Article

Discoveries

Tissue-specific evolution of protein coding genes in human and mouse.

Nadezda Kryuchkova-Mostacci[1,2], Marc Robinson-Rechavi[1,2,*]

[1]Department of Ecology and Evolution, University of Lausanne, Switzerland

[2]Swiss Institute of Bioinformatics, Lausanne, Switzerland

*Author for Correspondence: Marc Robinson-Rechavi, Department of Ecology and Evolution, University of Lausanne, Switzerland, +41 21 692 4220, marc.robinson-rechavi@unil.ch

Abstract

Protein-coding genes evolve at different rates, and the influence of different parameters, from gene size to expression level, has been extensively studied. While in yeast gene expression level is the major causal factor of gene evolutionary rate, the situation is more complex in animals. Here we investigate these relations further, especially taking in account gene expression in different organs as well as indirect correlations between parameters. We used RNA-seq data from two large datasets, covering 22 mouse tissues and 27 human tissues. Over all tissues, evolutionary rate only correlates weakly with levels and breadth of expression. The strongest explanatory factors of purifying selection are GC content, expression in many developmental stages, and expression in brain tissues. While the main component of evolutionary rate is purifying selection, we also find tissue-specific patterns for sites under neutral evolution and for positive selection. We observe fast evolution of genes expressed in testis, but also in other tissues, notably liver, which are explained by weak purifying selection rather than by positive selection.

Key words: gene expression, protein evolution, human, mouse, evolutionary rate, natural selection.

Introduction

Understanding the causes of variation in protein sequence evolutionary rates is one of the major aims of molecular evolution, and has even been called a "quest for the universals of protein evolution" (Rocha 2006). Studies in a variety of organisms have reported that protein evolutionary rates correlated with many parameters, structural and functional (Pál et al. 2006; Rocha & Danchin 2004). Most notably, expression level has been shown to be the best predictor of evolutionary rate in yeasts and bacteria: highly expressed proteins are generally more conserved (Drummond et al. 2005; Pál et al. 2001; Wall et al. 2005). In animals and plants, our understanding has been complicated by the fact that genes can have different expression levels depending on tissue or life history stage, and by correlations with multiple other factors such as recombination rate, gene length or compactness, and gene duplications (Larracuente et al. 2008; Makino et al. 2009; Yang & Gaut 2011; Liao et al. 2006; Li et al. 2007). In mammals, expression breadth has been suggested to be more important than expression level (Duret & Mouchiroud 2000; Park & Choi 2010). It has also been suggested that selection against protein misfolding is sufficient to explain covariation of gene expression and evolutionary rate across taxa, including mouse and human (Drummond & Wilke 2008). This notably explains the slower evolution of brain-expressed genes; the relation with the influence of breadth of expression is unclear. Moreover, it was shown that conserved sites and optimal codons are significantly correlated in many organisms, including mouse and human (Drummond & Wilke 2008).

These correlations of evolutionary rate to many other parameters, which are themselves correlated (e.g., gene length and GC content), poses problems to determining what is true and what is spurious correlation. To disentangle which factors could be determining evolutionary rates a solution is to use partial correlation, taking into account the relationship between gene

structure and other parameters when considering the correlations with evolutionary rate (Larracuente et al. 2008; Warnefors & Kaessmann 2013).

Here we aim to disentangle aspects of protein evolutionary rate and its explanatory factors in human and mouse. We use partial correlations, taking into account not only different structural parameters, but also different aspects of gene expression (level, tissue specificity), using expression in more than 20 tissues. We also used three measures of protein evolutionary rate estimated from the branch-site model (Zhang et al. 2005): strength of negative selection (value of dN/dS on sites under negative selection); proportion of neutrally evolving sites; and evidence for positive selection. This allows us to distinguish fast evolution due to weak purifying selection from that due to positive selection.

Materials and Methods

We used RNA-seq data for mouse from the ENCODE project (The ENCODE Project Consortium 2011) and for human from Fagerberg et al. (Fagerberg et al. 2013). For mouse, the raw reads in FASTQ format obtained from the ENCODE FTP server (ftp://hgdownload.cse.ucsc.edu/goldenPath/mm9/encodeDCC/wgEncodeCshlLongRnaSeq/) were processed with TopHat and Cufflinks (Trapnell et al. 2012), using the gene models from Ensembl version 69 (Flicek et al. 2013). For human, processed data from Fagerberg et al. (Fagerberg et al. 2013) were retrieved from the ArrayExpress database (E-MTAB-1733) (Rustici et al. 2013). 22 tissues for mouse and 27 tissues for human were analyzed, of which 16 are homologous between the two species. Processed RNA-seq expression was further treated as follows (R script in Supplementary Material): multiplied by 10^6 (to avoid values under 1, which are negative after log transformation); $\log_2$ transformation; and quantile normalization, replacing zero values by $\log_2(1.0001)$.

We used either global parameters of expression: median expression, maximal expression, and specificity among all tissues; or expression in each tissue separately. Expression specificity τ was calculated as follows (Yanai et al. 2005), where *x* is the vector of expression levels over all *n* tissues for a gene:

$$\tau = \frac{\sum_{i=1}^{n}(1-\hat{x_i})}{n-1};\ \hat{x_i} = \frac{x_i}{\max\limits_{1\le i\le n}(x_i)}$$

Values of expression specificity close to zero indicate that a gene is broadly expressed, and close to one that it is specific of one tissue.

Additional analysis was performed on microarray expression data from 22 human and 22 mouse tissues selected from the Bgee database (Bastian et al. 2008), as well as 8 human and 6 mouse tissues from Brawand et al. RNA-seq (Brawand et al. 2011). The corresponding results are presented in Supplementary Materials.

As measures of evolutionary rate of protein-coding genes, we used the estimates from the branch-site model (Zhang et al. 2005) as precomputed in Selectome (Moretti et al. 2014): purifying selection estimated by the dN/dS ratio ω_0 on the proportion of codons under purifying selection (noted "Omega" in the figures), evidence for positive selection estimated by the log-likelihood ratio ΔlnL of H_1 to H_0 (models with or without positive selection), and the proportion of neutrally evolving sites p_1. The evolutionary rate parameters were estimated from the Euteleostomi gene trees on the Murinae branch for mouse and the Homininae branch for human. We also present in Supplementary Materials another estimation of evolutionary rate, using the exon based MI score (Rodriguez et al. 2013; Ezkurdia et al. 2014).

For all parameters the longest coding transcript was chosen as a representative of the gene, as the evolutionary rate data were available only for this transcript. Analysis was also redone for mouse using the most expressed transcript; results are presented in Supplementary Materials.

Intron number, intron length, CDS length (coding sequence length) and GC content were taken from Ensembl 69 (Flicek et al. 2013). Essentiality data were manually mapped and

curated (Walid Gharib, personal communication) for human from the OMIM database (McKusick-Nathans Institute of Genetic Medicine 2014) and for mouse from the MGI database (Blake et al. 2014).

Data of expression at different developmental stages were obtained from Bgee (Bastian et al. 2008). The parameter stage number indicates the number of stages in which the gene was reported as expressed. Mouse development was divided in 10 stages: 1. Zygote 2. Cleavage 3. Blastula 4. Gastrula 5. Neurula 6. Organogenesis 7. Fetus 8. Infant 9. Adolescent 10. Adult; and human in 8 stages: 1. Cleavage 2. Blastula 3. Organogenesis 4. Fetus 5. Infant 6. Adolescent 7. Early adult 8. Late Adult.

