BRIEFINGS IN BIOINFORMATICS. VOL 12. NO 5. 436–441
Advance Access published on 15 June 2011
doi:10.1093/bib/bbr031

When orthologs diverge between human and mouse

Walid H. Gharib and Marc Robinson-Rechavi

Submitted: 18th March 2011; Received (in revised form): 11th May 2011

Abstract

Despite the common assumption that orthologs usually share the same function, there have been various reports of divergence between orthologs, even among species as close as mammals. The comparison of mouse and human is of special interest, because mouse is often used as a model organism to understand human biology. We review the literature on evidence for divergence between human and mouse orthologous genes, and discuss it in the context of biomedical research.

Keywords: *orthology; expression divergence; alternative splicing; copy number variants; phenotypic divergence*

INTRODUCTION

The mouse *Mus musculus* is the most widely used model organism to understand human biology. Relative to other mammals, and many other vertebrates, mice have fast reproduction, short life spans, are not expensive, easy to handle and can be manipulated at the molecular level [1]. There are almost 400 000 publications in PubMed with 'mouse' (or 'mice' or 'murine') in the title, second only to human (~700 000 publications with 'human' in the title). In addition to sharing the mammalian body plan, human and mouse have a median of 78.5% amino acid sequence identity [2]. In a first approximation, it seems reasonable to expect genes to have conserved function between human and mouse, both normal and pathological. This expectation is usually applied to orthologs. The definition of orthology is formally based on evolutionary criteria, but is often taken to imply functional conservation (discussed in Refs [3, 4]), especially for one-to-one orthologs.

The assumption of conserved function between orthologs has been supported even between relatively distant species, by observations of conserved phenotypic effects when orthologs were subject to knock-in experiments [5, 6] or *in situ* [7, 8], clarifying the role of genes involved in human diseases. Yet there is also some evidence of differential phenotypic effects [9]. In this review, we consider some sources of variation of ortholog function between human and mouse, especially in the context of biomedical research. We do not consider other sources of human–mouse differences, such as the emergence of novel genes [10].

In the specific case of humans and mice, while both species are placental mammals and share many common anatomical features and physiological processes, there are also a number of biological differences, which should be expected to translate into differences between orthologous genes, especially considering a divergence time of ~100 Mya [11, 12]. Rodents are notably small, specialized for gnawing, and have a high rate of reproduction [13], unlike primates. *Mus musculus* has an average weight of 12–30 g, sexual maturity at 1.5 months and up to 10 litters per year [14]. Probably related to the differences in life history, mice genomes have evolved faster than those of primates [2, 15].

Here, we first provide a few examples of experimentally determined divergence between human

Corresponding author. Marc Robinson-Rechavi, Department of Ecology and Evolution, Biophore, Swiss Institute of Bioinformatics, Lausanne University, CH-1015 Lausanne, Switzerland. Tel: +41 21 692 4220; Fax: +41 21 692 4165; E-mail: marc.robinson-rechavi@unil.ch

Walid H. Gharib is a PhD student at the Department of Ecology and Evolution in the University of Lausanne working in the laboratory of Professor Marc Robinson-Rechavi. His main interest is the detection of natural selection in coding sequences.

Marc Robinson-Rechavi is associate Professor at the Department of Ecology and Evolution in the University of Lausanne, and Group Leader at the Swiss Institute of Bioinformatics. His main interest is in the evolution of animal genomes in the context of organismal function and development.

and mouse orthologs, to illustrate that the existence of such differences are not to be dismissed as simply mistakes in genomic studies. We try to relate these examples to knowledge which can be derived from genomic databases. Then we discuss the evidence from comparative large-scale studies, concerning the frequency of differences between human and mouse orthologs. Of note, the 'function' of a gene does not have an unambiguous definition, and we have tried here to stay as close as possible to the aspects which are relevant to the use of mice as biomedical model organisms. Moreover, given that this question has been explicitly raised relatively recently, we are aware that we are presenting a still very incomplete view, which we hope will be enriched by future comparative studies.

