



# Similar gene expression profiles do not imply similar tissue functions

Itai Yanai<sup>1,2,\*</sup>, Jan O. Korbel<sup>3,4,\*</sup>, Stephanie Boue<sup>3,\*</sup>, Shannon K. McWeeney<sup>5</sup>, Peer Bork<sup>3,4</sup> and Martin J. Lercher<sup>3,6</sup>

<sup>1</sup>Department of Molecular Genetics, Weizmann Institute of Science, 76100 Rehovot, Israel

<sup>2</sup>Department of Molecular and Cellular Biology, Harvard University, MA 02138, USA

<sup>3</sup>European Molecular Biology Laboratory, Meyerhofstraße 1, 69117 Heidelberg, Germany

<sup>4</sup>Max-Delbrück-Center for Molecular Medicine, 13092 Berlin-Buch, Germany

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

<sup>6</sup>Department of Biology and Biochemistry, University of Bath, Bath, UK, BA2 7AY

**Although similarities in gene expression among tissues are commonly inferred to reflect functional constraints, this has never been formally tested. Furthermore, it is unclear which evolutionary processes are responsible for the observed similarities. When examining genome-wide expression data in mouse, we found that patterns of expression similarity between tissues extend to genes that are unlikely to function in the tissues. Thus, ectopic expression can seem coordinated across tissues. This indicates that knowledge of gene expression patterns *per se* is insufficient to infer gene function. Ectopic expression is possibly explained as expression leakage, caused by spreading of chromatin modifications or the transcription apparatus into neighboring genes.**

## Introduction

It is common practice to summarize gene expression data across different tissues by hierarchical clusters [1]. Although these clusters formally resemble phylogenetic trees, they are often assumed to represent functional relationships [1–6]. However, this functional assumption has never been formally tested. Zhang *et al.* [7] analyzed the similarity of expression among individual genes and found that groups of coexpressed genes are enriched in functionally similar genes, as classified by the Gene Ontology (GO) database. However, the expression patterns of individual genes alone (e.g. tissue-specificity [7]) seem to be of limited utility in predicting gene function – probably because a substantial proportion of gene expression is non-functional, a point of view supported by a multitude of recent observations (reviewed in Ref. [8]).

Thus, it currently remains unclear if expression similarity between tissues, measured across the complete transcriptome, reflects similar tissue functions. There are at least three plausible models that could explain these gene expression patterns. These models, outlined in detail below, are not mutually exclusive.

(i) Expression similarity among tissues might solely be driven by selection on gene function. Thus, tissues might express overlapping sets of genes because they perform similar functions, for example, skeletal and heart muscle. In this model, expression patterns arose independently in different tissues during evolution, and any similarities are a consequence of parallel selective forces. This process is termed convergent evolution.

(ii) Alternatively, expression similarity might reflect the evolutionary relatedness of tissues. The emergence of new tissues can be viewed akin to the emergence of new genes. New tissues (if they are not simply a conversion of a complete precursor tissue) arise by a process we term ‘tissue duplication’: either one part of the precursor tissue retains its original function while another part takes up a new function, or the functions of the ancestral tissue are split between two specialized descendants. For example, the tissue lineages of cerebellum and frontal cortex started to diverge from a primitive ancestral brain ~530 million years ago [9]. This evolutionary model rests on the assumption that gene expression patterns evolve independently in the descending tissues, and that expression evolution occurs sufficiently slowly to retain ancestral patterns over substantial timescales.

(iii) Expression similarity among tissues might be a consequence of expression leakage, where functionally important expression of one gene results in the ectopic expression of its neighbors. This can happen through: (i) direct neighborhood effects (e.g. overshooting transcription, possibly reflected in the massive transcripts observed in the mouse transcriptome [10]); (ii) modifications of the properties and accessibility of chromatin (e.g. histone modifications, which might be inherited epigenetically [11]); or (iii) combinatorics of tissue-factor (TF)-binding sites [8], where neighborhood is in regulatory rather than chromatin space.

Both the model of evolutionary relatedness and a special case of the transcriptional leakage model were

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

\* These authors contributed equally to this work.

suggested as possible explanations for the findings by Khaitovich *et al.* [12], who showed that expression differences among brain tissues in primates increased almost uniformly over time (i.e. in a clock-like manner).

We analyzed previously published mouse gene expression data to distinguish between the three models described (for details of the methods, see the supplementary material online). First, we constructed a tree of tissue relationships based on expression distances (i.e. on quantitative differences in gene expression measured across the complete transcriptome). We showed that tissue clusters are recovered from gene sets that are unlikely to have a function in the studied tissues, demonstrating that the functional convergent evolution model by itself is insufficient to explain the observations. Next, we attempted to differentiate between the two remaining models.

