Available online at www.sciencedirect.com
ScienceDirect
Genomics 89 (2007) 580–587
GENOMICS
www.elsevier.com/locate/ygeno

# Genomic organization of transcriptomes in mammals: Coregulation and cofunctionality

Antje Purmann [a], Joern Toedling [b], Markus Schueler [a], Piero Carninci [c,d], Hans Lehrach [a], Yoshihide Hayashizaki [c,d], Wolfgang Huber [b], Silke Sperling [a,*]

[a] *Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Ihnestrasse 73, 14195 Berlin, Germany*
[b] *EMBL European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK*
[c] *Genome Exploration Research Group (Genome Network Project Core Group), RIKEN Genomic Sciences Center, RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan*
[d] *Genome Science Laboratory, Discovery Research Institute, RIKEN Wako Institute, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan*

Received 29 August 2006; accepted 24 January 2007
Available online 21 March 2007

**Abstract**

In studies of their transcriptional activity, genomes have shown a high order of organization. We assessed the question of how genomically neighboring genes are transcriptionally coupled across tissues and what could be the driving force behind their coupling. We focused our analysis on the transcriptome information for 13 tissues of *Mus musculus* and 79 tissues of *Homo sapiens*. The analysis of coexpression patterns of genomically adjacent genes across tissues revealed 2619 and 1275 clusters of highly coexpressed genes, respectively. Most of these clusters consist of pairs and triplets of genes. They span a limited genomic length and are phylogenetically conserved between human and mouse. These clusters consist mainly of nonparalogous genes and show a decreased functional and similar regulatory relationship to one another compared to general genomic neighbors. We hypothesize that these clusters trace back to large-scale, qualitative, persistent reorganizations of the transcriptome, while transcription factor regulation is likely to handle fine-tuning of transcription on shorter time scales. Our data point to so far uncharacterized *cis*-acting units and reject cofunctionality as a driving force.


*Keywords:* Genomics; Gene expression; Transcription regulation; Regulatory elements; Mammals

Studies of genomes and transcriptomes have shown that genes are nonrandomly located in genomes and that genes of coordinated expression appear in clusters along the genome. This raises the question of how genomes have evolved and how they function. It is evident that the location of a gene in a genome affects its expression, for example, transgene activity can depend on the chromosomal integration site, or an intact gene in a different genomic location can have a pathological phenotype. Clusters of genes that are coexpressed were first identified on a genomic scale in *Saccharomyces cerevisiae* [1,2] and *Caenorhabditis elegans* [3,4]. In the latter, clusters could be attributed to the cotranscription of these genes in operons, a process that is unusual among eukaryotes. However, there is extensive evidence for clusters of coexpressed genes across all major eukaryotes. Using a less stringent definition of a cluster that allows for intervening genes with different expression patterns led to the identification of large groups of coexpressed genes in *Drosophila melanogaster* that span 10-30 genes or, on average, 125 kb of genomic DNA [5]. In *Homo sapiens*, genes with high expression levels tend to cluster in large domains [6,7]. Other reports indicate that genes coexpressed in a given tissue or cell state are clustered along the genome [8–14]. Further, it is well known that specific gene clusters, like for β-globin or HoxD genes, are regulated by a locus and global control region, respectively [15,16]. Here, the latter includes the control of several genes unrelated in structure and function and it is currently unclear whether this is an exceptional or a common feature for higher eukaryotes. Thus, the question arises

* Corresponding author. Fax: +49 30 8413 1699.
*E-mail address:* sperling@molgen.mpg.de (S. Sperling).

0888-7543/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2007.01.010

if there exist chromosomal transcriptional "hot spots" in a given tissue or if coexpressed clustered genes are mainly housekeeping genes that are expressed in many tissues [17,18]. Some reports indicate that clusters of coexpressed genes tend to be conserved through evolution, for example, coexpressed genes contain fewer breakpoints between human and mouse, indicating that they are held together by natural selection [19–21].

To date, however, we lack an understanding of how many such clusters may exist in mammals, how the transcriptional coupling of gene clusters is regulated in general, and what may be the driving force behind their formation. To address these issues, we assessed the conservation of gene clusters across tissues and species and investigated the functional and regulatory relationship between coexpressed clustered genes. We focused our analysis on two recent transcriptome datasets, namely the FANTOM3 data for 13 tissues of *Mus musculus*, obtained by cap-analysis gene expression (CAGE), and the GNF Symatlas data for 79 tissues of *H. sapiens*, obtained by microarray expression profiling [18,22].

## Results

### *Chromosomal clustering of transcriptomes*

We investigated the genomic organization of 13 *M. musculus* transcriptomes that had been extensively analyzed within the FANTOM3 project [22]. In particular, we were interested in the physical scale of coexpression of genes located adjacent to each other in the genome. We called a set of adjacent genes that are expressed in a particular tissue a cluster of coexpressed genes, independent of their expression levels. A cluster of two neighboring genes expressed in the same tissue is also called a pair, a cluster of three a triplet, and so on.

A large proportion of genes (approx 30-75%) were arranged in gene clusters along the genome without any prevalence for particular chromosomes. These clusters consisted mainly of pairs and triplets. Fig. 1A shows an example of the size distribution of the gene clusters identified in cerebellum, heart, macrophage, and muscle. To evaluate the significance of our observation, we compared the observed number of genes localized in clusters with permuted data, in which the ordering of genes on the DNA, but not their individual expression profiles across tissues, was permuted. This permutation scheme corresponds to a null hypothesis in which the coexpression of genes is independent of their genomic location, but follows the empirical correlations between tissues (Fig. 1B). Even though the number of clusters expected under the null hypotheses is high, the number of observed clusters is significantly larger.

