

Similar gene expression profiles do not imply similar tissue functions

Itai Yanai^{1,2}, Jan O. Korbel^{3,4}, Stephanie Boue³, Shannon K. McWeeney⁵, Peer Bork^{3,4} and Martin J. Lercher^{3,5}

¹Department of Molecular Genetics, Weizmann Institute of Science, 76100 Rehovot, Israel

²Department of Molecular and Cellular Biology, Harvard University, MA 02138, USA

³European Molecular Biology Laboratory, Meyerhofstraße 1, 69117 Heidelberg, Germany

⁴Max-Delbrück-Center for Molecular Medicine, 13092 Berlin-Buch, Germany

⁵Division of Biostatistics, Department of Public Health and Preventative Medicine, Oregon Health and Science University, Portland, OR 97239, USA

⁶Department of Biology and Biochemistry, University of Bath, Bath, UK, BA2 7AY

Corresponding author: Bork, P. (bork@embl-heidelberg.de).

*These authors contributed equally to this work.

Materials and methods

Mouse gene expression data

We used Affymetrix microarray data from a recent thorough analysis of the mouse and human transcriptomes [1]. We selected all 54 adult mouse non-cancer samples. The raw intensity data were transformed to normalized expression levels with the robust multi-array average (RMA) low-level algorithm [2] implemented in the BioConductor package [3]. We used standard settings, including perfect match (PM) only, model-based background and quantile normalization across experiments [4]. Similar results were obtained using the microarray analysis suite (MAS5) function followed by log-transformation to calculate expression levels (data not shown).

For the comparison of mouse and human transcriptomes, we selected a subset of 26 tissues surveyed in both species [1] (amygdala, hypothalamus, olfactory bulb, cerebellum, spinal cord, trigeminal, dorsal root ganglion, pituitary gland, lymph node, bone marrow, thymus, liver, kidney, heart, skeletal muscle, lung, trachea, tongue, adrenal gland, prostate, salivary gland, thyroid, uterus, ovary, placenta and testis).

Mappings of probe sets to Ensembl genes were obtained from UCSC (<http://genome.ucsc.edu>). Only probe sets that match exactly one Ensembl gene were retained. Approximately 10% of genes are represented by more than one probe set. To avoid biases in expression profiles owing to heterogeneity in the probe sets, we took an arbitrary sample as an exemplar of the gene (by selecting the first probe set in the published order of sets; for the two human chips, we preferentially chose probe sets on the gnGNF1Ba chip). To test if the latter procedure affected our results, we employed a second data set where we averaged over probe sets whenever more than one was available. The tree topology obtained from this data set was identical to the one shown in Figure 1.

Human gene expression data

Mouse-human one-to-one orthologs were obtained using SMART [5]. Different probe sets on microarray chips have different hybridization efficiencies. If results are compared across different chip designs, additional normalizations have to be performed to account for this (I. Yanai and M.J. Lercher, unpublished). Thus, for every gene, each replicate of the human tissues was multiplied by (mean across all human and mouse data sets for this gene) / (mean across all human data sets for this gene) and analogous for the corresponding mouse data.

This normalization had not been performed in a previous publication [6], which led to inflated distances between human and mouse expression profiles. Performing the additional normalization step does not change the major conclusions of Ref. [6] (that much of gene expression differences are neutral in nature). However, it does remove the dichotomy of expression between human and

mouse tissues. Thus, homologous human and mouse tissues do in fact cluster, contrary to Figure 2 of Ref. [6]: any human tissue is more similar to its corresponding mouse tissue than to any other human tissue.

These procedures resulted in expression levels for a set of $N = 16\,004$ Ensembl genes across two replicates in 54 mouse tissues, and a second set of $N = 8\,323$ orthologous Ensembl genes across two replicates in 26 homologous human and mouse tissues. Here, each 'tissue' represents an organ, a tissue sample or a cell type. Each organ can contain several tissues, and each tissue usually comprises several cell types, representing an unavoidable aggregation of data; however, this would tend to obscure rather than highlight any of the patterns found.

Expression distances, tree reconstruction and bootstrap analysis

We calculated Euclidean distances between tissue expression vectors, with each dimension corresponding to one gene. Except for the replicate analysis in Figure S1, distances were calculated after averaging expression values across replicates. Trees were constructed from these distances using neighbor joining as implemented in MEGA2 [7]. Similar results were obtained using squared Euclidean distances (data not shown).

We first performed a test on overall quality of the replicate experiments. For each tissue in turn, we constructed the tree based on the first replicate for this tissue, and replicate averages for all other tissues; the same was repeated for the second replicate. The experiments for this tissue passed our test if both replicates resulted in identical tissue tree topologies. Seven out of 54 mouse tissues failed this test (dorsal striatum, retina, olfactory bulb, pancreas, ovary, spleen and umbilical cord); these were excluded from all phylogenetic analyses (Figure 1, Tables 1, 2 and S1). Twelve out of 26 human tissues failed the test (olfactory bulb, trigeminal, lymph node, thymus, liver, kidney, heart, skeletal muscle, tongue, prostate, thyroid and uterus). Only one out of the remaining 14 tissues had biological replicates (i.e. samples taken from different individuals); all other tissues had replicates based on identical biological samples (technological replicates). The rejection rate for biological replicates was 80% (4/5), whereas it was 38% (8/21) for technological replicates. We concluded that most human tissue samples have large heterogeneity across individuals. This is not surprising, because human tissue donors differ by age, sex and cause of death [1]; conversely, mouse samples were always pooled across four male and three female mice, sacrificed at a fixed age. Consequently, the available human data appeared of insufficient quality to reconstruct a reliable human tree of tissue relationships.

To examine the robustness of the tree topology for mouse tissues in Figure 1, we created 1000 bootstrap samples of the expression matrix. This was done by randomly drawing genes (with replacement) from the original expression matrix until the sample had reached the same size as the original. We then calculated distance matrices from the bootstrap samples. These were converted to neighbor-joining trees, which were reconstructed and summarized in Phylip [8]. Each value in Figure 1 reports the percentage of bootstrap samples containing the tissue dichotomy represented by the branch.

Selection of gene subsets

In order to reject the hypothesis that the significant clustering of related tissues is only due to similarity of function, we selected and analyzed various subsets of genes. The sets were created following a multi-step procedure: first PubMed (<http://www.ncbi.nlm.nih.gov>) was checked to see whether a depository of genes exists for each selected subclass. If not, a keyword search was done on the Ensembl definition lines; these are based on either Swissprot or Refseq annotations, which are both manually curated. The resulting gene list was cleaned using a literature check for every gene, making sure that it indeed belongs to the selected gene set. Genes whose sole evidence for function in the tissue in question comes from expression studies were excluded to avoid circularity. Tables with Ensembl gene identifiers and annotations for the gene subsets are available as Table S2.

In order to avoid biases caused by differential abundances of the whole subset of genes in specific tissues, we applied a tissue-normalization to the oxidative phosphorylation and spliceosomal complex sets, respectively. We divided each tissue expression vector by the mean expression level of the subset in this tissue. Thus, in the normalized expression matrices, each tissue had the same total (scaled) intensity of oxidative phosphorylation genes and spliceosomal complex genes, respectively.

Estimation of cluster support for Table 1

We estimated the strength of support for each individual cluster highlighted in Figure 1 for the data set as a whole, and for each subset of genes. For each individual gene and each cluster in Figure 1, we calculated the mean absolute distance of expression (i) for tissue pairs taken from the

cluster (within-cluster distance), and (ii) for tissue pairs with one tissue inside the cluster and the other tissue outside (across-cluster distance). We then calculated the percentage f of genes that support the cluster, i.e., those genes for which within-cluster distance < across-cluster distance (Table 1). A one-sample binomial test with probability of 'success' at 0.5 was used to determine statistical significance.