EXAMPLES OF DIVERGENCE IN GENE FUNCTION BETWEEN HUMAN AND MOUSE

TDP1 is a gene that participates in the repair of Topo I–DNA complexes. The intra-cellular expression localizations of *TDP1* orthologs in human and mouse have been determined to be in the cytoplasm and in the nucleus, respectively [16]. The mutation *TDP1 1478A>G* in humans is linked to SCAN1 disorder, characterized by 'ataxia, cerebellar atrophy, and peripheral neuropathy', whereas there is no clear phenotype for this mutation in this mouse ortholog [16]. There are no obvious differences in gene expression patterns (as reported in Bgee, Ref. [17]), nor evidence of positive selection on the primary sequence (as reported in Selectome, Ref. [18]) between human and mouse. The intracellular expression localization of *TDP1* in human and mouse thus seems to result in different phenotypes.

While the molecular basis of inflammation is mostly conserved among mammals, the role of the two selectins, P and E, differs between human and mouse. The human ortholog of mouse P-selectin has lost the standard mammalian regulatory pathway. Notably, human P-selectin is not responsive to TNF (Tumor necrosis factor), a major inflammatory factor, a difference with major effects on the rolling of leukocytes *in vivo*, and on the contribution to inflammation [19]. There also seems to be a decreased role of human P-selectin in contact hypersensitivity. As Liu *et al.* [19] conclude, their 'results underscore the need for caution in extrapolating the functions of P-selectin obtained in mice to humans, particularly in the many models where mediators are generated that activate NF-κB– and ATF-2–dependent genes'. Interestingly, P-selectin is often associated in the biomedical literature to thymus activity [20], but the evidence seems derived from mouse models. Transcriptome data (as reported in Bgee) support expression of P-selectin in the thymus in mouse, but not in human, so it is possible that this role also is not conserved between the orthologs.

LEFTY is a locus that includes two genes, *LEFTY1* and *LEFTY2*, which arose by independent duplications in rodents and in primates (thus, human *LEFTY1* and mouse *lefty1* are not one-to-one orthologs, despite the names). In both mouse and human, the *LEFTY* genes are involved in the establishment of asymmetry during development. There is some evidence for positive selection on *Lefty1* in mouse and rat (reported in Selectome based on TreeFam 7), and there is experimental evidence that the molecular function is carried out differently in human and mouse [21]. Notably, it seems that the asymmetric expression patterns in development are controlled differently in human and mouse [21]. Thus, similar global functions are carried out by orthologs, but with differences in the specifics of protein sequence and expression pattern. Interestingly, Yashiro *et al.* [21] point out that there are also many specific differences in anatomical asymmetry between human and mouse, which might be related to these differences in *LEFTY*/*Lefty* function.

LARGE-SCALE QUANTITATIVE EVIDENCE FOR DIVERGENCE

Expression divergence

The examples above show that divergence of function between human and mouse orthologs can be mediated by gene expression regulation. While the same level of mechanistic details cannot be provided in genomic studies, it is interesting in this context to evaluate the scale of expression divergence between human and mouse orthologs.

The study of the evolution of gene expression is hampered by the difficulty of distinguishing experimental noise from *bona fide* functional divergence. In a careful study comparing relative expression profiles between human and mouse orthologs, Liao and Zhang [22] reanalyzed the GNF dataset of human and mouse microarrays [23]. They found that after correcting for experimental variation, only 16% of orthologs between human and mouse had expression

profiles as divergent as random pairs. Housekeeping orthologous genes appear to diverge more in expression than tissue-specific genes [22, 24]. Conservation of expression patterns between human–mouse tissue-specific orthologs has been confirmed by an alternative experimental approach [25], but without any specific quantification of divergent orthologs.

Three points should be noted about these results. Firstly, even 16% of orthologs is clearly above the 5% accepted false positive rate of the randomization method, which indicates that changes in expression pattern between human and mouse are not very rare (as previously noted in Ref. [4]). Secondly, the other 84% of genes are more conserved than a random expectation, but might still diverge in functionally relevant ways. Thirdly, Liao and Zhang [22] and other related studies have mostly used the Pearson's correlation coefficient as a measure of gene expression conservation, whereas this is biased especially for housekeeping genes [26] (B. Piasecka *et al.*, unpublished data). Of note, an alternative measure, the 'Gene expression barcode' [27], which detects organ specific overexpression of genes, recovers also a good conservation of organ-specific expression between human and mouse orthologs; but a more detailed quantification is not provided.