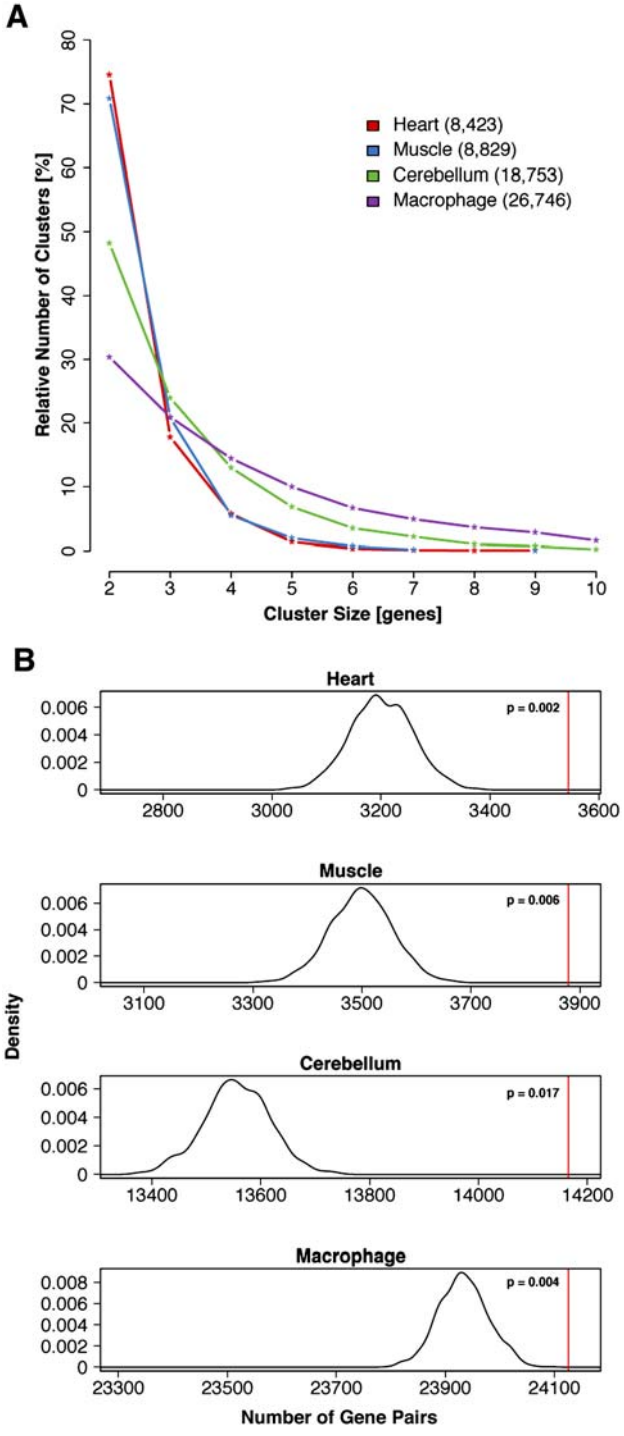


Fig. 1. Transcriptomes are organized in clusters. (A) Relative frequencies of observed gene clusters in the mouse transcriptome, for cerebellum, heart, macrophage, and muscle as examples. The total number of genes observed to be expressed in each tissue is given in parentheses. (B) To assess the statistical significance of the spatial clustering, we compared the number of observed clustered genes with the number of clusters from permuted data. Here, data are presented, as examples, for mouse cerebellum, heart, macrophage, and muscle.

### *Conservation of gene clusters across several tissues*

To quantify coexpression of a pair of genes in a set of $n$ tissues, we defined two coefficients. $\alpha$ is the proportion of tissues in which both genes are expressed, and $\Omega$ is the number of tissues in which either one or both genes are expressed divided by $n$. Both coefficients are numbers between 0 and 1, and $\alpha \leq \Omega$. If $\alpha = \Omega$, the two genes have an identical expression pattern across tissues, while a small ratio of $\alpha/\Omega$ indicates that their expression is not correlated. We computed these coefficients for each chromosomally neighboring pair of genes in the FANTOM3 data (Fig. 2A).

