

Poster A-24

A multi-species study of chromosomal domains of co-expression



Authors:

Aditya Ramani (*Department of Computer Science, University of Illinois at Urbana-Champaign*)

Kilannin Krysiak (*School of Molecular and Cell Biology, University of Illinois at Urbana-Champaign*)

Chengxiang Zhai (*Department of Computer Science, University of Illinois at Urbana-Champaign*)

Lei Liu (*W. M. Keck Center for Comparative and Functional Genomics, University of Illinois at Urbana-Champaign*)

Short Abstract: Chromosomal regions of co-expression are related to gene order and shared chromatin domains. We studied this phenomena across a number species with increased genome size and organism complexity. We found that most of localized co-expression can be explained by gene duplication and the clustering of housekeeping genes.

Long Abstract:

It was previously believed that Eukaryotic Gene Order was random [8], but, however it has been found to be false in practice. [1] looks at the gene expression profile in *Saccharomyces Cerevisiae* and shows that expression is more highly correlated across proximal genes. [2] shows that though there is significant clustering of genes in prokaryotes due to operonic structures, such effects are relatively reduced in eukaryotes, with clustering in *Homo Sapiens* limited to housekeeping genes. It has also been observed in [5] that the *Drosophila* genome has a large number of proximal groups of genes that show significant co-expression, and that most of these groups were non tissue specific. [6] makes a similar study of intestine expressed genes in *C. Elegans*, and finds a strong bias to clustering of housekeeping rather than specific intestine-enriched genes, suggesting a role of chromatin organization in the regulation of housekeeping gene expression. This is similar to the findings in [7]. In this paper we make a wide ranging study of chromosomal domains of expression across a number of species, and attempt to understand the underlying drivers. We attempt to answer a number of interesting questions:

- How does co-expression vary with distance between genes?
- How does this co-expression vary across species? Is it connected to organism complexity and genome size?
- Is it possible to characterize susceptible regions?
- Based on all this, is it possible to infer an underlying hypothesis?

The species that we consider in our study are - *Homo Sapiens*, *Mus Musculus*, *Arabidopsis Thaliana*, *Caenorhabditis Elegans*, *Drosophila Melanogaster*, *Saccharomyces Cerevisiae* and *Escherichia Coli*.

We first attempt to plot the co-expression with the extent of proximity between genes. As expected, we find co-expression to decrease with distance. In the next stage, we plot the proximity vs. co-expression curves for a number of species and find levels of co-expression

to decrease linearly with genome size. It is likely that, as an organism becomes more sophisticated, the regulatory mechanism becomes sufficiently complex to reduce the biological pressure to preserve proximity.

In the next stage, we attempt to plot the localized co-expression levels versus the density of housekeeping genes in that section of the genome. A list of human housekeeping genes was obtained from [4]. As we could not directly obtain lists of housekeeping genes for the other species, we mapped homologs from the human list. We notice that there is strong correlation in the peaks housekeeping gene density and localized co-expression levels. This trend was especially strong in higher species. When the peaks that could not be explained by housekeeping genes, it usually corresponded to gene duplicates. Considering the exceptionally high levels of conservation and clustering we observed in housekeeping genes, across a number of species, we are given to believe that they share chromatin structures that lead to this localized co-expression. Also, in the higher species, for tissue specific genes, localized co-expression was observed to be lower. For the most part they seem to keep to reasonably uniform levels. However, Yeast shows uniformly high levels of co-expression, extending even to tissue specific genes, hence showing very strong gene order.

In this study, we validate that most chromosomal domains of co-expression can be explained by housekeeping genes and gene duplicates. Hence, a lot of the order observed in higher species, stems from housekeeping genes, a set of genes we find to be strongly conserved, clustered and co-expressed across a wide number of species, thus indicating biological pressure to preserve function and proximity. Also, considering the fact that they are known to lack a complex regulatory system, it is likely that more newly evolved genes are less strongly ordered due to compensation from more sophisticated regulatory mechanisms. In lower species, the co-expression clustering effect was observed as a more broad ranging phenomena extending even to tissue specific genes. This raises an interesting question - does the presence of complex regulatory mechanisms reduce the biological pressure for chromosomal proximity and possibly shared chromatin structures? We find the answer to be true. From our study of housekeeping genes, we find it possible that most chromosomal domains present in higher species are relics of those necessitated in more primitive species. This paper is novel in the effort it has made to make a comparative study of trends derived from expression data of a number of species. It has identified some consistent drivers of chromosomal domains of co-expression and gene order. This work can be extended and further verified against more species. It would also be interesting to make a detailed curation of all observed domains in a specific species.

References:

1. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression, Barak A. Cohen, Robi D. Mitra¹, Jason D. Hughes and George M. Church, *Nature Genetics*, Volume no. 26 2, pp 183 - 186
2. Clustering of housekeeping genes provides a unified model of gene order in the human genome, Martin J. Lercher, Araxi Urrutia, Laurence D. Hurst, *Nature Genetics*, May 2002, Volume 31 no. 2 pp 180 - 183
3. Mammalian Housekeeping Genes Evolve More Slowly than Tissue-Specific Genes, Liqing Zhang and Wen-Hsiung Li, *Molecular Biology and Evolution*, 236-230 (2004)
4. Human housekeeping genes are compact, Eli Eisenberg and Erez Y Levanon, *Trends in Genetics*, Vol 19, 362-366, 2003
5. Evidence for large domains of similarly expressed genes in the *Drosophila* genome, Paul T Spellman and Gerald M Rubins, *Journal of Biology* 2002

6. Chromosomal clustering and GATA transcriptional regulation of intestine-expressed genes in *C. elegans*, Florencia Pauli, Yueyi Liu, Yoona A. Kim, Pei-Jiun Chen and Stuart K. Kim, *Development* 133, 287-295 (2006)
7. Large clusters of co-expressed genes in the *Drosophila* genome Alexander M. Boutanaev, Alla I. Kalmykova, Yuri Y. Shevelyov and Dmitry I. Nurminsky *Nature* 420, 666-669 (12 December 2002)
8. The Evolutionary Dynamics of Gene Order, Laurence D. Hurst, Csaba Pal and Martin J. Lercher (2004)