Co-regulation of human paralog genes in the three-dimensional chromatin architecture

Jonas Ibn-Salem Miguel A. Andrade-Navarro

Institute of Molecular Biology,
Faculty of Biology,
Johannes Gutenberg University Mainz,
Mainz, Germany
{j.ibn-salem, andrade}@uni-mainz.de

May 18, 2015

Abstract

Paralog genes arise from gene duplication events during evolution. The resulting sequence similarity between paralogs often leads to proteins of similar structures and functions, which may cooperate in common pathways and in protein complexes. Therefore, it can be useful for the cell to have paralogs co-regulated. In eukaryotes, genes are regulated in part by binding of transcription factors to distal enhancer elements, which perform looping interactions to contact the transcription machinery at gene promoters. These looping interactions can be measured by genome-wide chromatin conformation capture (Hi-C) experiments which revealed conserved megabase-sized self-interacting regions called topological association domains (TADs). We hypothesised that paralogs cluster in the three-dimensional chromatin architecture and share common regulatory mechanisms to enable coordinated expression.

To test this hypothesis, we integrated paralogy annotations with genome-wide data-sets of enhancer-promoter associations, Hi-C experiments, and gene expression in diverse human cell-types. As control we sampled random gene pairs by taking the linear distances of paralogs and the number of linked enhancers into account. We show that paralogs share significantly more common enhancer elements than expected. Furthermore, they are located significantly more often in the same TAD and contact each other more frequently than expected. Consequently, paralogs tend to show a positive correlation of gene expression over many cell-types.

Combined, our results indicated that human paralogs share common regulatory mechanisms and cluster not only in the linear genome but also in the three-dimensional chromatin architecture. This enables concerted expression of paralogs over diverse cell-types and indicate evolutionary constraints in functional genome organization.