Phyletic age and connectivity (protein-protein interactions) data were downloaded from the OGEE database (Chen et al. 2012), as ordinal data. Phyletic age stages used are: 1. Mammalia 2. Chordata 3. Metazoa 4. Fungi/Metazoa 5. Eukaryota 6. Cellular organism.

Recombination rate was calculated from Cox et al. 2009 data (Cox et al. 2009).

Correlation between the different parameters was performed in two ways: simple pairwise Spearman correlation and partial Spearman correlation (results for Pearson correlation are also presented in Supplementary Materials). For partial correlation each pair of parameters were compared taking into account all other parameters: first a linear model according to all other parameters for each of the two analyzed parameters was calculated; then the Spearman correlation was calculated on the corresponding residuals. All R code is available as Supplementary Materials.

Partial correlation was used to determine the correlation between two parameters excluding dependencies from other parameters. The principle of the partial correlation can be shown on a toy example (Supplementary table S1). As example data for human height and leg length were simulated, so that either a) the length of both legs is calculated depending on height, or b) the length of the left leg is calculated from height, and the length of the right leg is

calculated from the length of left leg. With simple correlation the two cases cannot be distinguished, as all three parameters correlate strongly with each other. With partial correlation we can distinguish the two cases: in case a) left leg and right leg length don't correlate with each other if we exclude influence of the height, but in case b) we see a strong correlation between them, as expected, while right leg length no longer correlates with height.

Expression, intron length, intron number, CDS length, τ, ω_0, paralog number were $\log_2$ transformed before calculations. p_1 and ΔlnL were normalized by taking the fourth root (Canal 2005; Roux et al. 2014). For parameters containing zeros a small value was added before log transformation, chosen as the minimal non zero value of the parameter (except for RNA-seq, see detailed treatment above). Altogether 9553 protein-coding genes for human and 9485 protein-coding genes for mouse were analyzed.

All the analysis was performed in R (R Core Team 2012) using Lattice (Sarcar 2008), plyr (Wickham 2011). For the representation of the data Cytoscape version 2.8.2 (Shannon et al. 2003) with library RCytoscape (Shannon et al. 2013) and Circos version 0.62-1 (Krzywinski et al. 2009) were used.

Results

We detail here the results of Spearman partial correlation analyses (table 1); standard Spearman and Pearson, as well as partial Pearson, correlations are provided in Supplementary Materials. Spearman correlation was preferred as most of the data analyzed are not normally distributed (supplementary fig. S1), even after transformation, and to avoid a large influence of outliers. It should be noted that parameters that are expected to have strong direct relations remain strongly correlated in the Spearman partial correlation. For example the correlation between coding sequence (CDS) length and intron number, in mouse, is ρ=0.683 for partial vs. ρ=0.760 for simple correlation, showing that longer genes have more introns. Similarly,

partial correlations still show that higher expressed genes are broadly expressed, and that specific genes have lower expression in general. Thus little relevant information is lost, while spurious correlations can be hopefully avoided.

Evolutionary rate: global influences on selection

Evolutionary rate is represented by three parameters in this study, taken from the branch-site model (see Methods): ω_0 = dN/dS, measures the intensity of purifying selection on the subset of sites determined to be under purifying selection; p_1 is the proportion of neutrally evolving sites; and ΔlnL measures the strength of evidence for positive selection.

In both mouse and human, none of the aspects of gene expression yields a strong partial correlation to any feature of evolutionary rate (table 1; fig. 1). There is a weak correlation of ω_0 to expression specificity τ in both human and mouse (ρ = 0.085 and 0.067 respectively), confirming that more broadly expressed genes evolve under stronger purifying selection. Purifying selection ω_0 is also negatively correlated to maximum expression, although this is weaker in human, indicating that genes with high expression in at least one tissue have a tendency to evolve under strong purifying selection. More surprisingly, purifying selection ω_0 is positively correlated to median expression. Note that these are partial correlations; without correcting for other parameters, as expected, ω_0 correlates negatively with median expression, i.e., highly expressed genes are under strong purifying selection. It appears that this negative correlation is driven by the effect of breadth of expression and of maximum expression, with the residual effect actually in the opposite direction.

Evolutionary features of the genes, paralog number and phyletic age, have a stronger partial correlation with ω_0 than expression: older genes, and genes with more paralogs, evolve under stronger purifying selection; again, this is after removing the effect of high levels of

expression, as well as the correlation between gene age and number of paralogs. In human, GC content also appears to have a strong influence on ω_0, but much less so in mouse.

It remains that none of these parameters can explain much of the differences in purifying selection. The total variance of ω_0 that they explain (using partial Pearson correlation, as Spearman ρ does not relate directly to variance) is 10.2% for human and 13.8% for mouse, thus leaving more than 85% of the variance unexplained.

The strongest correlation with ω_0 is for p_1, the proportion of sides evolving neutrally (fig. 1). This partial correlation is $\rho = 0.748$ in mouse and $\rho = 0.598$ in human; genes under strong purifying selection have a smaller proportion of sides evolving neutrally. This is not directly due to the way how these parameters are estimated in the branch-site test, since ω_0 is computed on a distinct set of codons from p_1. This proportion p_1 of neutrally evolving sites is otherwise mostly correlated with evolutionary features (phyletic age, paralog number) in human, and with structural features (intron length, GC content) in mouse, but correlations are weak (all $\rho < 0.09$).

Evidence for positive selection correlates negatively with median expression in both human and mouse (fig. 1), i.e. highly expressed genes are under weaker positive selection ($\rho = -0.105$ and -0.187 respectively). It should be noted that this correlation concerns relatively weak evidence for selection, since only 4 human and 23 mouse genes in the dataset have significant support for branch-site positive selection (using the false discovery rate of 10% cut-off of Selectome, see Methods).

Tissue-specific analysis

When the correlation between expression level, selective pressure, and other parameters, is analyzed for each tissue separately, there are large differences, notably in the correlation between expression and purifying selection between tissues (fig. 2).

In both human and mouse, the strongest correlation with purifying selection ω_0 is for level of expression in the brain, as expected from previous studies with less tissues (Kuma et al. 1995; Duret & Mouchiroud 2000; Khaitovich et al. 2006; Drummond & Wilke 2008; Tuller et al. 2008). After correcting for all other effects, the residual correlation is rather weak (ρ between -0.065 and -0.107 depending on species and brain part), but always in the direction of stronger purifying selection on genes with higher expression in brain. In human, there are also significant partial correlations for esophagus, prostate, adrenal, colon, and endometrium (fig. 2B). In mouse, there are correlations for all sampled tissues except liver, placenta and testis (fig. 2A); in human the homologous tissues to these three also have among the lowest partial correlations. Interestingly, the only positive partial correlation with ω_0 is for human testis expression, i.e. higher expression in testis correlates with weaker purifying selection.

The strongest correlations with the proportion of neutrally evolving sites are also for brain tissues, in human and in mouse. Again the correlation is negative, indicating less neutral evolution (i.e., more selection) for more highly expressed genes. There are almost no other tissues with significant partial correlation of expression and p_1, although for mouse large intestine the correlation is significantly positive.