### The mammalian tissue tree

We re-analyzed microarray data [13] that compared the expression patterns of 16 004 Ensembl genes across 47 adult mouse organs, tissues and cell types (for simplicity referred to as ‘tissues’). To measure expression differences among tissues, we calculated Euclidean distances between the gene expression vectors (i.e. the square-root of the summed squared expression differences per gene; similar results were obtained using mean squared differences, data not shown). From these distances, we constructed the tissue tree in Figure 1 with the neighbor joining method. This tree is a graphical representation of expression similarity among tissues. Its topology is remarkably robust, as shown by bootstrap (Figure 1) and replicate analyses (Figure S1 in the supplementary material online). Many groups in Figure 1 seem consistent with a clustering by tissue functions: for example, we recovered clusters comprising brain tissues (approximately resolved into front-, mid- and hind-brain), the immune system or the digestive tract. However, other clusters, for example, the grouping of adipose tissue, lung, trachea and adrenal gland, lack obvious functional similarities. Some clusters are consistent with ontological relationships (e.g. the grouping of central nervous system tissues; Figure 2), but this does not apply to others (e.g. the grouping of heart with skeletal muscle).

### Tissue clusters are recovered in the absence of gene function

The convergent evolution (or functional) model is based on the assumption that tissue similarities are caused by selection on gene expression in response to gene function. In this model, only genes that have a function in at least some of the tissues under study can support the clusters observed in Figure 1. Thus, if we can identify a set of functionally well-characterized genes, and then restrict the analysis to tissues where those genes have no annotated function, we expect the clusters observed in Figure 1 to disappear. In this context, it is important to distinguish between tissue-specific expression and tissue-specific function. Although these two are often implicitly equated, many genes might be expressed in tissues in which they have no selected function (ectopic expression) [8,14,15].

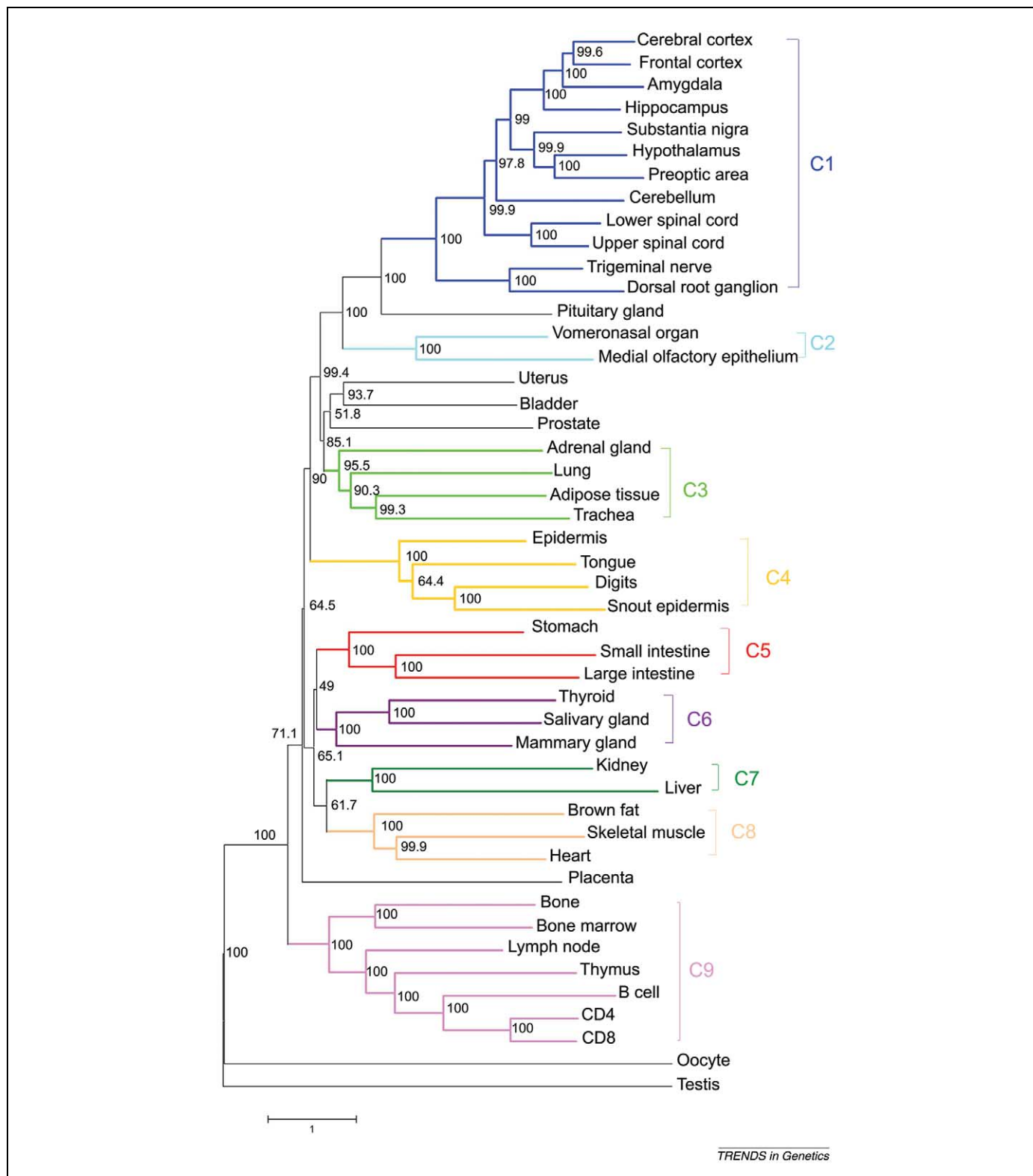
To test the validity of the functional assumption underlying the convergent evolution model, we analyzed two sets of genes with experimentally confirmed tissue-specific function: genes involved in germ cell formation (functioning in testis and oocyte), and genes with brain-specific function. Stringent literature checks were made for every gene, ensuring that functional specificity was experimentally supported (supplementary material online). After exclusion of the tissues in which these genes have a documented function, we found that each of the other clusters in Figure 1 was still supported by most genes (Table 1). A complementary measure based on Euclidean tissue distances gave similar results (Table S1 supplementary material online). Thus, many genes with a tissue-specific function are expressed non-randomly in other tissues, and this expression reproduces the tissue relationships in Figure 1. Although it is likely that gene annotations are incomplete, with some of the genes performing unknown additional functions in other tissues, this is unlikely to apply to more than a few genes in a few tissues. Because most genes in each subset support the clusters (Table 1), this should not affect our conclusions.