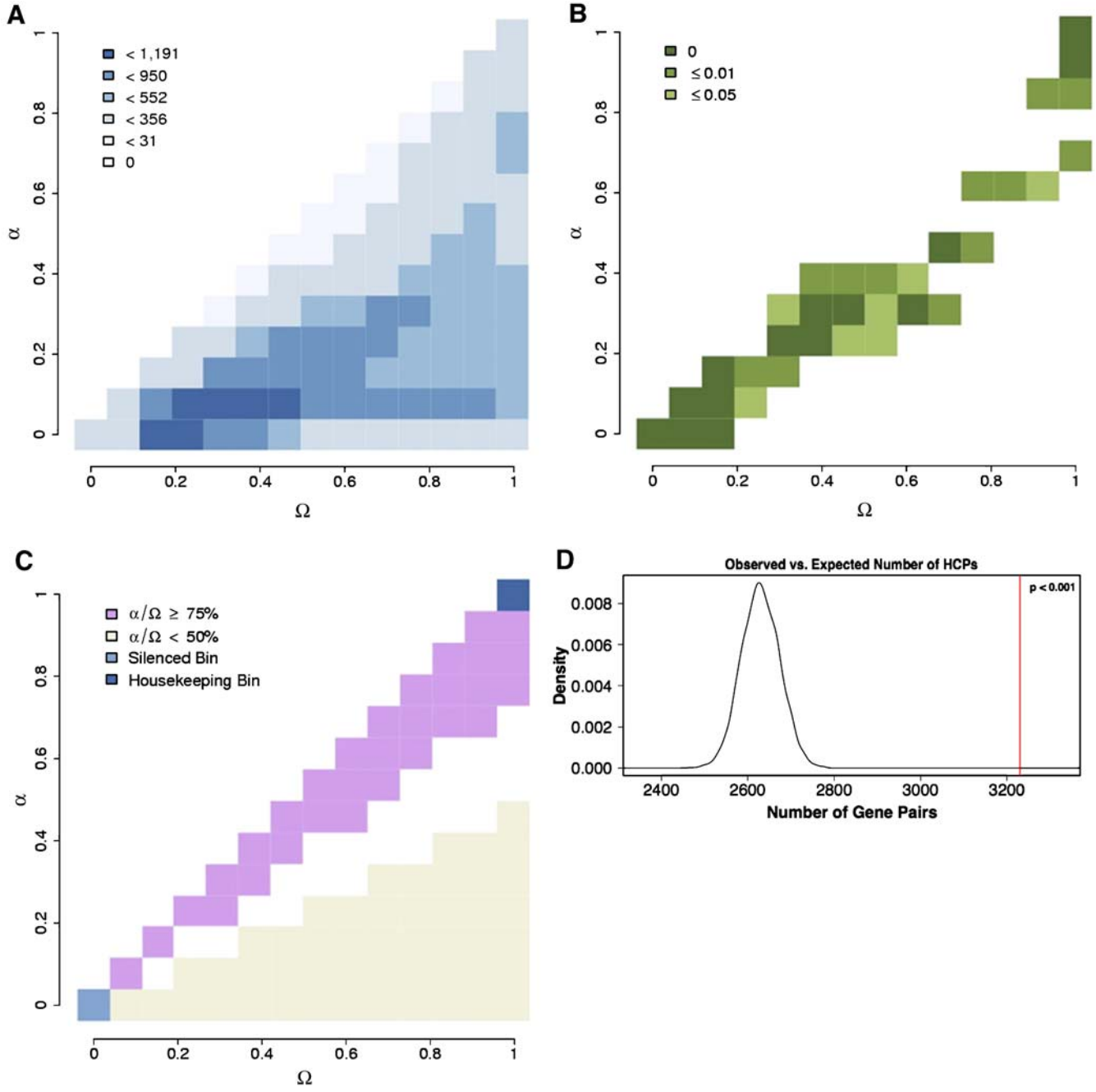


Fig. 2. Degree of coexpression of genomic neighbors defined by the coefficients α and Ω. We defined two coefficients, α and Ω, to quantify coexpression of a pair of genes in a set of *n* tissues. α is the proportion of tissues in which both genes are expressed, and Ω is the number of tissues in which either one or both genes are expressed divided by *n*. (A) Absolute bin occupancies, (B) empirical *p* values, and (C) defined coexpression categories based on the ratio between α and Ω. A high ratio between α and Ω indicates a high degree of coexpression. (D) Number of observed HCPs (red line) compared to the number of HCPs expected from permuted data.

The distribution of the bivariate coexpression measure (α, Ω) is nonrandom; certain combinations of α and Ω occur more frequently in the genome than expected if coexpression were independent of genomic location. Fig. 2B shows the *p* values for each combination of α and Ω, using the same permutation scheme as above. For each tuple (α, Ω), the empirical *p* value is given by the proportion of permutations in which equally many or more gene pairs display this coexpression pattern (α, Ω) than in the actual data.

Pairs were then assigned to one of the following coexpression categories, which depend on thresholds $\theta_{coex}$ and $\theta_{unc}$: (i) highly coexpressed, if $\alpha/\Omega \geq \theta_{coex}$ and $\alpha < 1$; (ii) housekeeping, if $\alpha = 1$; (iii) silenced, if $\Omega = 0$; (iv) uncorrelated, if $\alpha/\Omega \leq \theta_{unc}$. For the FANTOM3 data, we chose the thresholds $\theta_{coex} = 0.75$ and $\theta_{unc} = 0.5$ (Fig. 2C). This resulted in 3230 highly coexpressed pairs (HCPs), 154 housekeeping pairs, 36 silenced pairs, and 27,287 uncorrelated pairs (UCPs). Comparison of Fig. 1C with Fig. 1B shows that the number of

HCPs is larger than expected under the null hypothesis. Similarly, there are more housekeeping pairs, and more silenced pairs, while the frequency of UCPs is less than expected.

*Clusters decay at their flanks*

We considered highly coexpressed clusters (HCCs) of genes, which consist of one or several neighboring HCPs. In each tissue, either all of the genes in the HCC or a subset of them are expressed. A feature that goes along with our definition of HCC is that there is a preference for the central genes in the HCC to be expressed, while the flanking genes are more likely to get lost, meaning that genes clustering in one tissue are expressed in shorter clusters or not as clusters at all in other tissues. Thus they show a pronounced directionality in how they decay across tissues (66% of triplets decay directed). Conversely, in the case of unrelated transcriptional regulation, proposed for uncorrelated clusters (UCCs), genes at any position in the cluster get lost at the same rate (84% of triplets decay undirected).

*Highly coexpressed clusters and housekeeping functionality*

It has been suggested that housekeeping genes are often arranged in clusters along the genome [23]. However, the reverse is not true: most of our highly coexpressed pairs are expressed only in a limited fraction of tissues, hence these genes are not particularly housekeeping genes (Fig. 3).