Tissue associations for genes of different ages

We obtained estimates of phyletic distribution for Ensembl genes from Ref. [9], which classified genes as universal (present in eukaryotes and prokaryotes), eukaryote-specific (present in mammals and yeast), metazoan-specific, and mammal-specific. Phyletic annotations were mapped to the probesets as described above. For each phyletic group, we then calculated support values for the clusters in Figure 1 as described for Table 1. This was done first including all genes, and secondly including only those genes in the functionally well characterized subsets (Table S2). We restricted the latter analysis to non-functional variation in the examined tissues, by excluding cluster C1 (brain), and by scaling the oxidative phosphorylation and spliceosomal complex subsets as described above.

We used a two-way analysis of variance (ANOVA) test to examine the interaction between phyletic age and tissue for the genes from our literature derived sets (brain-specific and germ cell formation). Based on the extent of phyletic annotation, only the brain-specific set appeared to be adequately powered. There was a significant interaction between age and tissue ($P = 0.003$). Upon closer examination, it was seen that this interaction was driven by the fact that the significant difference in expression for these genes was only seen in the comparison of metazoan and mammalian ages and only for the brain-specific gene set's gene expression from tissues of the central nervous system. In other words, there was no difference in gene expression between genes of different phyletic age in non-brain tissues.

Ectopic expression and genomic neighbourhood

First, tissue specific genes were identified as those with a ratio of maximum to mean of expression of at least 1.5. Within such genes, tissue specificity was assigned to those tissues with expression at least one standard deviation greater than the mean expression for that profile. Next, for each gene/tissue combination we computed the genomic distance to the nearest gene with specificity for that tissue. To query for a correlation between genomic distances and gene expression across tissues, we normalized the expression across ectopic tissues (for the gene sets) by the mean. Here, ectopic tissues are all tissues except cluster C1 (the CNS) for brain-specific genes, and all tissues except testis and oocyte for germ cell formation genes.

Expression evolution and synteny

Gene positions were obtained from Ensembl for human (build 34) and mouse (build 32). Only genes with orthologs in both species (defined as above) were considered. If a gene had the same direct neighbors in both species, it was classified as located in a syntenic region; if both neighbors differed, it was classified as non-syntenic. For each gene, we calculated Pearson's correlation coefficient between human and mouse expression across the 26 tissues common to both data sets. Correlation coefficients were then averaged separately for syntenic ($N = 6703$) and non-syntenic ($N = 201$) genes. Correlation coefficients were compared with Fisher's Z transformation. The result was not significant ($P = 0.62$).

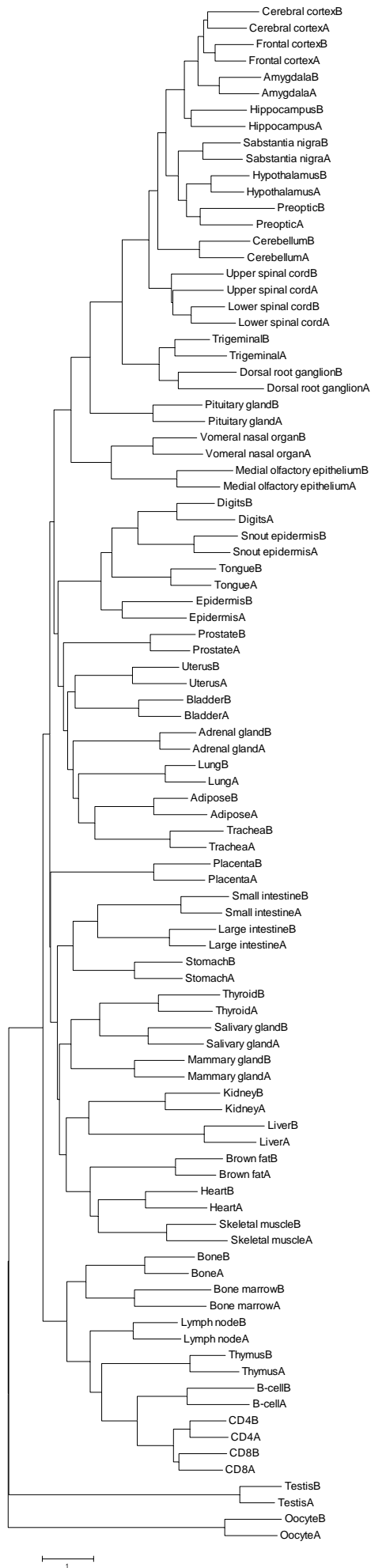


Figure S1. Neighbor joining tree constructed from Euclidean distances between gene expression vectors, including two replicates for each tissue. 7 tissues with inconsistent replicates have been removed.

Table S1. Support for tissue clusters in from gene sets

Gene set	<i>N</i>	1	2	3	4	5	6	7	8	9
All Genes	16 004	0.47****	0.59****	0.71****	0.61****	0.69****	0.67****	0.77****	0.69****	0.7****
Germ cell formation	39	0.6****	0.74	0.67	0.77**	0.85	1.01	0.78	0.88	0.94
Brain	29	-	0.52*	0.39*	0.45****	0.43***	0.5***	0.77	0.57**	0.52****
Oxidative phosphorylation	56	0.52****	0.63*	0.53**	0.78**	0.48****	0.72***	0.52**	0.65****	0.76****
Spliceosome	50	0.5****	0.53***	0.4****	0.72****	0.68****	0.7****	0.88	0.77**	0.68****

Gene expression distances within tissue clusters C1-C10 (cf. Figure 1) compared to the expression distances across clusters. The majority of the distances within clusters are significantly smaller than the distances across clusters. Displayed values represent ratios of (mean absolute expression distance among tissues in a cluster) / (mean absolute expression distance between tissues from a cluster and tissues outside the cluster). Asterisks and color intensity indicate significance levels from t-tests (*0.05; **0.01; ***0.001; ****0.0001). *N* represents the number of genes in the set.

Table S2. Functionally well characterised gene sets used in the analyses**1. Germ cell formation genes (meiosis + post-meiosis genes)**

Ensembl gene identifier	Description	Ref
ENSMUSG00000002324	Meiotic recombination protein REC8-like 1 (Cohesin Rec8p). [Source:Uniprot/SWISSPROT;Acc:Q8C5S7]	[10]
ENSMUSG00000004948	Zona pellucida sperm-binding protein 3 precursor (Zona pellucida glycoprotein ZP3) (Sperm receptor) (Zona pellucida protein C). [Source:Uniprot/SWISSPROT;Acc:P10761]	[11]
ENSMUSG00000005493	MutS protein homolog 4 (mMsh4). [Source:Uniprot/SWISSPROT;Acc:Q99MT2]	[12]
ENSMUSG00000005883	Meiotic recombination protein SPO11. [Source:Uniprot/SWISSPROT;Acc:Q9WTK8]	[13]
ENSMUSG00000010592	Deleted in azoospermia-like (DAZ-like autosomal) (Deleted in azoospermia-like 1). [Source:Uniprot/SWISSPROT;Acc:Q64368]	[14]
ENSMUSG00000018554	Y box protein 2; Y-box protein MSY2. [Source:RefSeq;Acc:NM_016875]	[15]
ENSMUSG00000020059	Synaptonemal complex protein 3 (SCP-3 protein). [Source:Uniprot/SWISSPROT;Acc:P70281]	[12]
ENSMUSG00000020380	RAD50 homolog. [Source: RefSeq (NM_009012)]	[16]
ENSMUSG00000021245	Similar to mutL (E. coli) homolog 3. [Source:Uniprot/SPTREMBL;Acc:Q99L38]	[10]
ENSMUSG00000021758	DEAD-box protein 4 (VASA homolog) (Mvh). [Source:Uniprot/SWISSPROT;Acc:Q61496]	[17]
ENSMUSG00000022039	ADAM 2 precursor (A disintegrin and metalloproteinase domain 2) (Fertilin beta subunit) (PH-30) (PH30) (PH30-beta). [Source:Uniprot/SWISSPROT;Acc:Q60718]	[18]
ENSMUSG00000022432	Structural maintenance of chromosomes 1-like 2 protein	[19]
ENSMUSG00000022501	Sperm protamine P1 (Cysteine-rich protamine). [Source:Uniprot/SWISSPROT;Acc:P02319]	[20]
ENSMUSG00000022622	Acrosin precursor (EC 3.4.21.10). [Source:Uniprot/SWISSPROT;Acc:P23578]	[21]
ENSMUSG00000022774	Nuclear cap binding protein subunit 2 (20 kDa nuclear cap binding protein) (NCBP 20 kDa subunit) (CBP20). [Source:Uniprot/SWISSPROT;Acc:Q9CQ49]	[22]