Thus, it appears that the changes of expression pattern found in small-scale studies do not represent very rare evolutionary events, but rather that divergence by expression is a relatively common phenomenon between human and mouse orthologs.

Gene isoforms

Alternative splicing is very frequent in human and mouse genes. A methodological consequence is that, as gene orthology prediction is mostly based on sequence similarity, orthologous genes can be erroneously inferred by grouping the wrong gene isoforms, which might have dissimilar functions. From a more fundamental perspective, many differences in splicing patterns have been reported between human and mouse orthologs [28, 29]. If a significant proportion of these splice forms have functional roles, then this provides a potential path for functional divergence between the orthologs.

In one study, >11% of human-mouse alternative cassette exons were found to be subject to exon skipping in one organism, yet consecutively spliced in the other [29]. Non-conserved exons between human and mouse are mostly found outside the coding sequences, suggesting that when non-conserved exons are localized within coding sequences, it might be due to species-specific functional effects [30]. In a more recent study, orthology at the gene level was distinguished from orthology at the transcript level (conservation of exon structure) [31]. Even using relaxed criteria for transcript orthology, 13% of human-mouse orthologous genes have non-orthologous transcripts [31]. This level of divergence, if it is confirmed, is of the same scale as the divergence observed at the expression level. The gain of splice forms has been shown to be a continuous process in human and mouse evolution [32], which certainly provides material for functional divergence.

The phenomenon of alternative promoters regulating different gene isoforms is related both to changes in expression and to changes in transcript structure. Sequence comparison between human-mouse alternative promoters shows not only rather low sequence conservation during evolution, but especially that the subsets of conserved and non-conserved alternative promoters can be distinguished clearly [33]. For example, the human *ACACB* gene has two alternative promoters. Only one of those promoters is highly conserved in rodents, while both promoters actively regulate the skeletal muscle *ACACB* gene function in humans [34].

Differences in gene copy number

Approximately 9% of orthologs are duplicated either in human, or mouse, or both independently, as was the case for *LEFTY* (estimated as the proportion of non one-to-one orthologs among orthologs in Ensembl Compara [35]). In most of these cases, identifying which ortholog is expected to share the function between species is difficult. Moreover, positive selection appears to affect strongly these lineage-specific duplicates [36], which might imply changes in biochemical function.

Not only are genes duplicated in the human and mouse lineages, but copy number variations (CNVs) are widely observed among human and mouse genomes. These can result from local alterations, such as duplications, deletions, translocations or inversions. In humans, CNVs have been shown to be medically relevant, e.g. linked to the reaction to cancer treatment [37]. In mice, CNVs have a significant impact on the measure of gene expression [38]. CNVs appear to affect a biased subset of the genome. Human CNVs are enriched in protein coding genes with high synonymous and non-synonymous

divergence to their mouse orthologs [39]. These genes are associated with olfaction, immunity and protein secretion. Mouse CNVs, on the other hand, seem to have decreased amino acid sequence divergence [39].

These variations, and the differences in the genes affected, render the definition of one-to-one orthology more complex between human and mouse. It is possible to have one-to-one orthologs for some individuals, but not for others. If the copy variants have differences in function (e.g. different expression levels), then orthologs might have functional conservation in some individuals but not others. The study of CNVs is mostly recent, and the functional and medical consequences remain to be elucidated in more detail. But we can already suggest that, parallel to the recently introduced concept of 'splicing orthology' [31], we might need to define a concept of 'copy number orthology', restricted to orthologs with the same number of copies in both organisms. Consistent with the original evolutionary definition of orthology, it would probably be best to restrict this further to the most probable ancestral copy number, whose function was probably conserved.

