



Published in final edited form as:

Science. 2013 November 29; 342(6162): 1100–1104. doi:10.1126/science.1242379.

Primate Transcript and Protein Expression Levels Evolve under Compensatory Selection Pressures

Zia Khan¹, Michael J. Ford², Darren A. Cusanovich¹, Amy Mitrano¹, Jonathan K. Pritchard^{1,3,4}, and Yoav Gilad¹

¹Department of Human Genetics, University of Chicago, Chicago, IL, 60637, USA

²MS Bioworks, LLC, 3950 Varsity Drive, Ann Arbor, MI 48108, USA

³Howard Hughes Medical Institute, University of Chicago, Chicago, IL, 60637, USA

Abstract

Changes in gene regulation have likely played an important role in the evolution of primates. Differences in mRNA expression levels across primates have often been documented, however, it is not yet known to what extent measurements of divergence in mRNA levels reflect divergence in protein expression levels, which are probably more important in determining phenotypic differences. We used high-resolution, quantitative mass spectrometry to collect protein expression measurements from human, chimpanzee, and rhesus macaque lymphoblastoid cell lines (LCLs) and compared them to transcript expression data from the same samples. We found dozens of genes with significant expression differences between species at the mRNA level yet little or no difference in protein expression. Overall, our data suggest that protein expression levels evolve under stronger evolutionary constraint than mRNA levels.

Measurements of mRNA levels have revealed substantial differences across primate transcriptomes (1–3) and have led to the identification of putatively adaptive changes in transcript expression (4). Traditionally, measurements of divergence in mRNA levels are assumed to be good proxies for divergence in protein levels. However, there are numerous mechanisms by which protein expression may be regulated independently of mRNA levels (5, 6). If transcript and protein expression levels are often uncoupled, mRNA levels may evolve under reduced constraint as changes at the transcript level could be buffered or compensated for at the protein level (7–9). To date, however, genome-wide studies of protein expression in primates have been limited (10, 11).

Correspondence to: zia@uchicago.edu (Z.K.); pritch@stanford.edu (J.K.P); gilad@uchicago.edu (Y.G.).

⁴Current Address: Departments of Genetics and Biology, Stanford University, Stanford, CA, 94305, USA

Author Contributions: ZK, JKP, and YG conceived of the study and designed it; MF acquired the mass spectrometry data; ZK conducted the computational analyses with input from DAC, JKP, and YG. MF acknowledges assistance from colleagues at MS Bioworks LLC. AM cultured cells and prepared protein samples. ZK, JKP, and YG wrote the paper with contributions from all authors.

Data availability: RNA-seq data have been deposited to the Gene Expression Omnibus (GSE49682). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (PXD000419).

Conflict of interest statement: JKP is on the scientific advisory boards for 23andMe and DNANexus with stock options.

We collected a comparative proteomic data set with SILAC (stable isotope labeling by amino acids in cell culture (12)). Using high-resolution, quantitative mass spectrometry (13), we measured peptide expression levels in LCLs from 5 human, 5 chimpanzee, and 5 rhesus macaque individuals (fig. S1; table S1). We analyzed the peptide expression data in the context of orthologous gene models (14) to obtain comparative protein expression measurements from all three species (table S2). We obtained measurements for 4,157 proteins in at least three human and three chimpanzee individuals and 3,688 proteins were quantified in at least three individuals from all three species (table S2; fig. S1). We also collected RNA-seq data from the same samples and estimated mRNA expression levels using reads that map to orthologous exons (fig. S1, table S3). We thus obtained both mRNA and protein expression levels for 3,390 genes in at least 3 individuals from each of the three species (fig. S2; table S4).

Focusing on differences between human and chimpanzee, we classified 1,151 genes as differentially expressed (DE) between species at the mRNA and/or protein expression levels, independently (LR test, FDR = 1%, table S5). The number of inter-species DE genes at the mRNA level was higher (815) than the number of DE proteins (571; fig. 1A, 1B). By accounting for incomplete power to detect inter-species differences in gene expression (15) we estimated that 266 genes (33%) are DE between humans and chimpanzees at the mRNA level but not at the protein level. We observed a similar pattern for comparisons that include the rhesus macaque data (table S5).

These observations may reflect a slower rate of divergence in protein levels or higher levels of within-species variation in protein than mRNA expression levels. To distinguish between these possibilities we compared estimates mRNA and protein divergence (fig. 1C). Among genes whose inter-species mRNA and protein divergence differ (FDR = 1%), inter-species variation at the mRNA level is higher than at the protein level much more often than the reverse pattern (fig. 1D). This indicates that protein expression levels might evolve under greater evolutionary constraint than mRNA expression levels.