Ensembl gene identifier	Description	Ref
ENSMUSG00000024734	Zona pellucida sperm-binding protein 1 precursor (Zona pellucida glycoprotein 1) (Zp-1). [Source:Uniprot/SWISSPROT;Acc:Q62005]	[11]
ENSMUSG00000027526	Synaptonemal complex protein 2 (SCP-2 protein) (Synaptonemal complex lateral element protein) (Fragment). [Source:Uniprot/SWISSPROT;Acc:Q9CUU3]	[23]
ENSMUSG00000027855	Synaptonemal complex protein 1 (SCP-1 protein). [Source:Uniprot/SWISSPROT;Acc:Q62209]	[23]
ENSMUSG00000028845	Tektin 2 (Tektin-t). [Source:Uniprot/SWISSPROT;Acc:Q922G7]	[24]
ENSMUSG00000029601	MEI5	[25]
ENSMUSG00000029613	PMS1 protein homolog 2 (DNA mismatch repair protein PMS2). [Source:Uniprot/SWISSPROT;Acc:P54279]	[26]
ENSMUSG00000029736	OG2 homeobox; newborn ovary homeobox gene. [Source:RefSeq;Acc:NM_130869]	[15]
ENSMUSG00000029848	Stimulated by retinoic acid gene 8. [Source:RefSeq;Acc:NM_009292]	[27]
ENSMUSG00000030001	Factor in the germline alpha; FIG alpha. [Source:RefSeq;Acc:NM_012013]	[28]
ENSMUSG00000030344	A-kinase anchor protein 3 (Protein kinase A anchoring protein 3) (PRKA3) (A-kinase anchor protein 110 kDa) (AKAP 110). [Source:Uniprot/SWISSPROT;Acc:O88987]	[29]
ENSMUSG00000030528	Bloom's syndrome protein homolog (EC 3.6.1.-) (mBLM). [Source:Uniprot/SWISSPROT;Acc:O88700]	[10]
ENSMUSG00000030911	Zona pellucida sperm-binding protein 2 precursor (Zona pellucida glycoprotein ZP2) (Zona pellucida protein A). [Source:Uniprot/SWISSPROT;Acc:P20239]	[11]
ENSMUSG00000031518	Testis spermatocyte apoptosis-related gene 2 protein (Testis and spermatogenesis cell related protein 2) (Spermatogenesis related gene 2 protein). [Source:Uniprot/SWISSPROT;Acc:Q8K3V1]	[30]
ENSMUSG00000031553	A disintegrin and metalloprotease domain 3 (cyritestin); cyritestin. [Source:RefSeq;Acc:NM_009619]	[31]
ENSMUSG00000032498	DNA mismatch repair protein Mlh1 (MutL protein homolog 1). [Source:Uniprot/SWISSPROT;Acc:Q9JK91]	[12]
ENSMUSG00000032937	Follicle stimulating hormone receptor precursor (FSH-R) (Follitropin receptor). [Source:Uniprot/SWISSPROT;Acc:P35378]	[32]

Ensembl gene identifier	Description	Ref
ENSMUSG00000033644	Piwi like homolog 2; Miwi like. [Source:RefSeq;Acc:NM_021308]	[33]
ENSMUSG00000034248	Mscp: mitochondrial solute carrier protein. [Source: RefSeq (NM_026331)]	[34]
ENSMUSG00000038015	Sperm histone P2 precursor (Protamine MP2). [Source:Uniprot/SWISSPROT;Acc:P07978]	[34]
ENSMUSG00000038498	Sperm ion channel. [Source:Uniprot/SWISSPROT, Acc:Q91ZR5]	[35]
ENSMUSG00000040013	FK506-binding protein 6 (EC 5.2.1.8) (Peptidyl-prolyl cis-trans isomerase) (PPIase) (Rotamase) (36 kDa FK506 binding protein) (FKBP- 36) (Immunophilin FKBP36). [Source:Uniprot/SWISSPROT;Acc:Q91XW8]	[10]
ENSMUSG00000043065	SPO13	[36]
ENSMUSG00000050089	A-kinase anchor protein 4 precursor (Major fibrous sheath protein) (FSC1) (AKAP82) (p82). [Source:Uniprot/SWISSPROT;Acc:Q60662]	[37]
ENSMUSG00000056436	Cytochrome c, testis-specific. [Source:Uniprot/SWISSPROT;Acc:P00015]	[38]

2. Oxidative phosphorylation complex genes

Complex	Ensembl gene identifier	Description	Refs
1	ENSMUSG00000000399	NADH-ubiquinone oxidoreductase 39 kDa subunit, mitochondrial precursor (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-39KD) (CI-39KD). [Source:Uniprot/SWISSPROT;Acc:Q9DC69]	[9]
1	ENSMUSG000000002416	NADH-ubiquinone oxidoreductase AGGG subunit, mitochondrial precursor (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-AGGG) (CI-AGGG). [Source:Uniprot/SWISSPROT;Acc:Q9CPU2]	[9]
1	ENSMUSG000000013593	NADH-ubiquinone oxidoreductase 49 kDa subunit, mitochondrial precursor (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-49KD) (CI-49KD). [Source:Uniprot/SWISSPROT;Acc:Q91WD5]	[9]
1	ENSMUSG000000014294	NADH-ubiquinone oxidoreductase B8 subunit (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-B8) (CI-B8). [Source:Uniprot/SWISSPROT;Acc:Q9CQ75]	[9]
1	ENSMUSG000000022354	NADH-ubiquinone oxidoreductase B22 subunit (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-B22) (CI-B22). [Source:Uniprot/SWISSPROT;Acc:Q9CQJ8]	[9]
1	ENSMUSG000000023089	NADH-ubiquinone oxidoreductase 13 kDa-B subunit (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-13Kd-B) (CI-13Kd-B) (Complex I subunit B13). [Source:Uniprot/SWISSPROT;Acc:Q9CPP6]	[9]
1	ENSMUSG000000024099	NADH-ubiquinone oxidoreductase 24 kDa subunit, mitochondrial precursor (EC 1.6.5.3) (EC 1.6.99.3). [Source:Uniprot/SWISSPROT;Acc:Q9D6J6]	[9]
1	ENSMUSG000000025968	NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial precursor (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-75Kd) (CI-75Kd). [Source:Uniprot/SWISSPROT;Acc:Q91VD9]	[9]
1	ENSMUSG000000026032	NADH-ubiquinone oxidoreductase B12 subunit (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-B12) (CI-B12). [Source:Uniprot/SWISSPROT;Acc:Q9CQZ6]	[9]
1	ENSMUSG000000026895	NADH-ubiquinone oxidoreductase 19 kDa subunit (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-19KD) (CI-19KD) (Complex I-PGIV) (CI-PGIV). [Source:Uniprot/SWISSPROT;Acc:Q9DCJ5]	[9]
1	ENSMUSG000000027673	NADH-ubiquinone oxidoreductase SGDH subunit, mitochondrial precursor (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-SGDH) (CI-SGDH). [Source:Uniprot/SWISSPROT;Acc:Q9CQH3]	[9]
1	ENSMUSG000000033938	NADH-ubiquinone oxidoreductase B18 subunit (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-B18) (CI-B18). [Source:Uniprot/SWISSPROT;Acc:Q9CR61]	[9]