Phenotypic divergence

Gene–phenotype relations can be complex, and different between species. For example, the alteration of GSK3 perturbs nutrient and stress signaling in yeast, anteroposterior patterning and segmentation in insects, dorsoventral patterning in frogs and craniofacial morphogenesis in mice [40, 41]. Obviously, predicting its phenotypic implication in human is not straightforward. Therefore, the relation of gene function to phenotype prediction between organisms is a difficult task.

Several cases of single genes linked to human diseases show apparently normal mouse phenotypes when experimentally manipulated. For example, *BCL10*, *SGCA* and *PKLR* are linked to different human diseases when mutated (from OMIM [42]), whereas they present no phenotypic effect in mouse. This indicates that there are several pathogenic human mutations that have become fixed in mouse evolution [43].

Liao and Zhang [44] showed that >20% of human essential genes are mouse non-essential, and that the rate of evolution of those 20% is significantly higher than for the human–mouse essential. Gene essentiality is an extreme case of phenotypic impact, yet orthologous human and mouse essential genes can result in different phenotypes. For example, *Adamts2*, *Acox1* and *Fancg* are essential for human [45, 46] and mouse [47, 48] but show different phenotypic effect when mutated (discussed in Ref. [44]). This finding shows a high rate of functional divergence between human-mouse orthologs. Recently, a review of 'phenologs', phenotypes associated to orthologous genes, showed that different phenotypes might correspond to deeper functional homology [49]. Such research might help to identify genes implicated in human disease, despite phenotypic divergence between orthologs.

CONCLUSION

This review is per force quite limited, because a systematic exploration of functional differences between orthologs has only come on the agenda of biological research recently [4, 50]. We believe that both small-scale and large-scale studies provide evidence that functional divergence between human and mouse orthologs, although a minority phenomenon, still affects a significant proportion of genes. Divergence of gene expression, of alternative splicing, and of mutant phenotypes, each affect of the order of 10–20% of ortholog pairs, under conservative estimates. If these and other different processes affect different genes, then it might be a majority of genes which are affected. But even if the same genes differ in expression pattern, splicing, etc., then having ~15% of human-mouse orthologs with strong differences will affect many pathways and biological processes of interest. We look forward to future explorations of this topic, preferably combining high-quality experimental data and large-scale approaches.

Key Points

- Significant divergence in expression between human and mouse orthologs.
- High divergence of alternative splicing between human and mouse orthologs.
- Fast evolution of genes with copy number variants in human.
- Significant divergence in gene-phenotype relations between human mouse orthologs.
- This divergence is relevant to biomedical research using mouse.

FUNDING

This work was supported by Etat de Vaud. Funding for open access charge: Swiss Institute of Bioinformatics.