The accuracy of SILAC has been established by biochemical means (16); yet, it is difficult to exclude all possible technical explanations for our observations. We thus conducted a large number of quality control analyses. First, we observed that the consistency of protein measurements is at least as good as that for mRNA (fig. S3). Additionally, biological variation associated with the mRNA and protein measurements, regardless of species, is comparable (fig. S4). We then proceeded to demonstrate that the protein measurements have a higher dynamic range than the mRNA measurements, and hence, our results are conservative with respect to this property of the data (fig. S5). We also confirmed that the observation of lower divergence of protein levels relative to mRNA levels could not be explained by insufficient quantification of protein expression (fig. S6), and is robust to differences in the approach used to summarize multiple peptide measurements into a single estimate of protein expression level (fig. S7). Finally, we established that our observations are robust by restricting our analysis only to the subset of genes with similar RNA-seq read depth across orthologous exons; only to genes with low inter-individual variation both at the mRNA and protein levels; only to genes whose protein and mRNA levels were measured in all 5 individuals from each species; only to genes whose protein expression levels were

measured by two peptides or more; by excluding the top 2% of most highly expressed genes at the transcript level; and by excluding all genes with RNA-seq RPKM less than 1. These analyses all resulted in consistent results (fig. S8–S13).

To gain further insight into the differences in evolutionary pressures acting on mRNA and protein expression in primates, we used data from all three species to identify genes whose regulation might have evolved under natural selection. We applied an empirical approach to identify expression patterns that are consistent with the action of stabilizing or directional selection on gene regulation (2, 17). The rationale of our approach is similar to that used in empirical scans of selection on nucleotide sequence data (18). We scanned for expression patterns on the basis of our expectations given different evolutionary scenarios. For example, patterns of low variation in expression levels, both within and between species, are consistent with a scenario of stabilizing selection on gene regulation (fig. S14A). In turn, a lineage-specific shift in expression level associated with high within-species variation is consistent with relaxation of evolutionary constraint (fig. S14B). A lineage-specific shift in expression level coupled with low within-species variation is consistent with directional selection acting on gene regulation in a particular lineage (fig. S14C).

We considered the transcript and protein comparative expression data independently. Among the 300 genes with the least varied protein expression levels within and between species, consistent with the action of stabilizing selection, we found enrichment of genes involved in conserved cellular processes including translation, splicing, and transcriptional regulation (table S6). Compared to genes not in this set (fig. 2), these 300 genes also evolve under stronger evolutionary constraint at the amino acid level (Wilcoxon rank sum, $P < 10^{-9}$), have higher expression levels ($P < 10^{-5}$), have shorter 3'UTRs ($P < 10^{-5}$), have more reported protein-protein interactions ($P < 10^{-15}$), and are expressed in more tissues ($P < 10^{-8}$). We found that these properties are also associated with the 300 genes with the least varied mRNA levels: stronger evolutionary constraint on amino acid sequence ($P < 0.003$); larger number of protein-protein interactions ($P < 10^{-4}$); and higher absolute expression levels ($P < 0.02$) as has been noted (1, 19). Yet, interestingly, all of these associations are stronger when genes are ranked by conservation of protein expression than when ranked by conservation of mRNA expression. Our observations are robust to arbitrary choices in cutoffs (fig. S15) and suggest that these regulatory and sequence properties are more coupled to protein expression levels.

We next focused on lineage-specific differences in gene regulation. We found that a subset of genes with lineage-specific expression differences were also associated with a lineage-specific increase in within-species variation in expression levels; this pattern is consistent with lineage-specific relaxation of evolutionary constraint on gene regulation. We classified 85 genes (one-sided F-test; $P < 0.05$) with expression patterns consistent with either human- or chimpanzee-specific relaxation of constraint on transcript expression levels but only 20 genes with regulatory patterns consistent with relaxation of constraint on protein expression levels. This observation supports that protein levels might evolve under greater evolutionary constraint than mRNA levels. Lineage-specific shifts in protein expression levels might also be associated with low within-species variation, consistent with directional selection on gene

regulation. We classified 196 and 161 such patterns in human or chimpanzee, respectively (table S7).

