Microarray data analysis

We identified all of the positive oligonucleotides with the threshold values $R = 13$ and $D = 12Q$ (ref. 21). R and D are threshold values for the ratio and the difference between perfect match intensity and mismatch intensity, respectively. Thus varying these values gives different measures of sensitivity and specificity. We then used BLAST to identify conserved blocks that corresponded to at least two positive oligonucleotides so as to reduce the number of false positives.

Received 16 September; accepted 30 October 2002; doi:10.1038/nature01251.

1. Hardison, R. C., Oeltjen, J. & Miller, W. Long human–mouse sequence alignments reveal novel regulatory elements: a reason to sequence the mouse genome. *Genome Res.* **7,** 959–966 (1997).
2. O'Brien, S. J. *et al.* The promise of comparative genomics in mammals. *Science* **286,** 458–462 (1999) 479–481.
3. Shabalina, S. A., Ogurtsov, A. Y., Kondrashov, V. A. & Kondrashov, A. S. Selective constraint in intergenic regions of human and mouse genomes. *Trends Genet.* **17,** 373–376 (2001).
4. Hardison, R. C. Conserved noncoding sequences are reliable guides to regulatory elements. *Trends Genet.* **16,** 369–372 (2000).
5. Dermitzakis, E. T. & Clark, A. G. Evolution of transcription factor binding sites in mammalian gene regulatory regions: conservation and turnover. *Mol. Biol. Evol.* **19,** 1114–1121 (2002).
6. Frazer, K. A. *et al.* Evolutionarily conserved sequences on human chromosome 21. *Genome Res.* **11,** 1651–1659 (2001).
7. Mural, R. J. *et al.* A comparison of whole-genome shotgun-derived mouse chromosome 16 and the human genome. *Science* **296,** 1661–1671 (2002).
8. Hattori, M. *et al.* The DNA sequence of human chromosome 21. *Nature* **405,** 311–319 (2000).
9. Antonarakis, S. E., Lyle, R., Deutsch, S. & Reymond, A. Chromosome 21: a small land of fascinating disorders with unknown pathophysiology. *Int. J. Dev. Biol.* **46,** 89–96 (2002).
10. Reymond, A. *et al.* Human chromosome 21 gene expression atlas in the mouse. *Nature* **420,** 582–586 (2002).
11. Schwartz, S. *et al.* PipMaker—a web server for aligning two genomic DNA sequences. *Genome Res.* **10,** 577–586 (2000).
12. Waterston, R. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420,** 520–562 (2002).
13. DeSilva, U. *et al.* Generation and comparative analysis of approximately 3.3 Mb of mouse genomic sequence orthologous to the region of human chromosome 7q11.23 implicated in Williams syndrome. *Genome Res.* **12,** 3–15 (2002).
14. Loots, G. G. *et al.* Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science* **288,** 136–140 (2000).
15. Davisson, M. T. *et al.* Evolutionary breakpoints on human chromosome 21. *Genomics* **78,** 99–106 (2001).
16. Gardiner, K., Slavov, D., Bechtel, L. & Davisson, M. Annotation of human chromosome 21 for relevance to down syndrome: gene structure and expression analysis. *Genomics* **79,** 833–843 (2002).
17. Reymond, A. *et al.* From PREDs and open reading frames to cDNA isolation: revisiting the human chromosome 21 transcription map. *Genomics* **78,** 46–54 (2001).
18. Reymond, A. *et al.* Nineteen additional unpredicted transcripts from human chromosome 21. *Genomics* **79,** 824–832 (2002).
19. Novichkov, P. S., Gelfand, M. S. & Mironov, A. A. Gene recognition in eukaryotic DNA by comparison of genomic sequences. *Bioinformatics* **17,** 1011–1018 (2001).
20. Rivas, E. & Eddy, S. R. Noncoding RNA gene detection using comparative sequence analysis. *BioMed Central Bioinformatics* **2,** 8 (2001).
21. Kapranov, P. *et al.* Large-scale transcriptional activity in chromosomes 21 and 22. *Science* **296,** 916–919 (2002).
22. Nekrutenko, A., Makova, K. D. & Li, W. H. The K(A)/K(S) ratio test for assessing the protein-coding potential of genomic regions: an empirical and simulation study. *Genome Res.* **12,** 198–202 (2002).
23. Madsen, O. *et al.* Parallel adaptive radiations in two major clades of placental mammals. *Nature* **409,** 610–614 (2001).
24. Murphy, W. J. *et al.* Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* **294,** 2348–2351 (2001).
25. Murphy, W. J. *et al.* Molecular phylogenetics and the origins of placental mammals. *Nature* **409,** 614–618 (2001).
26. Dubchak, I. *et al.* Active conservation of noncoding sequences revealed by three-way species comparisons. *Genome Res.* **10,** 1304–1306 (2000).
27. Enard, W. *et al.* Intra- and interspecific variation in primate gene expression patterns. *Science* **296,** 340–343 (2002).
28. Eyre-Walker, A. & Hurst, L. D. The evolution of isochores. *Nature Rev. Genet.* **2,** 549–555 (2001).
29. Kumar, S. & Subramanian, S. Mutation rates in mammalian genomes. *Proc. Natl Acad. Sci. USA* **99,** 803–808 (2002).

Supplementary Information accompanies the paper on *Nature*'s website (http://www.nature.com/nature).