References

1. Elso C, Lu X, Morrison S, *et al*. Germline translocations in mice: unique tools for analyzing gene function and long-distance regulatory mechanisms. *J Natl Cancer Inst Monogr* 2008;**39**:91–5.
2. Waterston RH, Lindblad-Toh K, Birney E, *et al*. Initial sequencing and comparative analysis of the mouse genome. *Nature* 2002;**420**:520–62.
3. Gabaldon T. Large-scale assignment of orthology: back to phylogenetics? *Genome Biol* 2008;**9**:235.
4. Studer RA, Robinson-Rechavi M. How confident can we be that orthologs are similar, but paralogs differ? *Trends Genet* 2009;**25**:210–6.
5. Franco AT, Malaguarnera R, Refetoff S, *et al*. Thyrotrophin receptor signaling dependence of Braf-induced thyroid tumor initiation in mice. *Proc Natl Acad Sci USA* 2011;**108**: 1615–20.
6. Goh AM, Coffill CR, Lane DP. The role of mutant p53 in human cancer. *J Pathol* 2011;**223**:116–26.
7. Quiring R, Walldorf U, Kloter U, *et al*. Homology of the eyeless gene of Drosophila to the Small eye gene in mice and Aniridia in humans. *Science* 1994;**265**:785–9.
8. Nagao T, Leuzinger S, Acampora D, *et al*. Developmental rescue of Drosophila cephalic defects by the human Otx genes. *Proc Natl Acad Sci USA* 1998;**95**:3737–42.
9. Liu H, Chen CH, Espinoza-Lewis RA, *et al*. Functional redundancy between human SHOX and mouse Shox2 in the regulation of Sinoatrial node formation and pacemaking function. *J Biol Chem* 2011;**286**:17029–38.
10. Knowles DG, McLysaght A. Recent de novo origin of human protein-coding genes. *Genome Res* 2009;**19**:1752–9.
11. Hedges SB, Dudley J, Kumar S. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* 2006;**22**:2971–2.
12. Benton MJ, Donoghue PC. Paleontological evidence to date the tree of life. *Mol Biol Evol* 2007;**24**:26–53.
13. Hartenberger JL. The order Rodentia: major questions on their evolutionary origin, relationships and suprafamilial systematics. In: Luckett WP, Hartenberger JL, (eds). *Evolutionary Relationships Among Rodents: a Multidisciplinary Analysis*. New York: Plenum, 1985;1–33.
14. Nowak RM, Walker EP. *Mammals of the World*. Baltimore (USA): Johns Hopkins University Press, 1991.
15. Nikolaev SI, Montoya-Burgos JI, Popadin K, *et al*. Life-history traits drive the evolutionary rates of mammalian coding and noncoding genomic elements. *Proc Natl Acad Sci USA* 2007;**104**:20443–8.
16. Hirano R, Interthal H, Huang C, *et al*. Spinocerebellar ataxia with axonal neuropathy: consequence of a Tdp1 recessive neomorphic mutation? *EMBO J* 2007;**26**: 4732–43.
17. Bastian F, Parmentier G, Roux J, *et al*. Bgee: integrating and comparing heterogeneous transcriptome data among species. *Data Integr Life Sci* 2008;124–31.
18. Proux E, Studer RA, Moretti S, *et al*. Selectome: a database of positive selection. *Nucleic Acids Res* 2009;**37**(Database issue):D404–7.
19. Liu Z, Miner JJ, Yago T, *et al*. Differential regulation of human and murine P-selectin expression and function in vivo. *J Exp Med* 2010;**207**:2975–87.
20. Gossens K, Naus S, Corbel SY, *et al*. Thymic progenitor homing and lymphocyte homeostasis are linked via S1P-controlled expression of thymic P-selectin/CCL25. *J Exp Med* 2009;**206**:761–78.
21. Yashiro K, Saijoh Y, Sakuma R, *et al*. Distinct transcriptional regulation and phylogenetic divergence of human LEFTY genes. *Genes Cells* 2000;**5**:343–57.
22. Liao BY, Zhang J. Evolutionary conservation of expression profiles between human and mouse orthologous genes. *Mol Biol Evol* 2006;**23**:530–40.
23. Su AI, Wiltshire T, Batalov S, *et al*. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci USA* 2004;**101**:6062–7.
24. Liao BY, Zhang J. Low rates of expression profile divergence in highly expressed genes and tissue-specific genes during mammalian evolution. *Mol Biol Evol* 2006;**23**: 1119–28.
25. Xing Y, Ouyang Z, Kapur K, *et al*. Assessing the Conservation of Mammalian Gene Expression Using High-Density Exon Arrays. *Mol Biol Evol* 2007;**24**:1283–5.
26. Pereira V, Waxman D, Eyre-Walker A. A problem with the correlation coefficient as a measure of gene expression divergence. *Genetics* 2009;**183**:1597–600.
27. McCall MN, Uppal K, Jaffee HA, *et al*. The gene expression barcode: leveraging public data repositories to begin cataloging the human and murine transcriptomes. *Nucleic Acids Res* 2011;**39**:D1011–5.
28. Nurtdinov RN, Artamonova II, Mironov AA, *et al*. Low conservation of alternative splicing patterns in the human and mouse genomes. *Hum Mol Genet* 2003;**12**: 1313–20.
29. Pan Q, Bakowski MA, Morris Q, *et al*. Alternative splicing of conserved exons is frequently species-specific in human and mouse. *Trends Genet* 2005;**21**:73–7.
30. Takeda J, Suzuki Y, Sakate R, *et al*. Low conservation and species-specific evolution of alternative splicing in humans and mice: comparative genomics analysis using well-annotated full-length cDNAs. *Nucleic Acids Res* 2008;**36**: 6386–95.
31. Zambelli F, Pavesi G, Gissi C, *et al*. Assessment of orthologous splicing isoforms in human and mouse orthologous genes. *BMC Genomics* 2010;**11**:534.
32. Roux J, Robinson-Rechavi M. Age-dependent gain of alternative splice forms and biased duplication explain the relation between splicing and duplication. *Genome Res* 2011;**21**:357–63.
33. Tsuritani K, Irie T, Yamashita R, *et al*. Distinct class of putative "non-conserved" promoters in humans: comparative studies of alternative promoters of human and mouse genes. *Genome Res* 2007;**17**:1005–14.
34. Oh SY, Lee MY, Kim JM, *et al*. Alternative usages of multiple promoters of the acetyl-CoA carboxylase beta gene are related to differential transcriptional regulation in human and rodent tissues. *J Biol Chem* 2005;**280**:5909–16.
35. Flicek P, Amode MR, Barrell D, *et al*. Ensembl 2011. *Nucleic Acids Res* 2010;**39**:D800–6.
36. Han MV, Demuth JP, McGrath CL, *et al*. Adaptive evolution of young gene duplicates in mammals. *Genome Res* 2009;**19**:859–67.
37. Cappuzzo F. Erlotinib in gliomas: should selection be based on EGFR and Akt analyses? *J Natl Cancer Inst* 2005;**97**: 868–9.