We then considered the protein and mRNA data jointly. As expected, in most cases, the patterns of mRNA and protein expression levels are consistent with the same evolutionary scenario. We found a few genes whose mRNA expression patterns are consistent with the action of stabilizing selection, while the patterns of their protein expression levels are consistent with lineage-specific directional selection in either human (14 genes, fig. 3A) or chimpanzee (10 genes). These patterns can potentially be explained by lineage-specific changes that specifically affect post-transcriptional regulation. Interestingly, we also identified 40 and 20 genes whose mRNA expression patterns consistent with the action of lineage-specific directional selection in human or chimpanzee, respectively, yet their protein levels are consistent with the action of stabilizing selection (fig. 3B). These observations may indicate that protein expression levels of these genes are buffered against changes in mRNA levels (20) or that these genes are evolving under compensatory selection pressures. Genes whose mRNA and protein expression levels are consistent with this pattern have slightly longer 5'UTRs (one-sided Wilcoxon rank sum; $P < 0.03$), a greater number of known ubiquitination sites ($P < 0.0002$), and, among those with a human-specific decrease in mRNA levels, more phosphorylation sites ($P < 0.006$). Put together, these are all properties typically common to genes that evolve under strong evolutionary constraint.

In summary, our data suggest that protein expression levels evolve under greater evolutionary constraint than mRNA levels. It seems likely that for many genes, evolutionary changes in mRNA levels may be effectively neutral, if buffered or compensated for at the protein level. As protein levels are presumably more relevant to understanding how the genotype give rise to the phenotype than mRNA levels of protein-coding genes, insight into the interplay between transcriptional and post-transcriptional regulatory differences may greatly advance our understanding of human-specific adaptations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank members of our labs for helpful discussions. Funded by NIH grant GM077959 to YG and by HHMI funds to JKP. ZK is supported by NRSA F32HG006972.

References and Notes

1. Khaitovich P, et al. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science*. 2005 Sep 16;309:1850. [PubMed: 16141373]
2. Gilad Y, Oshlack A, Smyth GK, Speed TP, White KP. Expression profiling in primates reveals a rapid evolution of human transcription factors. *Nature*. 2006 Mar 9;440:242. [PubMed: 16525476]
3. Brawand D, et al. The evolution of gene expression levels in mammalian organs. *Nature*. 2011; 478:343. [PubMed: 22012392]
4. Romero IG, Ruvinsky I, Gilad Y. Comparative studies of gene expression and the evolution of gene regulation. *Nat Rev Genet*. 2012; 13:505. [PubMed: 22705669]

5. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet.* 2012; 13:227. [PubMed: 22411467]
6. Cox B, et al. Integrated proteomic and transcriptomic profiling of mouse lung development and Nmyc target genes. *Molecular systems biology.* 2007; 3
7. Laurent J. Protein abundances are more conserved than mRNA abundances across diverse taxa. *Proteomics.* 2010; 10:4209. [PubMed: 21089048]
8. Schrimpf SP, et al. Comparative Functional Analysis of the *Caenorhabditis elegans* and *Drosophila melanogaster* Proteomes. *PLoS Biol.* 2009; 7:e1000048.
9. Wu L, et al. Variation and genetic control of protein abundance in humans. *Nature.* 2013; 499:79. [PubMed: 23676674]
10. Fu N, et al. Comparison of Protein and mRNA Expression Evolution in Humans and Chimpanzees. *PLoS ONE.* 2007; 2:e216. [PubMed: 17299596]
11. Enard W, et al. Intra- and Interspecific Variation in Primate Gene Expression Patterns. *Science.* 2002; 296:340. [PubMed: 11951044]
12. Ong S-E, et al. Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Molecular & Cellular Proteomics.* 2002; 1:376. [PubMed: 12118079]
13. Mann M, Kelleher NL. Precision proteomics: The case for high resolution and high mass accuracy. *Proceedings of the National Academy of Sciences.* 2008; 105:18132.
14. Blekhman R, Marioni JC, Zumbo P, Stephens M, Gilad Y. Sex-specific and lineage-specific alternative splicing in primates. *Genome Res.* 2010 Feb.20:180. [PubMed: 20009012]
15. Storey JD, Taylor JE, Siegmund D. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology).* 2004; 66:187.
16. Hanke S, Besir H, Oesterhelt D, Mann M. Absolute SILAC for Accurate Quantitation of Proteins in Complex Mixtures Down to the Attomole Level. *Journal of Proteome Research.* 2008; 7:1118. (2008/03/01. [PubMed: 18271523]
17. Meiklejohn CD, Parsch J, Ranz JM, Hartl DL. Rapid evolution of male-biased gene expression in *Drosophila*. *Proceedings of the National Academy of Sciences.* 2003; 100:9894.
18. Akey JM. Constructing genomic maps of positive selection in humans: where do we go from here? *Genome Res.* 2009; 19:711. [PubMed: 19411596]
19. Lemos B, Bettencourt BR, Meiklejohn CD, Hartl DL. Evolution of proteins and gene expression levels are coupled in *Drosophila* and are independently associated with mRNA abundance, protein length, and number of protein-protein interactions. *Mol Biol Evol.* 2005 May.22:1345. [PubMed: 15746013]
20. Rutherford SL. From genotype to phenotype: buffering mechanisms and the storage of genetic information. *BioEssays.* 2000; 22:1095. [PubMed: 11084625]