Concerning evidence of positive selection, on the other hand, there are significant negative partial correlations for all tissues, meaning that for each tissue genes with higher expression have less evidence of positive selection. Brain tissues again have some of the strongest correlations, although they stand out less than for ω_0 or p_1. In both mouse and human the correlation is weakest for testis expression, and also quite weak for placenta.

All these correlations include for each tissue both house-keeping and tissue-specific genes; the former might confuse tissue-specific patterns. Thus we repeated the analysis restricted to tissue-specific genes, defined as $\tau > 0.2$ (supplementary fig. S2). The global picture is similar, with notably significant negative partial correlations to ω_0 only for expression in human brain

and mouse cerebellum, significant negative correlation to p_1 only for mouse brain parts, and conversely significant positive correlations to expression in human and mouse testis.

Gene age and duplication

As expected, older genes have more paralogs (positive correlation in fig. 1) (Roux & Robinson-Rechavi 2011). Tissue specificity has a rather strong positive partial correlation with paralog number, and a significant weak negative correlation with phyletic age was detected; both correlations are stronger in human than in mouse. That means that, correcting for the correlation between gene age and paralog number, new genes and genes with more paralogs tend to have more tissue-specific expression. While in simple correlation, phyletic age and expression level (median or maximum) have a strong positive correlation (older genes have higher expression), this effect is almost completely lost in the partial correlation, and so is probably spurious.

The phyletic age of the genes correlates negatively with purifying selection but almost no correlation can be seen to neutral evolution or positive selection. This is consistent with previous observations that older genes evolve under stronger purifying selection ((Albà & Castresana 2005, 2007) but see (Elhaik et al. 2006)).

Paralog number correlates negatively with purifying selection in both organisms (-0.064 for mouse and -0.136 for human). This indicates a stronger effect of the biased preservation of duplicates under stronger purifying selection (Brunet et al. 2006; Davis & Petrov 2004; Jordan et al. 2004), than of the effect of faster evolution of duplicated genes (Satake et al. 2012).

Gene structure

Genes with higher GC content have higher expression level, as shown previously (Urrutia & Hurst 2003), although the effect is not very strong in partial correlation. Previous findings that highly expressed genes are shorter were only partly confirmed: there is a strong negative partial correlation between CDS length and maximal expression, but the partial correlation between median expression and CDS length is weakly positive. Curiously, the partial correlation with intron number is opposite, indicating that genes with high maximum expression tend to have more introns than expected given their CDS length.

Differences between human and mouse, and between datasets

In general correlations in human are slightly weaker than in mouse, but very consistent (supplementary fig. S3). The strongest difference is between the correlations of GC content and stage number; and of GC content and maximal expression.

There are also noticeable differences between mouse and human in the partial correlations among ω_0, p_1 and ΔlnL (evidence for positive selection). In human ΔlnL correlates negatively with p_1 and positively with ω_0, indicating that genes with high proportion of neutrally evolving sites and weak purifying selection show little evidence for positive selection. In mouse the correlations are not significant, and in the opposite directions, but the correlation between ω_0 and p_1 is much stronger. GC content and paralog number also have stronger correlations to purifying selection in human than in mouse.

We repeated our analyses with large microarray experiments (see Methods, and Supplementary Materials), to control for putative biases in RNA-seq data. There are a few differences, although they do not change our biological conclusions. First, with microarrays tissue-specificity τ appears overall lower, and the correlations between expression parameters (τ, maximal expression, median expression) are stronger. This might be due to the better

detection of lowly expressed genes by RNA-seq than by microarray, whereas there seems to be less difference for highly expressed genes (Wang et al. 2014). Conversely, correlations of expression parameters with all other parameters are much stronger for RNA-seq. The correlation between ω_0 and expression in each tissue separately is stronger with microarrays then with RNA-seq, and significant for all tissues, but the same tissues have the strongest (resp. weakest) correlation between ω_0 and expression with both techniques. Inversely, the evidence for positive selection has almost no significant correlations with expression in single tissues with microarray data.

We also reproduced our analysis using the precomputed "MI" score for most conserved exon (Rodriguez et al. 2013; Ezkurdia et al. 2014) instead of the branch-site model ω_0, and all results are similar despite the differences in multiple sequence alignment and in evolutionary model (supplementary fig. S4): e.g., phyletic age is the strongest correlation to MI and median expression has a weak positive partial correlation.

Finally, we repeated our analysis with the RNA-seq data for human and mouse 6 tissues from Brawand et al. (Brawand et al. 2011); results are extremely similar to those with the large RNA-seq experiments used in our main results (Supplementary Material), with less detail of tissues, and less resolution for τ, due to the smaller sampling.

Overall, our results appear quite robust across species and experimental techniques.

Discussion

Technical limitations and generality of observations

We use partial correlations to hopefully detect non-spurious relations. Of note, the lack of partial correlation between two parameters does not mean that they are not correlated in practice, but that the correlation is not directly informative, or insufficiently to be detected.

Our analysis was performed on approximately half of the known protein coding genes (9509 for human and 9471 for mouse), for which evolutionary rate could be computed reliably. While this may introduce some bias, it does not appear to have a large influence, since correlations other than to evolutionary rate are very similar on the other half of the coding genes (Supplementary Material).

Global study of evolutionary rate

Our aim is to understand the causes of variation in evolutionary rates among protein-coding genes in mammals. In yeast or bacteria, the major explanatory feature is the relation between the level of gene expression and purifying selection (Pál et al. 2006; Rocha & Danchin 2004; Rocha 2006). In mammals, firstly levels of expression are more complex to define, due to multicellularity and tissue-specificity, and secondly several other features have been reported to correlate as much or more with evolutionary rate, in studies which did not necessarily incorporate all alternative explanations.

In this study, we have focused on the dN/dS ratio, or ω, and distinguished further the three forces which affect this ratio, under the classical assumption that dS is overwhelmingly neutral (although see (Rubinstein et al. 2011; Macossay-Castillo et al. 2014; Dimitrieva & Anisimova 2014)). The intensity of purifying selection is clearly the main component of the overall ω: on average more than 85% of codons are in the purifying selection class of the evolutionary model used. Analyzing separately neutral evolution and purifying and positive selection, we find that (i) these three forces do not affect protein coding genes independently, and (ii) they have different relations to gene expression and to other features. Notably, genes which are under stronger purifying selection have less codons predicted under neutral evolution. Importantly, we computed evolutionary rates on filtered alignments (Moretti et al. 2014), which probably eliminates mostly neutrally evolving sites, thus underestimating p_1.

Still, it appears that to the best of our knowledge these two forces act in the same direction. The relation is less clear concerning the evidence for positive selection, with opposite correlations in human and mouse. But we are limited by the weak evidence for positive selection on the branches tested, at the human-chimpanzee and mouse-rat divergences. Overall, these relations between forces acting on ω_0 deserve further investigation with more elaborate evolutionary models (e.g., Murrell et al. 2012; Zaheri et al. 2014). Despite the limitations of the estimation of positive selection, this is the component of evolutionary rate which has the strongest partial correlation with the level of gene expression, both with the median expression over all tissues, and with expression in brain tissues. This implies that when expression patterns constrain the protein sequence, they also strongly limit adaptation (strong purifying selection and very low positive selection).