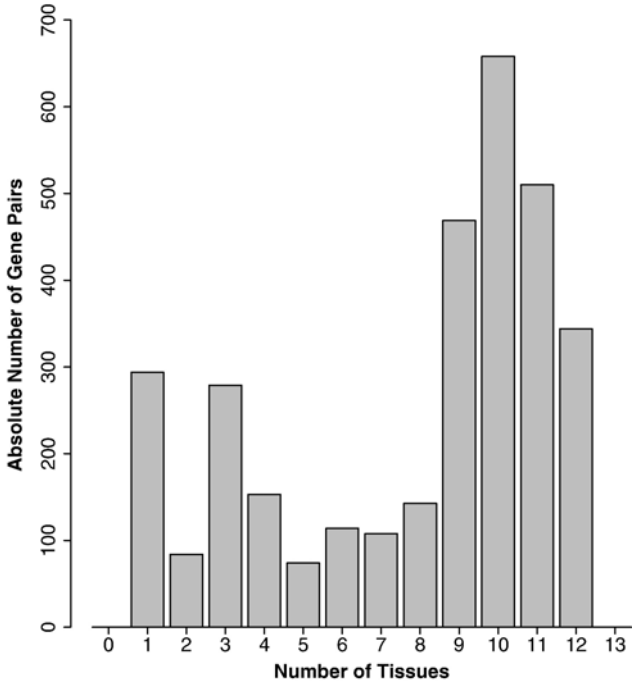


Fig. 3. Shared tissue expression of highly coexpressed gene pairs in mouse ($\alpha/\Omega \geq 0.75$ and $\alpha/\Omega < 1$). Shown is the number of gene pairs and their corresponding numbers of shared tissue expression. The number of tissues corresponds to α. Highly coexpressed gene pairs are not particularly housekeeping genes.

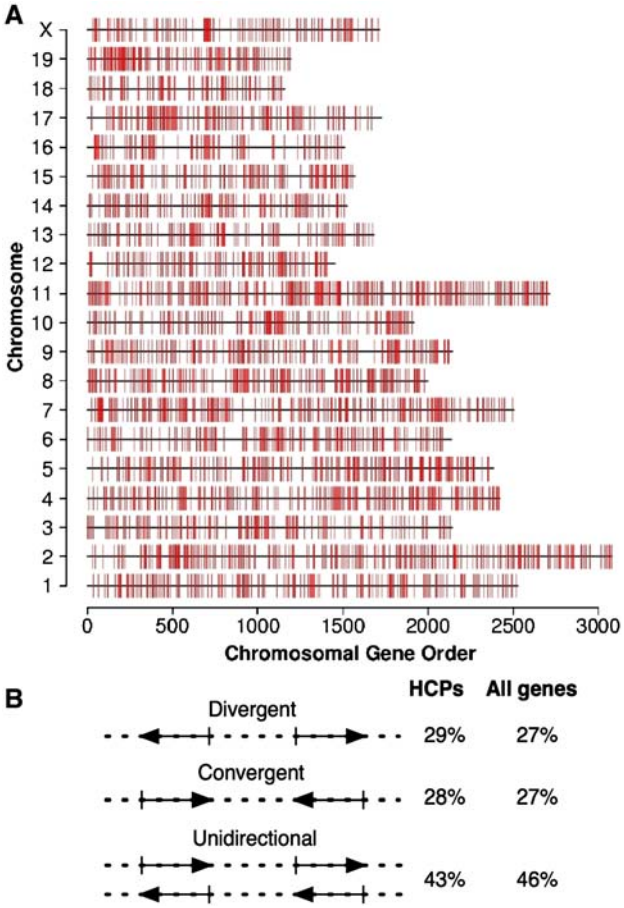


Fig. 4. Genomic distribution of highly coexpressed gene pairs. (A) Chromosomal gene order of HCPs (vertical red bars). Highly coexpressed gene pairs appear to distribute without notable prevalence among the chromosomes. (B) Distribution of strand orientations in HCPs compared to the overall distribution of orientations of adjacent genes in the mouse genome.

*Genomic location, orientation, and dimensions of HCPs*

Fig. 4A provides an overview of the spatial distribution of HCCs and shows certain regions with slightly higher concentrations of genes, but an almost homogeneous distribution of these regions across all chromosomes.

We addressed whether HCPs are characterized by a particular genomic orientation that might affect their transcriptional coupling. For example, a divergent orientation would enable the sharing of regulatory sequences between adjacent genes. Therefore, we divided the HCPs into three groups based on their relative orientation (divergent, convergent, and unidirectional) and compared this grouping with all the genomic pairs of FANTOM3 regardless of their level of coexpression. As shown in Fig. 4B, we found the distribution of genomic orientations to be similar between HCPs and all genomic pairs.

We assessed the intergenic and transcriptional start site distances for HCPs and all genomic pairs. HCPs have smaller intergenic (median of 7662 bp versus 18,665 bp for all pairs, $p = 3 \times 10^{-5}$, Wilcoxon rank sum test) and transcriptional start site distances (median of 28,781 bp versus 34,491 bp, $p = 8 \times 10^{-8}$, Wilcoxon rank sum test). We were then interested

in whether highly coexpressed clusters are characterized by limitations on their size, which could point to factors like chromatin remodeling contributing to the transcriptional coupling. We observed that the length of clusters measured in base pairs depends on the number of clustered genes. However, for HCCs the maximal observed number of adjacent genes within a cluster was limited to seven genes, and furthermore the 95% quantile of the cluster length was 320 kb (Fig. 5), which was much smaller compared to clusters of uncorrelated genes with 810 kb.

### *Functionality, paralogy, and transcriptional regulation of highly coexpressed gene clusters*

Paralogy, functional similarity, and the presence of common transcription factor binding sites (TFBSs) have been reported among genomic neighbors for a limited number of examples. We analyzed the frequency of shared Gene Ontology (GO) terms, protein domains, and TFBSs within adjacent gene pairs in FANTOM3 as a whole, as well as within HCPs. Naturally, our analysis was limited to only those genes annotated with Gene Ontology terms (36% FANTOM3 genes), protein domain information (42% FANTOM3 genes), or TFBS (33% FANTOM3 genes).