Complex	Ensembl gene identifier	Description	Refs
1	ENSMUSG00000035674	NADH-ubiquinone oxidoreductase B9 subunit (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-B9) (CI-B9). [Source:Uniprot/SWISSPROT;Acc:Q9CQ91]	[9]
1	ENSMUSG00000037916	NADH-ubiquinone oxidoreductase 51 kDa subunit, mitochondrial precursor (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-51KD) (CI-51KD). [Source:Uniprot/SWISSPROT;Acc:Q91YT0]	[9]
1	ENSMUSG00000043062	NADH-ubiquinone oxidoreductase 15 kDa subunit (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-15 kDa) (CI-15 kDa). [Source:Uniprot/SWISSPROT;Acc:Q99LY9]	[9]
1	ENSMUSG00000005510	NADH-ubiquinone oxidoreductase 30 kDa subunit, mitochondrial precursor (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-30KD) (CI-30KD). [Source:Uniprot/SWISSPROT;Acc:Q9DCT2]	[9]
1	ENSMUSG00000020153	NADH-ubiquinone oxidoreductase 20 kDa subunit, mitochondrial precursor (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-20KD) (CI-20KD) (PSST subunit). [Source:Uniprot/SWISSPROT;Acc:Q9DC70]	[9]
1	ENSMUSG00000021764	NADH-ubiquinone oxidoreductase 18 kDa subunit, mitochondrial precursor (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-18 kDa) (CI-18 kDa) (Complex I- AQDQ) (CI-AQDQ). [Source:Uniprot/SWISSPROT;Acc:Q9CXZ1]	[9]
1	ENSMUSG00000026260	NADH-ubiquinone oxidoreductase 42 kDa subunit, mitochondrial precursor (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-42KD) (CI-42KD). [Source:Uniprot/SWISSPROT;Acc:Q99LC3]	[9]
1	ENSMUSG00000040048	NADH-ubiquinone oxidoreductase PDSW subunit (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-PDSW) (CI-PDSW). [Source:Uniprot/SWISSPROT;Acc:Q9DCS9]	[9]
1	ENSMUSG00000020022	NADH-ubiquinone oxidoreductase subunit B17.2 (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-B17.2) (CI-B17.2) (CIB17.2). [Source:Uniprot/SWISSPROT;Acc:Q7TMF3]	[9]
1	ENSMUSG00000041881	NADH-ubiquinone oxidoreductase subunit B14.5a (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-B14.5a) (CI-B14.5a). [Source:Uniprot/SWISSPROT;Acc:Q9Z1P6]	[9]
1	ENSMUSG00000030647	NADH-ubiquinone oxidoreductase subunit B14.5b (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-B14.5b) (CI-B14.5b). [Source:Uniprot/SWISSPROT;Acc:Q9CQ54]	[9]
1	ENSMUSG00000022450	NADH-ubiquinone oxidoreductase B14 subunit (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-B14) (CI-B14). [Source:Uniprot/SWISSPROT;Acc:Q9CQZ5]	[9]
1	ENSMUSG00000016427	NADH-ubiquinone oxidoreductase MWFE subunit (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-MWFE) (CI-MWFE). [Source:Uniprot/SWISSPROT;Acc:O35683]	[9]

Complex	Ensembl gene identifier	Description	Refs
2	ENSMUSG00000000171	Succinate dehydrogenase complex, subunit D, integral membrane protein. [Source:RefSeq;Acc:NM_025848]	[9]
2	ENSMUSG000000021577	Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial [Precursor] Q8K2B3	[9]
2	ENSMUSG000000009863	Succinate dehydrogenase lp subunit. [Source:RefSeq;Acc:NM_023374]	[9]
3	ENSMUSG000000021520	Ubiquinol-cytochrome C reductase complex 14 kDa protein (EC 1.10.2.2) (Complex III subunit VI). [Source:Uniprot/SWISSPROT;Acc:Q9D855]	[9]
3	ENSMUSG000000022551	Cytochrome c1, heme protein, mitochondrial precursor (Cytochrome c-1). [Source:Uniprot/SWISSPROT;Acc:Q9D0M3]	[9]
3	ENSMUSG000000025651	Ubiquinol-cytochrome C reductase complex core protein I, mitochondrial precursor (EC 1.10.2.2). [Source:Uniprot/SWISSPROT;Acc:Q9CZ13]	[9]
3	ENSMUSG000000030884	Ubiquinol-cytochrome C reductase complex core protein 2, mitochondrial precursor (EC 1.10.2.2) (Complex III subunit II). [Source:Uniprot/SWISSPROT;Acc:Q9DB77]	[9]
3	ENSMUSG000000044894	Ubiquinol-cytochrome C reductase complex ubiquinone-binding protein QP-C (EC 1.10.2.2) (Ubiquinol-cytochrome C reductase complex 9.5 kDa protein) (Complex III subunit VII). [Source:Uniprot/SWISSPROT;Acc:Q9CQ69]	[9]
3	ENSMUSG000000038462	Ubiquinol-cytochrome c reductase, Rieske iron-sulfur polypeptide 1. [Source:RefSeq;Acc:NM_025710]	[9]
3	ENSMUSG000000020163	Ubiquinol-cytochrome c reductase subunit. [Source: RefSeq (NM_025650)]	[9]
4	ENSMUSG000000030785	Cytochrome c oxidase polypeptide VIa-heart, mitochondrial precursor (EC 1.9.3.1) (COXVIAH). [Source:Uniprot/SWISSPROT;Acc:P43023]	[9]
4	ENSMUSG000000032330	Cytochrome c oxidase polypeptide VIIa-liver/heart, mitochondrial precursor (EC 1.9.3.1) (Cytochrome c oxidase subunit VIIa-L). [Source:Uniprot/SWISSPROT;Acc:P48771]	[9]
4	ENSMUSG000000037359	Cytochrome c oxidase polypeptide Vb, mitochondrial precursor (EC 1.9.3.1). [Source:Uniprot/SWISSPROT;Acc:P19536]	[9]
4	ENSMUSG000000000088	Cytochrome c oxidase polypeptide Va, mitochondrial precursor (EC 1.9.3.1). [Source:Uniprot/SWISSPROT;Acc:P12787]	[9]
4	ENSMUSG000000014313	Cytochrome c oxidase polypeptide VIc (EC 1.9.3.1). [Source:Uniprot/SWISSPROT;Acc:Q9CPQ1]	[9]