Acknowledgements This project was supported by grants from the Swiss National Science Foundation, National Center for Competence in Research 'Frontiers in Genetics', the European Union/Federal office of Education and Health 'Child Care' foundation (to S.E.A.), and a Swiss National Science Foundation grant (to P.B.). We thank E. Lander for advice and support, C. Rossier for core sequencing support and J. Yang for providing programs.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.E.A. (e-mail: stylianos.antonarakis@medecine.unige.ch).

Human chromosome 21 gene expression atlas in the mouse

Alexandre Reymond*†, Valeria Marigo†‡, Murat B. Yaylaoglu†§, Antonio Leoni‡, Catherine Ucla*, Nathalie Scamuffa*, Cristina Caccioppoli‡, Emmanouil T. Dermitzakis*, Robert Lyle*, Sandro Banfi‡, Gregor Eichele§, Stylianos E. Antonarakis* & Andrea Ballabio‡||

** Division of Medical Genetics, University of Geneva Medical School and University Hospital of Geneva, CMU, 1, rue Michel Servet, 1211 Geneva, Switzerland*
‡ Telethon Institute of Genetics and Medicine, Via Pietro Castellino 111, 80131 Naples, Italy
§ Max Planck Institute of Experimental Endocrinology, Feodor-Lynen-Str. 7, D-30625 Hannover, Germany
|| Medical Genetics, Second University of Naples, Naples, Italy
† These authors contributed equally to this work

Genome-wide expression analyses have a crucial role in functional genomics. High resolution methods, such as RNA *in situ* hybridization provide an accurate description of the spatiotemporal distribution of transcripts as well as a three-dimensional '*in vivo*' gene expression overview[1–5]. We set out to analyse systematically the expression patterns of genes from an entire chromosome. We chose human chromosome 21 because of the medical relevance of trisomy 21 (Down's syndrome)[6]. Here we show the expression analysis of all identifiable murine orthologues of human chromosome 21 genes (161 out of 178 confirmed human genes) by RNA *in situ* hybridization on whole mounts and tissue sections, and by polymerase chain reaction with reverse transcription on adult tissues. We observed patterned expression in several tissues including those affected in trisomy 21 phenotypes (that is, central nervous system, heart, gastrointestinal tract, and limbs). Furthermore, statistical analysis suggests the presence of some regions of the chromosome with genes showing

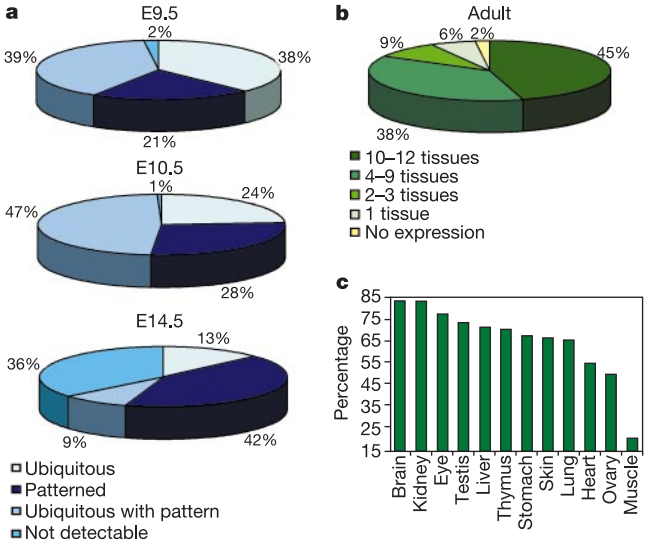

Figure 1 Distribution of expression patterns and transcriptome complexity. **a**, Each slice corresponds to the percentage of genes belonging to the four categories of expression pattern observed by ISH at E9.5 (whole mount), E10.5 (whole mount) and E14.5 (sections). **b**, Each slice represents the percentage of genes expressed in 0, 1, 2–3, 4–9 and 10–12 adult tissues by RT–PCR. **c**, Percentage of the analysed 161 human chromosome 21 murine orthologues identified in each murine adult tissue.

either lack of expression or, to a lesser extent, co-expression in specific tissues. This high resolution expression 'atlas' of an entire human chromosome is an important step towards the understanding of gene function and of the pathogenetic mechanisms in Down's syndrome.

So far 178 confirmed genes and 36 predicted genes have been identified on human chromosome 21 (refs 7–11). The mouse syntenic regions (segments of mouse chromosomes 10, 16 and 17) harbour 170 orthologues. We isolated 237 complementary DNA fragments representing 158 mouse orthologues (93%) for *in situ* hybridization (ISH) experiments and designed primer pairs for 161 genes for polymerase chain reaction with reverse transcription (RT–PCR). To generate the human chromosome 21 gene expression atlas, these orthologues were studied by normalized RT–PCR in 4 developmental stages and 12 adult tissues; whole-mount ISH of embryonic day E9.5 and E10.5 embryos; and ISH on serial sagittal sections of E14.5 embryos (see Supplementary Information for detailed Methods; see also http://www.tigem.it/ch21exp/). For ISH of sections, we developed an ISH robot and an automated microscope that permitted the analysis of about 6,500 tissue sections generated for this atlas[12].

Expression was detected for 98% of the tested genes by at least one of the selected methods. The results are compiled in the Supplementary Information and at http://www.tigem.it/ch21exp/, and are also summarized in Fig. 1a–c. Supplementary Information consists of three items: (1) atlas expression map tables, containing a list of the genes ordered by their position on chromosome 21 and all original ISH images, annotation tables and details on probes; (2) some examples of patterned expressions in gut, cerebellum, heart, thymus, pancreas and limbs; and (3) a list of Methods. Previously described expression patterns are in agreement with our data (for example, *Sh3bgr*[13]).

