin dependent protein kinase 2
4003		0.03					Human	TALLA-1
4016		0.03	2	BR	1		Mouse	teg267
4026		0.02				2p23-qter	Human	dipeptidyl peptidase 4
4045		0.02				19q13.3	Human	glandular kallikrein 2
4047		0.02	1	HL	15		Yeast	60S ribosomal protein L30A (RP29)
4083		0.03	1	HG			Human	transcription factor BTF3b
4084	0.04	0.01			17		Chicken	endoplasmin
4090		0.02					Human	protein phosphatase 2C alpha
4092		0.02	1	FB			Rat	cytoplasmic dynein 74kd intermediate chain
4093	0.04	0.02			14		Mouse	MSEC66
4098		0.03	8	HI,BR	11		Mouse	sodium/potassium-transporting ATPase gamma chain
4100		0.05	3	HI,BR,HA	7		Chicken	amphiphysin (synaptic vesicle-associated protein)
4120		0.02	1	HA	9		Rat	postsynaptic density protein 95
4129		0.03	3	HL,BR,CC	18		Brugia malayi	potentially protective 63 kd antigen
4131		0.03			12		Rat	sodium-dependent neurotransmitter transporter
4136		0.03					C. elegans	hypothetical 29.0kd protein ZK632.12 in chromosome III
4142		0.02			X		Chicken	basic-leucine zipper-like protein
4169		0.03	1	HL	6+11		Mouse	zinc finger protein MFG3
4170		0.02					Yeast	nuclear protein SNF4
4171		0.02			19		Human	zinc finger protein ZNF132
4176		0.02	1	BR			Candida albicans	carboxypepdidase Y
4181	0.04	0.01	1	HE		20q11.2	Human	pp60-cSrc
4203		0.02					Mouse	GABA transporter (GAT4)
4214		0.02			19	11q23.1-q23.2	Human	zinc finger ZNF128
4280		0.02					Mouse	serine/threonine-protein kinase TIK
4283		0.02			5		Sheep	ovis aries secretory protein
4312		0.02	1	BR			Rat	neuronal protein P25
4333		0.03			10		Yeast	probable ATP-dependent RNA helicase
4337		0.05	3	HI,BR			Rat	stannin
4342		0.02	2	BR	14		Yeast	FUN20
4344	0.04				3+5	6p21.3	Human	HLA class-2 66kb-region
4381	0.04	0.03			12+17		Manduca sexta	vacuolar ATP synthase 14kD subunit
4386		0.02					Drosophilia	cadherin-related tumor súppressor protein FAT
4400		0.03	2	BR,PI	1		Mouse	mRNA clone pMAT1 for new transforming gene
4436	0.04	0.01			3		Yeast	mitochondrial carrier protein YMC1
4445		0.02	1	BR	15		Yeast	hypothetical 21.9kD protein in MRPL6 5' region
4484	0.04				2		Rabbit	titin
4488	0.04						Enterococcus h.	sodium-proton antiporter
4494	0.04				14		E. coli	ATP-dependent RNA helicase SRMB
4507	0.57				2		Drosophilia	ring canal protein KELCH
4520	0.04				1		Mouse	transcriptional control element
4542	0.04		1	HL	1		Mouse	DLA-91 mRNA
4552	0.04				X	13q34	Human	NAD+ ADP ribosyltransferase pseudogene
4565	0.04					2q31-q32	Human	nebulin
4568	0.04					11q11	Human	oxysterol-binding protein
4569	0.04				10		Human	TPA-inducible c54 mRNA
4587	0.04						Drosophilia	trithorax
4603	0.04						Human	putative serine/threonine protein kinase p78