Complex	Ensembl gene identifier	Description	Refs
4	ENSMUSG00000025488	Cytochrome c oxidase polypeptide VIII-heart, mitochondrial precursor (EC 1.9.3.1) (Cytochrome c oxidase subunit 8-1). [Source:Uniprot/SWISSPROT;Acc:P48772]	[9]
4	ENSMUSG00000031231	Cytochrome c oxidase polypeptide VIIb, mitochondrial precursor (EC 1.9.3.1). [Source:Uniprot/SWISSPROT;Acc:P56393]	[9]
4	ENSMUSG00000035885	Cytochrome c oxidase polypeptide VIII-liver, mitochondrial precursor (EC 1.9.3.1) (Cytochrome c oxidase subunit 8-2). [Source:Uniprot/SWISSPROT;Acc:Q64445]	[9]
4	ENSMUSG00000036751	Cytochrome c oxidase polypeptide VIb (EC 1.9.3.1) (Cytochrome c oxidase subunit AED). [Source:Uniprot/SWISSPROT;Acc:P56391]	[9]
4	ENSMUSG00000041697	Cytochrome c oxidase polypeptide VIa-liver, mitochondrial precursor (EC 1.9.3.1). [Source:Uniprot/SWISSPROT;Acc:P43024]	[9]
5	ENSMUSG00000022956	ATP synthase oligomycin sensitivity conferral protein, mitochondrial precursor (EC 3.6.3.14) (OSCP). [Source:Uniprot/SWISSPROT;Acc:Q9DB20]	[9]
5	ENSMUSG00000025781	ATP synthase gamma chain, mitochondrial precursor (EC 3.6.3.14). [Source:Uniprot/SWISSPROT;Acc:Q91VR2]	[9]
5	ENSMUSG00000025393	ATP synthase beta chain, mitochondrial precursor (EC 3.6.3.14). [Source:Uniprot/SWISSPROT;Acc:P56480]	[9]
5	ENSMUSG00000000563	ATP synthase B chain, mitochondrial precursor (EC 3.6.3.14). [Source:Uniprot/SWISSPROT;Acc:Q9CQQ7]	[9]
5	ENSMUSG00000034566	ATP synthase D chain, mitochondrial (EC 3.6.3.14). [Source:Uniprot/SWISSPROT;Acc:Q9DCX2]	[9]
5	ENSMUSG00000003072	ATP synthase, H ⁺ transporting, mitochondrial F1 complex, delta subunit. [Source:RefSeq;Acc:NM_025313]	[9]
5	ENSMUSG00000038690	ATP synthase f chain, mitochondrial (EC 3.6.3.14). [Source:Uniprot/SWISSPROT;Acc:P56135]	[9]
5	ENSMUSG00000038717	ATP synthase, H ⁺ transporting, mitochondrial F0 complex, subunit g; F1F0-ATP synthase g subunit. [Source:RefSeq;Acc:NM_013795]	[9]
5	ENSMUSG00000050856	ATP synthase e chain, mitochondrial (EC 3.6.3.14). [Source:Uniprot/SWISSPROT;Acc:Q06185]	[9]
5	ENSMUSG00000016252	ATP synthase epsilon chain, mitochondrial (EC 3.6.3.14). [Source:Uniprot/SWISSPROT;Acc:P56382]	[9]

Complex	Ensembl gene identifier	Description	Refs
5	ENSMUSG00000006057	ATP synthase lipid-binding protein, mitochondrial precursor (EC 3.6.3.14) (ATP synthase proteolipid P1) (ATPase protein 9) (ATPase subunit C). [Source:Uniprot/SWISSPROT;Acc:P48202]	[9]

3. Spliceosome

Ensembl gene identifier	Description	Ref
ENSMUSG00000002129	Splicing factor 3a, subunit 1. [Source:RefSeq;Acc:NM_026175]	[40]
ENSMUSG00000002455	U5 snRNP-associated 102 kDa protein (U5-102 kDa protein). [Source:Uniprot/SWISSPROT;Acc:Q91YR7]	[40]
ENSMUSG00000002477	Small nuclear ribonucleoprotein Sm D1 (snRNP core protein D1) (Sm-D1) (Sm-D autoantigen). [Source:Uniprot/SWISSPROT;Acc:P62314]	[40]
ENSMUSG00000004980	Heterogeneous nuclear ribonucleoproteins A2/B1 (hnRNP A2 / hnRNP B1). [Source:Uniprot/SWISSPROT;Acc:O88569]	[40]
ENSMUSG00000008333	U2 small nuclear ribonucleoprotein B. [Source:RefSeq;Acc:NM_021335]	[40]
ENSMUSG00000016921	Arginine/serine-rich splicing factor 6. [Source:RefSeq;Acc:NM_026499]	[40]
ENSMUSG00000018379	Splicing factor, arginine/serine-rich 1 (ASF/SF2). [Source:RefSeq;Acc:NM_173374]	[40]
ENSMUSG00000020018	Small nuclear ribonucleoprotein F (snRNP-F) (Sm protein F) (Sm-F) (SmF). [Source:Uniprot/SWISSPROT;Acc:P62306]	[40]
ENSMUSG00000020180	Small nuclear ribonucleoprotein Sm D3 (snRNP core protein D3) (Sm-D3). [Source:Uniprot/SWISSPROT;Acc:P62318]	[40]
ENSMUSG00000020211	Splicing factor 3A subunit 2 (Spliceosome associated protein 62) (SAP 62) (SF3a66). [Source:Uniprot/SWISSPROT;Acc:Q62203]	[40]
ENSMUSG00000020719	Probable RNA-dependent helicase p68 (DEAD-box protein p68) (DEAD-box protein 5) (DEAD-box RNA helicase DEAD1) (mDEAD1). [Source:Uniprot/SWISSPROT;Acc:Q61656]	[40]
ENSMUSG00000020929	116 kDa U5 small nuclear ribonucleoprotein component (U5 snRNP- specific protein, 116 kDa) (U5-116 kDa). [Source:Uniprot/SWISSPROT;Acc:O08810]	[40]
ENSMUSG00000021134	Splicing factor, arginine/serine-rich 5 (Pre-mRNA splicing factor SRP40) (Delayed-early protein HRS). [Source:Uniprot/SWISSPROT;Acc:O35326]	[40]
ENSMUSG00000021546	Heterogeneous nuclear ribonucleoprotein K. [Source:Uniprot/SWISSPROT;Acc:P61979]	[40]
ENSMUSG00000022019	Tudor domain containing protein 3. [Source:Uniprot/SWISSPROT;Acc:Q91W18]	[40]

Ensembl gene identifier	Description	Ref
ENSMUSG00000022478 (ENSMUSG00000061360)	PHD finger-like domain protein 5A (Splicing factor 3B associated 14 kDa protein) (SF3b14b). [Source:Uniprot/SWISSPROT;Acc:P83870]	[40]
ENSMUSG00000022774	Nuclear cap binding protein subunit 2 (20 kDa nuclear cap binding protein) (NCBP 20 kDa subunit) (CBP20). [Source:Uniprot/SWISSPROT;Acc:Q9CQ49]	[40]
ENSMUSG00000024040 (ENSMUSG00000061613)	Splicing factor U2AF 35 kDa subunit (U2 auxiliary factor 35 kDa subunit) (U2 snRNP auxiliary factor small subunit). [Source:Uniprot/SWISSPROT;Acc:Q9D883]	[40]
ENSMUSG00000024097	Splicing factor, arginine/serine-rich 7; splicing factor, arginine /serine-rich 7 (35kD). [Source:RefSeq;Acc:NM_146083]	[40]
ENSMUSG00000024217	U1 small nuclear ribonucleoprotein C (U1 snRNP protein C) (U1C protein) (U1-C). [Source:Uniprot/SWISSPROT;Acc:Q62241]	[40]
ENSMUSG00000024735	Nuclear matrix protein SNEV. [Source:RefSeq;Acc:NM_134129]	[40]
ENSMUSG00000025024	Survival of motor neuron-related splicing factor 30 (SMN-related protein) (30 kDa splicing factor SMNrp) (Survival motor neuron domain containing protein 1). [Source:Uniprot/SWISSPROT;Acc:Q8BGT7]	[40]
ENSMUSG00000025134	THO complex subunit 4 (Tho4) (RNA and export factor binding protein 1) (REF1-I) (Ally of AML-1 and LEF-1) (Aly/REF). [Source:Uniprot/SWISSPROT;Acc:O08583]	[40]
ENSMUSG00000025982	Splicing factor 3B subunit 1 (Spliceosome associated protein 155) (SAP 155) (SF3b155) (Pre-mRNA splicing factor SF3b 155 kDa subunit). [Source:Uniprot/SWISSPROT;Acc:Q99NB9]	[40]
ENSMUSG00000027620	RNA-binding region containing protein 2 (Coactivator of activating protein-1 and estrogen receptors) (Coactivator of AP-1 and ERs) (Transcription coactivator CAPER). [Source:Uniprot/SWISSPROT;Acc:Q8VH51]	[40]
ENSMUSG00000028639	Nuclease sensitive element binding protein 1 (Y box binding protein-1) (Y-box transcription factor) (YB-1) (CCAAT-binding transcription factor I subunit A) (CBF-A) (Enhancer factor I subunit A) (EFI-A) (DNA-binding protein B) (DBPB). [Source:Uniprot/SWISSPROT;Acc:P27817]	[40]
ENSMUSG00000028666 (ENSMUSG00000057124)	Heterogeneous nuclear ribonucleoprotein R. [Source: RefSeq (NM_028871)]	[40]
ENSMUSG00000028902	Splicing factor 3A subunit 3 (Spliceosome associated protein 61) (SAP 61) (SF3a60). [Source:Uniprot/SWISSPROT;Acc:Q9D554]	[40]
ENSMUSG00000029169	Putative pre-mRNA splicing factor RNA helicase (DEAH box protein 15). [Source:Uniprot/SWISSPROT;Acc:O35286]	[40]

