Application of Quantitative Proteomics and RNA-Seq to Study the Evolution of Primate Transcipt and Protein Expression Levels

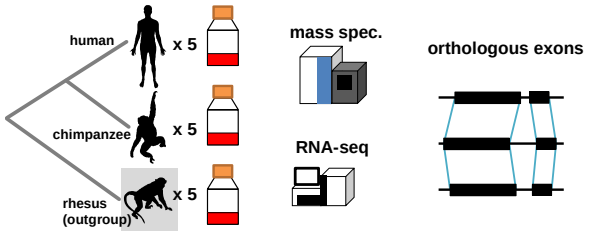
Zia Khan[1], Michael Ford[2], Darren Cusanovich[3], Amy Mitrano[3], Jonathan Pritchard[4], Yoav Gilad[3]

1. University of Maryland, College Park, MD. 2. MS Bioworks, Ann Arbor, MI 48108, 3. University of Chicago. IL, 4. Stanford University, Stanford, CA.

1. Abstract and introduction

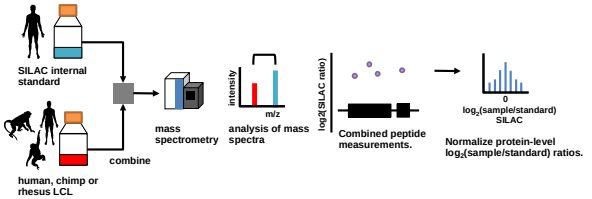
Variation in gene regulation is thought to have played an important role in the evolution of primates, and many studies have documented differences in mRNA expression levels across primate species. However, it is not yet known to what extent measurements of divergence in mRNA levels reflect divergence in protein expression levels, which are more directly tied to phenotypic differences. To address this question, we used high-resolution, quantitative mass spectrometry to collect thousands of protein expression measurements from human, chimpanzee, and rhesus macaque lymphoblastoid cell lines (LCLs). We also used RNA sequencing to collect transcript expression data from the same samples. Considering the two datasets jointly, we found more inter-species divergence at the mRNA level than at the protein level. Remarkably, 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 a much stronger evolutionary constraint on protein expression levels than on mRNA levels. We conclude that inter-species mRNA expression differences often may not have functional consequences due to either buffering or compensatory changes in post-transcriptional or post-translational regulation.

2. Study Design



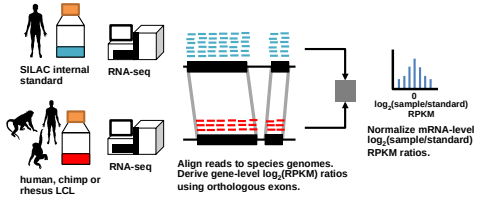
- We collected a large-scale data set where mRNA levels were measured by RNA-seq and protein levels were quantified by high-resolution quantitative mass spectrometry from lymphoblastoid cell lines derived from five genetically distinct human, chimpanzee, and rhesus individuals.
- Data was analyzed in the context of orthologous exonic regions between each species.

3. Protein Quantification Strategy



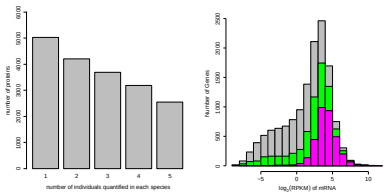
- A single batch of a SILAC lymphoblastoid cell line (LCL) GM19238 was grown in a large batch culture to provide sufficient labeled protein material for analyses of 15 unlabeled human, chimpanzee, or rhesus LCLs.
- For each cell line equal protein amounts were combined and separated at the protein level by SDS-PAGE. Gel lanes were excised into forty equal sized fragments. In-gel digestion was performed on each segment with trypsin. Digested samples were analyzed using 1-hour LC-MS/MS gradients using a NanoAcquity combined with an Orbitrap Velos Pro.
- Protein quantification ratios between each sample and the internal standard cell line were derived from peak heights across extracted ion chromatograms from the unlabeled LCL sample and labeled internal standard.

