doi:10.1016/j.jmb.2006.03.051 J. Mol. Biol. (2006) 359, 825–831

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Co-expressed Yeast Genes Cluster Over a Long Range but are not Regularly Spaced

Martin J. Lercher[1,2]* and Laurence D. Hurst[1]

[1]*Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK*

[2]*European Molecular Biology Laboratory, 69012 Heidelberg Germany*

**Corresponding author*

Analysis in yeast of the relationship between a gene's genomic position and its expression profile, derived from chip array data, suggests that both closely linked genes and genes spaced at regular intervals show correlated expression profiles. Unfortunately, yeast arrays are often printed in genomic order. The above results may hence reflect little more than known spatial biases within arrays. To circumvent this problem, we analyse spatially unbiased expression data derived from a large Northern blot study. We find that local domains of co-expressed genes range up to 30 genes (100 kb), and are thus much larger than previously considered. There is, by contrast, no evidence for periodicity of co-expression in yeast. We likewise find no convincing evidence for periodicity in the human or mouse genome. Further, analysis of yeast transcription factor binding data sets suggests that there is currently no statistical evidence for chromosomal periodicity of co-regulation.

Keywords: gene order; nuclear functional organization; transcription regulation; regular spacing; yeast

Introduction

Recently, there has been interest in the problem of non-random gene order, with considerable attention being paid to the possibility that genes are arranged in genomes with respect to their expression.[1] In yeast, for example, there is evidence from microarray chip-based assays that neighbouring genes can have expression profiles more similar than expected by chance.[2–4] Were this owing solely to bidirectional promoters, we should expect the effects to be limited to pairs of neighbouring genes, but in a few cases the regulatory mechanisms appear somewhat more complex and small groupings of genes appear to be co-expressed.[3,5] Moreover, many co-expressed genes are not in the correct orientation for control by bidirectional promoters.[4,5] Others have noted, again using microarray data, an apparent periodicity in expression parameters;[3,6] genes that are regularly spaced, but not immediate neighbours, show more similar expression profiles than expected. In apparent confirmation, a periodicity in location of genes under the control of certain transcription factors (as inferred from ChIP-Chip data) was reported.[7] Finally, genes up-regulated or down-regulated in the response of haploid *Saccharomyces cerevisisae* to DNA damage may be regularly spaced.[8] As we can find no statistical support for this last possibility (see Supplementary Data, Table 1) we shall not consider it further.

Even where statistically significant, the above evidence for non-random gene order is far from definitive. Most problematically, chips are often biased in ways that can automatically generate a signal of local similarity and periodicity as an artefact. There exists, for example, a "spotting effect" that introduces a periodicity into the signal.[9] A given spot in a given row on the array is similar to spots that are an integer number (X) rows apart. This most likely represents a plate bias, i.e. within each block, cDNA printed in row $N+X$ is taken from the same plate as that printed in row N. A similar pattern is observed in columns.[9] Likewise, there exists a "print tip effect" that introduces a further periodicity, which depends on the printing configuration (e.g. of two and 24 genes).[10] Unfortunately, cDNAs on the yeast chips are printed in genomic order, and so these biases do not only add unwanted noise, but resolve to an artefacts of genomic periodicity and local similarity.[10]

That such biases exist seems unequivocal (we confirmed their existence in the data from Spellman *et al.*[11] (see also Balazsi *et al.*[10]) and Gasch *et al.*,[12]

Abbreviations used: ChIP-Chip, chromatin immunoprecipitation analysed by microarray chips; EST, expressed sequence tags; MAS5, Affymetrix Microarray Suite 5; SAGE, serial analysis of gene expression.