So what explains evolutionary rate? The strongest partial correlation of ω_0 is with phyletic age: older genes evolve under stronger purifying selection. While the use of partial correlation allows us to correct for some obvious biases in detecting distant orthologs, such as gene length, we cannot exclude that results be partially caused by the easier detection of orthologs in distant species for proteins with more conserved sequences (Elhaik et al. 2006; Albà & Castresana 2007; Moyers & Zhang 2014). I.e., genes with weak purifying selection may be reported as younger than they are, because the orthologs were not detected by sequence similarity. We obtain similar results with an exon-based index of sequence conservation, MI (supplementary fig. S4). Whatever the contributions of methodological bias and biological effect, this correlation is not very informative about causality, since stronger selection will not be caused by the age of the gene.

The next strongest partial correlation with ω_0 is the GC content of the gene. In mammals, the variation in GC content of genes seems mostly due to GC-biased gene recombination (Montoya-Burgos et al. 2003), and this in turn has been show to impact estimation of dN/dS

(Galtier et al. 2009). But while GC-biased gene recombination is expected to lead to high GC and an overestimation of ω, we find a negative correlation between ω_0 and GC content, consistent with previous observations in Primates (Bullaughey et al. 2008). Of note, estimating the actual biased recombination rate rather than GC content is limited by the rapid turn-over of recombination hotspots (Glémin et al. 2014), and recombination rate appears to have only a very weak effect on dN/dS in Primates once GC content is taken into account (Bullaughey et al. 2008). This could be seen in our study, as recombination rate did not show any significant correlation to any of the parameters, except GC content (Supplementary Material). The previously reported relation between dN/dS and intron length seems to be mostly an indirect effect of the strong correlation between GC content and intron length (Montoya-Burgos et al. 2003; Duret et al. 1995).

The significant, although weaker, partial correlation of ω_0 to paralog number is consistent with previous observations that genes under stronger purifying selection are more kept in duplicate (Davis & Petrov 2004; Yang & Gaut 2011; Jordan et al. 2004; Brunet et al. 2006).

The level of gene expression has been reported repeatedly to be the main explanatory variable for dN/dS (Subramanian & Kumar 2004; Liao & Zhang 2006; Pál et al. 2001; Drummond et al. 2005; Wall et al. 2005), notably in *S. cerevisiae*. Our first observation is that no aspect of expression in human and mouse adult tissues is as strong an explanatory factor for any component of evolutionary rate as what was reported in yeast. Our second observation is that three aspects of expression influence evolutionary rate most strongly: breadth of expression τ; number of developmental stages (fig. 1; table 1); and expression in brain tissues (fig. 2). The third, surprising, observation is that median expression is positively correlated with ω_0: taking into account other parameters, genes which have higher expression on average are under weaker purifying selection; whereas the correlation with maximal expression is negative, as expected. Thus in mammals the negative correlation between median expression and

evolutionary rate appears to be an indirect effect of stronger selection on broadly expressed genes and on genes with high maximal expression in at least one tissue (this is also true if we take the mean instead of median expression, see Supplementary Materials).

We confirm previously reported observations that expression breadth is more important then expression level itself in mammals (Park & Choi 2010). dN was previously found to be threefold lower in ubiquitous than in tissue-specific genes, while dS did not vary with expression specificity (Duret & Mouchiroud 2000). Other studies indicate that genes expressed in few tissues evolve faster then genes expressed in a wide range of tissues (Liao & Zhang 2006; Gu & Su 2007; Park & Choi 2010), or that tissue-specific genes have more evidence for positive selection (Haygood et al. 2010). In mouse, but not human, τ is weakly negatively correlated to evidence for positive selection: broadly expressed genes seem to be more affected by positive selection, *contra* Haygood et al (Haygood et al. 2010). We also notice that tissue-specificity and maximal expression are correlated, i.e. more tissue-specific genes have higher maximum expression in one tissue. Thus these two forces appear to act on different genes: some genes are under strong purifying selection because they are broadly expressed, suggesting an important role of pleiotropy, while other genes are under strong purifying selection because they are highly expressed in few tissues, suggesting an important role of the tissue-specific optimization of protein sequences. Of note, analyzing separately only brain expression relative to maximal and breadth of expression in other tissues gave similar results, thus brain expression alone is not driving these patterns (not shown).

Some studies have reported that expression level and tissue specificity are less important than gene compactness and essentiality in mammals (Liao et al. 2006). Liao et al. (Liao et al. 2006) reported that compact genes evolve faster, but this correlation is very weak in our study. We could not either confirm that highly expressed genes are shorter (Li et al. 2007; Chen et al. 2005; Urrutia & Hurst 2003). We have used the longest transcript for each protein coding

gene, as evolutionary parameters (ΔlnL, ω_0, p_1) were calculated for the transcript. But this might not be the transcript most expressed and used in all tissues (Gonzalez-Porta et al. 2013). We repeated calculations with the most expressed transcript (Supplementary Material), but results were unchanged; we show these results only in supplementary materials, as the estimation of transcript-level expression does not yet appear to be very reliable (Lahens et al. 2014; Cho et al. 2014). Finally, we tried to investigate the impact of essentiality, but we found no significant effect (supplementary fig. S5); we note that we have very low power to test this effect, especially in human.

The analysis was also performed adding connectivity and recombination data, for mouse only. This reduced the number of analyzed genes to 4599 (Supplementary Material). Correlations were mostly unchanged, with the largest difference being for the correlation between stage number and phyletic age, from 0.11 to 0.093. No notable change of correlations with ω_0 was detected, and connectivity and recombination rate do not show any significant correlation to evolutionary rate.

The largest partial correlations that we observed for components of evolutionary rate are between brain expression level and evidence for positive selection, at -0.203 to -0.188 in mouse (for different brain parts), and -0.168 in human (whole brain). For purifying selection we find weaker but significant partial correlations with brain expression and with the number of stages, between 0.065 and 0.119. And brain tissues also have the strongest partial correlation over expression in tissues for neutral evolution (fig. 2). It has been previously reported that brain expression is a major component of evolutionary rate in mammals and other animals (Khaitovich et al. 2006; Duret & Mouchiroud 2000; Kuma et al. 1995; Drummond & Wilke 2008), and here we confirm the dominance of this component, even taking other effects into account. Importantly we show that this affects all forces acting on protein evolutionary rate: purifying selection, neutral evolution, and positive selection. Thus

the median expression of genes over more than 20 tissues is a poor explanation of protein evolutionary rate, relative to brain expression.

Tissue specific patterns

There are striking differences between tissues in the extent of the correlations with structural and evolutionary parameters. As already mentioned, brain tissues present the strongest partial correlations with evolutionary rate; results are consistent when only tissue-specific genes are used. We observe this for the three evolutionary forces estimated. In most comparisons, the correlation is stronger for brain expression than for any global measure of expression. This is consistent with the translational robustness hypothesis, which proposes that highly expressed genes are under stronger pressure to avoid misfolding caused by translational errors, thus these genes are more conserved in evolution (Drummond et al. 2005), and that neural tissues are the most sensitive to protein misfolding (Drummond & Wilke 2008). This slow evolution of genes expressed in neural tissues has been repeatedly reported (Duret & Mouchiroud 2000; Kuma et al. 1995; Necsulea & Kaessmann 2014), especially for the brain (Park & Choi 2010); it has also been related to higher complexity of biochemical networks in the brain than in other tissues (Kuma et al. 1995).

