

Organismal complexity, cell differentiation and gene expression: human over mouse

Alexander E. Vinogradov* and Olga V. Anatskaya

Institute of Cytology, Russian Academy of Sciences, Tikhoretsky Avenue 4, St. Petersburg 194064, Russia

Received July 2, 2007; Revised August 12, 2007; Accepted September 1, 2007

ABSTRACT

We present a molecular and cellular phenomenon underlying the intriguing increase in phenotypic organizational complexity. For the same set of human–mouse orthologous genes (11 534 gene pairs) and homologous tissues (32 tissue pairs), human shows a greater fraction of tissue-specific genes and a greater ratio of the total expression of tissue-specific genes to housekeeping genes in each studied tissue, which suggests a generally higher level of evolutionary cell differentiation (specialization). This phenomenon is spectacularly more pronounced in those human tissues that are more directly involved in the increase of complexity, longevity and body size (i.e. it is reflected on the organismal level as well). Genes with a change in expression breadth show a greater human–mouse divergence of promoter regions and encoded proteins (i.e. the functional genomics data are supported by the structural analysis). Human also shows the higher expression of translation machinery. The upstream untranslated regions (5'UTRs) of human mRNAs are longer than mouse 5'UTRs (even after correction for the difference in genome sizes) and contain more uAUG codons, which suggest a more complex regulation at the translational level in human cells (and agrees well with the augmented cell specialization).

INTRODUCTION

The growth of organizational complexity (called sometimes 'progress') is an intriguing and probably most important biological and social phenomenon. The measures and the driving forces of this process are still debated (1–12). A prominent feature of 'progressive' evolution in biological and social systems is the increase of division of labor among system elements (1–5). In multicellular organisms, it is reflected in cell differentiation accompanied by enhancement of specialized cell function and

appearance of tissue-specific genes with the corresponding increase in the number of different cell types in the organism. The number of cell types (or other organismal parts), which reflects the degree of division of labor, is an agreeable indicator of biological complexity (3,4,6,11). However, the count of morphologically discernable cell types is ambiguous in higher multicellulars (6,11). Furthermore, there can be cell types that are not distinguishable morphologically. Similarly, no universal (suitable for all cell types) measure of cell differentiation is known. We introduce here a general molecular-level indicator that might be suitable for this purpose: the fraction of tissue-specific genes and the ratio of the total expression of tissue-specific genes to the total expression of housekeeping genes. We studied the genome-wide differences in expression of orthologous genes between human and mouse from this perspective and found that human has the higher cell differentiation in each studied homologous tissue. This genomic and cellular phenomenon is spectacularly more pronounced in those human tissues that are more directly involved in the increase of complexity, longevity and body size (i.e. it is reflected on the organismal level as well).

The preliminary results on the fraction of housekeeping genes (using a much lower number of homologous genes and tissues) were reported for the older versions of human and mouse microarray platforms (10). It should also be noted that human–mouse comparison of gene expression was performed previously in a number of works but from other angles [e.g. (13–17), and references therein].

MATERIALS AND METHODS

The data on expression of human and mouse genes were taken from the Novartis Gene Expression Atlas (14). They present the results of high-density oligonucleotide microarray experiments performed uniformly for all tissues. The uniform platforms were used for all human (U133A + GNF1B) and mouse (GNF1M) tissues. The signals from probes on the chip corresponding to the same gene were averaged; the samples representing the same tissue were also averaged. Only probes that presented the characterized genes, i.e. with links to the Entrez Gene (18)

*To whom correspondence should be addressed. Tel: +7-812-2971452; Fax: +7-812-2970341; Email: aevin@mail.cytspb.rssi.ru

and RefSeq (19) databases, and that were presumably orthologous between human and mouse were used. The orthology was established using the HomoloGene database (18) (11 534 gene pairs with reciprocal best hits were found). Only normal tissues that were homologous for human and mouse were used (32 tissues, listed in Figure 1).

The modern microarrays show good reproducibility and portability across platforms (20–22). Even if there can be problems in regard to individual genes, the phenomena described in the present article are based on the expression of hundreds and thousands of genes. Furthermore, discrepancies in regard to individual genes are usually due to poorly designed probes (because of incorrect gene annotations, i.e. poorly known genes) (20,22), whereas we used only well-characterized genes (those with links to Entrez Gene, RefSeq and HomoloGene databases). To prove the main phenomena, we used very diverse tests (see below). Moreover, the preliminary results on the fraction of housekeeping genes were obtained with the older human and mouse platforms (U95A and U74A) standardized with the older algorithm (10). Thus, the effect is reproducible and portable across platforms. It also should be noted that putative artifacts due to variation in gene GC content [such as higher stability or affinity of GC-rich RNAs (23,24)], if any, would act in the direction opposite to the revealed effect because housekeeping genes are more GC-rich than tissue-specific ones, and this difference is higher in human compared to mouse (10). Furthermore, all RNAs samples used in the Gene Expression Atlas were obtained by the same procedure from the frozen tissues, and the quality of all samples was checked with the Agilent Bioanalyzer (14). As for the probe affinity, the effect of GC content is unlikely in oligonucleotide microarrays because the so-called ‘mismatch’ probe used as control contains the same 25-nt sequence as the ‘perfect match’ probe, except for one central nucleotide (i.e. it has nearly the same GC content) (25).