By ISH, patterned (regional) gene expression was observed for 21% (E9.5), 28% (E10.5) and 42% (E14.5) of genes (see examples in Fig. 2). The highest numbers of genes with a restricted expression pattern were observed in the brain, the eye and the gut at all stages. Ubiquitous expression was observed in 38% (E9.5), 24% (E10.5) and 13% (E14.5) of cases. Genes with both weak ubiquitous expression and strong regional expression were also observed (39% at E9.5, 47% at E10.5, and 9% at E14.5; for example, *Pfkl*, Fig. 2). No expression was detected for 2% (E9.5), 1% (E10.5) and 36% (E14.5) of genes.

In addition to ISH, we carried out 2,576 RT–PCR reactions covering 95% of the human chromosome 21 murine orthologues (RT–PCR results are documented on Supplementary Information and http://www.tigem.it/ch21exp/). The 161 orthologues analysed were found to be expressed on average in 8 of 12 adult tissues tested (s.d. = 3.6). The transcriptomes of brain and kidney showed the highest complexity, each tissue expressing 85% of the 161 genes. Other tissues are less complex (muscle, 21%; heart, 56%; ovary, 51%; lung 67%; skin, 68%; stomach, 69%; thymus, 72%; liver, 73%; testis 75%; and eye, 79%; Fig. 1c). Forty-five per cent of all genes were widely expressed (>9 tissues out of 12 tissues were positive; Fig. 1b). Thirty-eight per cent of the genes show expression in 4–9 tissues. Restricted expression (2–3 tissues) and single-tissue expression was observed in 9% and 6% of the genes, respectively. Only 2% of the genes were not detectable by RT–PCR.

There is good concordance between the whole-mount data at E9.5 and E10.5 (Fig. 1a). It is, however difficult to compare these results with the data obtained by the two other methods, (that is, sections at E14.5 and RT–PCR on adult tissues). This is due to differences in sensitivity (RT–PCR is more sensitive), resolution (sections have a cellular resolution), and ascertaining background signal compared with weak ubiquitous expression (easier in whole