38. Henrichsen CN, Vinckenbosch N, Zollner S, *et al*. Segmental copy number variation shapes tissue transcriptomes. *Nat Genet* 2009;**41**:424–9.

39. Nguyen DQ, Webber C, Ponting CP. Bias of selection on human copy-number variants. *PLoS Genet* 2006;**2**:e20.

40. Kassir Y, Rubin-Bejerano I, Mandel-Gutfreund Y. The Saccharomyces cerevisiae GSK-3 beta homologs. *Curr Drug Target* 2006;**7**:1455–65.

41. Liu KJ, Arron JR, Stankunas K, *et al*. Chemical rescue of cleft palate and midline defects in conditional GSK-3beta mice. *Nature* 2007;**446**:79–82.

42. Amberger J, Bocchini CA, Scott AF, *et al*. McKusick's Online Mendelian Inheritance in Man (OMIM). *Nucleic Acids Res* 2009;**37**:D793–6.

43. Gao L, Zhang J. Why are some human disease-associated mutations fixed in mice? *Trends Genet* 2003;**19**:678–81.

44. Liao BY, Zhang J. Null mutations in human and mouse orthologs frequently result in different phenotypes. *Proc Natl Acad Sci USA* 2008;**105**:6987–92.

45. Jimenez-Sanchez G, Childs B, Valle D. Human disease genes. *Nature* 2001;**409**:853–5.

46. Suzuki Y, Iai M, Kamei A, *et al*. Peroxisomal acyl CoA oxidase deficiency. *J Pediatrics* 2002;**140**:128–30.

47. Fan CY, Pan J, Chu R, *et al*. Hepatocellular and hepatic peroxisomal alterations in mice with a disrupted peroxisomal fatty acyl-coenzyme A oxidase gene. *J Biol Chem* 1996;**271**:24698–710.

48. Yang Y, Kuang Y, Montes De Oca R, *et al*. Targeted disruption of the murine Fanconi anemia gene, Fancg/Xrcc9. *Blood* 2001;**98**:3435–40.

49. McGary KL, Park TJ, Woods JO, *et al*. Systematic discovery of nonobvious human disease models through orthologous phenotypes. *Proc Natl Acad Sci USA* 2010;**107**:6544–9.

50. Markov G, Lecointre G, Demeneix B, *et al*. The 'street light syndrome', or how protein taxonomy can bias experimental manipulations. *Bioessays* 2008;**30**:349–57.