Ensembl gene identifier	Description	Ref
ENSMUSG00000029538	Splicing factor, arginine/serine rich 9; splicing factor, arginine/serine rich 9 (25 kDa). [Source:RefSeq;Acc:NM_025573]	[40]
ENSMUSG00000030435 (ENSMUSG00000049919)	Splicing factor U2AF 65 kDa subunit (U2 auxiliary factor 65 kDa subunit) (U2 snRNP auxiliary factor large subunit). [Source:Uniprot/SWISSPROT;Acc:P26369]	[40]
ENSMUSG00000030512	U2 small nuclear ribonucleoprotein A' (U2 snRNP-A'). [Source:Uniprot/SWISSPROT;Acc:P57784]	[40]
ENSMUSG00000030795	RNA-binding protein FUS (Pigpen protein). [Source:Uniprot/SWISSPROT;Acc:P56959]	[40]
ENSMUSG00000030810 (ENSMUSG00000063511)	U1 small nuclear ribonucleoprotein 70 kDa (U1 SNRNP 70 kDa) (snRNP70) (Fragment). [Source:Uniprot/SWISSPROT;Acc:Q62376]	[40]
ENSMUSG00000031134	RNA binding motif protein, X chromosome. [Source:RefSeq;Acc:NM_011252]	[40]
ENSMUSG00000032178	Interleukin enhancer-binding factor 3. [Source:Uniprot/SWISSPROT;Acc:Q9Z1X4]	[40]
ENSMUSG00000032580	RNA binding motif protein 5. [Source:RefSeq;Acc:NM_148930]	[40]
ENSMUSG00000033732	Splicing factor 3b, subunit 3, 130kDa. [Source:RefSeq;Acc:NM_133953]	[40]
ENSMUSG00000034437	Splicing factor. [Source:Uniprot/SPTREMBL;Acc:Q8C3H6]	[40]
ENSMUSG00000037361	Pre-mRNA branch site protein p14. [Source:Uniprot/SWISSPROT;Acc:P59708]	[40]
ENSMUSG00000039148	Squamous cell carcinoma antigen recognized by T-cells 1. [Source:RefSeq;Acc:NM_016882]	[40]
ENSMUSG00000039630	Heterogenous nuclear ribonucleoprotein U; nuclear matrix protein sp120; scaffold attachment factor A. [Source:RefSeq;Acc:NM_016805]	[40]
ENSMUSG00000040725	EIB-55kDa associated protein 5. [Source:RefSeq;Acc:NM_144922]	[40]
ENSMUSG00000041133	Structural maintenance of chromosome 1-like 1 protein (SMC1alpha protein) (Chromosome segregation protein SmcB) (Sb1.8). [Source:Uniprot/SWISSPROT;Acc:Q9CU62]	[40]
ENSMUSG00000042502	CD2 antigen cytoplasmic tail-binding protein 2 (CD2 cytoplasmic domain binding protein) (CD2 tail binding protein). [Source:Uniprot/SWISSPROT;Acc:Q9CWK3]	[40]

Ensembl gene identifier	Description	Ref
ENSMUSG00000042699	ATP-dependent RNA helicase A (Nuclear DNA helicase II) (NDH II) (DEAH-box protein 9) (mHEL-5). [Source:Uniprot/SWISSPROT;Acc:O70133]	[40]
ENSMUSG00000045548 (ENSMUSG00000045728)	Splicing factor 3b, subunit 4; splicing factor 3b, subunit 4, 49kD; spliceosome associated protein 49. [Source: RefSeq (NM_153053)]	[40]
ENSMUSG00000049681 (ENSMUSG00000061029)	Splicing factor 3B subunit 5 (SF3b5) (Pre-mRNA splicing factor SF3b 10 kDa subunit). [Source:Uniprot/SWISSPROT;Acc:Q923D4]	[40]
ENSMUSG00000051210 (ENSMUSG00000049124)	Small nuclear ribonucleoprotein G (snRNP-G) (Sm protein G) (Sm-G) (SmG). [Source:Uniprot/SWISSPROT;Acc:P62308]	[40]
ENSMUSG00000054405	DnaJ (Hsp40) homolog, subfamily C, member 8. [Source:RefSeq;Acc:NM_172400]	[40]

4. Brain-specific genes

Ensembl gene identifier	Description	Refs
ENSMUSG00000003378	Glutamate receptor, ionotropic kainate 5 precursor (Glutamate receptor KA-2) (KA2) (Glutamate receptor gamma-2) (GluR gamma-2). [Source:Uniprot/SWISSPROT;Acc:Q61626]	[41]
ENSMUSG00000004892	Brevican core protein precursor. [Source:Uniprot/SWISSPROT;Acc:Q61361]	[42]
ENSMUSG00000004894	Hyaluronan and proteoglycan link protein 2 precursor (Brain link protein 1). [Source:Uniprot/SWISSPROT;Acc:Q9ESM3]	[43]
ENSMUSG00000007021	Synaptogyrin 3; Sgyr III; synaptogyrin III. [Source:RefSeq;Acc:NM_011522]	[44]
ENSMUSG00000007440	Protocadherin alpha 12; cadherin-related neuronal receptor 5; protocadherin alpha 13. [Source:RefSeq;Acc:NM_138663]	[45]
ENSMUSG000000020436	Gamma-aminobutyric-acid receptor gamma-2 subunit precursor (GABA(A) receptor). [Source:Uniprot/SWISSPROT;Acc:P22723]	[46]
ENSMUSG000000020466	Calmodulin-dependent protein kinase II beta M isoform (Fragment). [Source:Uniprot/SPTREMBL;Acc:Q9WVJ1]	[47]
ENSMUSG000000020591	Neurotensin receptor type 2 (NT-R-2) (Low-affinity levocabastine- sensitive neurotensin receptor) (NTRL). [Source:Uniprot/SWISSPROT;Acc:P70310]	[48]
ENSMUSG000000020932	Glial fibrillary acidic protein, astrocyte (GFAP). [Source:Uniprot/SWISSPROT;Acc:P03995]	[49]
ENSMUSG000000025777	Ganglioside-induced differentiation associated protein 1 (GDAP1). [Source:Uniprot/SWISSPROT;Acc:O88741]	[50]
ENSMUSG000000026587	Astrotactin 1 (Neuronal migration protein GC14) (Fragment). [Source:Uniprot/SWISSPROT;Acc:Q61137]	[51]
ENSMUSG000000026833	Noelin precursor (Neuronal olfactomedin-related ER localized protein) (Olfactomedin 1) (Pancortin). [Source:Uniprot/SWISSPROT;Acc:O88998]	[52]
ENSMUSG000000027096	Glutamate decarboxylase, 67 kDa isoform (EC 4.1.1.15) (GAD-67) (67 kDa glutamic acid decarboxylase). [Source:Uniprot/SWISSPROT;Acc:P48318]	[53]
ENSMUSG000000028020	Glycine receptor beta chain precursor (Glycine receptor 58 kDa subunit). [Source:Uniprot/SWISSPROT;Acc:P48168]	[54]
ENSMUSG000000028785	Neuron specific calcium-binding protein hippocalcin (P23K) (Calcium-binding protein BDR-2). [Source:Uniprot/SWISSPROT;Acc:P32076]	[55]

