Methods

Mouse cDNA orthologs to HSA21genes

EST selection. The 238 HSA21genes were used as a reference set for isolating the corresponding mouse orthologs by running similarity searches with mouse sequences using BLASTn against nrdb, htgs and mouse dbEST databases (Genbank or EMBL), setting a threshold of 85% identity; human entries that failed matching a mouse sequence were re-analysed with a threshold of 50% identity, which generally did not identify any more *bona fide* matches. We could associate a mouse mRNA to 168 HSA21 genes, facilitating direct EST searches by identity matches in mouse dbEST. Publicly available EST clones were selected for 138/168 mmu21genes and we initially retained one or two ESTs/gene. 20 additional cDNAs were cloned by RT-PCR using information from available mouse cDNA or genomic sequences. 299 mouse clones were pre-selected and sequence-verified. Finally, we retained 187 *bona fide* clones matching 158 unique HSA 21 genes (9 clones were provided by S. Antonarakis). We fail to clone cDNAs for 10/168 mmu21 genes and we could not find any ortholog for 70 HSA21 genes so far. Most of the missing orthologs correspond to yet uncharacterised gene predictions identified only by short exons or by 3´ untranslated sequences complicating cross-species similarity searches.

EST sequencing. cDNAs were available as directionally cloned in pT7T3-PacMod1, pBluescript SK-, pSPORT1, pBluescribe modified, pCMV-Sport6, pSVSport1, pYX or pME18S-FL3. Inserts were PCR-amplified with M13 primers, except for those cloned in pSVSport1 (Sp6/T7), pYX (T3/T7) and pME18S-FL3 (F: 5'-CTTCTGCTCTAAAAGCTGCG-3'; R: 5'-CGACCTGCAGCTCGAGCACA-3') using standard conditions and 200 ng plasmid template; sequencing was from the 5’ end by cycle sequencing (MPIMG service). 22% of the pre-selected clones were eliminated

because their sequence did not match that of the Genbank entry, and those were subsequently replaced by alternative clones. All sequences are in Table 1.

Cloning of mRNAs by RT-PCR. PCRs were performed on 1-50 ng first strand cDNAs originating from adult brain, heart, spleen, liver, 7 days, 11 days, 17 days whole mouse embryo and 9-11 weeks pooled mouse brains (Clontech). Reactions were carried out using standard conditions, cycled 40 times (45 sec at 94°C, annealing at primer-specific temp for 45 sec, extension for 3 min at 72°C). Gene-specific primers were designed for spanning an intron (mmu21_primers_mapping in SI). PCR products were sequence-verified and cloned into pCRII-TOPO or pCR2.1-TOPO vector (Invitrogen).

RT-PCR on whole mouse brain at E15.5 and P2. First strand reverse transcription was performed by random priming from 2 μg total RNA using 200 units of Superscript II reverse transcriptase (Invitrogen) at 42°C for 50 min. RT-PCR was performed in standard conditions with 50 ng first strand cDNA and cycled 30 times (94°C for 1 min, annealing at primer-specific temperature for 1 min, extension at 72°C for 2 min).

Chromosome mapping. EST mapping was verified by PCR on DNA from somatic cell hybrids containing respectively: chromosome MMU16 (line 96aZ2) [1], MMU17+MMU3 (line 167EJ) [2] and MMU3 (line SN11C5-3 [3] used as negative control), and MMU10 (lines N9C2/93 and SN17C3) [3] using 100 ng of DNA template. Reactions were cycled 30 times (94°C for 1 min, annealing at primer-specific temperature for 1 min, extension at 72°C for 2 min). When possible, we performed *in silico* mapping on the mouse genomic sequence. Mapping results and primers sequences are available in the table mmu21_primers_mapping in SI. Mouse ESTs were cross-blasted against the human genome verifying that HSA21 was ranked as best hit over other human chromosomes.

Riboprobe synthesis

Antisense (and control sense) riboprobes were synthesised by *in vitro* transcription from 1-2 µg of PCR-amplified cDNAs (insert size range of 500 – 1,500 bp). RNA labelling was carried out in (1 mM of ATP, CTP and GTP; 0.65 mM TTP/0.35 mM DIG-11-UTP pH 7.5; 40 units RNAse inhibitor, 2 µl 10x transcription buffer), and 40 units of appropriate RNA polymerase (Sp6, T7, or T3) in a 20 µl reaction (DIG RNA labelling kit, Roche) at 37°C for 2 h. After RNA labelling, the DNA was degraded by adding of 20 units of RNAse-free-DNAse I for 30 min at 37°C and the reaction was stopped with SDS 0,04% and EDTA 1.6 mM final concentration. Riboprobes were precipitated with 0.1 vol.3M Sodium acetate/2,5 vol. absolute ethanol and stored at –20°C until use.