4617	0.04		1	BR	16		Human	cytosolic alanine aminotransferase
4621	0.04						C. elegans	cosmid C50C3
4622	0.04					9q22	Human	tropomodulin
4639	0.04				2		S. pompe	protein kinase BYR2
4651	0.09				1+5		Yeast	chromosome 5 right arm sequence proximal to GLC7
4652	0.04				7	19q13.1	Human	zinc finger ZFP-36
4655	0.04					2q31-q32	Human	nebulin
4668	0.04				4		Rat	calmodulin-dependent protein kinase 2 delta
4692	0.22		1	PI	X		Rat	peroxisomal acetyl-CoA acyltransferase
4701	0.04				6		Bovine	endozepine related protein
4704	0.04				2	2q31-q32	Human	titin
4712	0.04				5		Mouse	secreted isoform neural cell adhesion molecule (N-CAM 120)
4719	0.04						Human	MAP kinase activated protein kinase-2
4722	0.04						Rabbit	titin
4749	0.04				10	Xp11.3-p11.23	Human	properdin
4767	0.04					3q22-q25	Human	polyadenylate binding protein
4773	0.04						Chicken	smooth muscle myosin light chain kinase
4774	0.04				2		Heliantus a.	pollen specific protein SF3
4804	0.13				22		Rat	mitochondrial carnitine palmitoyltransferase 1
4806		0.01				Xq26.1	Human	hypoxanthine phosphoribosyltransferase
4823		0.01	1	BR	16		Yeast	general negative regulator of transcription subunit 1 (NOT1)
4843		0.01	2	BR		9	Human	J kappa recombinaison signal binding protein pseudogene 1
4939		0.01					Mouse	protein-tyrosine phosphatase kappa
4957		0.01				16p13.1	Human	GST1-HS GTP-binding protein
4962		0.01	3	BR		17q24-qter	Human	nuclear p68 protein
4986		0.01					Yeast	squalene monooxygenase
4993		0.01					Mouse	necdin
5003		0.01			19		E. coli	acetolactate synthase isozyme III, I chain
5017		0.01	1	HL			Yeast	protein SLY1
5026		0.01					Human	putative ribosomal protein L13
5062		0.01			3		Human	metabotropic glutamate receptor 5a
5064		0.02	1	BR	6		Yeast	possible 1-acyl-sn-glycerol-3-phosphate acyltransferase
5095		0.01					Human	glycine cleavage system H protein
5133		0.01			22		C. elegans	hypothetical 58.3kd protein F42H10.7 in chromosome III
5134		0.01			6		Mouse	zinc-finger protein Blimp-1
5169		0.01	1	CC			E. coli	isoleucyl-tRNA synthetase
5172		0.01					Drosophilia	ring canal protein KELCH
5179		0.01	3	BR			Human	CDC4 related protein
5263		0.01					Drosophilia	REF (2)P protein
5293		0.01	2	FLS	22		Rat	phosphatidylinositol transfer protein
5311		0.01			2		Human	GP36b glycoprotein
5336		0.01					Human	randomly sampled cDNA D13763
5343		0.01					Vaccinia virus	protein C3 precursor
5351		0.01			7		Mouse	putative EC-8
5409		0.01					Human	PINCH (a new LIM protein)
5436		0.02	2	BR			Bovine	cellular retinaldehyde-binding protein
5447		0.01	1	BR	4		Human	ribosomal protein L23
5474		0.01			12		Mouse	cytotoxic T-cell membrane glycoprotein Ly-3