Ensembl gene identifier	Description	Refs
ENSMUSG00000029212	Gamma-aminobutyric-acid receptor beta-1 subunit precursor (GABA(A) receptor). [Source:Uniprot/SWISSPROT;Acc:P50571]	[56]
ENSMUSG00000029819	neuropeptide Y [Source:MarkerSymbol;Acc:MGI:97374]	[56]
ENSMUSG00000030519	Amyloid beta A4 precursor protein-binding family A member 2 (Neuron-specific X11L protein) (Neuronal Munc18-1-interacting protein 2) (Mint-2) (Adapter protein X11beta) (Fragment). [Source:Uniprot/SWISSPROT;Acc:P98084]	[57]
ENSMUSG00000031425	Myelin proteolipid protein (PLP) (Lipophilin). [Source:Uniprot/SWISSPROT;Acc:P60202]	[58]
ENSMUSG00000031517	Neuronal membrane glycoprotein M6-a (M6a). [Source:Uniprot/SWISSPROT;Acc:P35802]	[59]
ENSMUSG00000031760	Metallothionein-III (MT-III) (Growth inhibitory factor) (GIF). [Source:Uniprot/SWISSPROT;Acc:P28184]	[60]
ENSMUSG00000032174	Intercellular adhesion molecule-5 precursor (ICAM-5) (Telencephalin). [Source:Uniprot/SWISSPROT;Acc:Q60625]	[61]
ENSMUSG00000032517	Myelin-associated oligodendrocytic basic protein. [Source:RefSeq;Acc:NM_008614]	[62]
ENSMUSG00000039037	Alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase V (GD1 alpha synthase) (GalNAc alpha-2,6-sialyltransferase V) (ST6GalNAc V) (Sialyltransferase 7E). [Source:Uniprot/SWISSPROT;Acc:Q9QYJ1]	[63]
ENSMUSG00000040276	Protein kinase C and casein kinase substrate in neurons protein 1. [Source:Uniprot/SWISSPROT;Acc:Q61644]	[64]
ENSMUSG00000041607	Myelin basic protein (MBP) (Myelin A1 protein). [Source:Uniprot/SWISSPROT;Acc:P04370]	[65]
ENSMUSG00000046159	Muscarinic acetylcholine receptor M3 (Mm3 mAChR). [Source:Uniprot/SWISSPROT;Acc:Q9ERZ3]	[66]
ENSMUSG00000047261	Neuromodulin (Axonal membrane protein GAP-43) (PP46) (B-50) (Protein F1) (Calmodulin-binding protein P-57). [Source:Uniprot/SWISSPROT;Acc:P06837]	[67]
ENSMUSG00000049612	Oligodendrocyte-myelin glycoprotein precursor. [Source:Uniprot/SWISSPROT;Acc:Q63912]	[68]

References

1. Su, A.I., *et al.* (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci U S A* 101, 6062-6067
2. Irizarry, R.A., *et al.* (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4, 249-264
3. Gentleman, R.C., *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5, R80
4. Bolstad, B.M., *et al.* (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19, 185-193
5. Letunic, I., *et al.* (2004) SMART 4.0: towards genomic data integration. *Nucleic Acids Res* 32 Database issue, D142-144
6. Yanai, I., *et al.* (2004) Incongruent Expression Profiles between Human and Mouse Orthologous Genes Suggest Widespread Neutral Evolution of Transcription Control. *OMICS* 8, 15
7. Kumar, S., *et al.* (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17, 1244-1245
8. Felsenstein, J. (2002) PHYLIP (Phylogeny Inference Package) version 3.6a3. *Distributed by the author. Department of Genome Sciences, University of Washington, Seattle*
9. Freilich, S., *et al.* (2005) Relationship between the tissue-specificity of mouse gene expression and the evolutionary origin and function of the proteins. *Genome Biol* 6, R56
10. Svetlanov, A., and Cohen, P.E. (2004) Mismatch repair proteins, meiosis, and mice: understanding the complexities of mammalian meiosis. *Experimental cell research* 296, 71-79
11. Liang, L.F., *et al.* (1997) FIGalpha, a germ cell specific transcription factor involved in the coordinate expression of the zona pellucida genes. *Development* 124, 4939-4947
12. Escalier, D., *et al.* (2002) Sex-specific gene expression during meiotic prophase I: Xlr (X linked, lymphocyte regulated), not its male homologue Xmr (Xlr related, meiosis regulated), is expressed in oocytes. *Biology of reproduction* 67, 1646-1652
13. Hwang, S.Y., *et al.* (2001) Expression of genes involved in mammalian meiosis during the transition from egg to embryo. *Molecular Reproduction and Development* 59, 144-158
14. Ruggiu, M., *et al.* (1997) The mouse Dazla gene encodes a cytoplasmic protein essential for gametogenesis. *Nature* 389, 73-77
15. Rajkovic, A., *et al.* (2004) NOBOX deficiency disrupts early folliculogenesis and oocyte-specific gene expression. *Science* 305, 1157-1159
16. Eijpe, M., *et al.* (2003) Meiotic cohesin REC8 marks the axial elements of rat synaptonemal complexes before cohesins SMC1 β and SMC3. *The journal of Cell Biology* 160, 657-670
17. Shibata, N., *et al.* (2004) Mouse RanBPM is a partner gene to a germline specific RNA helicase, mouse vasa homolog protein. *Mol Reprod Dev* 67, 1-7
18. Cho, C., *et al.* (1998) Fertilization defects in mice from mice lacking fertilin β . *Science* 281, 1857-1859
19. Revenkova, E., *et al.* (2001) A novel meiosis-specific isoform of mammalian SMC1. *Molecular and cellular biology* 21, 6984-6998
20. Steger, K., *et al.* (2001) Round spermatids from infertile men exhibit decreased protamine-1 and -2 mRNA. *Human reproduction* 16, 709-716
21. Nayernia, K., *et al.* (2002) Proacrosin-deficient mice and zona pellucida modifications in an experimental model of multifactorial infertility. *Molecular human reproduction* 8, 434-440
22. McKee, A.H.Z., and Kleckner, N. (1997) Mutations in *Saccharomyces cerevisiae* that block meiotic prophase chromosome metabolism and coner cell cycle arrest at pachytene identify new meiosis-specific genes SAE1 and SAE3. *Genetics* 146, 817-834
23. Pelttari, J., *et al.* (2001) A metiotic chromosomal core consisting of cohesin complex proteins recruits DNA recombination proteins and promotes synapsis in the absence of an axial element in mammalian meiotic cells. *Molecular and cellular biology* 21, 5667-5677
24. Tanaka, H., *et al.* (2004) Mice deficient in the axonemal protein Tektin-t exhibit male infertility and immotile-cilium syndrome due to impaired inner arm dynein function. *molecular and cellular biology* 24, 7958-7964
25. Hayase, A., *et al.* (2004) A protein complex containing Mei5 and Sae3 promotes the assembly of the meiosis-specific RecA homolog Dmc1. *Cell* 119, 927-940
26. Baker, S.M., *et al.* (1995) Male mice defective in the DNA mismatch repair gene PMS2 exhibit abnormal chromosome synapsis in meiosis. *Cell* 82, 309-319
27. Menke, D.B., *et al.* (2003) Sexual differentiation of germ cells in XX mouse gonads occurs in an anterior-to-posterior wave. *Developmental Biology* 262, 303-312
28. Soyol, S.M., *et al.* (2000) FIGalpha, a germ cell-specific transcription factor required for ovarian follicle formation. *Development* 127, 4645-4654
29. Vijayaraghavan, S., *et al.* (1999) Isolation and molecular characterization of AKAP110, a novel sperm-specific protein kinase A-anchoring protein. *Mol Endocrinol* 13, 705-717
30. Shen, W., *et al.* (2002) Molecular cloning of SRG2, a mouse testis spermatocyte apoptosis-related gene. 34, 796-799
31. Shamsadin, R., *et al.* (1999) Male mice deficient for germ-cell cyritestin are infertile. *Biol Reprod* 61, 1445-1451
32. Tena-Sempere, M., *et al.* (1999) Molecular cloning of the mouse follicle-stimulating hormone receptor complementary deoxyribonucleic acid: functional expression of alternatively spliced variants and receptor inactivation by a C566T transition in exon 7 of the coding sequence. *Biology of reproduction* 60, 1515-1527
33. Kuramochi-Miyagawa, S., *et al.* (2004) Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development* 131, 839-849
34. Nayernia, K., *et al.* (2002) Asthenozoospermia in mice with targeted deletion of the sperm mitochondrion-associated cysteine-rich protein (Smcp) gene. *molecular and cellular biology* 22, 3046-3052
35. Ren, D., *et al.* (2001) A sperm ion channel required for sperm motility and male fertility. *Nature* 413, 603-609
36. Primig, M., *et al.* (2000) The core meiotic transcriptome in budding yeasts. *Nature genetics* 26, 415-423