Automated in situ hybridization on floating brain sections

Mice (Swiss Webster mice, NYU-TGESF) brains were either fixed (E15) or perfused (P2) using 4% paraformaldehyde. ISH on 60µm floating cryosections were performed as previously described [4-7] (all reagents from Sigma). The fixed brains were quickly dissected out in sterile, ice-cold Phosphate-Buffered Saline (PBS) and cryoprotected by an overnight immersion in 30% sucrose/PBS at 4°C. The brains were then embedded in OCT's Tissue-Tek solution and frozen on dry ice. Floating cryosections (60 µm thick) were collected in ice-cold PBS, dehydrated in 50% methanol series and stored at –20°C. Pre-treatment of the sections included serial re-hydration in methanol, followed by permeabilization in 0.1% Tween-20 in PBS (PBTW) prior to incubation with 10µg/ml Proteinase K at 37°C for 20 min. Proteolysis was monitored and stopped by large dilution in ice-cold 4% PFA/0.1% glutaraldehyde/0.1% Tween-20, for 20 min on ice. The sections were extensively rinsed in PBTW on ice, and

preincubated in a 1/1 mix of PBTW and hybridization buffer (50% deionised formamide, 5X SSC pH7.2, 1mg/ml yeast Torula RNA, 0.1mg/ml Heparin, 1x Denhardt's solution, 0.1% Tween-20, 0.1% CHAPS, 5mM EDTA) for 10 min at room temperature. The sections were then saturated by at least 3 h preincubation at 65°C in hybridization solution and conserved at –20°C. Prior to each ISH run, the sections were equilibrated for 10 min at 65°C. *In situ* hybridizations (using 1µg/ml digoxigenine-labelled riboprobes) were carried out on floating sections using the ABIMED *in situ* machine. The sections were distributed into a 30 wells plate and the *in situ* procedure was as described by Wilkinson (1992) [5] with the following changes: the washes after washing with solution 3 were done with maleic acid buffer [4]. All steps prior to staining were done in the ABIMED machine. Staining was done in microtiter plate at room temperature. After staining, reactions were stopped and sections were washed several times in PBTW, post-fixed in 4% PFA and mounted in 50% glycerol.

Automated whole mount in situ hybridization (WISH)

The automated WISH procedure was performed essentially as described in Neidhardt *et al*. 2000 [6]. All probes were initially analysed in high-throughput format on E9.5 mouse embryos (NMRI outbred). Probes showing restricted expression were selected and analysed further on E8.5 to E11.5 embryos.

EST mining and *in silico* gene expression profiles in the mouse

Qualitative analysis. For the 168 mouse orthologs, we could associate a complete mouse mRNA to 146 entries and a partial mRNA to 22 entries. The longest available nucleotide sequences of the mmu21 genes were used as a reference set for

extracting all matching EST entries from normal mouse tissues. Identity searches against mouse dbEST (2,253,129 sequences distributed in 632 cDNA libraries after removing cancer derived libraries; EMBL download 26/06/2002) were run using BLASTn, setting criteria at 98% identity over more than 100 nucleotides to compensate for sequencing errors and polymorphisms. Libraries described with exact matching terms regarding the tissue origin were considered as "identical" in our analysis and merged together; and this pooling scheme reduced the number of libraries to 427. We obtained a blast output containing a total of 5,454 matching EST entries for 159 mmu21 genes (9 mRNAs did not match any EST), distributed in 271 non-redundant libraries for which the original description was parsed for extracting the tissue origin and developmental stage information. For each gene entry, ESTs were classified by tissue, representing the equivalent of an "electronic RT-PCR" (Table 1).