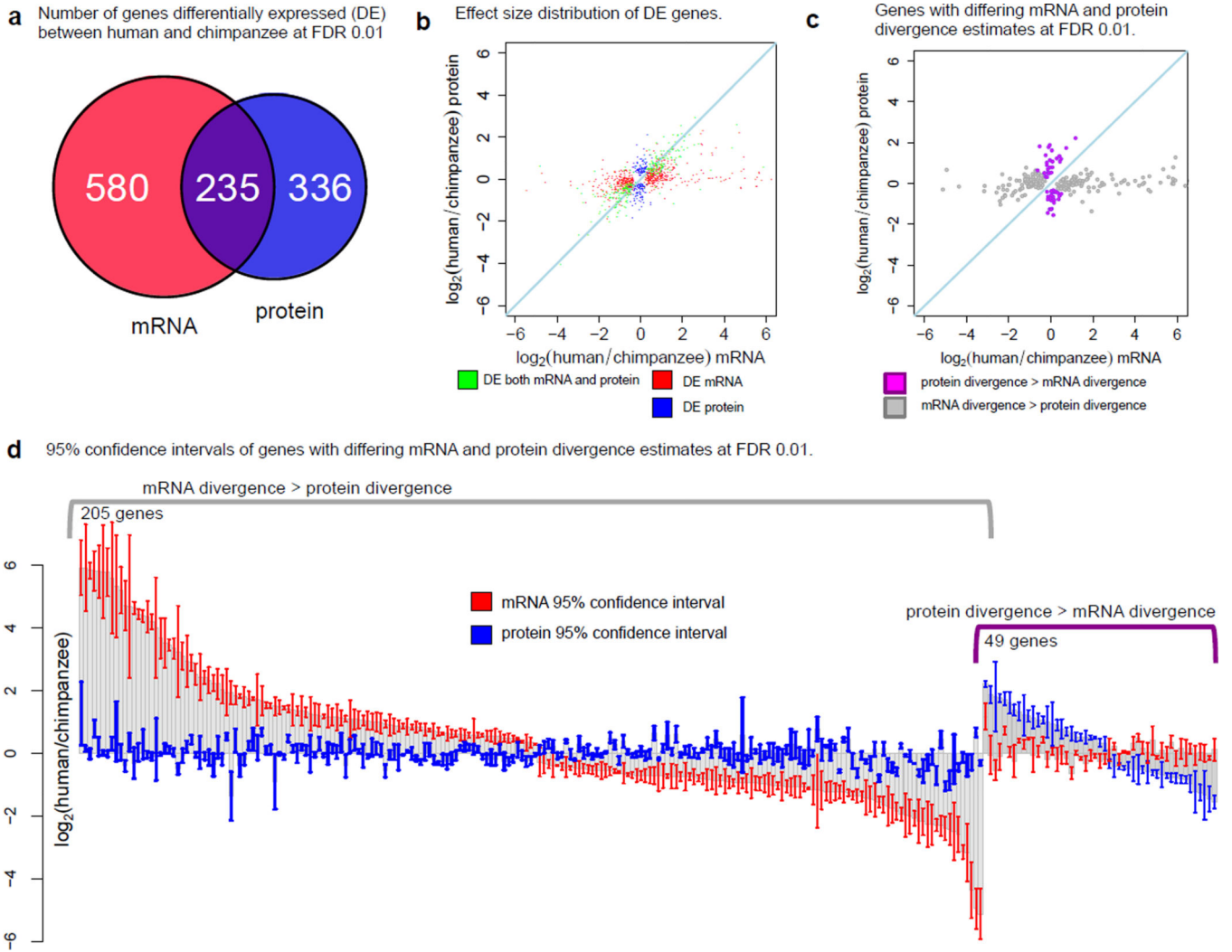


Fig. 1.