We can derive a second, related prediction from the convergent evolution model. If we restrict our analysis to a set of genes that together perform the same function in all tissues, then selection should act in exactly the same way on each gene in the set. Although convergent evolution can cause the set as a whole to be similarly up- or down-regulated in tissues with related functions, we do not expect any systematic variation in the relative abundance of different transcripts. Thus, after controlling for abundance of the set as a whole, any remaining variation should not support the clusters in Figure 1. To test this prediction, we first selected 56 genes encoding oxidative phosphorylation proteins [16,17], which are essential for energy production in the mitochondria. To correct for variations in mitochondria number and morphology across tissues, we normalized the expression data from the set by dividing the expression intensities of each tissue by the mean of the 56 genes for this tissue (supplementary material online; similar results are obtained without this normalization). Each of the clusters in Figure 1 was supported by most of the genes in this normalized set (Table 1), again contradicting the assumption underlying the convergent evolution model. Similar results were obtained from a set of genes with products that are part of the spliceosomal complex [18] (Table 1).

We conclude that even genes that do not experience distinct selective pressures across tissues recover the clusters in Figure 1. Thus, convergent evolution of gene expression owing to parallel selection pressures is not sufficient to explain the observed similarities in gene expression patterns among tissues.

### Tissue clusters are recovered from expression of both old and new proteins

Could the evolutionary relatedness of tissues be responsible for the non-functional expression similarity that we observed? It is conceivable that two tissues (e.g. cerebellum and frontal cortex) developed from a common precursor, and retained part of the ancestral expression