Table 1 shows that the sharing of domains and GO terms is less frequent in HCPs than in general genomic neighbors, whereas sharing of common TFBSs occurs at a similar rate. However, all of these observations are still more frequent than between nonneighboring genes, as is indicated by random permutations (as described above).

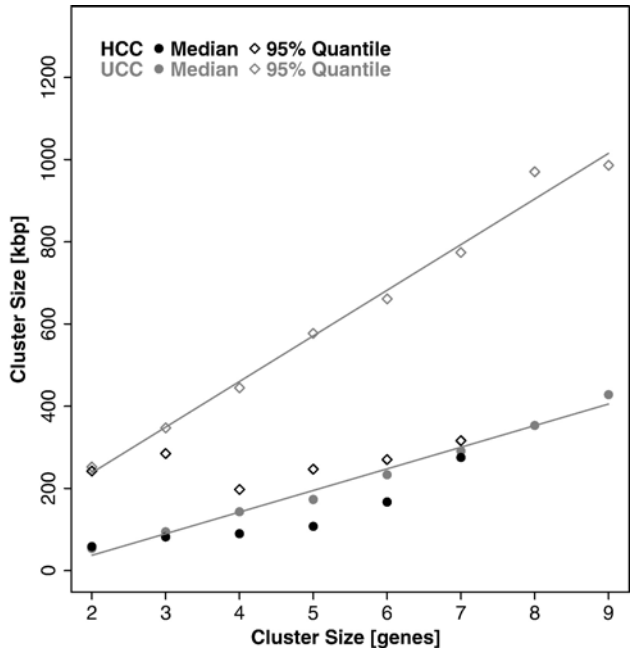


Fig. 5. Cluster sizes in base pairs and genes. Shown is the comparison of cluster sizes between HCCs and uncorrelated clusters measured in base pair length (*y* axis) as well as by the number of genes located in a particular cluster (*x* axis). The number of adjacent genes located in HCCs is limited to seven genes and their cluster length in base pairs is smaller compared to uncorrelated gene clusters.

Table 1
Functional and transcriptional properties of genomic neighbors in mouse

| | No. of annotated genomic pairs | Genomic pairs sharing similar annotations in % | No. of annotated HCPs | HCPs sharing similar annotations in % |
|---|---|---|---|---|
| GO terms | 5586 | 17.1 | 1272 | 8.8 |
| Protein domains | 7335 | 18.1 | 1567 | 10.8 |
| TF/TFBS | 4800 | 36.8/27.4 | 770 | 38.2/29.7 |

Genomic neighbors, irrespective of their coexpression, share Gene Ontology (GO) terms and protein domains to a much higher extent than highly coexpressed gene pairs (HCPs), whereas similar numbers of both groups of neighbors are potentially regulated by common transcription factors (TF) through their respective binding sites (TFBS).

To investigate the relationship between coexpression and paralogy, we analyzed gene pairs that had highly similar protein domains but showed only weak coexpression (in total 1307 gene pairs). Among them, we found members of well-known gene families that have previously been described to be clustered at certain genomic locations but to display tissue-specific expression nonetheless. For example, a family of S100 –calcium-binding proteins with its FANTOM3 tissue expression is depicted in Fig. 6 [24].

The 100 most frequently shared GO terms, protein domains, and transcription factors within general genomic neighbors are listed in the supplementary material.

### *Comparing FANTOM3 with microarray data of human*

To verify that our observations are not limited to one single dataset and organism, we performed the same analyses as described above for the 79 tissues in the *H. sapiens* part of the GNF Symatlas dataset [18]. We used slightly relaxed thresholds for the definition of HCPs, $\theta_{coex}=0.50$ and $\theta_{unc}=0.33$, to account for the lower coverage and higher false negative rate of these data.

We observed highly similar results for all analyses performed with the FANTOM3 dataset, such as the chromosomal clustering of genes expressed in human tissues, the conservation of gene clusters across tissues, and the tissue distribution of observed HCCs, as well as for the relationship of functional similarity, paralogy, and transcriptional regulation. From these results, which are presented in the supplementary material, we conclude that our observations are not biased toward the FANTOM3 data and may hold true for mammals in general.

### *Phylogenetic conservation of highly coexpressed gene clusters*

To assess if the observed transcriptional coregulation of neighbored genes is phylogenetically conserved between *H. sapiens* and *M. musculus*, we extracted the human homologs of all FANTOM3 genes, resulting in 2245 gene pairs consisting of direct genomic neighbors in both species. We found that HCPs in mice also tend to be highly coexpressed in human, as the frequency of human HCPs among mouse