Protein expression levels evolve under greater evolutionary constraint than mRNA expression levels. (a) A Venn diagram of the numbers of mRNAs (red) and proteins (blue) classified as differentially expressed (DE). (b) Mean effect size of the inter-species difference in expression for genes classified as DE as mRNA-only, protein-only, or both. Each point corresponds to a single DE gene. (c) Scatterplot of median mRNA and protein divergence of genes where estimates of mRNA and protein divergence between human and chimpanzee differed significantly (FDR = 1%). (d) 95% confidence intervals around estimates of mean mRNA and protein divergence of genes in (c).

Differing properties of genes with least varied intra- and inter-species mRNA and protein levels.

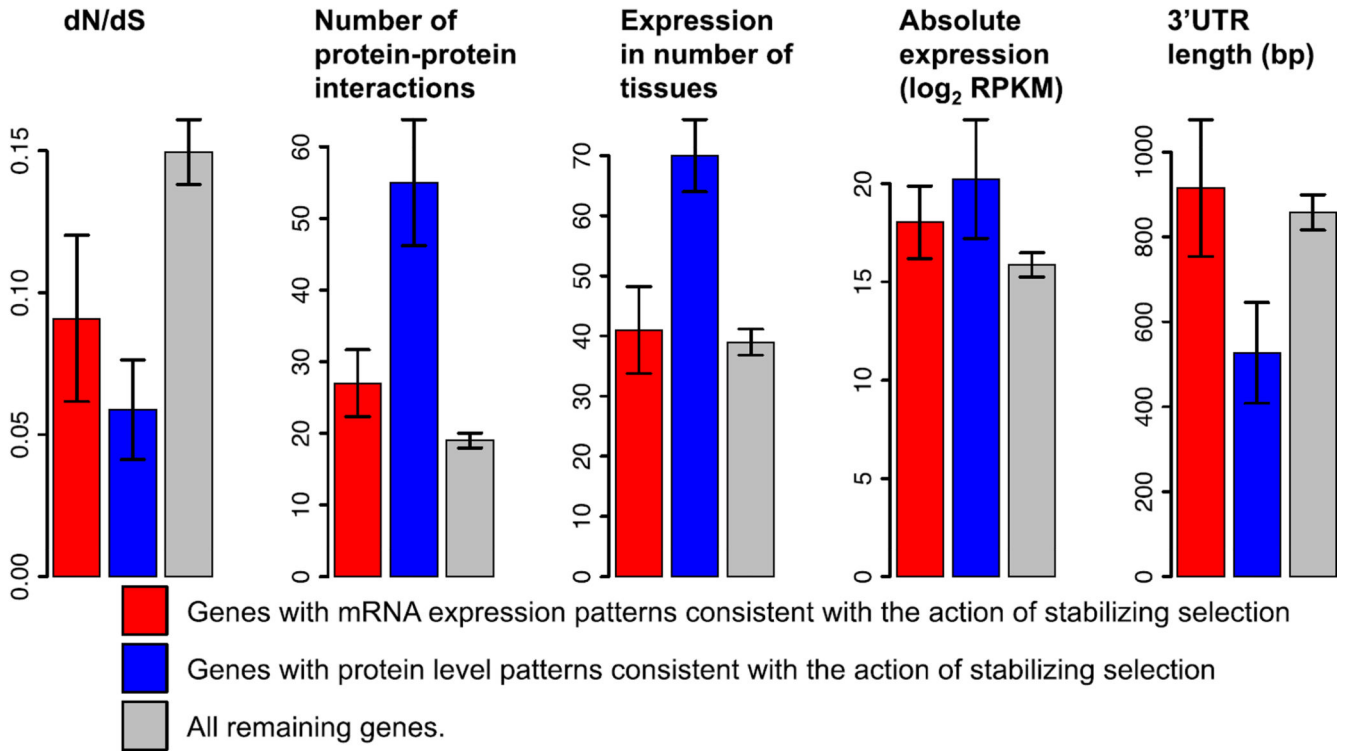


Fig. 2.

Properties of genes whose protein and mRNA expression levels are inferred to have evolved under stabilizing selection. Error bars represent the 95% confidence intervals around the mean. Data are shown for the top 300 genes with the least varied mRNA (red) or protein (blue) expression levels between and within species. Gray bars correspond to all other genes.