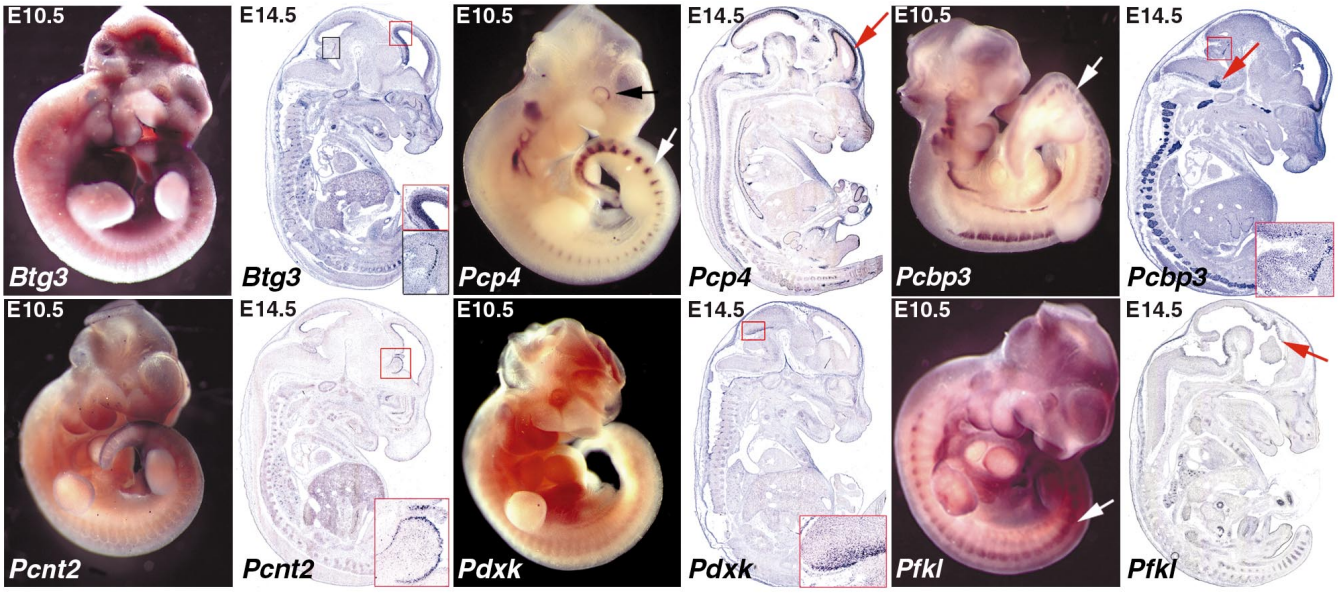

Figure 2 Representative examples of *in situ* hybridization data of E10.5 (whole mount) and E14.5 (sections) embryos. See supplementary Information for the list of genes studied and images of their expression patterns. *Btg3* messenger RNA is ubiquitously expressed at E10.5, with higher abundance in brain and gut. At E14.5 expression in the brain is restricted to the ventricular zone (red insert) and to a group of cells migrating into the developing cerebellum (black insert). *Pcp4* is transcribed in brain, eye (black arrow) and dorsal root ganglia (white arrow) at E10.5. At E14.5 cells transcribing this gene are found in numerous tissues including the cortical plate (red arrow), midbrain, cerebellum, spinal chord, intestine, heart and dorsal root ganglia. *Pcbp3* transcripts are restricted to the central and peripheral nervous systems both at E10.5 and E14.5 (white arrow, dorsal root ganglia; red arrow, facial nerve nucleus). Insert shows a group of neuroblasts migrating from the midbrain into the cerebellar anlage. Expression of *Pcnt2* at E10.5 is ubiquitous but stronger in brain, eye, limbs, branchial arches and gut. At E14.5 *Pcnt2* brain expression is restricted to proliferating cells of the ventricular zone (red insert). At E10.5 *Pdxk* mRNA is ubiquitous but is more strongly expressed in brain and eye. At E14.5 a group of cells in the inferior colliculus express *Pdxk* (red insert). *Pfkl* transcripts are found throughout the E10.5 embryo with higher levels in brain, eye and dorsal root ganglia (white arrow). At E14.5 this transcript is widely expressed but exhibits regional upregulation in the ventricular zone (red arrow), axial and cranial cartilage, thymus, gut and lung.

mounts). Moreover, E14.5 has a more complex tissue architecture than earlier stages. A longitudinal analysis of 147 of the analysed genes at E9.5, E10.5 and E14.5 revealed that 59% of the genes retain their expression pattern during these developmental stages, whereas 17% acquire new, specific expression at E14.5. For the remaining 24%, we are uncertain owing to low expression in some stages.

Figure 2 presents examples of high-resolution expression patterns in the developing nervous system. Some of the brain-expressed genes may contribute to the Down's syndrome cognitive defects. At E14.5 *Btg3*, *Pcnt2* and *Pfkl* transcripts are detected in the ventricular zone, whereas *Pcp4* staining is observed in the cortical plate and the mantel layer of the midbrain. *Btg3* and *Pcbp3* staining was also observed in neuronal precursors that migrate from the inferior colliculus into the cerebellum. At the same developmental stage, *Pdxk* is expressed in a group of cells in the inferior colliculus. The KH-domain (maxi-K Homology domain) encoding *Pcbp3* transcript is restricted to the central and peripheral nervous systems both at E10.5 and E14.5. Finally, *Pcp4*, *Pcbp3* and *Pfkl* are strongly expressed in dorsal root ganglia, as visible in whole-mount ISH at E10.5.

Down's syndrome is associated with congenital heart disease, and provides an important model to link individual genes to pathways controlling heart development. In trisomy 21 the most frequent and specific heart abnormalities are atrioventricular canal and atrial septal defects. *Pwp2h* and *C21orf11* show elevated expression in the developing atria (Fig. 3), and *Pfkl* is strongly expressed in the ventricular wall and atrium. Two other heart-specific expression patterns are noteworthy: *Adarb1* and *C21orf18* are both expressed in E10.5 aortic sac; the precursor of the ascending aorta and the pulmonary artery. Furthermore, *C21orf18* is expressed in the bulbus cordis, which develops into the right ventricle. At E14.5, *Kcnj15* and *Adarb1* are expressed in the aortic valve and trunk, while *Kcnj15* is also expressed along the outflow track of the heart and in the superior vena cava. *Atp50* and *Sh3bgr* transcripts are detected throughout the heart, whereas *Cldn8* is regionally expressed in the primitive ventricle. The human genes *COL6A1, COL6A2, COL18A1* and *KCNE2* are mutated in Bethlem myopathy, Ullrich's disease, long QT6, and Knobloch syndrome, respectively[14–18]. We found that *Kcne2* is expressed in the entire developing heart including its vessels, whereas *Col18a1* is detected in cardiac vessels only. *Col6a1* and *Col6a2* are strongly expressed in the mitral valve and along the pericardium[16,18] (Fig. 3).