For discerning housekeeping and tissue-specific genes, we used both the cutoff-based criteria of expression detection (with different cutoff values, as indicated in Table 1) and the Affymetrix calls (presence/absence) provided in the Gene Expression Atlas. The detection of gene expression with cutoff value equal to dataset median was recommended by the authors of the Gene Expression Atlas; it is based on the extensive PCR-validation of oligonucleotide microarray data (13–15). The Affymetrix call for a given gene is a local feature, which is not depended on microarray scaling (i.e. on signal from probes for other genes) (25). Also, we used various parameters of the among-tissues distribution of expression signal, which are related to the degree of tissue-specificity (and which were used previously for other purposes or introduced here): entropy (26), ratio of maximum to average value (27), index of tissue-specificity (16,28), skew (distribution of tissue-specific genes should be more right-skewed), coefficient of variation (should be higher for tissue-specific genes), first absolute central moment (should be higher for tissue-specific genes), ratio of the total expression of tissue-specific genes to the total expression of housekeeping genes.

For comparison with the novel exon microarray platforms, the gene expression data available for six homologous human–mouse tissues were taken from the Exon Array dataset normalized with the GeneBASE program (17). The tissues were heart, kidney, liver, muscle, spleen and testis. (For comparison with the Gene Expression Atlas, lymphnode was taken in the Gene Expression Atlas as analogous of spleen in the Exon Array dataset.) There were 8229 genes common for the Gene Expression Atlas and the Exon Array dataset. We used the definition of housekeeping genes obtained with the Gene Expression Atlas (because it contains the much larger number of tissues) to analyze the expression of housekeeping and non-housekeeping genes in six tissues of the Exon Array dataset. [The cutoff ‘median expression’, as recommended in refs (13–15), was used for definition of housekeeping genes. The results were similar for definition based on the Affymetrix calls.]

For human–mouse comparison of gene promoter regions, we extracted their sequences from the database of experimentally determined exact transcriptional start sites (DBTSS) (29). The comparison was made as described (30). Briefly, we first masked these sequences for lineage-specific repeats (that were inserted after the human–mouse split) using the standalone RepeatMasker and DateRepeats programs (Smit,A.F.A., Hubley,R., Green,P., <http://repeatmasker.org>). Then, the matching of human–mouse promoter regions (200 nt-long sequences upstream transcription start site) was done using a very rigorous Huang–Miller algorithm for local sequence alignment, implemented in the Lalign program (31). (The significance level for spurious match was set to a conservative threshold $P < 10^{-6}$.)

The analysis of functional gene modules was done as described (32). Briefly, we checked the average human–mouse difference of a tested parameter for a predefined gene set against the average difference for the total dataset. The predefined gene sets were prepared using Gene Ontology (GO) categories (33), and pathways compilations from KEGG (34) and Reactome (35) databases (using Entrez Gene mapping), and HumanCyc (36). In the case of Gene Ontology categories, we collected for each category all its subcategories using GO graphs, and a gene was regarded as belonging to a given category if it was mapped to any of its subcategories in Entrez Gene. For estimation of significance level, we did 20 000 random samplings from the total dataset (of a size equal to the size of a tested gene group). After obtaining two-tailed significance level (P -value), we estimated false discovery rate (q -value) for correction for multiple comparisons (37). As the main parameter, the (log-transformed) expression level averaged among all tissues was used.

For analysis of upstream untranslated regions (5'UTR), the human and mouse mRNAs were extracted from the RefSeq database (19). To ensure the completeness of 5'UTRs, we used the database of experimentally determined exact transcriptional start sites (DBTSS) (29). We used only those 5'UTRs whose lengths were non-zero in both species and equal (for a given species) in both databases. There were 1578 orthologous gene pairs whose

5'UTRs of processed mRNAs satisfied these conditions. (If zero-length 5'UTRs were included in the analysis, the results were similar.) If there were several mRNAs for a gene, the longest 5'UTR was taken. (The shortest and the average 5'UTR lengths were also compared; the results were qualitatively the same.) For correction for the difference in genome sizes, the lengths of mouse 5'UTRs were multiplied by factor of 1.16 (38).