Fast evolution of genes expressed in testis is also well documented (Khaitovich et al. 2006; Brawand et al. 2011; Necsulea & Kaessmann 2014), and could be due to lower purifying selection, an excess of young genes and leaky expression, or to positive selection due to sexual conflict. We observe neither a stronger correlation between expression in sexual tissues and evidence for positive selection, nor a stronger correlation between expression in sexual tissues and the proportion of sites evolving neutrally. What we do observe is that the weakest partial correlation between expression in a tissue and purifying selection is for testis, and that it is also quite weak for placenta, with even a surprising positive correlation between

ω_0 and expression in human testis, which remains when only tissue-specific genes are used. This is consistent with the "leaky expression" model: being expressed in the testis does not appear to be an indicator of function carried by the protein sequence. Interestingly, expression in testis is negatively correlated with the number of paralogs, significantly so in mouse: genes which are more expressed in testis have less paralogs, after correcting for other effects.

While the strong correlation of ω_0 with expression in the brain, and the weak correlation with expression in testis are expected, we also observe less expected patterns. Most notably, liver expression has the next weakest correlation with ω_0 after testis (and placenta in mouse). Although it was reported before that liver expressed genes are evolving faster (Khaitovich et al. 2006; Duret & Mouchiroud 2000), it was reported with much fewer tissues, and not highlighted. Liver expression is also positively correlated with the proportion of neutral sites, unlike brain or testis expression, although this is not significant. Interestingly, liver has the strongest correlation of expression with phyletic age, implying that despite low purifying selection, old genes are more expressed in liver. In any case, this outlier position of liver has important practical implications, since liver is often used as a "typical" tissue in studies of gene expression for molecular evolution (e.g., (Gilad et al. 2006; Blekhman et al. 2010; Enard et al. 2002)).

Conclusion

The main result of our study is that average adult gene expression is quite lowly informative about protein evolutionary rate, while purifying selection on genes highly expressed in the brain and breadth of expression are our best bets for a causal factor explaining evolutionary rates. A practical consequence is that great care should be taken before using expression from other tissues, including widely used ones such as liver, as proxies for the functional importance of mammalian genes.

Finally, all calculations were performed with expression in adult tissues. It is possible that expression in embryonic development be more important for evolutionary constraints in mammals, and this should be explored further.

Supplementary Material

The most important Supplementary Materials are available at Genome Biology and Evolution online (http://gbe.oxfordjournals.org/). Data sets and other supplementary figures are available at: http://dx.doi.org/10.6084/m9.figshare.1221771.

Acknowledgments

We thank Julien Roux for helpful comments on the manuscript. This work was supported by the Swiss National Science Foundation (grants number 31003A 133011/1 and 31003A_153341/1) and Etat de Vaud. The computations were performed at the Vital-IT Center (http://www.vital-it.ch) for high-performance computing of the SIB Swiss Institute of Bioinformatics.

References

Albà MM, Castresana J. 2005. Inverse relationship between evolutionary rate and age of mammalian genes. Mol. Biol. Evol. 22:598–606. doi: 10.1093/molbev/msi045.

Albà MM, Castresana J. 2007. On homology searches by protein Blast and the characterization of the age of genes. BMC Evol. Biol. 7:53. doi: 10.1186/1471-2148-7-53.

Bastian F, Parmentier G, Roux J. 2008. Bgee: integrating and comparing heterogeneous transcriptome data among species. Data Integr. …. 124–131. http://link.springer.com/chapter/10.1007/978-3-540-69828-9_12 (Accessed March 28, 2013).

Blake J, Bult C, Eppig J, Kadin J, Richardson J, et al. 2014. The Mouse Genome Database. Nucleic Acids Res. 42(D1):D810–D817.

Blekhman R, Marioni JC, Zumbo P, Stephens M, Gilad Y. 2010. Sex-specific and lineage-specific alternative splicing in primates. Genome Res. 20:180–9. doi: 10.1101/gr.099226.109.

Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, et al. 2011. The evolution of gene expression levels in mammalian organs. Nature. 478:343–8. doi: 10.1038/nature10532.

Brunet FG, Roest Crollius H, Paris M, Aury J-M, Gibert P, et al. 2006. Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. Mol. Biol. Evol. 23:1808–16. doi: 10.1093/molbev/msl049.

Bullaughey K, Przeworski M, Coop G. 2008. No effect of recombination on the efficacy of natural selection in primates. Genome Res. 18:544–54. doi: 10.1101/gr.071548.107.

Canal L. 2005. A normal approximation for the chi-square distribution. Comput. Stat. Data Anal. 48:803–808. doi: 10.1016/j.csda.2004.04.001.

Chen J, Sun M, Rowley JD, Hurst LD. 2005. The small introns of antisense genes are better explained by selection for rapid transcription than by “genomic design”. Genetics. 171:2151–5. doi: 10.1534/genetics.105.048066.

Chen W-H, Minguez P, Lercher MJ, Bork P. 2012. OGEE: an online gene essentiality database. Nucleic Acids Res. 40:D901–6. doi: 10.1093/nar/gkr986.

Cho H, Davis J, Li X, Smith KS, Battle A, et al. 2014. High-Resolution Transcriptome Analysis with Long-Read RNA Sequencing. PLoS One. 9:e108095. doi: 10.1371/journal.pone.0108095.

Cox A, Ackert-Bicknell CL, Dumont BL, Ding Y, Bell JT, et al. 2009. A new standard genetic map for the laboratory mouse. Genetics. 182:1335–44. doi: 10.1534/genetics.109.105486.

Davis JC, Petrov D a. 2004. Preferential duplication of conserved proteins in eukaryotic genomes. PLoS Biol. 2:E55. doi: 10.1371/journal.pbio.0020055.

Dimitrieva S, Anisimova M. 2014. Unraveling Patterns of Site-to-Site Synonymous Rates Variation and Associated Gene Properties of Protein Domains and Families. PLoS One. 9:e95034. doi: 10.1371/journal.pone.0095034.

Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. 2005. Why highly expressed proteins evolve slowly. Proc. Natl. Acad. Sci. U. S. A. 102:14338–43. doi: 10.1073/pnas.0504070102.

Drummond DA, Wilke CO. 2008. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. Cell. 134:341–52. doi: 10.1016/j.cell.2008.05.042.

Duret L, Mouchiroud D. 2000. Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. Mol. Biol. Evol. 17:68–74. http://www.ncbi.nlm.nih.gov/pubmed/10666707.

Duret L, Mouchiroud D, Gautier C. 1995. Statistical analysis of vertebrate sequences reveals that long genes are scarce in GC-rich isochores. J. Mol. Evol. 40:308–17. http://www.ncbi.nlm.nih.gov/pubmed/7723057 (Accessed October 8, 2014).

Elhaik E, Sabath N, Graur D. 2006. The “inverse relationship between evolutionary rate and age of mammalian genes” is an artifact of increased genetic distance with rate of evolution and time of divergence. Mol. Biol. Evol. 23:1–3. doi: 10.1093/molbev/msj006.