Down's syndrome fetuses exhibit reduced growth rate of long

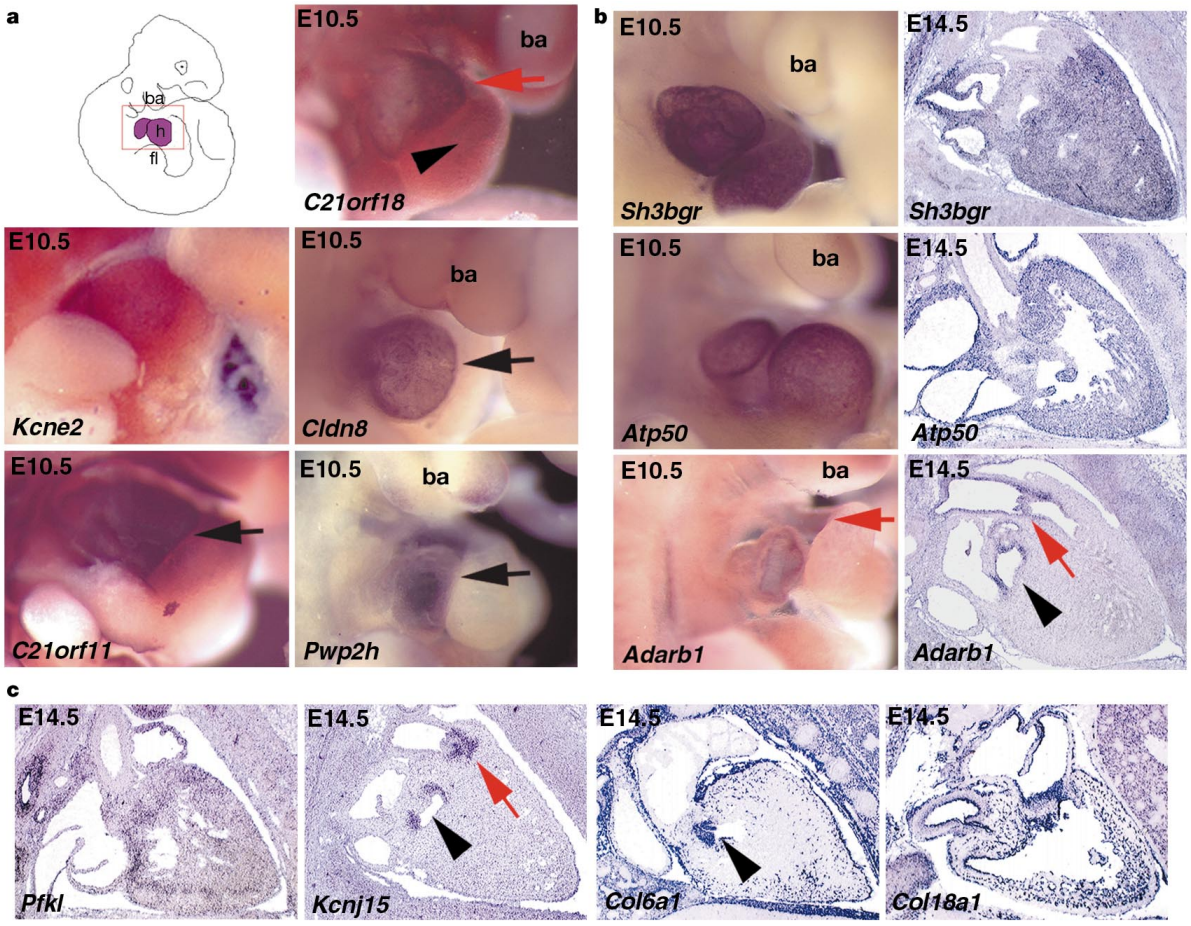

Figure 3 Expression analysis in the developing heart. E10.5 whole embryos and E14.5 sections were probed for *Adarb1*, *Atp50*, *C21orf11*, *C21orf18*, *Col6a1*, *Col18a1*, *Cldn8*, *Kcnj15*, *Kcne2*, *Pfkl*, *Pwp2h* and *Sh3bgr* genes. **a**, The region portrayed from the whole-mount embryos is schematically represented in the red box of the top left panel. h, heart; ba, branchial arch; fl, forelimb. At E10.5, *C21orf18* transcripts are present in the aortic sac (red arrow) and the bulbus cordis (black arrowhead). *Kcne2* is expressed in atria and ventricles. mRNA for *C21orf11* and *Pwp2h* is restricted to the atria (black arrows), and *Cldn8* transcripts are restricted to the primitive ventricle (black arrow). **b,** *Sh3bgr* and *Atp50* transcripts are detected throughout the heart. *Adarb1* mRNA is present in the aortic sac (red arrow) at E10.5, and is restricted to the mitral valve (black arrowhead), aortic valve (red arrow) and the endothelium of the aortic trunk at E14.5. **c**, At E14.5 *Pfkl* is expressed throughout the heart. *Kcnj15* transcripts are found in the aortic and mitral valve (red arrow and black arrowhead, respectively). *Col6a1* is strongly expressed in the mitral valve (black arrowhead) and along the pericardium, whereas *Col18a1* expression is seen in the vessels of the heart.