To compare nucleotide composition between silent sites of coding DNA and background, we calculated chi-squares of 4-fold degenerate third codon positions using intronic and nearby intergenic sequences for estimation of background nucleotide probabilities. In a variant of analysis, introns were preliminarily masked for (human or mouse) lineage-specific repeats. (This was done because repeat-insertion is a special mode of mutation pressure, which might change nucleotide composition and which is suppressed in coding regions.) Then, we compared the obtained chi-squares in pairwise way for human-mouse orthologous genes (using the Mann-Whitney test).

RESULTS AND DISCUSSION

Evolutionary cell differentiation

The conclusion about the difference in evolutionary cell differentiation between two mammals may seem surprising. Therefore, we checked it extensively using the diverse tests (both dependent and independent of expression threshold criteria). By any expression threshold criterion, there are ~3-fold more orthologous genes expressed in all studied tissues (i.e. housekeeping genes) in the mouse than in the human (Table 1). Both the mean and the median of among-tissues breadth of expression (number of tissues

where a given gene is expressed) are also higher in the mouse (Table 1). Similarly, other parameters (which are independent of expression threshold criteria) show the consistent differences indicating a relative bias towards housekeeping genes in mouse [entropy (26)] or towards tissue-specific genes in human (ratio of maximum to average expression value, index of tissue-specificity, skew, coefficient of variation, first absolute central moment; see 'Methods' section) (Table 1).

Moreover, we found changes at the structural level in those genes that changed their expression breadth after the human-mouse split. Thus, the human-mouse identity of encoded protein sequence (compared at amino acid level) is lower in those genes that are housekeeping in mouse but non-housekeeping in human compared to genes that are housekeeping in both species (0.890 ± 0.003 versus 0.933 ± 0.006 , Mann-Whitney $P < 10^{-12}$). Similarly, those genes that are housekeeping in human but non-housekeeping in mouse show the lower protein sequence identity than genes that are housekeeping in both species (0.898 ± 0.010 versus 0.933 ± 0.006 , $P < 10^{-7}$). [The cutoff 'median expression', as recommended in refs (13–15), was used for definition of housekeeping genes. The results were similar for definition based on the Affymetrix calls.]

Furthermore, the human-mouse conservation of nucleotide sequence in the promoter region (200 nt upstream transcription start site) is lower in those genes that are housekeeping in mouse but non-housekeeping in human compared to genes that are housekeeping in both species (for the length of conserved sequence multiplied by its human-mouse identity: 70.5 ± 2.7 versus 87.7 ± 5.2 , Mann-Whitney $P < 10^{-7}$). Similarly, those genes that are housekeeping in human but non-housekeeping in mouse show the lower conservation of promoter regions than

Table 1. Parameters indicative of the degree of tissue-specificity in mouse and human (integrative for 32 homologous tissues studied, based on 11 534 orthologous genes)

Parameter	Mouse	Human	<i>P</i> for difference*
Fraction of genes expressed in all 32 tissues (i.e. housekeeping genes) according to:			
(1) Affymetrix calls	9.0%	2.5%	$P < 10^{-12}$
(2) Cutoff '1/2 of median expression'	40.6%	16.4%	$P < 10^{-12}$
(3) Cutoff 'median expression' (as recommended in refs. 13–15)	17.3%	5.3%	$P < 10^{-12}$
(4) Cutoff '2 median expression'	4.01%	1.38%	$P < 10^{-12}$
Mean number of tissues where a gene is expressed (Affymetrix calls)	13.24 ± 0.11	9.06 ± 0.10	$P < 10^{-12}$
Median number of tissues where a gene is expressed (Affymetrix calls)	9.0	4.0	$P < 10^{-12}$
Entropy sensu (26)	0.1455 ± 0.0002	0.1416 ± 0.0001	$P < 10^{-12}$
Ratio of maximum to average value (log-transformed)	0.564 ± 0.004	0.725 ± 0.003	$P < 10^{-12}$
Index of tissue-specificity sensu refs. 16, 28	0.2025 ± 0.0011	0.2592 ± 0.0009	$P < 10^{-12}$
Skew	1.742 ± 0.013	1.962 ± 0.010	$P < 10^{-12}$
Coefficient of variation, %	70.54 ± 0.66	89.52 ± 0.51	$P < 10^{-12}$
First absolute central moment, %	43.65 ± 0.28	58.58 ± 0.20	$P < 10^{-12}$
Mean ratio of the total expression of tissue-specific genes (expressed in <32 tissues, according to Affymetrix calls) to the total expression of housekeeping genes (expressed in all 32 tissues) in a tissue	2.89 ± 0.08	5.12 ± 0.26	$P < 10^{-9}$
Mean ratio of the total expression of tissue-specific genes (expressed in <32 tissues, according to cutoff 'median expression') to the total expression of housekeeping genes (expressed in all 32 tissues) in a tissue	1.35 ± 0.03	2.79 ± 0.13	$P < 10^{-10}$

*Binomial test for fractions, Mann-Whitney for other parameters.