37. Miki, K., *et al.* (2002) Targeted disruption of the Akap4 gene causes defects in sperm flagellum and motility. *Developmental Biology* 248, 331-342
38. Narisawa, S., *et al.* (2002) Testis-specific cytochrome c-null mice produce functional sperm but undergo early testicular atrophy. *molecular and cellular biology* 22
39. Devreese, B., *et al.* (2002) Mass spectrometric identification of mitochondrial oxidative phosphorylation subunits separated by two-dimensional blue-native polyacrylamide gel electrophoresis. *Electrophoresis* 23, 2525-2533
40. Hartmuth, K., *et al.* (2002) Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. *PNAS* 99, 16719-16724
41. Sakimura, K., *et al.* (1992) Primary structure and expression of the gamma 2 subunit of the glutamate receptor channel selective for kainate. *Neuron* 8, 267-274
42. Rauch, U., *et al.* (1997) Sequence and chromosomal localization of the mouse brevican gene. *Genomics* 44, 15-21
43. Hirakawa, S., *et al.* (2000) The brain link protein-1 (BRAL1): cDNA cloning, genomic structure, and characterization as a novel link protein expressed in adult brain. *Biochem Biophys Res Commun* 276, 982-989
44. Belizaire, R., *et al.* (2004) Characterization of synaptogyrin 3 as a new synaptic vesicle protein. *Journal of comparative neurology* 470, 266-281
45. Kohmura, N., *et al.* (1998) Diversity revealed by a novel family of cadherins expressed in neurons at a synaptic complex. *Neuron* 20, 1137-1151
46. Kofuji, P., *et al.* (1991) Generation of two forms of the gamma-aminobutyric acidA receptor gamma 2-subunit in mice by alternative splicing. *Journal of Neurochemistry* 56, 713-715
47. Karls, U., *et al.* (1992) Structure, expression, and chromosome location of the gene for the beta subunit of brain-specific Ca2+/calmodulin-dependent protein kinase II identified by transgene integration in an embryonic lethal mouse mutant. *molecular and cellular biology* 12, 3644-3652
48. Mazella, J., *et al.* (1996) Structure, functional expression, and cerebral localization of the levocabastine-sensitive neurotensin/neuromedin N receptor from mouse brain. *Journal of Neurosciences* 16, 5613-5620
49. Balcarek, J.M., and Cowan, N.J. (1985) Structure of the mouse glial fibrillary acidic protein gene: implications for the evolution of the intermediate filament multigene family. *Nucleic acids research* 13, 5527-5543
50. Liu, H., *et al.* (1999) Isolation of 10 differentially expressed cDNAs in differentiated Neuro2a cells induced through controlled expression of the GD3 synthase gene. *Journal of Neurochemistry* 72, 1781-1790
51. Zheng, C., *et al.* (1996) CNS gene encoding astrotactin, which supports neuronal migration along glial fibers. *Science* 272, 417-419
52. Nagano, T., *et al.* (1998) Differentially expressed olfactomedin-related glycoproteins (Pancortins) in the brain. *Brain Res Mol Brain Res* 53, 13-23
53. Katarova, Z., *et al.* (1990) Molecular identification of the 62 kd form of glutamic acid decarboxylase from the mouse. *European Journal of Neurosciences* 2, 190-202
54. Kingsmore, S.F., *et al.* (1994) Glycine receptor beta-subunit gene mutation in spastic mouse associated with LINE-1 element insertion. *Nature genetics* 7, 136-141
55. Takamatsu, K., *et al.* (1994) Molecular cloning of human hippocalcin cDNA and chromosomal mapping of its gene. *Biochem Biophys Res Commun*. 200, 606-611
56. Kamatchi, G.L., *et al.* (1995) GABAA receptor beta 1, beta 2, and beta 3 subunits: comparisons in DBA/2J and C57BL/6J mice. *Biochimica et biophysica acta* 1261, 134-142
57. Duclos, F., and Koenig, M. (1995) Comparison of primary structure of a neuron-specific protein, X11, between human and mouse. *Mammalian Genome* 6, 57-58
58. Ikenaka, K., *et al.* (1988) Myelin proteolipid protein gene structure and its regulation of expression in normal and jimpy mutant mice. *Journal of molecular biology* 199, 587-596
59. Yan, Y., *et al.* (1993) Molecular cloning of M6: identification of a PLP/DM20 gene family. *Neuron* 11, 423-431
60. Palmiter, R.D., *et al.* (1992) MT-III, a brain-specific member of the metallothionein gene family. *Proc Natl Acad Sci USA* 89, 6333-6337
61. Yoshihara, Y., *et al.* (1994) An ICAM-related neuronal glycoprotein, telencephalin, with brain segment-specific expression. *Neuron* 12, 541-553
62. Holz, A., *et al.* (1997) Chromosomal localization of the myelin-associated oligodendrocytic basic protein and expression in the genetically linked neurological mouse mutants ducky and tippy. *Journal of Neurochemistry* 69, 1801-1809
63. Okajima, T., *et al.* (1999) Molecular cloning of brain-specific GD1alpha synthase (ST6GalNAc V) containing CAG/Glutamine repeats. *Journal of biological chemistry* 274, 30557-30562
64. Plomann, M., *et al.* (1998) PACSIN, a brain protein that is upregulated upon differentiation into neuronal cells. *European Journal of Neurosciences* 256, 201-211
65. Campagnoni, A.T., *et al.* (1993) Structure and developmental regulation of Golli-mbp, a 105-kilobase gene that encompasses the myelin basic protein gene and is expressed in cells in the oligodendrocyte lineage in the brain. *Journal of biological chemistry* 268, 4930-4938
66. Andre, C., *et al.* (1994) Cultured neurons from mouse brain reproduce the muscarinic receptor profile of their tissue of origin. *European Journal of Neurosciences* 6, 1691-1701
67. Cimler, B.M., *et al.* (1987) Characterization of murine cDNAs encoding P-57, a neural-specific calmodulin-binding protein. *Journal of biological chemistry* 262, 12158-12163
68. Wang, K.C., *et al.* (2002) Oligodendrocyte-myelin glycoprotein is a Nogo receptor ligand that inhibits neurite outgrowth. *Nature* 417, 941-944