4. mRNA Quantification Strategy



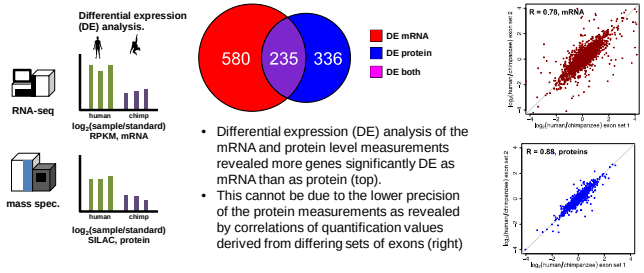
- RNA-seq data was collected on an Illumina Hi-Seq 2000
- We used length 50bp unpaired reads.
- We used a curated set of orthologous exons (http://giladlab.uchicago.edu/orthoExon) to derive expression measurements for each gene.
- Reads overlapping a single orthologous exons in each species were counted to derive reads per kilo-base per million mapped reads (RPKM) values for a gene in each species per cell line. We also obtained RNA-seq data for the SILAC labeled internal standard cell line GM19238.
- From these data, we computed $\log_2$-ratio of the RPKM values obtained between an unlabeled LCL and the GM19238 internal standard to normalize the expression measurements.

5. Protein Quantification Data



- 4,157 proteins were quantified in at least three human and three chimpanzee individuals and 3,688 proteins were quantified in at least three individuals from each of the three species.
- Gray bars illustrate the RPKM distribution of all transcripts. Green bars designate the RPKM distribution of proteins detected and quantified in any one of the 15 cell lines. Magenta bars designate the RPKM distribution of proteins quantified in 3 or more individuals across all 3 species.

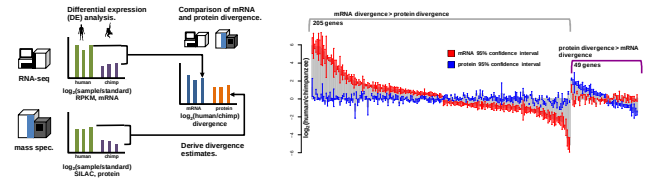
6. Differential Expression Analysis



- Differential expression (DE) analysis of the mRNA and protein level measurements revealed more genes significantly DE as mRNA than as protein (top).
- This cannot be due to the lower precision of the protein measurements as revealed by correlations of quantification values derived from differing sets of exons (right)

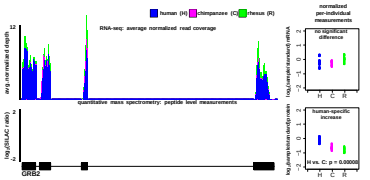
Reprints available from info@msbioworks.com

7. Integrated Differential Expression Analysis



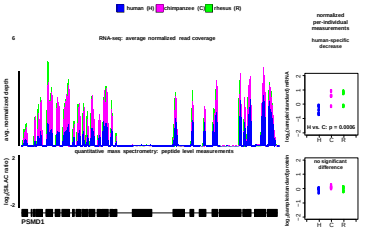
- We next examined the divergence, or the ratio $\log_2$(human/chimpanzee), measured using RNA-seq and measured using quantitative mass spectrometry.
- Using the biological replication in our study, we asked for which genes did divergence differ significantly between mRNA and protein (at FDR 1%).
- We found that more of these genes had higher divergence than protein divergence (error bars designate standard errors).

8. Change in human protein expression but not mRNA



- Our data allowed us to identify genes where protein levels changed in the human linage, but mRNA levels remained the same between human and chimpanzee.
- This could be due to an adaptive change in the post-translational or post-transcriptional regulation of this gene GRB2.

9. Change in human mRNA but not protein expression



- We also found many patterns or we observed an mRNA level change specific to the human linage, yet no difference at the protein level.
- This pattern could be due to buffering at the protein level or, more interestingly, a compensatory mutation that acted to restore similar protein levels.

10. Summary

- The proteome is under greater evolutionary constraint than the transcriptome.
- Many differences in mRNA levels are effectively neutral if buffered or compensated as protein.
- Our data allow us to identify genes of possible adaptive significance where regulatory changes act post-transcriptionally or post-translationally.