Appendix 1C: Related gene transcripts (*continued*)

5490		0.01					Mouse	brain protein DN38
5500		0.01	2	HL,PI	6		Yeast	P24B protein
5507		0.01			20		Rat	phospholipase C beta 4
5519		0.01			16	16p13.3	Human	polycystic kidney disease 1 protein
5550		0.01					Rat	golgi apparatus sialoglycoprotein MG-160
5576		0.01			17		Mouse	brain protein H5
5600		0.01			2		Yeast	NIFS-like 54.5kd protein
5640		0.01			16		Human	DNAJ protein homolog
5664		0.01				4q25-q27	Human	brain ankyrin
5695		0.01	5	HG,BR,CC			Bombyx mori	alanyl-tRNA synthetase
5706		0.01					Mouse	zinc-finger protein ZFP-27
5714		0.01					Human	hnRNP-E2
5717		0.01					Mouse	Rab-18
5722		0.01					C. elegans	hypothetical protein F44E2.6 in chromosome 3
5723		0.01					E. coli	aminopeptidase A/I
5724		0.01	2	BR,FB	17		Human	gamma tubulin
5811		0.01				3p21-p22	Human	zinc finger protein
5816		0.01			17		Yeast	hypothetical 34.9 kd protein in URA1
5831		0.01			1	9q34.1	Human	adenylate kinase isoenzyme 1
5839		0.01				1q32.1	Human	MAP kinase phosphatase 1
5847		0.01	1	BR			Yeast	ubiquitin carboxyl-terminal hydrolase 4
5850		0.01	3	BR			Dog	sec61 homologue
5870		0.01			19	11q23	Human	zinc finger protein HRX
5875		0.01					Dictyostelium	myosin light chain kinase
5882		0.01	1	BR			Yeast	NAM7 protein (probable helicase)
5929	0.09				2		Amphoxius	CAVP-target protein
6078	0.04				6		Rabbit	P59
6084	0.04				X	Xp11.4-p11.3	Human	monoamine oxidase B
6091	0.04				12		Yeast	transcriptional activator GCN3
6095	0.04				19		Human	Kruppel-type zinc finger
6123		0.01			X		C. elegans	hypothetical 22.6kd protein F42H10.3 in chromosome III
6153	0.04				3	4p16.3	Human	cosmid clone HDAB (1S149)
6159	0.04				10		Xenopus	ubiquitin-like fusion protein
6192	0.04					1p31	Human	glucose transporter type 5
6199	0.04						Rat	prohibitin
6209	0.04				17		Wheat	ubiquitin-conjugating enzyme E2-20 kD
6237	0.04		1	PI			Human	ribosomal protein L4
6240	0.04						Human	liver 6-phosphofructokinase
6282	0.04				22		Rat	mitochondrial carnitine palmitoyltransferase 1
6289	0.04					17pter-p12	Human	perinatal cardiac myosin heavy chain
6335	0.04						E. coli	peptide chain release factor 2
6370	0.04		1	HL	19		Human	sequence M77024 isolated by cross-reactivity of ARF sera
6371	0.04						Yeast	ABC1 protein
6374	0.04						Human	ubiquitin
6375	0.04		1	PI			Rat	prohibitin
6383	0.04					9q22	Human	tropomodulin
6391	0.04					2q31-q32	Human	titin
6404	0.04		1	BR			Drosophilia	ubiquitin-conjugating enzyme E2-16 kD
6414	0.04		1	BR			Drosophilia	BRAHMA protein
6417	0.04					17pter-p11	Human	embryonic fast skeletal muscle myosin heavy chain
6428	0.04						E. coli	DNAJ protein
6441	0.04				17	17pter-p12	Human	perinatal cardiac myosin heavy chain
6453	0.04						E. coli	glutamyl-tRNA synthetase
6472	0.04						Human	UDP-glucose pyrophosphorylase
6478	0.04				20		Yeast	hypothetical 44.1 kDa protein
6482		0.02					Vaccinia virus	protein K4
6496	0.04				1		Yeast	ABC1 protein
6511	0.04				3		Mouse	cytoplasmic glycerol-3-phosphate dehydrogenase
6543	0.04						Rabbit	sarcolumenin
6545	0.04				3	9q22	Human	tropomodulin
6550	0.04						Human	pregnancy-specific beta-1-glycoprotein-11
6551	0.04						Bovine	mitochondrial adenylate kinase 3
6557	0.04				12		Hamster	mevalonate transporter
6558	0.04						Human	peptidylproline cis-trans isomerase
6573	0.09	0.01	1	BR	8		Mouse	muscle-brain cAMP dependent protein kinase inhibitor (PKI-alpha)
6577	0.04				14+16	3q22-q25	Human	polyadenylate binding protein
6583	0.04				2		Chicken	titin
6586	0.04						Xenopus	transcription factor 3A
6591	0.04				17		Human	short-chain specific acyl-CoA dehydrogenase
6592	0.04				1	22q13.31-qter	Human	NADH-cytochrome B5 reductase
6618	0.04				9		Dictyostelium	vegetative specific protein H7
6630	0.04						Schistosoma m.	hemoglobinase
6633	0.04				22	12	Human	tubulin alpha 4
6648	0.04						E. coli	insertion element 5
6669		0.01					C. elegans	hypothetical 29.0kd protein ZK632.12 in chromosome III
6706		0.01					Human	ADP-ribosylation factor 1
6739		0.01					Human	coatomer beta' subunit
6743		0.01				4	Human	long-chain-fatty-acid-CoA ligase
6760		0.01					Dictyostelium	protein kinase 2
6771		0.01			3		Bovine	vacuolar H+ATPase subunit A
6776		0.01					Human	HSP binding immunophilin P59
6779		0.01					C. elegans	hypothetical 64.5 kda protein ZK652.9
6792		0.01					Maize	homeotic protein knotted-1
6809		0.01				21q22.3	Human	cystathionine beta-synthase
6814		0.01			4		Rat	calmodulin-dependent protein kinase 2 delta
6824		0.01	1	HA	12		C. elegans	putative ATP-dependent RNA helicase T26G10.1 in chromosome III
6838		0.01			19	19q13.2-q13.4	Human	zinc finger protein ZNF42
6849		0.01					Rat	brain alpha-tropomyosin TMBr-3
6869		0.01			19	11q23.1-q23.2	Human	zinc finger ZNF128
6870		0.01					Yeast	delta-1-pyrroline-5-carboxylate dehydrogenase
6871		0.01			19	4q12-q21	Human	krueppel related DNA binding protein (PF4)
6876		0.01	1	HA			Rat	CaBP1 calcium binding protein