Statistical analysis. We analysed gene expression profiles as estimated from EST counts using a previously validated method based on Pearson's correlation coefficient [8]. This method is little sensitive to the size of the different libraries since it compares profiles shared by genes rather than EST ratios. Normalised libraries and those containing less than 1,000 ESTs were removed for the statistical analysis. We have finally analysed the EST counts for 159 mouse orthologs to HSA21, representing a total of 4,186 hits distributed in 190 "pooled", non-normalized, and relatively large libraries (table Fig4_Libraries in SI). This raw information was tabulated for establishing a matrix displaying the 159 mmu21 genes and their associated EST counts originating from the 190 libraries. This initial table was used to compute the gene pairwise Pearson's linear correlation coefficients between genes leading to a symetrical matrix (159x159) of correlation values subsequently used to derive a second square matrix measuring the Euclidian distance between genes for building the associated dendogram according to the UPGMA algorithm[9]. The initial Pearson's gene square matrix was reordered according to this gene dendogram, allowing to visualise eight statistically significant

clusters of genes with similar expression profiles (correlation coefficient values are ranging from -1 to 1; genes that are significantly correlated by their expression profiles show the higher correlation coefficient). Gene clusters with a mean correlation coefficient greater than 0.4 and P-values below or equal to 0,001 were considered statistically significant. The cutoff of 0.4 was chosen upon published reference tables giving the critical values of the Pearson correlation coefficient ([10,11] or http://cyberlearn.fau.edu/sta6113/PearsonT.htm). Similarly, we computed a dendogram of libraries from their Euclidian distances. Thus, we obtained two square correlation matrices, one for genes and one for libraries (not shown), used for calculating two Euclidian distance matrices. The respective UPGMA trees of genes and libraries determined the ordering of genes and libraries (Genes_Libraries_dendograms in SI). In order to represent a hierarchical classification of clusters and libraries, the order provided by the dendograms were used for rearranging the original data table, used as a basis for plotting the clustered correlation colour map shown in Figure 4 (using Matlab v.6.0.0.88 released 12; MathWorks, Inc).

Orthology of HSA genes with yeast, worm and fly proteins.

Orthology relationships of HSA21 genes were computed after masking low complexity sequences [12] using PSI-BLAST with 2 iterations against the protein catalogs of S. *cerevisiae* (Y; SCE: 6,917 entries), C. *elegans* (C; CEL: 17,135 entries), and D. *melanogaster* (D; DME: 14,335 entries) in Genbank. We selected as best hit the longest match among all with a score higher than the top score minus 100. Selected hits were cross-blasted against all human proteins in ENSEMBL (v.1.1.1; 23,867 peptides) for verifying that the "best hit" did mach HSA21. Orthologs were aligned to the HSA21 proteins using CLUSTAL W [13] (Table 1).

References for the methods

1. Kozak, C. A., Strauss, P. G. & Tsichlis, P. N. Genetic mapping of a cellular DNA region involved in induction of thymic lymphomas (Mlvi-1) to mouse chromosome 15. *Mol Cell Biol* **5**, 894-7. (1985).

2. Cox, R. D., Stubbs, L., Evans, T. & Lehrach, H. A mouse specific polymerase chain reaction (PCR) primer: probe generation from somatic cell hybrids. *Nucleic Acids Res* **19**, 2503. (1991).

3. Sabile, A., Poras, I., Cherif, D., Goodfellow, P. & Avner, P. Isolation of monochromosomal hybrids for mouse chromosomes 3, 6, 10, 12, 14, and 18. *Mamm Genome* **8**, 81-5. (1997).

4. Harland, R. M. In situ hybridization: an improved whole-mount method for Xenopus embryos. *Methods Cell Biol* **36**, 685-95 (1991).

5. Wilkinson, D. G. in *In Situ Hybridization: A Practical Approach* (ed. Wilkinson, D. G.) 75-83 (Oxford University Press, New York, 1992).

6. Neidhardt, L. et al. Large-scale screen for genes controlling mammalian embryogenesis, using high-throughput gene expression analysis in mouse embryos. *Mech Dev* **98**, 77-94. (2000).

7. Dahmane, N. et al. The Sonic Hedgehog-Gli pathway regulates dorsal brain growth and tumorigenesis. *Development* **128**, 5201-12. (2001).

8. Ewing, R. M. et al. Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression. *Genome Res* **9**, 950-9. (1999).

9. Sokal, R. & Michener, C. A statistical method for evaluating systematic relationships. *Univ. Kansas Sci. Bull.* **28**, 1409-1438 (1958).

10. Sachs, L. *Angewandte Statistik* (ed. Springer-Verlag) (Berlin, 1983).

11. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* **95**, 14863-8. (1998).

12. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-402. (1997).

13. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673-80. (1994).