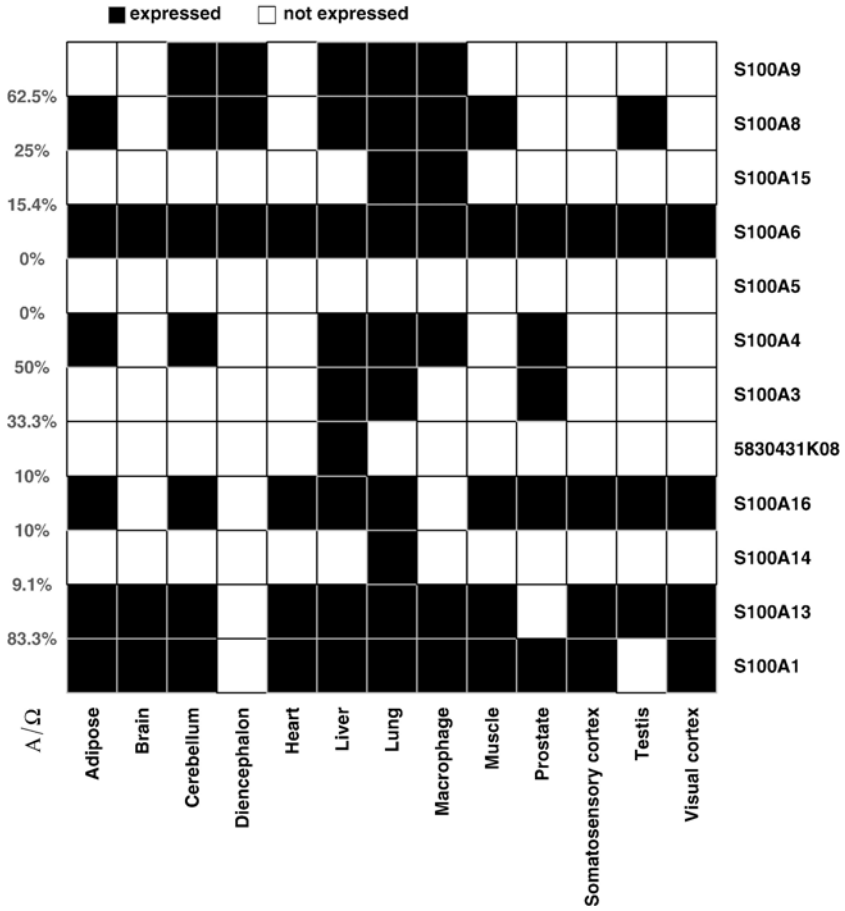


Fig. 6. Tissue expression of the S100 calcium-binding protein gene cluster. This cluster is likely to have originated by duplication of a common ancestor gene, as indicated by the genes' highly similar protein domain structures. Rows hold the genes, columns the FANTOM3 tissues. Black boxes indicate gene expression in a tissue and white boxes indicate no expression.

HCPs is 31% compared to 8% of human HCPs among mouse UCPs.

## Discussion

There is striking evidence that eukaryotic genomes show a high degree of gene organization. We focused on the large-scale, qualitative features of transcriptional regulation rather than on the fine-tuning. Thus we considered neighboring genes expressed in a particular tissue as coexpressed gene clusters regardless of expression levels of individual genes. We analyzed mouse transcriptomes of the FANTOM3 dataset derived by cloning techniques and confirmed our results on human data obtained through microarrays (GNF Symatlas).

Observing a surprisingly large number of gene clusters (sets of genomically neighboring, expressed genes), we suggested that only a subgroup of those clusters is actively transcriptionally coupled, whereas for a proportion of clusters, this observation would probably just be an effect of crowding of a given number of genes into a given genomic size. Therefore, we assessed the coexpression of gene clusters across tissues and observed a significant proportion of HCPs in addition to a small number of housekeeping gene clusters. This allowed us to extract sets of gene clusters (HCCs) that are characterized by a mainly directed decay across tissues, where the gene loss originates from only one end of the cluster and that may be indeed transcriptionally coupled. HCCs are characterized by clear-cut upper limits for physical cluster size and the number of genes making up these clusters, possibly reflecting the underlying mechanism. This finding may point to so far uncharacterized *cis*-acting units regulating the coexpression of certain sets of genes. To uncover such features further, our observation that HCCs are phylogenetically conserved between *M. musculus* and *H. sapiens* should provide the basis to extract potentially interesting conserved sequence features in the future. The coupling of highly coexpressed clusters could for example be controlled by histone modifications that are mediated by specific proteins initiating the opening or closing of chromatin and that spread along a chromosomal region until they meet a boundary element [23,25,26]. On the other hand, uncorrelated clusters probably arise as a consequence of intervening genes being transcriptionally silenced, for example, during cell differentiation. It has been shown that stem cells have a largely open chromatin formation and each step toward specialization is accompanied by down-regulation of genes in specific chromosomal regions [27]. These modifications could be stably inherited through cell division by DNA methylation, slowly reversed by silencing by histone lysine methylation or rapidly

modulated by histone acetylation. For *S. cerevisiae* it has been reported that genes that are regulated by the same sequence-specific transcription factor tend to be regularly spaced across the genome [28]. Other reports suggested that the transcriptional regulation has shaped the organization of transcriptional units on the chromosome [29], and recently, genes controlled by the transcription factor *aire* were shown to be clustered along the genome [30]. These reports are in line with our finding that sharing of TFBSs is a general phenomenon among genomic neighbors, but furthermore, we saw that this does not simultaneously result in their coexpression, as TFBSs are shared at a similar rate between highly coexpressed and general genomic neighbors.

Considering a broader cluster definition of genomic genes independent of their expression, it has been shown that genes coding for proteins involved in the same metabolic pathways tend to appear in chromosomal clusters [31]. Also genes that are involved in stable protein-protein complexes tend to be more tightly linked than expected [32]. We assessed paralogy and functional similarity as potential driving forces for the arrangement of clusters. Previous reports have demonstrated the cofunctionality of coexpressed gene clusters [2,5,33–35] but did not investigate cofunctionality of genomic neighbors in general for comparison. However, highly coexpressed neighbors do not seem to have a higher degree of co-functionality than general genomic neighbors. This unexpected finding may be explained in light of models of gene duplication in which duplication leads to neofunctionalization and subfunctionalization. Neofunctionalization can result in expression of duplicate genes in tissues lacking expression of the ancestral gene, while subfunctionalization can result in division of the ancestral expression pattern onto duplicates [36–39].