Appendix 1D: Unknown gene transcripts (*partial*)

GENX	% Muscle	% Brain	Seq	Tissues	eSTS
4		0.02			7
5		0.01			
7		0.01			X
8		0.01			5
9		0.01	1	PI	14
11		0.01			X
13		0.01	1	BR	2
14		0.01			6
15		0.02			4
16		0.01			
17		0.01	1	HL	
18		0.02	1	BR	12

GENX	% Muscle	% Brain	Seq	Tissues	eSTS
2208		0.01			
2210		0.01			
2212		0.02	1	UN	6
2213		0.01	2	BR	
2214		0.01			
2215		0.01			
2216		0.03	1	FB	
2217		0.01	1	BR	
2218		0.01			
2219		0.01			
2220		0.01			
2221		0.02	1	HI	

GENX	% Muscle	% Brain	Seq	Tissues	eSTS
4627		0.02			
4628	0.04				19
4629	0.04				
4630	0.04				17
4631	0.04				13
4632	0.04		1	BR	8
4633	0.04				1
4634	0.04				1
4635	0.04				5
4636	0.04		6	HI,FB,BR	14
4637	0.04				3
4638	0.04				8

Appendix 1E: New gene transcripts in known regions

GENX	% Muscle	% Brain	Seq	Tissues	eSTS	Map	DNA region	Similarity
44		0.01				12p13	gamma enolase	Unknown
556		0.02				17q11.2	neurofibromin	Unknown
1138		0.01				19q13.3	DNA 106kb from cosmid MMDA	Unknown
1202		0.01			16	16q22.1	haptoglobin	Unknown
1251		0.01	1	BR			glycogene synthase gene 5' flank	Unknown
2313		0.01			4	4p16.3	cosmid HDAC	Unknown
2408		0.01				6p21.3	MHC class III HLA-RP1	Unknown
3830		0.02				19q13.3	MMDB/MMDC cosmid from chromosome 19q13.3	Unknown
4023		0.02	1	BR	20+22	22q11	BCR	Unknown
4079		0.02			22	22	CYP2D7BP pseudogene for cytochrome P4502D6	Unknown
4221		0.02			15	15	genomic YAC-end left arm of chromosome 15	Unknown
4318		0.05				22q11	BCR	Unknown
5283		0.01				6p21.3	HLA class-2 66kb-region	Unknown
5592		0.01				11q13	rod outer segment membrane protein 1	Unknown
6191	0.04		1	CM			DNA found 5' to NCA	Unknown
6896		0.01				9q34.1	c-ABL	Unknown