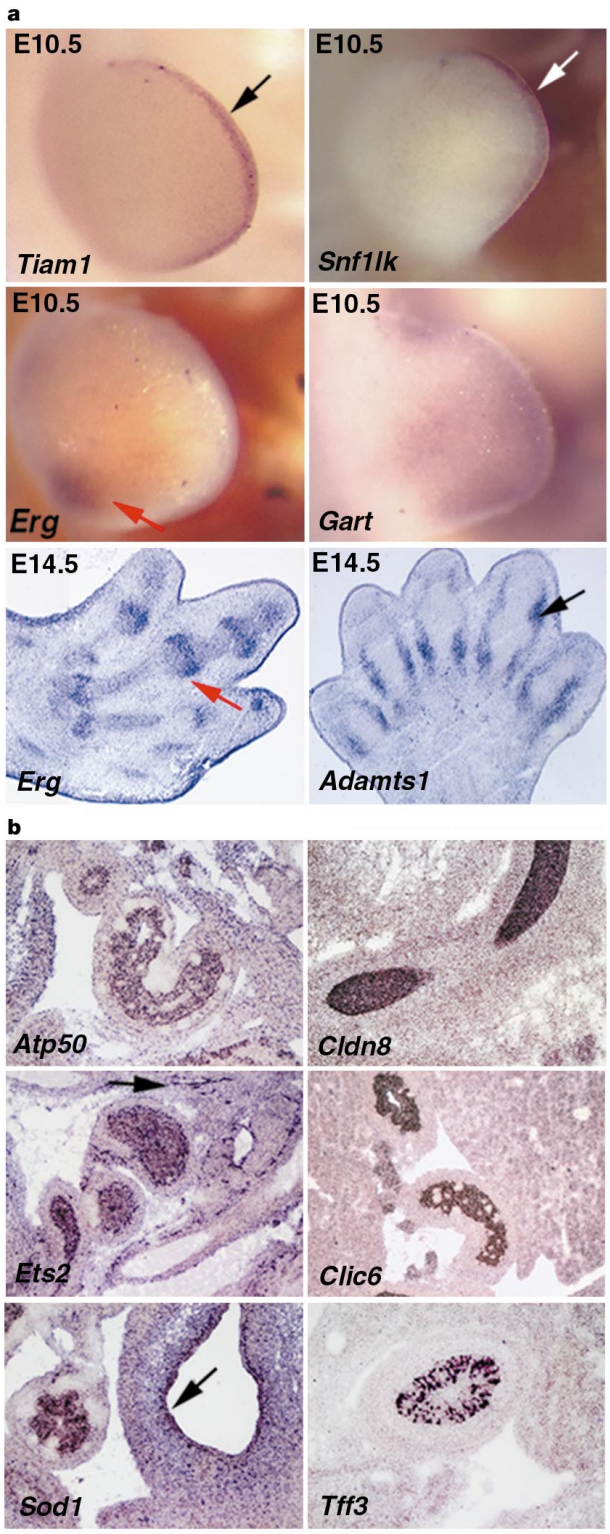

Figure 4 Expression analysis in the developing limb and gastrointestinal tract. **a**, At E10.5 *Tiam1* and *Snf1lk* are expressed specifically in the apical ectodermal ridge, which is a specialized ectodermal thickening regulating proliferation of the underlying mesoderm (arrows). A more diffuse distribution in the limb bud mesoderm was found for *Gart*. *Erg* transcripts are present in the posterior proximal mesoderm at E10.5 and in the joints (red arrows) at E14.5. *Adamts1* showed expression in the perichondrium of the developing digits (black arrow). **b**, *Atp50*, *Cldn8*, *Ets2* and *Clic6* are expressed within the duodenum endoderm. *Ets2* transcripts are also detected in the vasculature surrounding the duodenum (arrow). *Sod1* mRNA is present in the midgut endoderm and in the stomach endothelium (arrow). *Tff3* is transcribed by a subgroup of cells within the gastroduodenal junction region of the pyloric sphincter.

limb bones during the third trimester of pregnancy. Furthermore, a non-ossified or hypoplastic middle phalanx of the fifth digit is present in these fetuses[6]. Numerous genes are expressed in the limb buds (Fig. 4a). *Tiam1* and *Snf1lk* transcripts were found specifically in the apical ectodermal ridge; a specialized ectodermal thickening regulating proliferation of the underlying mesoderm (Fig. 4a). *Erg* showed strong expression in the posterior proximal mesoderm at E10.5 and in joints at E14.5. Finally, *Adamts1* was expressed at E14.5 in the perichondrium of developing bones.

Gastrointestinal abnormalities such as duodenal stenosis, Hirschsprung's disease, gastro-oesophageal reflux and imperforate anus are frequent in Down's syndrome patients[6]. At E14.5 the digestive tract is well differentiated and the expression of many genes can be detected. *Atp50*, *Cldn8*, *Clic6* and *Ets2* are expressed within the endothelium of the duodenum (Fig. 4b). Interestingly, the latter is also expressed in the skeletal system and previously found to result in skeletal abnormalities reminiscent of Down's syndrome when overexpressed in transgenic mice[19]. *Tff3* and *Sod1* transcripts are present in a subgroup of cells within the gastroduodenal junction region of the pyloric sphincter, and in the endothelium of the stomach and the intraperitoneal portion of the midgut, respectively.

Previous studies suggested the presence of genomic regions containing clusters of genes with similar expression patterns[20–24]. To search for such clusters on chromosome 21, all RT–PCR expression data underwent a test for clustering[25] (see Methods in Supplementary Information). In four regions we found significant clustering of genes showing absence of expression in specific tissues. These include genes not expressed in heart (from *B3galt5* to *Hsf2bp*; 3.9 Megabases (Mb), 31 genes, $P < 0.001$), lung (from *Dscam* to *Slc37a1*; 2.7 Mb, 19 genes, $P = 0.007$), testis (from *C21orf25* to *Pde9a*; 1.0 Mb, 12 genes, $P = 0.028$), and muscle (from *Hsf2bp* to *Hrmt1l1*; 3.4 Mb, 43 genes, $P = 0.02$). The following duplicated genes were found: *Mx1* and *Mx2*; *Tff1*, *Tff2* and *Tff3*; *Ifnar1* and *Ifnar2*; and *Col6a1* and *Col6a2*. Notably, each of these regions is fully contained within a syntenic block on the mouse chromosomes. Consistent clustering results were obtained when the cluster analysis was extended to expression data collected by ISH. In addition, E14.5 ISH data suggested a cluster for presence of expression within fore-, mid- and hindbrain, which extended from *Hunk* to *Atp50* (22 genes, $P = 0.004$). Significant clustering was observed even when applying two additional statistical analyses: the Bonferoni correction and Fisher's combined probability (see Supplementary Information). Albeit preliminary, these data suggest that regions containing either co-silenced or co-expressed genes exist on chromosome 21, and their presence should be considered in future expression studies of large genomic regions.