Enard W, Khaitovich P, Klose J, Zöllner S. 2002. Intra-and interspecific variation in primate gene expression patterns. Science (80-.). 296:340–343. http://www.sciencemag.org/content/296/5566/340.short (Accessed October 13, 2014).

Ezkurdia I, Juan D, Rodriguez JM, Frankish A, Diekhans M, et al. 2014. Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. Hum. Mol. …. 1–45. http://hmg.oxfordjournals.org/content/early/2014/06/16/hmg.ddu309.short (Accessed June 30, 2014).

Fagerberg L, Hallstrom BM, Oksvold P, Kampf C, Djureinovic D, et al. 2013. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Mol. Cell. Proteomics. doi: 10.1074/mcp.M113.035600.

Flicek P, Ahmed I, Amode MR, Barrell D, Beal K, et al. 2013. Ensembl 2013. Nucleic Acids Res. 41:D48–55. doi: 10.1093/nar/gks1236.

Galtier N, Duret L, Glémin S, Ranwez V. 2009. GC-biased gene conversion promotes the fixation of deleterious amino acid changes in primates. Trends Genet. 25:1–5. http://www.sciencedirect.com/science/article/pii/S0168952508003004 (Accessed October 7, 2014).

Gilad Y, Oshlack A, Smyth GK, Speed TP, White KP. 2006. Expression profiling in primates reveals a rapid evolution of human transcription factors. Nature. 440:242–5. doi: 10.1038/nature04559.

Glémin S, Arndt PF, Messer PW, Petrov D, Galtier N. 2014. Quantification of GC-biased gene conversion in the human genome Quantification of GC-biased gene conversion in the human genome.

Gonzalez-Porta M, Frankish A, Rung J, Harrow J, Brazma A. 2013. Transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene. Genome Biol. 14:R70. doi: 10.1186/gb-2013-14-7-r70.

Gu X, Su Z. 2007. Tissue-driven hypothesis of genomic evolution and sequence-expression correlations. Proc. Natl. Acad. Sci. U. S. A. 104:2779–84. doi: 10.1073/pnas.0610797104.

Haygood R, Babbitt CC, Fedrigo O, Wray G a. 2010. Contrasts between adaptive coding and noncoding changes during human evolution. Proc. Natl. Acad. Sci. U. S. A. 107:7853–7. doi: 10.1073/pnas.0911249107.

Jordan IK, Wolf YI, Koonin E V. 2004. Duplicated genes evolve slower than singletons despite the initial rate increase. BMC Evol. Biol. 4:22. doi: 10.1186/1471-2148-4-22.

Khaitovich P, Enard W, Lachmann M, Pääbo S. 2006. Evolution of primate gene expression. Nat. Rev. Genet. 7:693–702. doi: 10.1038/nrg1940.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, et al. 2009. Circos: an information aesthetic for comparative genomics. Genome Res. 19:1639–45. doi: 10.1101/gr.092759.109.

Kuma K, Iwabe N, Miyata T. 1995. Functional Constraints against Variations on Molecules from the Tissue Level: Slowly Evolving Brain-Specific Genes Demonstrated by Protein Kinase and Immunoglobulin SupergeneFamilies. Mol. Biol. Evol. 12:123–130. http://mbe.oxfordjournals.org/content/12/1/123.short (Accessed February 4, 2014).

Lahens NF, Kavakli IH, Zhang R, Hayer K, Black MB, et al. 2014. IVT-seq reveals extreme bias in RNA-sequencing. Genome Biol. 15:R86. doi: 10.1186/gb-2014-15-6-r86.

Larracuente AM, Sackton TB, Greenberg AJ, Wong A, Singh ND, et al. 2008. Evolution of protein-coding genes in Drosophila. Trends Genet. 24:114–23. doi: 10.1016/j.tig.2007.12.001.

Li S-W, Feng L, Niu D-K. 2007. Selection for the miniaturization of highly expressed genes. Biochem. Biophys. Res. Commun. 360:586–92. doi: 10.1016/j.bbrc.2007.06.085.

Liao B-Y, Scott NM, Zhang J. 2006. Impacts of gene essentiality, expression pattern, and gene compactness on the evolutionary rate of mammalian proteins. Mol. Biol. Evol. 23:2072–80. doi: 10.1093/molbev/msl076.

Liao B-Y, Zhang J. 2006. Low rates of expression profile divergence in highly expressed genes and tissue-specific genes during mammalian evolution. Mol. Biol. Evol. 23:1119–28. doi: 10.1093/molbev/msj119.

Macossay-Castillo M, Kosol S, Tompa P, Pancsa R. 2014. Synonymous constraint elements show a tendency to encode intrinsically disordered protein segments. PLoS Comput. Biol. 10:e1003607. doi: 10.1371/journal.pcbi.1003607.

Makino T, Hokamp K, McLysaght A. 2009. The complex relationship of gene duplication and essentiality. Trends Genet. 25:147–52. doi: 10.1016/j.tig.2009.02.001.

McKusick-Nathans Institute of Genetic Medicine. 2014. Online Mendelian Inheritance in Man, OMIM®. John Hopkins Univ. (Baltimore MD). http://omim.org/.

Montoya-Burgos JI, Boursot P, Galtier N. 2003. Recombination explains isochores in mammalian genomes. Trends Genet. 19:128–30. doi: DOI: 10.1016/S0168-9525(03)00021-0.

Moretti S, Laurenczy B, Gharib WH, Castella B, Kuzniar A, et al. 2014. Selectome update: quality control and computational improvements to a database of positive selection. Nucleic Acids Res. 42:D917–21. doi: 10.1093/nar/gkt1065.

Moyers BA, Zhang J. 2014. Phylostratigraphic bias creates spurious patterns of genome evolution. Mol. Biol. Evol. 734–763.

Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, et al. 2012. Detecting individual sites subject to episodic diversifying selection. PLoS Genet. 8:e1002764. doi: 10.1371/journal.pgen.1002764.

Necsulea A, Kaessmann H. 2014. Evolutionary dynamics of coding and non-coding transcriptomes. Nat. Rev. Genet. doi: 10.1038/nrg3802.

Pál C, Papp B, Hurst LD. 2001. Highly Expressed Genes in Yeast Evolve Slowly. Genetics. 158:927–931. doi: 10.2460/javma.242.4.458.

Pál C, Papp B, Lercher MJ. 2006. An integrated view of protein evolution. Nat. Rev. Genet. 7:337–48. doi: 10.1038/nrg1838.

Park SG, Choi SS. 2010. Expression breadth and expression abundance behave differently in correlations with evolutionary rates. BMC Evol. Biol. 10:241. doi: 10.1186/1471-2148-10-241.

R Core Team. 2012. R: A Language and Environment for Statistical Computing. http://www.r-project.org/.

Rocha EPC. 2006. The quest for the universals of protein evolution. Trends Genet. 22:412–6. doi: 10.1016/j.tig.2006.06.004.

Rocha EPC, Danchin A. 2004. An analysis of determinants of amino acids substitution rates in bacterial proteins. Mol. Biol. Evol. 21:108–16. doi: 10.1093/molbev/msh004.

Rodriguez JM, Maietta P, Ezkurdia I, Pietrelli A, Wesselink J-J, et al. 2013. APPRIS: annotation of principal and alternative splice isoforms. Nucleic Acids Res. 41:D110–7. doi: 10.1093/nar/gks1058.