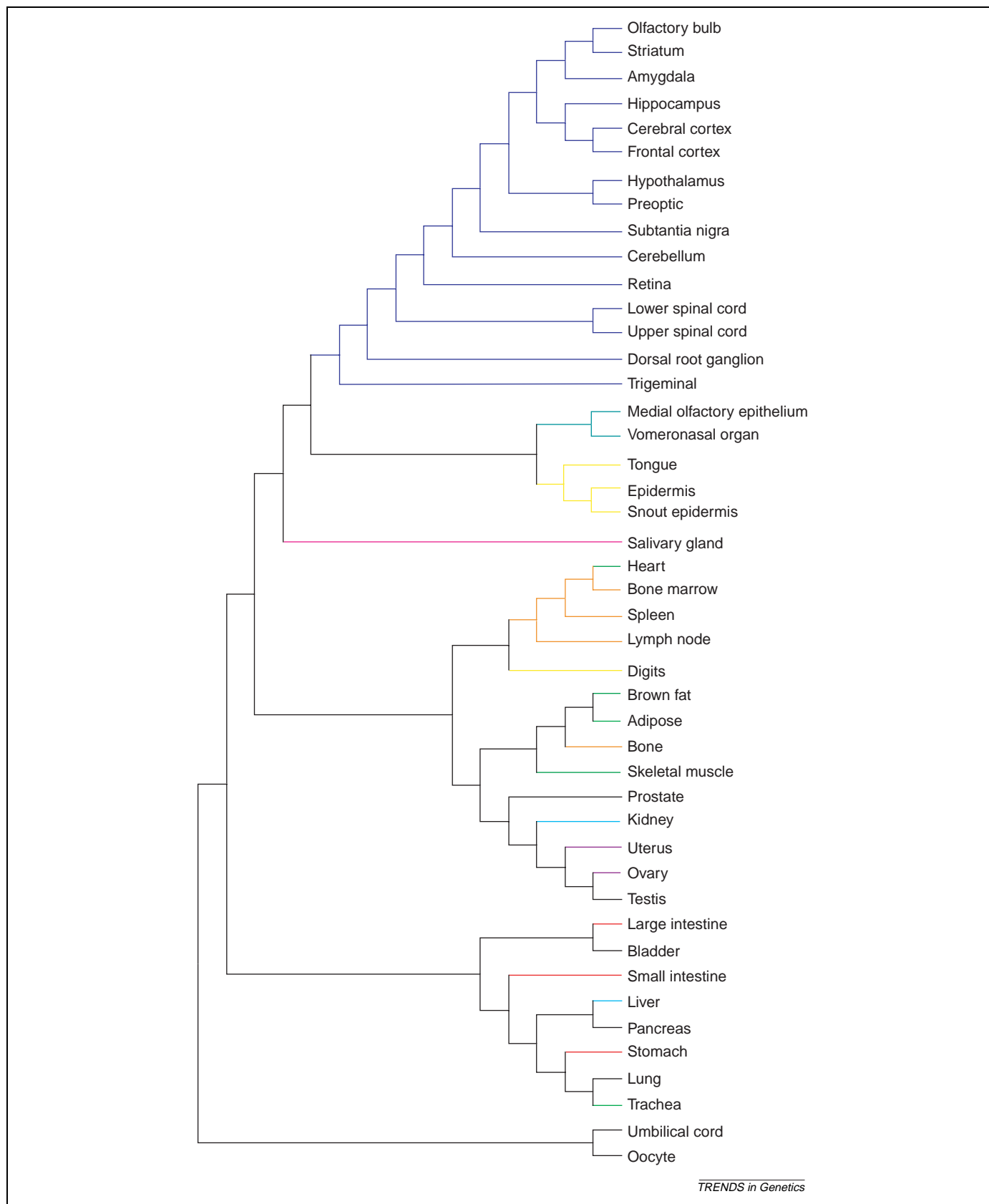


**Figure 1.** A tissue tree constructed from Euclidean distances between expression vectors with the neighbor-joining method. Bootstrap support for branches is given as percentages. The major clusters (each with bootstrap support > 95%) are colored; see Table 1 for the proportion of genes that support each cluster.

pattern even in the absence of a relevant selection pressure. In the light of previous evidence that gene expression tends to evolve rapidly [19–21], it might seem unlikely that this evolution still proceeds sufficiently slowly to retain remnants of ancestral expression patterns over the timescale of tissue evolution, which spans several

hundred million years. However, we cannot dismiss this model without further examination.

Proteins can be classified according to their phylogenetic age [22]. Although in most cases new proteins were the product of gene duplication, rapid divergence of amino acid sequence and expression patterns after



TRENDS in Genetics

**Figure 2.** Ontogenetic relationships between mammalian tissues [30]. Colored branches indicate clusters of tissues identified through common gene expression patterns (see Figure 1).

duplication [19–21] are likely to have obscured any ancestral patterns in those genes that show a restricted phylogenetic distribution. Thus, we expect that genes can

record only the evolutionary history of tissues that arose after their own emergence. Under the evolutionary relatedness model, this predicts that whereas universal