We hypothesize that HCCs trace back to large-scale, persistent reorganizations of the transcriptome, while TF regulation is likely to handle the fine-tuning of transcription on shorter time scales. To date, the underlying mechanism of transcriptional coupling between genomic neighbors is a matter of speculation. Our data point to so far unknown conserved *cis*-acting units involved in this regulatory process in mammals. It is hoped that studies addressing the chromatin remodeling process, e.g., through histone modifications, modifying transcription factors, or the nuclear spacing of transcriptional events, will provide further insights.

## Methods

### *FANTOM3 dataset and chromosomal clustering*

We considered the gene set and gene ordering defined by FANTOM3, consisting of 39,593 genes mapped to build mm5of the mouse genome [22]. In FANTOM3 genes are defined as transcriptional units (TU), representing discontinuous genomic regions from which one mature mRNA is derived. We used the term "gene" synonymously for TUs throughout this report. For cluster analyses, we concentrated on genes expressed in the following tissues, from which the expression information was obtained using CAGE technology (numbers of expressed genes): adipose (19,166), brain (13,766), cerebellum (18,753), diencephalon (6567), heart (8423), liver (30,721), lung (30,560), macrophage (26,746), muscle (8829), prostate gland (10,795), somatosensory cortex (17,193), testis (13,347), visual cortex (17,216). A set of physically neighboring genes coexpressed in a particular tissue is called a cluster of coexpressed genes.

### *Transcription factor binding sites*

TFBSs conserved in human/mouse/rat alignments were considered as annotated by the UCSC Genome Browser (http://genome.ucsc.edu/). Based on the ENSEMBL gene IDs of FANTOM3, we extracted all TFBSs annotated in the 10-kb upstream region of each gene. Taking into account that the TFBS annotation of the UCSC Genome Browser is conservative, we considered two genes to have common cis-acting regulatory units if they shared one or more TFBSs. The numbers of shared TFBSs per coexpressed gene pair were compared to a randomly built dataset (1000 permutations).

### *Gene ontology*

For each gene in the FANTOM3 data, we extracted the most specific GO terms to which that gene had been annotated and disregarded their parental terms [40]. We defined that two genes have a similar GO annotation if they shared at least 50% of these most specific terms as annotation. The GO terms for the genes were obtained from the given RefSeq and LocusLink identifiers using the Bioconductor software package "biomaRt" to query the Ensembl database (build 33, May 2005) [41–43]. We compared the concordance of GO annotation for gene pairs of interest with the one expected by random permutation of gene order [44].

### *Protein domain information*

We extracted the available domain information for the encoded protein of each gene in the FANTOM3 data. We considered two genes sharing at least 50% of their domains to have a similar domain annotation. Again, the annotations were obtained from the given RefSeq and LocusLink identifiers using the Bioconductor software package biomaRt to query the Ensembl database (build 33, May 2005). To compare observed similarities in protein domain annotation for gene pairs of interest with the one expected by chance, we employed the same permutation method as for the Gene Ontology annotation (see above).

## Acknowledgments

We thank Anja von Heydebreck, Steffen Grossmann, Stefanie Hammer, and Martin Vingron for helpful discussions. We also thank Claudio Lottaz, Steffen Durinck, and Raffaello Galli for their assistance with GNF Symatlas annotation, the biomaRt package, and transcription factor binding site annotation. This work was funded by the Max Planck Society for the Advancement of Science. The preparation of the FANTOM3 dataset was funded by a research grant for the RIKEN Genome Exploration Research Project from the Ministry of Education, Culture, Sports, Science, and Technology of the Japanese Government to Y.H.; a grant from the Genome Network Project of the Ministry of Education, Culture, Sports, Science, and Technology, Japan; and The Strategic Programs for R&D of RIKEN. We also thank Chikatoshi Kai and Jun Kawai for their contribution to FANTOM3.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2007.01.010.

## References


[1] R.J. Cho, et al., A genome-wide transcriptional analysis of the mitotic cell cycle, Mol. Cell 2 (1998) 65–73.

[2] B.A. Cohen, R.D. Mitra, J.D. Hughes, G.M. Church, A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression, Nat. Genet. 26 (2000) 183–186.

[3] T. Blumenthal, et al., A global analysis of *Caenorhabditis elegans* operons, Nature 417 (2002) 851–854.

[4] P.J. Roy, J.M. Stuart, J. Lund, S.K. Kim, Chromosomal clustering of muscle-expressed genes in Caenorhabditis elegans, Nature 418 (2002) 975–979.

[5] P.T. Spellman, G.M. Rubin, Evidence for large domains of similarly expressed genes in the Drosophila genome, J. Biol. 1 (2002) 5.

[6] H. Caron, et al., The human transcriptome map: clustering of highly expressed genes in chromosomal domains, Science 291 (2001) 1289–1292.

[7] R. Versteeg, et al., The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes, Genome Res. 13 (2003) 1998–2004.

[8] J.D. Barrans, et al., Chromosomal distribution of the human cardiovascular transcriptome, Genomics 81 (2003) 519–524.

[9] S. Bortoluzzi, et al., A comprehensive, high-resolution genomic transcript map of human skeletal muscle, Genome Res. 8 (1998) 817–825.

[10] B.L. Gabrielsson, B. Carlsson, L.M. Carlsson, Partial genome scale analysis of gene expression in human adipose tissue using DNA array, Obes. Res. 8 (2000) 374–384.

[11] S. Minagawa, K. Nakabayashi, M. Fujii, S.W. Scherer, D. Ayusawa, Functional and chromosomal clustering of genes responsive to 5-bromodeoxyuridine in human cells, Exp. Gerontol. 39 (2004) 1069–1078.

[12] E. Soury, et al., Chromosomal assignments of mammalian genes with an acute inflammation-regulated expression in liver, Immunogenetics 53 (2001) 634–642.