Appendix 1F: Overlapping gene transcripts

GENX	% Muscle	% Brain	Seq	Tissues	eSTS	Map	Overlapped gene	Similarity	Species	Gene product
31		0.01	2	BR,HE	22		GENX-790	Unknown		
181		0.02	2	FB	16	16q13-q21	DNA-directed RNA polymerase II 33kd polypeptide	Unknown		
280		0.02	1	BR	19		GENX-5139	Unknown		
673		0.01				12	cation-dependent mannose 6-phosphate receptor	Unknown		
703		0.01					phospholipid transfer protein	Identical	Human	lysosomal protective protein
973		0.01					selenoprotein P	Unknown		
1545		0.02			17	17	acetylcholine receptor epsilon subunit	Unknown		
1680		0.01	1	FB			amino acid transport protein	Unknown		
1825		0.01			5	5	histidyl tRNA synthetase	Unknown		
2176		0.01	2	FB,BR	6	6	T-complex-associated testes expressed 1 homolog	Unknown		
2438		0.01					neuronal acetylcholine receptor protein alpha-7	Unknown		
2666		0.01				17q11-q21	estradiol 17 beta-dehydrogenase 1	Unknown		
2736		0.02	5	FB,BR		22	glycoprotein Ib beta	Related	Mouse	brain protein H5
2785		0.03	3	HG,BR,CC		4q23-q35	aspartylglucosaminidase	Homologous	Mouse	T complex protein 1 delta subunit
2824		0.01	1	HL		6p23	dek mRNA	Unknown		
3331		0.06	1	BR			serine-kinase	Unknown		
3397		0.05	2	BR,PI		1q35	rhesus polypeptide Rh6dd	Unknown		
3967		0.06			7	7q21	multidrug resistance protein 3	Unknown		
4020		0.02			20	20q13.1	lysosomal protective protein	Identical	Human	phospholipid transfert protein
4231		0.05	2	HL,PI		6q25-q27	T-complex protein 1 alpha subunit	Related	Zoogloea r.	acetyl-CoA acetyltransferase
4356	0.04	0.02			17		D38	Unknown		
5024		0.01			5	5q13	beta-hexosaminidase beta-subunit	Related	Thermus a.	elongation factor G
5139		0.01			19		GENX-280	Unknown		
5306		0.01	1	BR	19		RSU-1	Unknown		
5437		0.01	1	HI			rapamycin-binding protein	Unknown		
5681		0.01	2	BR	3	3p25	c-raf-1	Unknown		
6104		0.01					GENX-4318	Unknown		

Appendix 2A: eSTS markers (*partial*)

2	D14S658E	CGCACCTTCTTCTCTTGCTC	ACCAACAGGCTCTTGGATGAC	128	1300
4	D7S2134E	CCTGACGGCTCCCTAAAC	TCTTACGCATTCTTGGCACAC	205	205
7	DXS7012E	CTACTATCAATCCTACCCAGAG	GGTTTTCCTTTGGCGTATGAC	101	101
8	D5S1640E	CAGCCTGTGTTTGAACCACC	GTTTATAGCACCCGTCTCTCC	110	110
9	D14S694E	CCCGACAGTTTCCAAATGTTC	CATCTCTTACATCTCTCACCC	107	107
11	DXS7034E	CCAAACCCCAAGTCCTCAC	GAAAGGTCTTCCCAGAGACAC	159	159
13	D2S1656E	CACCTCCTACTCCCACAC	AGAAAGCCGAGAGGAAATTGC	196	196
14	D6S1187E	GCTTTGGTTGTTTGCTTTCTGG	CTGAAGATGGATGGGATTGTG	88	88
15	D4S2553E	CCCTGCCCTCACTCTAAG	ACTGCTGATACATGCAACAACC	168	168
18	D12S1246E	CCTGTTGATGGAGTGGTGTG	GCTGCTGATGCTGAGAAAGTG	221	221
19	D12S1205E	GGAAGAGACAATCACAACTCAC	CACTCTGCTCTGATCTTTTAGG	105	105
20	D2S1723E	AACTTGGAATTACTGGGTGGG	ACTAGGATGGGGACTGGG	174	174
21	D17S1406E	AGCCTCACCCATCAGACC	TCCCATTTGTTTCTCACAGCC	134	134

Appendix 2B: Accession number cross-index (*partial*)

1	F00009	F00483	F01121		
2	F00039	F00516	F00635	F00990	F01167
2	#Z17415	#Z21830	#Z21832		
3	F02961	F02985	F00590	F00705	F06280
3	F06696	F06716	F08458	#M62187	#Z17832
4	F01528	Z39764	F05275	Z43704	
5	Z39362	Z43290			
6	Z39364	F06437			
7	Z39365	Z43292			
8	Z39368	Z43293			
9	Z39370	Z43295	#T10740		
10	F02728	F06443	#D20842		

2306	#T16746				
2307	Z41275	Z45589	#T03897		
2308	Z41278	Z45590			
2309	Z41280	Z45591			
2310	Z41281	F08269			
2311	Z45594				
2312	F01467	F02587	F02676	Z40181	Z41950
2312	F06297	Z43223	Z44209	#T03617	#T17387
2313	F04500	F08271			
2314	Z41284	Z45596			
2316	Z39060	F04503	Z45597	#T08590	
2317	Z41285	Z45598			