**Table 1. Support for the tissue clusters from gene subsets<sup>a,b</sup>**

Gene set	Number of genes	Tissue clusters								
		C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	C <sub>4</sub>	C <sub>5</sub>	C <sub>6</sub>	C <sub>7</sub>	C <sub>8</sub>	C <sub>9</sub>
All genes	16 004	0.97 <sup>c</sup>	0.81 <sup>c</sup>	0.82 <sup>c</sup>	0.83 <sup>c</sup>	0.76 <sup>c</sup>	0.81 <sup>c</sup>	0.71 <sup>c</sup>	0.76 <sup>c</sup>	0.75 <sup>c</sup>
Germ cell formation	39	0.92 <sup>c</sup>	0.69 <sup>e</sup>	0.74 <sup>d</sup>	0.72 <sup>e</sup>	0.74 <sup>d</sup>	0.51	0.64 <sup>f</sup>	0.59	0.59
Brain	29	–	0.86 <sup>c</sup>	0.97 <sup>c</sup>	0.93 <sup>c</sup>	1 <sup>c</sup>	0.9 <sup>c</sup>	0.79 <sup>d</sup>	0.9 <sup>c</sup>	0.9 <sup>c</sup>
Oxidative phosphorylation	56	1 <sup>c</sup>	0.89 <sup>c</sup>	0.84 <sup>c</sup>	0.82 <sup>c</sup>	0.95 <sup>c</sup>	0.82 <sup>c</sup>	0.86 <sup>c</sup>	0.8 <sup>c</sup>	0.84 <sup>c</sup>
Spliceosome	50	1 <sup>c</sup>	0.88 <sup>c</sup>	0.9 <sup>c</sup>	0.92 <sup>c</sup>	0.82 <sup>c</sup>	0.82 <sup>c</sup>	0.56	0.68 <sup>e</sup>	0.84 <sup>c</sup>

<sup>a</sup>For a list of the tissue clusters, see Figure 1.

<sup>b</sup>Individual values report the fraction of genes that support each cluster (i.e. that show greater similarity in expression within the cluster compared with across cluster).

<sup>c</sup> $P < 0.0001$ .

<sup>d</sup> $P < 0.001$ .

<sup>e</sup> $P < 0.01$ .

<sup>f</sup> $P < 0.05$ .

or eukaryote-specific genes (those mammalian genes with orthologs in yeast) should be informative for the history of all tissues, the expression patterns of mammalian-specific genes should have recorded only the history of mammalian-specific tissues. Contrary to these expectations, we found that genes of all age classes recover the tissue clusters in Figure 1 equally well, both when using all genes and when restricting to gene subsets that are unlikely to have a selected function in the studied tissues (Table 2). Expression similarities of brain-specific genes in the non-brain tissues were not dependant on the age of the genes (supplementary material online).

Expression evolution seems to proceed at a regular pace [12,19,20]. If gene expression similarities reflect evolutionary relatedness, then – in close analogy to the evolution of duplicated genes – the overall expression difference between tissues resulting from the duplication of a tissue should be independent of later speciation events. To test this prediction, we examined 8323 genes across 26 tissues for which expression profiles were available in both mouse and human [13]. Contrary to expectations, we found only a weak (although statistically significant) correlation between tissue expression distances measured within human and within mouse ( $r^2 = 0.229$ ,  $P < 10^{-15}$ ). This is also consistent with the finding of shorter within- than between-species distances for primate brain regions [12]. Taken together, these results suggest that similarity in tissue expression profiles is unlikely to represent remnants of ancestral expression patterns.

### Is transcriptional leakage responsible?

What then might cause the systematic tissue associations? Recently, genomic analyses in a wide range of eukaryotes have identified chromosomal clusters of coexpressed genes (reviewed in Ref. [23]). Although such clusters can reflect the association of housekeeping genes [24] or essential genes [25], there are often no obvious functional associations among genes comprising a cluster [26]. This has led to the suggestion that clusters of coexpressed genes are a consequence of transcriptional leakage [26] (e.g. by chromatin modifications granting the transcription machinery access [11,23] to a genomic region containing both target and non-target genes).

To test this idea, we turned again to the sets of functionally well-characterized genes. For each gene in our brain-specific set, we examined its expression outside the central nervous system (CNS). For each non-brain tissue, we classified the brain-specific genes into those that were expressed above average in that tissue (compared with all genes) and those that were below average. Above-average brain genes were closer to their nearest genomic neighbor expressed in that tissue compared with below-average genes (3.9 Mb versus 4.7 Mb,  $P = 0.045$ ). Thus, ectopic expression seems to be associated with the tissue-specific expression of neighboring genes. However, when we applied the same test to genes involved in germ-cell formation, we did not find any significant difference between above- and below-average gene-tissue combinations (3.9 Mb versus 3.8 Mb,