[13] H. Zhang, K.H. Pan, S.N. Cohen, Senescence-specific gene expression fingerprints reveal cell-type-dependent physical clustering of up-regulated chromosomal loci, Proc. Natl. Acad. Sci. USA 100 (2003) 3251–3256.

[14] J.H. Vogel, A. von Heydebreck, A. Purmann, S. Sperling, Chromosomal clustering of a human transcriptome reveals regulatory background, BMC Bioinformatics 6 (2005) 230.

[15] D. Schubeler, et al., Nuclear localization and histone acetylation: a pathway for chromatin opening and transcriptional activation of the human beta-globin locus, Genes Dev. 14 (2000) 940–950.

[16] F. Spitz, F. Gonzalez, D. Duboule, A global control region defines a chromosomal regulatory landscape containing the HoxD cluster, Cell 113 (2003) 405–417.

[17] M.J. Lercher, A.O. Urrutia, L.D. Hurst, Clustering of housekeeping genes provides a unified model of gene order in the human genome, Nat. Genet. 31 (2002) 180–183.

[18] A.I. Su, et al., A gene atlas of the mouse and human protein-encoding transcriptomes, Proc. Natl. Acad. Sci. USA 101 (2004) 6062–6067.

[19] C. Pal, L.D. Hurst, Evidence for co-evolution of gene order and recombination rate, Nat. Genet. 33 (2003) 392–395.

[20] J.W. Pepper, The evolution of evolvability in genetic linkage patterns, Biosystems 69 (2003) 115–126.

[21] G.A. Singer, A.T. Lloyd, L.B. Huminiecki, K.H. Wolfe, Clusters of co-expressed genes in mammalian genomes are conserved by natural selection, Mol. Biol. Evol. 22 (2005) 767–775.

[22] P. Carninci, et al., The transcriptional landscape of the mammalian genome, Science 309 (2005) 1559–1563.

[23] L.D. Hurst, C. Pal, M.J. Lercher, The evolutionary dynamics of eukaryotic gene order, Nat. Rev. Genet. 5 (2004) 299–310.

[24] I. Marenholz, C.W. Heizmann, G. Fritz, S100 proteins in mouse and man: from evolution to function and pathology (including an update of the nomenclature), Biochem. Biophys. Res. Commun. 322 (2004) 1111–1122.

[25] S. Cai, H.J. Han, T. Kohwi-Shigematsu, Tissue-specific nuclear architecture and gene expression regulated by SATB1, Nat. Genet. 34 (2003) 42–51.

[26] R. van Driel, P.F. Fransz, P.J. Verschure, et al., The eukaryotic genome: a system regulated at different hierarchical levels, J. Cell Sci. 116 (2003) 4067–4075.

[27] K. Akashi, et al., Transcriptional accessibility for genes of multiple tissues and hematopoietic lineages is hierarchically controlled during early hematopoiesis, Blood 101 (2003) 383–389.

[28] F. Kepes, Periodic epi-organization of the yeast genome revealed by the distribution of promoter sites, J. Mol. Biol. 329 (2003) 859–865.

[29] R. Hershberg, E. Yeger-Lotem, H. Margalit, Chromosomal organization is shaped by the transcription regulatory network, Trends Genet. 21 (2005) 138–142.

[30] J.B. Johnnidis, et al., Chromosomal clustering of genes controlled by the aire transcription factor, Proc. Natl. Acad. Sci. USA 102 (2005) 7233–7238.

[31] J.M. Lee, E.L. Sonnhammer, Genomic gene clustering analysis of pathways in eukaryotes, Genome Res. 13 (2003) 875–882.

[32] S.A. Teichmann, R.A. Veitia, Genes encoding subunits of stable complexes are clustered on the yeast chromosomes: an interpretation from a dosage balance perspective, Genetics 167 (2004) 2121–2125.

[33] Y. Fukuoka, H. Inaoka, I.S. Kohane, Inter-species differences of co-expression of neighboring genes in eukaryotic genomes, BMC Genomics 5 (2004) 4.

[34] H.K. Lee, A.K. Hsu, J. Sajdak, J. Qin, P. Pavlidis, Coexpression analysis of human genes across many microarray data sets, Genome Res. 14 (2004) 1085–1094.

[35] H.R. Ueda, et al., Genome-wide transcriptional orchestration of circadian rhythms in Drosophila, J. Biol. Chem. 277 (2002) 14048–14052.

[36] L. Huminiecki, K.H. Wolfe, Divergence of spatial gene expression profiles following species-specific gene duplications in human and mouse, Genome Res. 14 (2004) 1870–1879.

[37] M. Lynch, Gene duplication and evolution, in: A. Moya, E. Font (Eds.), Evolution: from Molecules to Ecosystems, Oxford Univ. Press, Oxford, 2004, pp. 33–47.

[38] M. Lynch, J.S. Conery, The evolutionary fate and consequences of duplicate genes, Science 290 (2000) 1151–1155.

[39] S. Ohno, Evolution by Gene and Genome Duplication, Springer-Verlag, Berlin, 1970.

[40] M. Ashburner, et al., Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, Nat. Genet. 25 (2000) 25–29.

[41] S. Durinck, et al., BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis, Bioinformatics 21 (2005) 3439–3440.

[42] R.C. Gentleman, et al., Bioconductor: open software development for computational biology and bioinformatics, Genome Biol. 5 (2004) R80.

[43] E. Birney, et al., Ensembl 2006, Nucleic Acids Res. 34 (2006) D556–D561.

[44] R.C. Gentleman, D. Scholtens, B. Ding, V.J. Carey, W. Huber, Case studies using graphs on biological data, Bioinformatics and Computational Biology Solutions Using R and Bioconductor, Springer-Verlag, New York, 2005.