Roux J, Privman E, Moretti S, Daub JT, Robinson-Rechavi M, et al. 2014. Patterns of positive selection in seven ant genomes. Mol. Biol. Evol. 31:1661–85. doi: 10.1093/molbev/msu141.

Roux J, Robinson-Rechavi M. 2011. Age-dependent gain of alternative splice forms and biased duplication explain the relation between splicing and duplication. Genome Res. 21:357–63. doi: 10.1101/gr.113803.110.

Rubinstein ND, Doron-Faigenboim A, Mayrose I, Pupko T. 2011. Evolutionary models accounting for layers of selection in protein-coding genes and their impact on the inference of positive selection. Mol. Biol. Evol. 28:3297–308. doi: 10.1093/molbev/msr162.

Rustici G, Kolesnikov N, Brandizi M, Burdett T, Dylag M, et al. 2013. ArrayExpress update--trends in database growth and links to data analysis tools. Nucleic Acids Res. 41:D987–90. doi: 10.1093/nar/gks1174.

Sarcar D. 2008. *Lattice: Multivariate data visualization with R*. Springer: New York http://lmdvr.r-forge.r-project.org.

Satake M, Kawata M, McLysaght A, Makino T. 2012. Evolution of Vertebrate Tissues Driven by Differential Modes of Gene Duplication. DNA Res. 1–12. doi: 10.1093/dnares/dss012.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13:2498–504. doi: 10.1101/gr.1239303.

Shannon PT, Grimes M, Kutlu B, Bot JJ, Galas DJ. 2013. RCytoscape: tools for exploratory network analysis. BMC Bioinformatics. 14:217. doi: 10.1186/1471-2105-14-217.

Subramanian S, Kumar S. 2004. Gene expression intensity shapes evolutionary rates of the proteins encoded by the vertebrate genome. Genetics. 168:373–81. doi: 10.1534/genetics.104.028944.

The ENCODE Project Consortium. 2011. A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol. 9:e1001046. doi: 10.1371/journal.pbio.1001046.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, et al. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. 7:562–78. doi: 10.1038/nprot.2012.016.

Tuller T, Kupiec M, Ruppin E. 2008. Evolutionary rate and gene expression across different brain regions. Genome Biol. 9:R142. doi: 10.1186/gb-2008-9-9-r142.

Urrutia A, Hurst L. 2003. The signature of selection mediated by expression on human genes. Genome Res. 2260–2264. doi: 10.1101/gr.641103.

Wall DP, Hirsh AE, Fraser HB, Kumm J, Giaever G, et al. 2005. Functional genomic analysis of the rates of protein evolution. Proc. Natl. Acad. Sci. U. S. A. 102:5483–8. doi: 10.1073/pnas.0501761102.

Wang C, Gong B, Bushel PR, Thierry-Mieg J, Thierry-Mieg D, et al. 2014. The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. Nat. Biotechnol. doi: 10.1038/nbt.3001.

Warnefors M, Kaessmann H. 2013. Evolution of the Correlation Between Expression Divergence and Protein Divergence in Mammals. Genome Biol. Evol. 5:1324–1335. doi: 10.1093/gbe/evt093.

Wickham H. 2011. The Split-Apply-Combine Strategy for Data. J. Stat. Softw. 40:1–29. http://www.jstatsoft.org/v40/i01/.

Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, et al. 2005. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics. 21:650–9. doi: 10.1093/bioinformatics/bti042.

Yang L, Gaut BS. 2011. Factors that contribute to variation in evolutionary rate among Arabidopsis genes. Mol. Biol. Evol. 28:2359–69. doi: 10.1093/molbev/msr058.

Zaheri M, Dib L, Salamin N. 2014. A Generalized Mechanistic Codon Model. Mol. Biol. Evol. 31:2528–2541. doi: 10.1093/molbev/msu196.

Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol. Biol. Evol. 22:2472–9. doi: 10.1093/molbev/msi237.

Table 1. Values of partial Spearman correlations between parameters, over all tissues
Top right of table: values for mouse (corresponding to Fig 1A); bottom left of table: values for human (corresponding to Fig 1B).
Not significant (p-value>0.0005) are in italics.

mouse / human	ω_0	ΔlnL	p_1	τ	Median expression	Maximal expression	Stage Number	GC content	Intron length	Intron number	CDS length	Paralogs number	Phyletic age
ω_0		*-0.031*	0.748	0.067	0.051	-0.074	*-0.012*	-0.055	*-0.030*	0.052	-0.062	-0.064	-0.163
ΔlnL	0.133		*0.014*	-0.048	-0.187	0.079	-0.043	*-0.020*	-0.038	-0.060	0.042	*-0.017*	*0.000*
p_1	0.598	-0.037		*0.024*	*-0.005*	*0.029*	-0.047	-0.066	-0.080	*-0.034*	0.042	-0.069	*-0.017*
τ	0.085	*-0.006*	*0.018*		-0.803	0.468	-0.374	-0.039	0.094	-0.061	0.074	0.125	-0.069
Median expression	0.049	0.105	*0.015*	-0.790		0.553	0.012	0.088	0.135	*0.002*	0.061	0.070	*-0.025*
Maximal expression	-0.041	*0.033*	*-0.005*	0.406	0.530		0.164	0.231	-0.076	0.168	-0.283	0.075	*0.010*
Stage Number	-0.041	-0.055	-0.043	-0.381	0.063	0.163		-0.287	-0.048	*-0.006*	*0.015*	-0.087	0.108
GC content	-0.159	-0.049	-0.042	0.121	0.139	-0.039	*-0.020*		-0.518	-0.190	0.113	0.071	-0.090
Intron length	-0.088	-0.074	-0.047	0.163	0.165	-0.137	0.103	-0.517		-0.103	0.044	0.123	-0.105
Intron number	0.039	-0.039	*0.002*	-0.091	*-0.029*	0.205	-0.065	-0.093	-0.037		0.683	*0.029*	0.141
CDS length	*-0.014*	*-0.011*	*0.008*	0.135	0.109	-0.312	0.101	*-0.011*	-0.041	0.629		0.181	0.111
Paralogs number	-0.136	-0.019	-0.069	0.155	0.104	0.092	-0.052	0.070	0.103	0.065	0.173		0.175
Phyletic age	-0.213	-0.034	-0.084	-0.136	-0.081	0.048	0.091	-0.065	-0.078	0.151	0.122	0.188	