**Table 2. Young and old genes support the same tissue associations<sup>a</sup>**

Gene set	Number of genes	Tissue clusters								
		C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	C <sub>4</sub>	C <sub>5</sub>	C <sub>6</sub>	C <sub>7</sub>	C <sub>8</sub>	C <sub>9</sub>
Universal	981	0.99 <sup>b</sup>	0.84 <sup>b</sup>	0.86 <sup>b</sup>	0.87 <sup>b</sup>	0.76 <sup>b</sup>	0.85 <sup>b</sup>	0.75 <sup>b</sup>	0.81 <sup>b</sup>	0.83 <sup>b</sup>
Eukaryotic	1411	0.98 <sup>b</sup>	0.85 <sup>b</sup>	0.84 <sup>b</sup>	0.86 <sup>b</sup>	0.79 <sup>b</sup>	0.85 <sup>b</sup>	0.76 <sup>b</sup>	0.81 <sup>b</sup>	0.8 <sup>b</sup>
Metazoan	1070	0.95 <sup>b</sup>	0.85 <sup>b</sup>	0.83 <sup>b</sup>	0.85 <sup>b</sup>	0.76 <sup>b</sup>	0.83 <sup>b</sup>	0.72 <sup>b</sup>	0.81 <sup>b</sup>	0.78 <sup>b</sup>
Mammalian	736	0.97 <sup>b</sup>	0.83 <sup>b</sup>	0.83 <sup>b</sup>	0.9 <sup>b</sup>	0.8 <sup>b</sup>	0.87 <sup>b</sup>	0.76 <sup>b</sup>	0.81 <sup>b</sup>	0.79 <sup>b</sup>
<b>Functionally well-characterized genes without annotated function in the tissues examined</b>										
Universal	30	–	0.9 <sup>b</sup>	0.97 <sup>b</sup>	0.87 <sup>b</sup>	0.97 <sup>b</sup>	0.83 <sup>b</sup>	0.77 <sup>c</sup>	0.8 <sup>c</sup>	0.87 <sup>b</sup>
Eukaryotic	32	–	0.81 <sup>b</sup>	0.84 <sup>b</sup>	0.84 <sup>b</sup>	0.88 <sup>b</sup>	0.88 <sup>b</sup>	0.78 <sup>c</sup>	0.81 <sup>b</sup>	0.81 <sup>b</sup>
Metazoan	16	–	0.81 <sup>d</sup>	0.75 <sup>e</sup>	0.81 <sup>d</sup>	0.88 <sup>c</sup>	0.75 <sup>e</sup>	0.81 <sup>d</sup>	0.81 <sup>d</sup>	0.81 <sup>d</sup>
Mammalian	15	–	0.8 <sup>d</sup>	0.87 <sup>c</sup>	0.93 <sup>b</sup>	0.87 <sup>c</sup>	0.87 <sup>c</sup>	0.93 <sup>b</sup>	0.87 <sup>c</sup>	0.87 <sup>c</sup>

<sup>a</sup>Values are derived as detailed in Table 1; only genes of known age are included. The shaded part includes only functionally well-characterized subgroups (Table 1); here cluster C1 is excluded because one of the gene sets is functionally specific to the brain.

<sup>b</sup> $P < 0.0001$ .

<sup>c</sup> $P < 0.001$ .

<sup>d</sup> $P < 0.01$ .

<sup>e</sup> $P < 0.05$ .

$P=0.33$ ); this could reflect a lack of power, because only five genes in this set had ectopic expression of  $>20\%$  above average.

If transcriptional leakage was an important source of ectopic expression, we would expect that genes would change their ectopic expression patterns if they move into a new genomic neighborhood. Thus, when comparing two species, we would expect to find that expression patterns among orthologs located in syntenic regions were more similar than among those located in different genomic environments. We found that the correlation between human and mouse gene expression vectors [13] in conserved neighborhoods ( $r=0.231 \pm 0.004$ ,  $N=6703$ ) was greater than in non-conserved neighborhoods

( $r=0.197 \pm 0.021$ ,  $N=201$ ); however, this difference was not statistically significant ( $P=0.107$  from  $t$ -test).