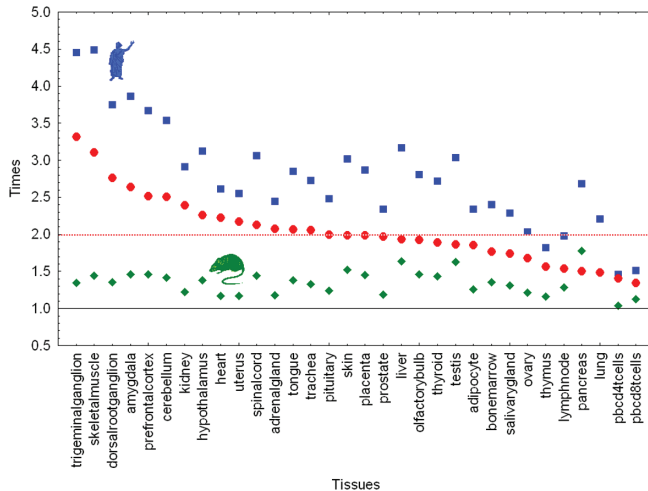


Figure 1. The ratio of the total expression of tissue-specific genes to the total expression of housekeeping genes in human (blue squares), mouse (green diamonds) and the ratio of the former to the latter, i.e. the human/mouse ratio of both ratios (red circles) in 32 homologous tissues (based on 11534 orthologous genes). Note that all blue squares are higher than the corresponding green diamonds, and, as a consequence, all red circles are above unity. (Symbol size is greater than error. Red dotted line indicates median for all 32 tissues. Tissue-specific genes are those expressed in <32 tissues; housekeeping genes are those expressed in all 32 tissues, according to cutoff 'median expression' (as recommended in refs 13–15). If the Affymetrix calls were used, the picture was qualitatively the same with a similar ranking of the tissues.)

genes that are housekeeping in both species (66.7 ± 6.7 versus 87.7 ± 5.2 , $P < 10^{-5}$). These facts suggest that the revealed changes in gene expression breadth are not due to some artifact of the functional genomics dataset because they correlate with the corresponding alterations in the structure of gene regulatory regions.

Strength of the effect in different tissues

For comparing different tissues, we used a combined parameter, which estimated both the among-tissues breadth and the level of expression: the ratio of the total expression of tissue-specific (non-housekeeping) genes to the total expression of housekeeping genes in a tissue (Table 1). Noteworthy, this parameter allows the estimation of the relative expression of tissue-specific genes in a tissue using the universal internal standard for all tissues (expression of the same set of housekeeping genes), thus removing possible problems with the variation of mRNAs amounts in different tissues, among-samples data standardization, etc. We used all non-housekeeping genes as tissue-specific genes because of the following reasons. First, on the histogram of genes expressed in each particular tissue there is no clear-cut peak of tissue-specific genes, only the peak of housekeeping genes and the plateau of genes with gradually changing tissue-specificity [Figure 2B in ref. (10); the picture was similar with the present dataset]. Therefore, any other threshold would be arbitrary. Second, genes with the intermediate expression breadth (i.e. expressed in more than one tissue

but less than all tissues) show the highest informational load and probably contribute in a greater extent to multicellular complexity (39). Third, the uniqueness of gene expression pattern is determined by all non-housekeeping genes, which are expressed in a given tissue.

The ratio of the total expression of tissue-specific genes to housekeeping genes is higher in human compared to mouse in all studied homologous tissues (Figure 1: all red circles are above unity). This phenomenon is more pronounced in the nervous system (except for olfactory bulb), skeletal muscle, kidney and heart (Figure 1). In regard to the nervous system, the relation to organismal complexity is obvious. Of special interest are the trigeminal ganglion (TGG) and the dorsal root ganglion (DRG), where the effect is even higher than in the other parts of the nervous system. The TGG where the ophthalmic, maxillary and mandibular nerves converge is involved in visual information processing. Also, it controls face and mouth movements (speech), which are so important for human social organization. The DRG contains cell bodies of incoming sensory fibers from the rest of the body. Among other things, the precise finger activity, which is so important for human, depends on the DRG. Thus, the prominent human–mouse differences in these two tissues might reflect not only the increase in organismal complexity but also the transition to the qualitatively higher complexity level: social organization (TGG) and tool making (DRG). The amygdala is involved in emotional learning. The cerebellum is responsible for complex movement in 3D space (especially important for primates) and, in particular, in the maintenance of posture.

The higher cell differentiation in human skeletal muscle might be related to the larger body size [because of allometry of muscle strength to body mass known since Galileo (11)] and the upright posture. The heart is important for increased body size, upright posture and longevity. The heart is a known bottleneck in human constitution, which limits lifespan and works near the ceiling of its capability (32). The higher cell differentiation in human kidney is probably related to increased longevity (maintenance of homeostasis). The kidney disease is among the leading causes of human death (40), which indicates that human kidney works near the limit of its capability.