E-mail address of the corresponding author: m.j.lercher@bath.ac.uk

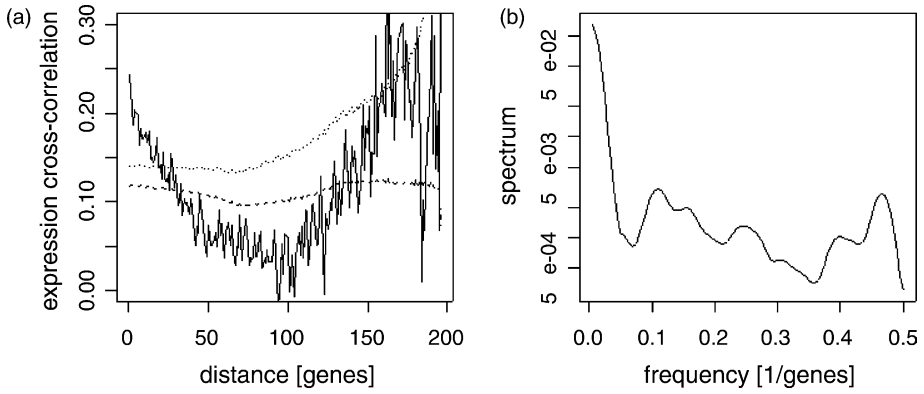

Figure 1. (a) Co-expression of yeast genes as a function of the distance between them, and (b) the corresponding frequency spectrum, the periodogram. Data from Northern blots.[13] Broken and dotted lines in (a) indicate mean and 95th percentile, respectively, calculated from 1000 randomised genomes. The low-frequency peak in (b) corresponds to the correlation between subtelomeric genes, seen at distances of ~170 genes in (a). Distances are measured in numbers of intervening genes (the corresponding results for distances measured in kb are shown in Supplementary Data, Figure 3).

Supplementary Data, Figures 1 and 2). Whether they fully explain local similarity and periodicity of expression is, however, unresolved. Resolution of this issue will require analysis of unbiased data sets. One approach is to apply normalization methods to array data that attempt to account for these biases. While this seems possible for some biases,[10] it is not obviously possible for others.[9] Even if possible, there would always exist the possibility that any bias has not been fully accounted for. Convincing resolution of the issue will therefore require extensive datasets that avoid the bias inherent in chips printed in genomic order. Such an unbiased expression dataset has been produced from Northern blots,[13] covering approximately one-sixth of the yeast genome. Here, we employ these data to ask whether in yeast we can indeed find evidence for local similarity and for periodicity in expression profiles, and to quantify these effects. Additionally, to better understand the generality of any conclusions, we ask the same questions in the human and mouse genomes using a variety of data sources.

Results

Northern data confirm clusters but not periodicity of co-expression

To circumvent the problems associated with biased microarray chip designs, we employed the Northern blot data reported by Brown *et al.*†.[13] As can be seen in Figure 1, there is a general trend for genes in close proximity to be highly co-expressed (for the comparable figure using base-pairs rather than intervening genes as distance measure, see Supplementary Data, Figure 3). This trend continues over a surprisingly large range, up to approximately 30 genes or 100 kb; prior analyses had suggested the effects to be very much more local.[5]

Figure 1(b) shows the frequency spectrum (periodogram) of the expression correlation. The strongest signal of periodicity is seen for a period of nine genes, corresponding to a peak at frequency 0.11/gene. However, from comparison to 1000 randomised genomes, we find no evidence that a peak of this height is greater than expected by chance ($P=0.87$). This holds equally true when distances are measured in base-pairs rather than number of genes ($P=0.66$; Supplementary Data, Figure 3). The absence of periodicity in our unbiased Northern dataset contrasts markedly with the analysis of biased microarray data,[11,12] in which the periodic signal is always much stronger than expected by chance ($P<0.002$).

It has been suggested that putative coiled chromatin conformations might result in periodicity that is specific to individual chromosome arms.[7] To explore this possibility, we repeated the above analysis for each available chromosome arm separately. As before, we find no periodicity that is stronger than expected by chance (Table 1; and Supplementary Data, Table 2). Thus, whether analysing by genome, chromosome arm, or (data not shown) whole chromosomes, we find no statistical support for any periodic organisation of co-expressed genes.

Yeast regulator binding data demonstrates clusters but not periodicity

Clusters (or periodicity) of co-expression can result from specific chromatin conformations or co-ordinated binding of transcription factors, or both. While no large-scale data on chromatin conformations are available, we can examine the distribution of a wide range of different regulatory binding sites.[14,15] When analysing micro-array based ChIP-Chip data,[15] we observe strong signals of local similarity and of periodicity: many pairs of

† http://www.yeastresearch.man.ac.uk/Publications/emboj/default.asp