The combination of gene mapping with expression analysis is a helpful tool in the identification of positional candidate genes for human diseases[26]. Thus the human chromosome 21 expression atlas provides a rich resource for candidate genes for both monogenic and multifactorial diseases mapping to chromosome 21. We anticipate, however, that the greatest impact of the atlas lies in assessing the contribution of specific genes to Down's syndrome traits and phenotypes. The correlation of the atlas expression patterns with specific features of Down's syndrome may lead to the identification of candidate genes for these features of the disease. In particular, *ADARB1*, *KCNJ15*, *PFKL*, *PWP2H* and *SH3BGR* are promising candidate genes for Down's syndrome heart defects, as they map to the Down's syndrome congenital heart disease critical region[27], and because of the expression patterns of their murine orthologues in the developing heart. Similarly, the human orthologues of the gut-expressed genes *Atp50*, *Cldn8*, *Clic6*, *Ets2*, *Hmgn1*, *Sh3bgr*, *Sod1* and *Wrb* may have a role in the Down's syndrome gastrointestinal abnormalities[6]. Mapping gene expression of an entire chromosome at high resolution defines a new level of gene annotation, which is anticipated to advance our knowledge on gene function and

regulation, and our understanding of human aneuploidies, such as Down's syndrome. □

Received 11 May; accepted 19 September 2002; doi:10.1038/nature01178.

1. Strachan, T., Abitbol, M., Davidson, D. & Beckmann, J. S. A new dimension for the human genome project: towards comprehensive expression maps. *Nature Genet.* **16,** 126–132 (1997).
2. 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).
3. Bulfone, A. *et al.* The embryonic expression pattern of 40 murine cDNAs homologous to *Drosophila* mutant genes (Dres): a comparative and topographic approach to predict gene function. *Hum. Mol. Genet.* **7,** 1997–2006 (1998).
4. Reymond, A. *et al.* The tripartite motif family identifies cell compartments. *EMBO J.* **20,** 2140–2151 (2001).
5. Gawantka, V. *et al.* Gene expression screening in *Xenopus* identifies molecular pathways, predicts gene function and provides a global view of embryonic patterning. *Mech. Dev.* **77,** 95–141 (1998).
6. Epstein, C. J. *The Metabolic and Molecular Bases of Inherited Disease* (eds Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D.) 749–794 (McGraw Hill, New York, 1995).
7. Hattori, M. *et al.* The DNA sequence of human chromosome 21. *Nature* **405,** 311–319 (2000).
8. Davisson, M. T. *et al.* Evolutionary breakpoints on human chromosome 21. *Genomics* **78,** 99–106 (2001).
9. Pletcher, M. T., Wiltshire, T., Cabin, D. E., Villanueva, M. & Reeves, R. H. Use of comparative physical and sequence mapping to annotate mouse chromosome 16 and human chromosome 21. *Genomics* **74,** 45–54 (2001).
10. Reymond, A. *et al.* Nineteen additional unpredicted transcripts from the human chromosome 21. *Genomics* **79,** 824–832 (2002).
11. Reymond, A. *et al.* From PREDs and open reading frames to cDNA isolation: revisiting the human chromosome 21 transcription map. *Genomics* **78,** 46–54 (2001).
12. Herzig, U. *et al.* Development of high-throughput tools to unravel the complexity of gene expression patterns in the mammalian brain. *Novartis Found. Symp.* **239,** 129–146 (2001).
13. Egeo, A. *et al.* Developmental expression of the SH3BGR gene, mapping to the Down syndrome heart critical region. *Mech. Dev.* **90,** 313–316 (2000).
14. Abbott, G. W. *et al.* MiRP1 forms Ikr potassium channels with HERG and is associated with cardiac arrhythmia. *Cell* **97,** 175–187 (1999).
15. Sertie, A. L. *et al.* Collagen XVIII containing an endogenous inhibitor of angiogenesis and tumour growth, plays a critical role in the maintenance of retinal structure and in neural tube closure (Knobloch syndrome). *Hum. Mol. Genet.* **9,** 2051–2058 (2000).
16. Camacho Vanegas, O. *et al.* Ullrich scleroatonic muscular dystrophy is caused by recessive mutations in collagen type VI. *Proc. Natl. Acad. Sci. USA* **98,** 7516–7521 (2001).
17. Higuchi, I. *et al.* Frameshift mutation in the collagen VI gene causes Ulrich's disease. *Ann. Neurol.* **50,** 261–265 (2001).
18. Jobsis, G. J. *et al.* Type VI collagen mutations in Bethlem myopathy, an autosomal dominant myopathy with contractures. *Nature Genet.* **14,** 113–115 (1996).
19. Sumarsono, S. H. *et al.* Down's syndrome-like skeletal abnormalities in Ets2 transgenic mice. *Nature* **379,** 534–537 (1996).
20. Kim, S. K. *et al.* A gene expression map for *Caenorhabditis elegans*. *Science* **293,** 2087–2092 (2001).
21. Lercher, M. J., Urrutia, A. O. & Hurst, L. D. Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nature Genet.* **6,** 6 (2002).
22. Cohen, B. A., Mitra, R. D., Hughes, J. D. & Church, G. M. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nature Genet.* **26,** 183–186 (2000).
23. Bortoluzzi, S. *et al.* A comprehensive, high-resolution genomic transcript map of human skeletal muscle. *Genome Res.* **8,** 817–825 (1998).
24. Spellman, P. T. & Rubin, G. M. Evidence for large domains of similarly expressed genes in the *Drosophila* genome. *J. Biol.* **1,** 5 (2002).
25. Tang, H. & Lewontin, R. C. Locating regions of differential variability in DNA and protein sequences. *Genetics* **153,** 485–495 (1999).
26. Ballabio, A. The rise and fall of positional cloning? *Nature Genet.* **3,** 277–279 (1993).
27. Barlow, G. M. *et al.* Down syndrome congenital heart disease: a narrowed region and a candidate gene. *Genet. Med.* **3,** 91–101 (2001).