Of special interest is the olfactory bulb where the effect is below the median of studied tissues (and especially weaker if compared to other parts of the nervous system) (Figure 1). It is well known that rodents are olfactory specialists while primates are visual specialists (41). In other words, the case of olfactory bulb is just that kind of exception that confirms the rule.

Comparison with the Exon Array dataset

There is now the data on gene expression in a few homologous human–mouse tissues obtained with the novel exon microarrays (designed mainly for studying alternative splicing) (17). We determined the ratio of the total expression of tissue-specific (non-housekeeping) genes to the total expression of housekeeping genes for

Table 2. The functional gene modules with the higher (or lower) level of expression in human compared to mouse (based on the human/mouse ratio of the among-tissues average log-transformed levels of expression)

Module ID	Module name	Contrast ^a	N genes in module ^b	q-value ^c
GO:0005853	Eukaryotic translation elongation factor 1 complex	1.77	2	0.007
KEGG:03010	Ribosome	0.70	40	0
GO:0005830	Cytosolic ribosome (sensu Eukaryota)	0.55	20	0
Reactome:72766	Translation	0.55	66	0
GO:0015935	Small ribosomal subunit	0.39	22	0.019
GO:0005840	Ribosome	0.37	86	0
GO:0003735	Structural constituent of ribosome	0.33	90	0
Reactome:74160	Gene expression	0.20	179	0
GO:0030529	Ribonucleoprotein complex	0.15	200	0
GO:0006412	Protein biosynthesis	0.14	325	0
GO:0009059	Macromolecule biosynthesis	0.12	378	0
GO:0043227	Membrane-bound organelle	-0.03	3352	0.007
GO:0043231	Intracellular membrane-bound organelle	-0.03	3350	0

^aThe positive contrast indicates the relatively higher expression in the human, the negative ones, in the mouse.

^bThe gene modules are Gene Ontology (GO) categories, KEGG, Reactome and HumanCyc pathways.

^cThe 'q-value'—false discovery rate.

six tissues provided in the Exon Array dataset (heart, kidney, liver, muscle, spleen and testis) (17). In consistence with the results described in the previous section, these ratios were higher in all human tissues compared to mouse ones. The human–mouse ratio of these ratios was even seemingly higher than that obtained using the Gene Expression Atlas but did not differ significantly (2.51 ± 0.23 versus 2.18 ± 0.48 , Mann–Whitney $P > 0.1$). Also, there is a good correlation between the among-tissues rankings of these ratios in the Gene Expression Atlas and the Exon Array dataset (with the exception of one outlier—testis). If testis was excluded, Pearson $r = 0.94$, $P < 0.02$ (for log-transformed ratios), Spearman $r = 1.0$, $P < 10^{-4}$. The exception of testis might be related to a very high level of alternative splicing in this tissue (42).

Functional gene modules

A surprising (and most pronounced) difference on the modular level is the higher expression of human translation machinery (Table 2). Probably, it is not due to more intensive protein synthesis because human has a lower metabolic rate with a lower protein turnover (43). Then, how could the higher expression of human translation machinery be explained?

More complex regulation at translational level: a hypothesis

The regulation of gene expression at the level of translation is an emerging theme now (44–46).

Noteworthy, it endows local sites with independent decision-making authority, which is especially important for highly specialized cells with complex cellular architectures (45). Most of the translational regulatory mechanisms are inhibitory (44,46). Therefore, the more complex regulation at translational level should involve the more developed translation machinery. This situation is similar with the regulation at transcriptional level where the increase in complexity (e.g. after transition from prokaryotes to eukaryotes) was associated with the enlargement of the genome and a general reduction of transcription rate because of mostly suppressive regulatory means (e.g. chromatin condensation), which switch-off genes whose expression should not be allowed in a given cell. The untranslated regions of processed mRNAs, especially located upstream (5'UTRs), are known to serve for translational regulation (44,46–49). The longer 5'UTRs usually occur in those genes that should be more strongly and finely controlled (47). The AUG (potentially start) codons in the 5'UTRs (named uAUGs) are also involved in translational regulation and generally attenuate protein synthesis (49). If human has a more complex regulation at the translational level, it should be reflected in the structure of human mRNAs.

We found that human 5'UTRs are on average 37% longer than the corresponding mouse ones (difference of log-transformed lengths in orthologous genes compared in pairwise way: 0.138 ± 0.024 , Mann–Whitney $P < 10^{-12}$). This difference remains significant after correction for 16% difference in genome sizes (0.073 ± 0.024 , $P < 10^{-7}$). The latter fact suggests that the difference in 5'UTR lengths cannot be just due to mutation pressure (if one assumes that the difference in genome sizes was caused by neutral drift), and should have functional significance. The number of uAUGs is 30% greater in human compared to mouse (difference of counts compared in pairwise way: 0.152 ± 0.081 , $P < 10^{-5}$).