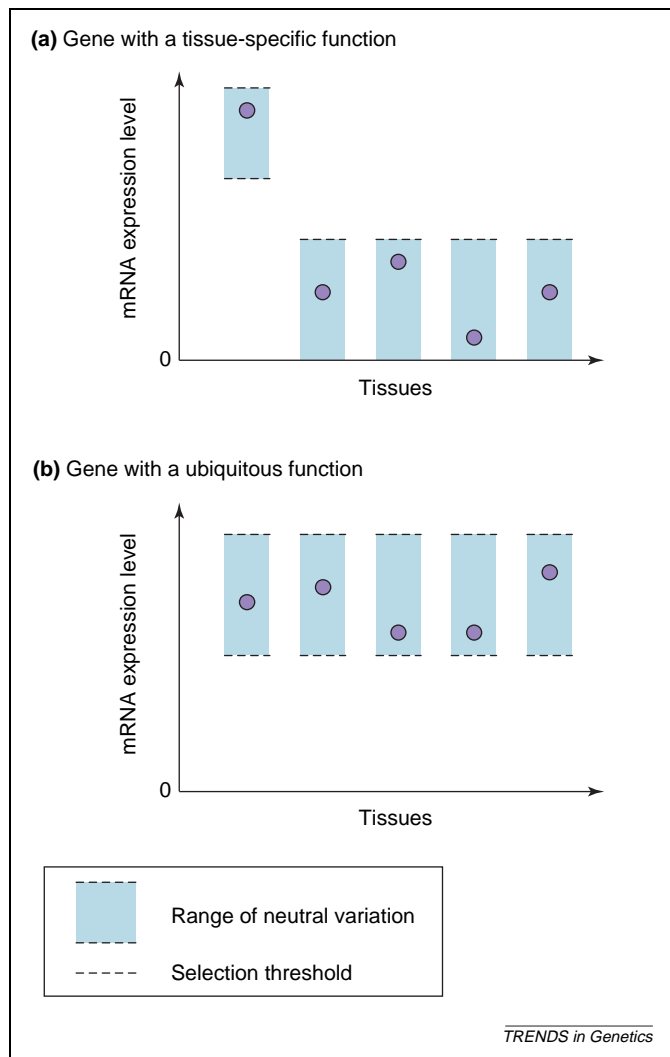
In sum, current data seem consistent with the notion of transcriptional leakage: expression of tissue-specific genes can 'overflow' into neighboring genes that have no function in the respective tissue. Accordingly, ectopic expression patterns might represent the 'shadows' of functional associations. However, the results are only suggestive, and further studies are necessary to confirm or reject this model. Alternative ideas must also be considered, such as the overlapping specificities of combinations of distinct transcription-factor-binding sites [8].

### Selection and neutrality in gene expression profiles

The observation of systematic non-functional gene expression is consistent with several recent studies that have questioned the functional impact of variability in gene expression [12,27,28]. How can we interpret such neutral variability in gene expression, in particular, in the light of the model of transcriptional leakage favored by our tests? Let us assume that for optimal fitness ( $\pm$  the range  $1/(2N_e)$  that cannot be seen by selection, where  $N_e$  is the effective population size), the expression pattern of a gene has to lie within a certain range. Selection would act only on extreme expression patterns outside of this range, permitting a range of selectively neutral variation (Figure 3). In some cases, the allowable range might have a lower limit of zero (i.e. the gene can be turned on or off in a tissue without consequence). Our results then suggest that some bias leads to a systematic use of the available expression space, such that neutral expression patterns might mirror functional expression of other (probably physically neighboring) genes.

### Concluding remarks

The transcriptional leakage model, and, in particular, our observation that expression patterns are informative even for tissues where the analyzed genes have no function, further suggests that there is little selection against unnecessary transcription in many mouse genes; this was noted previously based on PCR of individual genes [14,15]. Although unnecessary transcription seems wasteful, repression of transcription can also be energetically costly. More-stringent selection can operate on protein expression, mediated through post-transcriptional control mechanisms [29]. Thus, we must be cautious before inferring selected function from context-specific mRNA expression. This is consistent with recent findings that tissue-specific gene expression *per se* is a poor predictor of gene function [7]. Leaky transcription is likely to occur gene by gene. Therefore, coordinated transcription of a group of genes across different tissues might be more informative; this was confirmed by Zhang *et al.* [7]. However, analogous to sequence evolution, selected gene expression can only be inferred rigorously if conservation of expression patterns across species exceeds neutral levels; estimation of such neutral levels must include the effects of transcriptional leakage.



**Figure 3.** A schematic representation of neutral variability in gene expression. The expression levels across different tissues are depicted. (a) A gene that has a selected function only in the first tissue. (b) A gene performing the same selected function in all tissues. The colored areas indicate the allowed ranges of neutral variation, with the dashed lines indicating approximately where stabilizing selection starts to set in. In the first tissue in (a), the lower limit for neutral variation is more than zero (i.e. selection favors expression of the gene). In all other tissues in (a), the lower limit is zero, such that the gene can be either off or expressed at low levels. It is conceivable that the upper limit of neutral variation for tissues without a selected function might overlap with the range of neutral variation for the tissue with a selected function; expression levels could then be greater in tissues without a function than in those with a function. The simplified representation in this figure ignores temporal variation and variation across individuals, which could be included as additional dimensions of neutral expression.