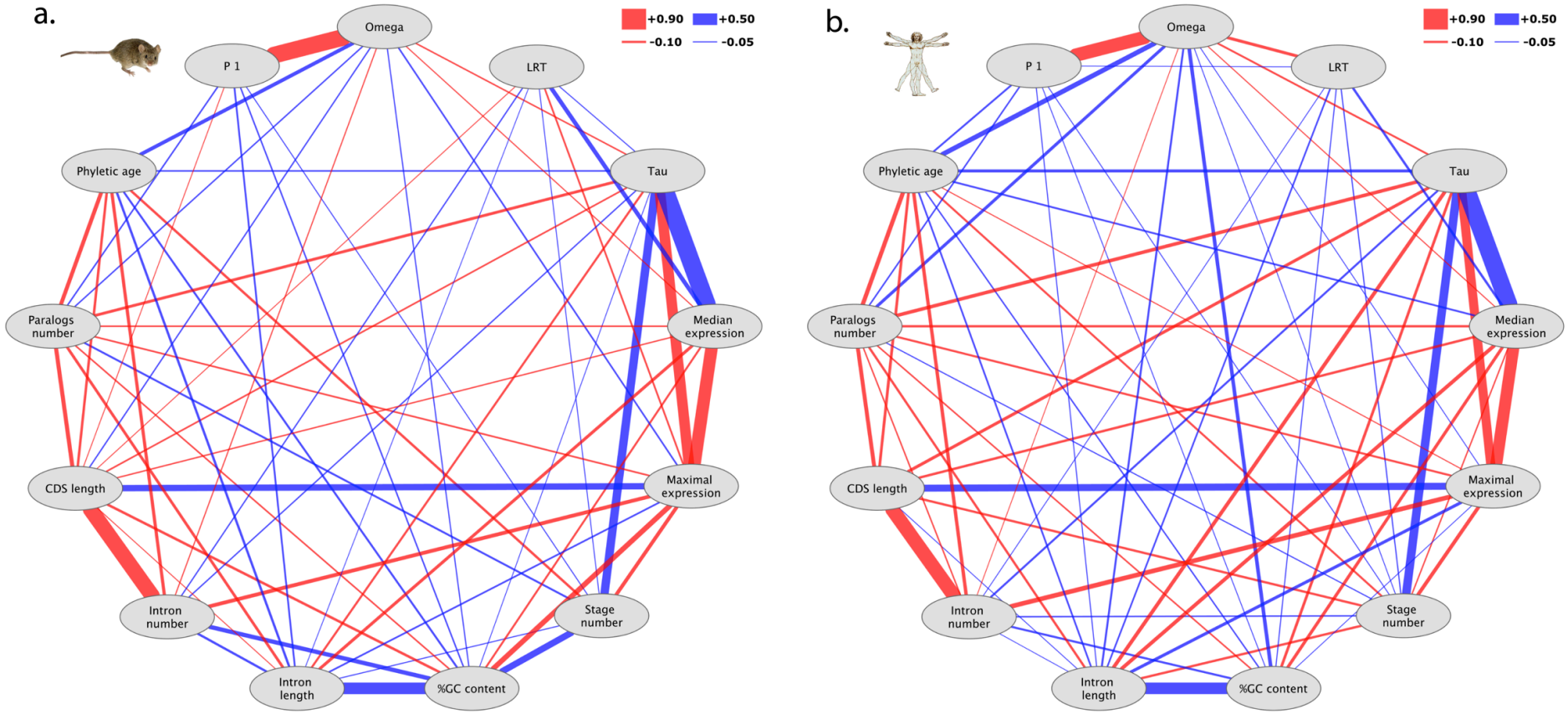

Fig. 1. Spearman partial correlations in a) mouse and b) human. The width of the lines shows the strength of correlations. Red lines show positive correlations, blue lines show negative correlations. Only significant correlations ($p<0.0005$) are shown.

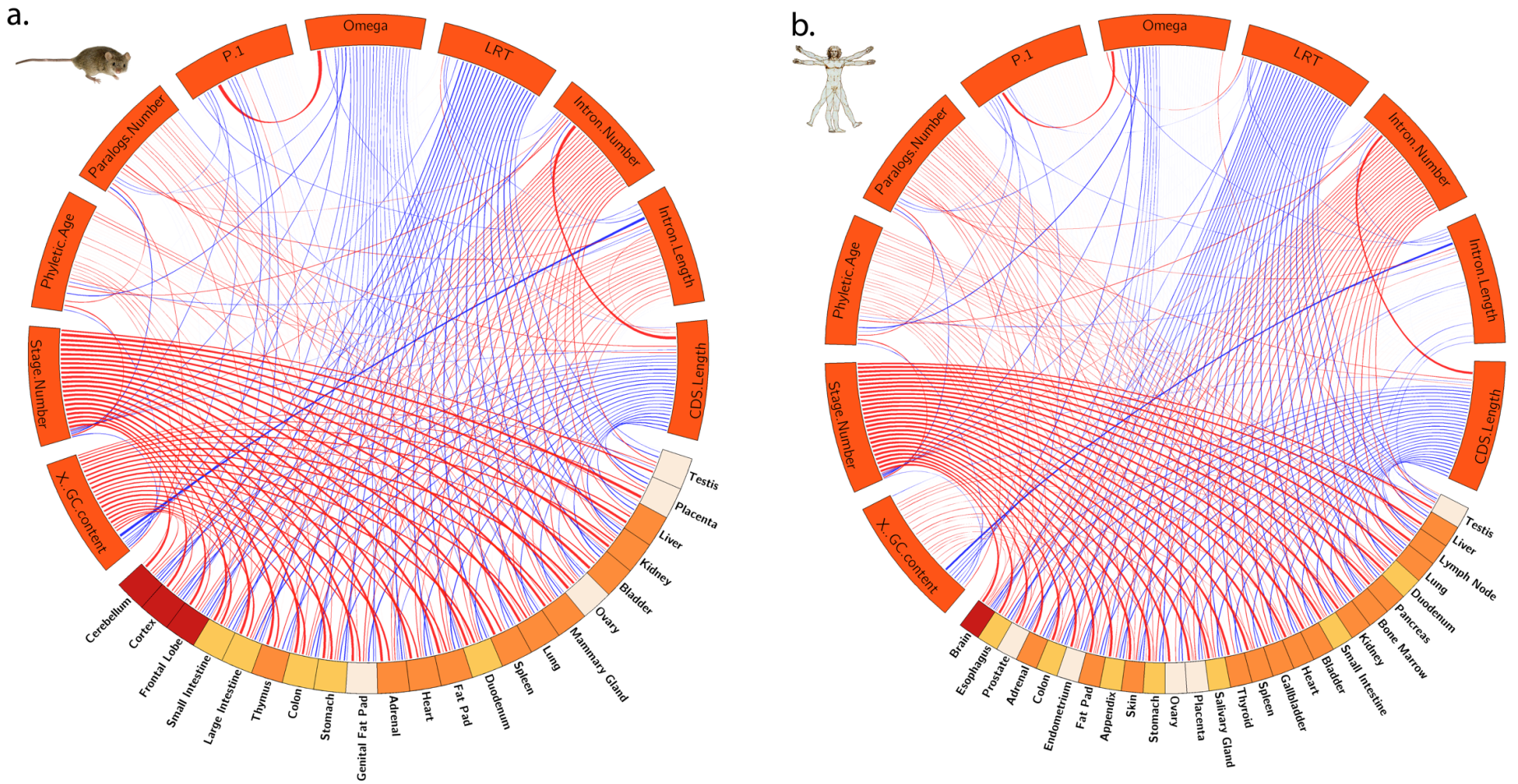

Fig. 2. Spearman partial correlation with expression values for each tissue separately for a) mouse and b) human. The width of the lines shows the strength of correlations. Red lines show positive correlations, blue shows negative correlations. Only significant correlations ($p<0.0005$) are shown. Color of the tissue bands represents different groups of tissues (gastrointestinal system, central nervous system, reproductive system and misc).