The recent data suggest that synonymous codon usage can be involved in translational regulation in mammals (50–52). If translational regulation is more complex in human, the nucleotide composition of silent sites in human coding sequences (presumably determined by translation-related selection) might differ stronger from background nucleotide composition (assumably determined by mutation pressure) compared to mouse. We calculated chi-squares of 4-fold degenerate third codon positions against intronic and nearby intergenic sequences (used for estimation of background nucleotide probabilities) and found that these chi-squares are higher in human (differences for human–mouse orthologous genes compared in pairwise way: for intergenic sequences, 16.4 ± 1.5 , Mann–Whitney $P < 10^{-12}$; for intronic sequences, 11.8 ± 0.7 , $P < 10^{-12}$; for intronic sequences masked for human- or mouse-specific repeats, 12.6 ± 0.7 , $P < 10^{-12}$).

CONCLUSION

For the same set of human–mouse orthologous genes and homologous tissues, human shows a greater fraction of tissue-specific genes. From the neutralist standpoint,

the main difference between human and mouse is in effective population size, which may result in a weaker purifying selection in human. However, its effect (if any) should be in the direction opposite to our finding. Thus, a relaxed purifying selection could lead to the higher non-functional gene expression, i.e. higher transcriptional background noise (e.g. 53,54). This would result in a seemingly higher fraction of broadly expressed genes in human compared to mouse (opposite to what we found). Therefore, we believe that the selectionist interpretation related to human–mouse difference in complexity is more likely. It seems that the higher complexity of human exists not only on the phenotypic organismal level but also on the genomic and cellular levels.

In any studied tissue, human has a greater ratio of the total expression of tissue-specific genes to the total expression of housekeeping genes, which indicates the intensification of specialized cell function. How the homologous cells of two mammals (having the similar sets of organs) can differ in the degree of specialization? The biochemical diversification (specialization) of cells of seemingly the same type (hepatocytes) within the complex organ architecture was reported even for such relatively ‘homogenous’ organ as the liver (albeit it may not be discerned morphologically) (55,56). The degree of such specialization can be higher in human organs. Thus, there can be a higher number of different ‘biochemical’ cell types (determined by uniqueness of gene expression pattern).

On the organismal level, there is a striking correlation of the strength of this phenomenon with a type of tissue. It is more pronounced in those tissues that are more directly involved in the increase of organizational complexity, longevity and body size. Noteworthy, human cells are more resistant to transformation compared to mouse cells (57,58), i.e. the greater human longevity can indeed be reflected on the cellular level [to say nothing of the questionable ‘Hayflick limit’ (59)].

The higher expression of human translation machinery is in good agreement with the longer 5'UTRs of human mRNAs and the greater number of uAUG codons, which suggest a more complexly regulated translation. The translational regulation can be more important for the more specialized cells because it allows a more rapid and economical (and localized within the cell with complex architecture) response to stimuli, albeit in a narrower range (specified by the current transcriptome), compared to regulation at the level of transcription. (As an utmost example, compare the nucleus-free mammalian erythrocytes, which do not have transcriptional regulation at all, with the nucleated erythrocytes of lower vertebrates.)

Because of universality of the principle of division of labor in the ‘progressive’ evolution of biological and social systems, the level of cell differentiation (estimated by the relative number and expression of tissue-specific versus housekeeping genes) might become a general indicator of multicellular organismal complexity. In a practical sense, as mouse is a paramount model for biomedical research, it is important to understand its principal differences from human.

ACKNOWLEDGEMENTS

We thank the anonymous reviewers for helpful comments. This work was supported by the Russian Foundation for Basic Research (RFBR). The Open Access publication charges for this article were waived by Oxford University Press.

Conflict of interest statement. None declared.