## Acknowledgements

We thank John Walker for providing the raw data from Su *et al.* [13], and Jonathan Slack, Laurence Hurst, Lars Jensen, Christian von Mering, Detlev Arendt, Evgeny Zdobnov, Doron Lancet and Ron Shamir for helpful discussions. I.Y. is a Koshland Scholar. We are grateful for financial support from EMBO (I.Y.), the EU (J.S.B.), the BMBF (J.K. and P.B.), NIH NHLBI (S.K.M.), the Royal Society and the DFG (M.J.L.).

## Supplementary data

Supplementary data associated with this article can be found at doi:10.1016/j.tig.2006.01.006

## References

- Eisen, M.B. *et al.* (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U. S. A.* 95, 14863–14868
- Alon, U. *et al.* (1999) Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Natl. Acad. Sci. U. S. A.* 96, 6745–6750
- Hsiao, L.L. *et al.* (2001) A compendium of gene expression in normal human tissues. *Physiol. Genomics* 7, 97–104
- Saito-Hisaminato, A. *et al.* (2002) Genome-wide profiling of gene expression in 29 normal human tissues with a cDNA microarray. *DNA Res.* 9, 35–45
- Su, A.I. *et al.* (2002) Large-scale analysis of the human and mouse transcriptomes. *Proc. Natl. Acad. Sci. U. S. A.* 99, 4465–4470
- Walker, J.R. *et al.* (2004) Applications of a rat multiple tissue gene expression data set. *Genome Res.* 14, 742–749
- Zhang, W. *et al.* (2004) The functional landscape of mouse gene expression. *J. Biol.* 3, 21
- Rodriguez-Trelles, F. *et al.* (2005) Is ectopic expression caused by deregulatory mutations or due to gene-regulation leaks with evolutionary potential? *BioEssays* 27, 592–601
- Butler, A.B. and Hodos, W. (1996) *Comparative vertebrate neuroanatomy: evolution and adaptation*, Wiley-Liss
- Carninci, P. *et al.* (2005) The transcriptional landscape of the mammalian genome. *Science* 309, 1559–1563
- Turner, B.M. (2002) Cellular memory and the histone code. *Cell* 111, 285–291
- Khaitovich, P. *et al.* (2004) A neutral model of transcriptome evolution. *PLoS Biol.* DOI:10.1371/journal.pbio.0020132 (<http://biology.plosjournals.org/>)
- 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
- Sarkar, G. and Sommer, S.S. (1989) Access to a messenger RNA sequence or its protein product is not limited by tissue or species specificity. *Science* 244, 331–334
- Chelly, J. *et al.* (1989) Illegitimate transcription: transcription of any gene in any cell type. *Proc. Natl. Acad. Sci. U. S. A.* 86, 2617–2621
- Mootha, V.K. *et al.* (2003) Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria. *Cell* 115, 629–640
- 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
- Gaever, G. *et al.* (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418, 387–391
- Gu, Z. *et al.* (2002) Rapid divergence in expression between duplicate genes inferred from microarray data. *Trends Genet.* 18, 609–613
- Makova, K.D. and Li, W.H. (2003) Divergence in the spatial pattern of gene expression between human duplicate genes. *Genome Res.* 13, 1638–1645
- Wagner, A. (2000) Decoupled evolution of coding region and mRNA expression patterns after gene duplication: implications for the neutralist–selectionist debate. *Proc. Natl. Acad. Sci. U. S. A.* 97, 6579–6584
- 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, DOI:10.1186/gb-2005-6-7-r56 (<http://genomebiology.com/2005/6/7/R56>)
- Hurst, L.D. *et al.* (2004) The evolutionary dynamics of eukaryotic gene order. *Nat. Rev. Genet.* 5, 299–310
- Lercher, M.J. *et al.* (2002) Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nat. Genet.* 31, 180–183
- Pal, C. and Hurst, L.D. (2003) Evidence for co-evolution of gene order and recombination rate. *Nat. Genet.* 33, 392–395
- Spellman, P.T. and Rubin, G.M. (2002) Evidence for large domains of similarly expressed genes in the *Drosophila* genome. *J. Biol.* 1, 5
- Oleksiak, M.F. *et al.* (2002) Variation in gene expression within and among natural populations. *Nat. Genet.* 32, 261–266
- Yanai, I. *et al.* (2004) Incongruent expression profiles between human and mouse orthologous genes suggest widespread neutral evolution of transcription Control. *OMICS* 8, 15–24
- Keene, J.D. and Tenenbaum, S.A. (2002) Eukaryotic mRNPs may represent posttranscriptional operons. *Mol. Cell* 9, 1161–1167
- Gilbert, S.F. (2003) *Developmental biology*, Sinauer Associates