Supplementary Information accompanies the paper on *Nature*'s website (◗ **http://www.nature.com/nature**).

Acknowledgements We thank M. Traditi and G. Lago for the design of the website, and B. Fischer for preparation of the specimens. We are grateful to F. Chapot, S. Deutsch, M. Guipponi, K. Hashimoto, P. Kahlem, J. Michaud, H. S. Scott and M. L. Yaspo for plasmids and reagents, and to J. Ahidan, M. Friedli, C. Rossier and the Telethon Institute of Genetics and Medicine (TIGEM) RNA *in situ* hybridization core for core assistance. This work was supported by grants from the Jérôme Lejeune Foundation to R.L. and A.R.; from the Swiss Fonds National Suisse de la Recherche Scientifique, the European Union/Office Fédéral de l'Education et de la Santé and ChildCare foundation to S.E.A.; from the German Ministry of Research to G.E.; from the EC Fifth Framework Program to A.B. and G.E.; from the Italian Telethon Foundation to TIGEM; from The National Center for Competence in Research-Frontiers in Genetics to S.E.A.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.B. (e-mail: ballabio@tigem.it), S.E.A. (e-mail: stylianos.antonarakis@medecine.unige.ch) or G.E. (e-mail: gregor.eichele@mpihan.mpg.de).

Authors' contributions The three laboratories of A.B., S.E.A. and G.E. contributed equally to this work.

A gene expression map of human chromosome 21 orthologues in the mouse

The HSA21 expression map initiative
***Group 1:Yorick Gitton†, Nadia Dahmane‡, Sonya Baik†**
Ariel Ruiz i Altaba†
***Group 2:Lorenz Neidhardt§, Manuela Scholze§, Bernhard G. Herrmann§**
***Group 3:Pascal Kahlem‖, Alia Benkahla‖, Sabine Schrinner‖,**
Reha Yildirimman‖, Ralf Herwig‖, Hans Lehrach‖ & Marie-Laure Yaspo‖

** These author groups contributed equally to the work.*
† Skirball Institute, Developmental Genetics Program and Department of Cell Biology, New York University School of Medicine, 540 First Avenue, New York, New York 10016, USA
‡ Université de la Méditerranée – CNRS UMR 6156, Institut de Biologie du Développement, Campus de Luminy, Case 907, 13 288 Marseille Cedex 09, France
§ Max Planck Institute for Immunology, Department of Developmental Biology, Stübeweg 51, D-79108 Freiburg, Germany
‖ Max Planck Institute for Molecular Genetics, Department of Vertebrate Genomics, Ihnestrasse 73, D-14195 Berlin, Germany

The DNA sequence of human chromosome 21 (HSA21)[1] has opened the route for a systematic molecular characterization of all of its genes. Trisomy 21 is associated with Down's syndrome, the most common genetic cause of mental retardation in humans. The phenotype includes various organ dysmorphies, stereotypic craniofacial anomalies and brain malformations[2]. Molecular analysis of congenital aneuploidies poses a particular challenge because the aneuploid region contains many protein-coding genes whose function is unknown. One essential step towards understanding their function is to analyse mRNA expression patterns at key stages of organism development. Seminal works in flies, frogs and mice showed that genes whose expression is restricted spatially and/or temporally are often linked with specific ontogenic processes. Here we describe expression profiles of mouse orthologues to HSA21 genes by a combination of large-scale mRNA *in situ* hybridization at critical stages of embryonic and brain development and *in silico* (computed) mining of expressed sequence tags. This chromosome-scale expression annotation associates many of the genes tested with a potential biological role and suggests candidates for the pathogenesis of Down's syndrome.

The current HSA21 gene catalogue[1] (http://chr21.molgen.mpg.de) contains 238 entries for which we have identified 168 cognate mouse orthologues (see Methods in the Supplementary Information). We isolated 187 mouse cDNA clones matching 158 unique genes (referred to as mmu21 genes; see Supplementary Table 1) whose orthology was confirmed on the basis of the known synteny between HSA21 and segments of mouse chromosomes MMU16, MMU10 and MMU17 (http://www.informatics.jax.org/) (see Supplementary Information; this database is also available at http://chr21.molgen.mpg.de/hsa21/).

To identify potential candidates with a role in patterning and organ development, we have explored the expression of the 158 mmu21 genes by systematic whole-mount *in situ* hybridization[3] (WISH) at mid-gestation (embryonic day 9.5; E9.5), a stage covering a wide range of embryonic processes. We also analysed a subset of clones at two other stages (Supplementary Table 1). We found 111 of 158 genes expressed at E9.5; 78 genes showed widespread expression and 33 genes a restricted pattern (49% and 21%, respectively, of the genes examined at E9.5). In addition, 12 widespread genes also defined particular embryonic structures (for example, *Prkcbp2*; Supplementary Table 1). Among the mmu21 genes conserved in *Saccharomyces cerevisiae* (Y), *Caenorhabditis*