REFERENCES

- Smith, A. (1776) *The Wealth of Nations*. Edinburgh.
- Spencer, H. (1891) *Essays: Scientific, Political, and Speculative*, Vol. 1. Williams and Norgate, London.
- Bonner, J.T. (1988) *The Evolution of Complexity*. Princeton University Press, Princeton, NJ.
- McShea, D.W. (1996) Metazoan complexity and evolution: Is there a trend? *Evolution*, **50**, 477–492.
- Gould, S.J. (2002) *The Structure of Evolutionary Theory*. Harvard University Press, Cambridge, Massachusetts.
- Carroll, S.B. (2001) Chance and necessity: the evolution of morphological complexity and diversity. *Nature*, **409**, 1102–1109.
- McShea, D.W. (2001) The hierarchical structure of organisms: a scale and documentation of a trend in the maximum. *Paleobiology*, **27**, 405–423.
- Szathmary, E., Jordan, F. and Pal, C. (2001) Can genes explain biological complexity? *Science*, **292**, 1315–1316.
- Wolfe, K.H. (2001) Yesterday's polyploids and the mystery of diploidization. *Nat. Rev. Genet.*, **2**, 333–341.
- Vinogradov, A.E. (2003) Isochores and tissue-specificity. *Nucleic Acids Res.*, **31**, 5212–5220.
- Bonner, J.T. (2004) Perspective: the size-complexity rule. *Evolution*, **58**, 1883–1890.
- Taft, R.J., Pheasant, M. and Mattick, J.S. (2007) The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays*, **29**, 288–299.
- Su, A.I., Cooke, M.P., Ching, K.A., Hakak, Y., Walker, J.R., Wiltshire, T., Orth, A.P., Vega, R.G., Sapinoso, L.M. et al. (2002) Large-scale analysis of the human and mouse transcriptomes. *Proc. Natl Acad. Sci. USA*, **99**, 4465–4470.
- Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M. et al. (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA*, **101**, 6062–6067.
- Yang, J., Su, A.I. and Li, W.H. (2005) Gene expression evolves faster in narrowly than in broadly expressed mammalian genes. *Mol. Biol. Evol.*, **22**, 2113–2118.
- Liao, B.Y. and Zhang, J. (2006) Low rates of expression profile divergence in highly expressed genes and tissue-specific genes during mammalian evolution. *Mol. Biol. Evol.*, **23**, 1119–1128.
- Xing, Y., Ouyang, Z., Kapur, K., Scott, M.P. and Wong, W.H. (2007) Assessing the conservation of mammalian gene expression using high-density exon arrays. *Mol. Biol. Evol.*, **24**, 1283–1285.
- Maglott, D., Ostell, J., Pruitt, K.D. and Tatusova, T. (2005) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.*, **35**, D26–D31.
- Pruitt, K.D., Tatusova, T. and Maglott, D.R. (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.*, **35**, D61–D65.
- Larkin, J.E., Frank, B.C., Gavras, H., Sultana, R. and Quackenbush, J. (2005) Independence and reproducibility across microarray platforms. *Nat. Methods*, **2**, 337–344.
- Irizarry, R.A., Warren, D., Spencer, F., Kim, I.F., Biswal, S., Frank, B.C., Gabrielson, E., Garcia, J.G., Geoghegan, J. et al. (2005) Multiple-laboratory comparison of microarray platforms. *Nat. Methods*, **2**, 345–350.
- Quackenbush, J. and Irizarry, R.A. (2006) Response to Shields: ‘MIAME, we have a problem’. *Trends Genet.*, **22**, 471–472.
- Margulies, E.H., Kardia, S.L. and Innis, J.W. (2001) Identification and prevention of a GC content bias in SAGE libraries. *Nucleic Acids Res.*, **29**, E60.