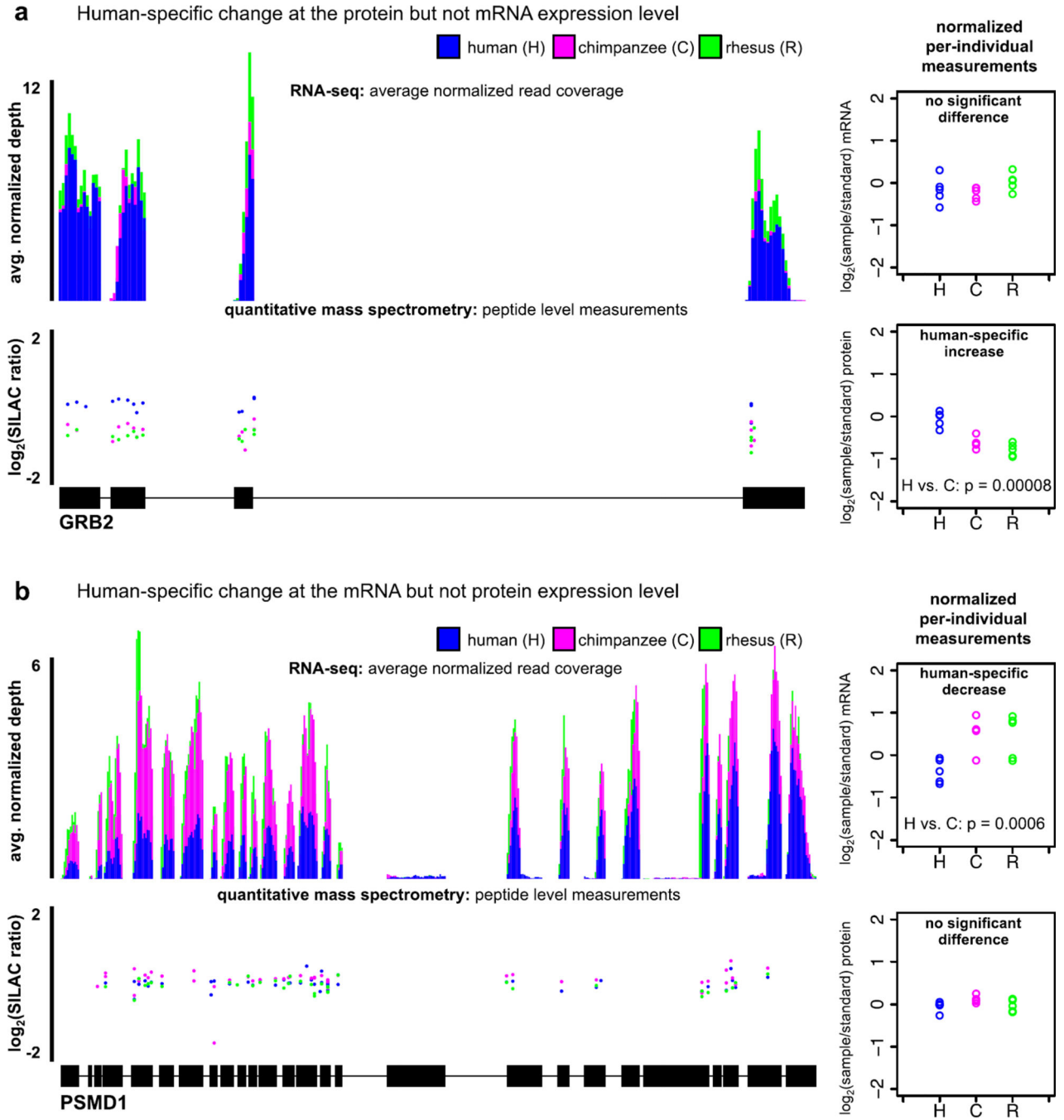


Fig. 3.

Examples of genes whose mRNA and protein expression levels are consistent with different evolutionary scenarios. (a) A gene whose mRNA and protein expression levels are consistent with a lineage-specific change in posttranscriptional regulation. (b) A gene whose inter-species mRNA levels are consistent with buffering or compensation at the protein expression level. In both cases, RNA-seq coverage is standardized to per million mapped reads and averaged across all 5 individuals. Protein measurements are plotted at the starting genomic position of the peptides. The plots on the right are of mRNA and protein expression levels from all individuals, normalized relative to the internal standard cell line.