24. Barreau, C., Paillard, L. and Osborne, H.B. (2006) AU-rich elements and associated factors: are there unifying principles? *Nucleic Acids Res.*, **33**, 7138–7150.
25. Affymetrix inc. (2004) GeneChip Expression Analysis. Data Analysis Fundamentals.
26. Schug, J., Schuller, W.P., Kappen, C., Salbaum, J.M., Bucan, M. and Stoeckert, C.J.Jr (2005) Promoter features related to tissue specificity as measured by Shannon entropy. *Genome Biol.*, **6**, R33.
27. Vinogradov, A.E. (2005) Dualism of gene GC content and CpG pattern in regard to expression in the human genome: magnitude versus breadth. *Trends Genet.*, **21**, 639–643.
28. Yanai, I., Benjamin, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., Bar-Even, A., Horn-Saban, S., Safran, M. *et al.* (2005) Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics*, **21**, 650–659.
29. Yamashita, R., Suzuki, Y., Wakaguri, H., Tsuritani, K., Nakai, K. and Sugano, S. (2006) DBTSS: DataBase of Human Transcription Start Sites, progress report 2006. *Nucleic Acids Res.*, **34**, D86–D89.
30. Vinogradov, A.E. (2006) “Genome design” model: evidence from conserved intronic sequence in human-mouse comparison. *Genome Res.*, **16**, 347–354.
31. Pearson, W.R. (1999) Flexible similarity searching with the FASTA3 program package. In Misener, S. and Krawetz, S.A. (eds), *Bioinformatics Methods and Protocols*. Humana Press, Totowa, NJ, pp. 185–219.
32. Anatskaya, O.V. and Vinogradov, A.E. (2004) Heart and liver as developmental bottlenecks of mammal design: evidence from cell polyploidization. *Biol. J. Linn. Soc.*, **83**, 175–186.
33. The Gene Ontology Consortium (2006) The Gene Ontology (GO) project in 2006. *Nucleic Acids Res.*, **34**, D322–D326.
34. Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M. and Hirakawa, M. (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.*, **34**, D354–D357.
35. Vastrik, I., D'Eustachio, P., Schmidt, E., Joshi-Tope, G., Gopinath, G., Croft, D., de Bono, B., Gillespie, M., Jassal, B. *et al.* (2007) Reactome: a knowledgebase of biological pathways and processes. *Genome Biol.*, **8**, R39.
36. Romero, P., Wagg, J., Green, M.L., Kaiser, D., Krummenacker, M. and Karp, P.D. (2005) Computational prediction of human metabolic pathways from the complete human genome. *Genome Biol.*, **6**, R2.
37. Storey, J.D. and Tibshirani, R. (2003) Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA*, **100**, 9440–9445.
38. Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M. *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, **420**, 520–562.
39. Vinogradov, A.E. (2006) ‘Genome design’ model and multicellular complexity: golden middle. *Nucleic Acids Res.*, **34**, 5906–5914.
40. Eknoyan, G., Lameire, N., Barsoum, R., Eckardt, K.U., Levin, A., Levin, N., Locatelli, F., MacLeod, A., Vanholder, R. *et al.* (2004) The burden of kidney disease: improving global outcomes. *Kidney Int.*, **66**, 1310–1314.
41. Ache, B.W. and Young, J.M. (2005) Olfaction: diverse species, conserved principles. *Neuron*, **48**, 417–430.
42. Elliott, D.J. and Grellscheid, S.N. (2006) Alternative RNA splicing regulation in the testis. *Reproduction*, **132**, 811–819.
43. Hulbert, A.J. and Else, P.L. (2004) Basal metabolic rate: history, composition, regulation, and usefulness. *Physiol. Biochem. Zool.*, **77**, 869–876.
44. Gebauer, F. and Hentze, M.W. (2004) Molecular mechanisms of translational control. *Nat. Rev. Mol. Cell Biol.*, **5**, 827–835.
45. Kindler, S., Wang, H., Richter, D. and Tiedge, H. (2005) RNA transport and local control of translation. *Annu. Rev. Cell Dev. Biol.*, **21**, 223–245.
46. Richter, J.D. and Sonenberg, N. (2005) Regulation of cap-dependent translation by eIF4E inhibitory proteins. *Nature*, **433**, 477–480.
47. Mignone, F., Gissi, C., Liuni, S. and Pesole, G. (2002) Untranslated regions of mRNAs. *Genome Biol.*, **3**, REVIEWS0004.
48. Iacono, M., Mignone, F. and Pesole, G. (2005) uAUG and uORFs in human and rodent 5′ untranslated mRNAs. *Gene*, **349**, 97–105.
49. Churbanov, A., Rogozin, I.B., Babenko, V.N., Ali, H. and Koonin, E.V. (2005) Evolutionary conservation suggests a regulatory function of AUG triplets in 5′-UTRs of eukaryotic genes. *Nucleic Acids Res.*, **33**, 5512–5520.
50. Chamary, J.V., Parmley, J.L. and Hurst, L.D. (2006) Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nat. Rev. Genet.*, **7**, 98–108.
51. Nackley, A.G., Shabalina, S.A., Tchivileva, I.E., Satterfield, K., Korchynskiy, O., Makarov, S.S., Maixner, W. and Diatchenko, L. (2006) Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. *Science*, **314**, 1930–1933.
52. Kimchi-Sarfaty, C., Oh, J.M., Kim, I.W., Sauna, Z.E., Calcagno, A.M., Ambudkar, S.V. and Gottesman, M.M. (2007) A “silent” polymorphism in the MDR1 gene changes substrate specificity. *Science*, **315**, 525–528.
53. Bird, A.P. (1995) Gene number, noise reduction and biological complexity. *Trends Genet.*, **11**, 94–100.
54. Hurst, L.D. (1995) Evolutionary genetics. The silence of the genes. *Curr. Biol.*, **5**, 459–461.
55. Benhamouche, S., Decaens, T., Godard, C., Chambrey, R., Rickman, D.S., Moinard, C., Vasseur-Cognet, M., Kuo, C.J., Kahn, A. *et al.* (2006) Apc tumor suppressor gene is the ‘zonation-keeper’ of mouse liver. *Dev. Cell*, **10**, 759–770.
56. Braeuning, A., Ittrich, C., Kohle, C., Hailfinger, S., Bonin, M., Buchmann, A. and Schwarz, M. (2006) Differential gene expression in periportal and perivenous mouse hepatocytes. *FEBS J.*, **273**, 5051–5061.
57. Rangarajan, A. and Weinberg, R.A. (2003) Comparative biology of mouse versus human cells: modelling human cancer in mice. *Nat. Rev. Cancer*, **3**, 952–959.
58. Rangarajan, A., Hong, S.J., Gifford, A. and Weinberg, R.A. (2004) Species- and cell type-specific requirements for cellular transformation. *Cancer Cell*, **6**, 171–183.
59. Rubin, H. (2002) The disparity between human cell senescence in vitro and lifelong replication in vivo. *Nat. Biotechnol.*, **20**, 675–681.