

Two by Two: Co-Regulating Adjacent Gene Pairs in Yeast and Beyond

Michael A McAlear*

Department of Molecular Biology and Biochemistry, Wesleyan University, Church St. Middletown, CT, USA

Ever since Jacob, Monod and their colleagues described how the multiple, tandemly arranged genes of the *Lac* operon from *E. coli* are co-regulated in response to different sugar substrates [1], one of the main pursuits of molecular biology has been to dissect the myriads of mechanisms whereby cells control and effect gene regulation. The importance of this question is evident across diverse areas of cellular and organismal biology, from the control of progression through the cell cycle in unicellular organisms, to the processes of development and differentiation in complex metazoans. Indeed, the 2012 Nobel Prize for Physiology or Medicine was recently awarded, in part, for research aimed at understanding and manipulating the regulatory circuits that can be used to reprogram the gene expression patterns of differentiated adult cells into those of so-called 'induced' pluripotent stem cells [2]. If anything, the significance of understanding how organisms 'use' their genes appropriately is as important as ever given that advancing, massively parallel DNA sequencing technologies continue to provide an ever-increasing catalog of new and diverse genomes.

While there are many strategies that cells use to regulate the activities of their genes and gene products, one of the most critical strategies remains at the level of gene transcription. Our appreciation for how important and widespread this level of control is has increased immensely over the last decade, in part, because of the development of technologies such as micro-array analyses: a technique that allows for the quantitative assessment of mRNA levels under various conditions at a genome-wide scale [3]. So extensive has been this kind of analysis, that in some systems one can access genome-wide mRNA expression profiles from hundreds of different samples, conditions and time courses. The analysis of these datasets has allowed for the identification of groups of genes that share common regulatory responses to a given condition or stimulus, and subsequent analysis of their associated regulatory sequences has often revealed common promoter sequence motifs that represent targets for *trans*-acting regulatory factors [4]. In many ways, the control of these co-regulated gene sets, or regulons, has been found to echo the mechanism originally described for the control of the genes of the *Lac* operon. That is, mRNA expression levels are determined, in part, by the activities of transcription factors (i.e. the Lac repressor) that bind to promoter sequences (i.e. the Lac operator), thereby influencing the activity of RNA polymerase.

Our own foray into this type of analysis began with the discovery that there was a large set of transcriptionally co-regulated genes in *S. cerevisiae* that contained a high proportion of members that were known to function in the ribosome and rRNA biogenesis (RRB) pathways [5]. As a group, the genes of the RRB regulon were transcriptionally repressed under stressful conditions (i.e. heat shock, nutrient deprivation etc.), and they were activated under conditions (i.e. glucose replenishment, release from alpha factor block etc.) that favoured cell growth and division. Independently and concurrently, a similar set of co-regulated genes was recognized also via micro-array analysis and it was given the related moniker of the Ribi regulon [4]. We found that the promoters of the RRB genes were enriched for the PAC and RRPE promoter motifs, and substitutions in these motifs disturbed the regulated expression of candidate RRB genes *in vivo* [5]. By analyzing genome-wide expression profiles, and cross referencing with the presence of the conserved promoter motifs, we were able to suggest that the RRB regulon contained some 200 members, even though many of the genes had at the time, no ascribed function [5,6].

These predictions have turned out to be fairly accurate, as since then some 90% of the unknown members have been subsequently identified as contributing towards ribosome and rRNA biogenesis. The factors that recognize the PAC and RRPE motifs have also been subsequently identified (the Stb3 and Tod6/Dot6 proteins respectively), and they have been shown to play key roles in RRB gene regulation [7].

One of the most unusual properties associated with the genes of the RRB regulon was their genomic distribution: some 15% of the RRB genes were located on the chromosomes as immediately adjacent gene pairs [6]. Whereas one might assume that paired, co-regulated genes from within the same metabolic pathway might be driven by bi-directional promoters- as is known for the *GAL1* and *GAL10* genes [8] the RRB gene pairs were frequently found in the tandem, and even convergent orientations. A highly statistical enrichment for adjacent gene pairing could also be recognized in the genes of the sister ribosomal protein (RP) regulon, and across widely divergent yeast species. For example, in *C. albicans* roughly 27% of the 168 RRB genes and 21% of the 118 RP genes are found as discrete, immediate gene pairs across their respective chromosomes. Again, the adjacent gene pairs were found in all three possible divergent, convergent, and tandem orientations (Table 1).

This striking enrichment for adjacent pairing of co-regulated genes was not limited to the ribosome biogenesis pathways. Significant gene pairing can be seen across numerous functional pathways in *S. cerevisiae* including the Gene Ontology (GO) groupings for DNA damage response (16/175 genes), carbohydrate metabolism (9/91 genes), nitrogen metabolism (8/86 genes), heat shock response (4/18 genes) and more. Furthermore, the paired members of the respective gene sets tended to represent those genes that were the most highly expressed, and they were more likely to remain as paired genes across divergent yeasts when compared to random pairings [9]. There is even evidence that elevated levels of RP gene pairing exists across other eukaryotes, including in *C. elegans*, and *D. melanogaster* (Table 1). Therefore, it appears that the strategy for positioning transcriptionally co-regulated members of distinct functional pathways as adjacent gene pairs (regardless of relative gene orientations) is widespread across eukaryotes.

Following the lead of Jacob and Monod, we sought to identify the *cis* and *trans*-acting elements that contribute to adjacent gene co-regulation. To do this, we dissected the workings of the convergently oriented, adjacent *MPP10* and *YJR003C* RRB gene pair from *S. cerevisiae* (Figure 1). Perhaps not surprisingly, we found that mutations in the conserved PAC and RRPE motifs from within the *MPP10*

***Corresponding author:** Dr. Michael A McAlear, Department of Molecular Biology and Biochemistry, Wesleyan University, 228 Hall-Atwater Laboratory, 237 Church St. Middletown, CT, USA, Tel: 06459-0175; E-mail: MMcAlear@wesleyan.edu

Received October 31, 2012; Accepted November 01, 2012; Published November 03, 2012

Citation: McAlear MA (2013) Two by Two: Co-Regulating Adjacent Gene Pairs in Yeast and Beyond. Mol Biol 2: e105. doi:[10.4172/2168-9547.1000e105](https://doi.org/10.4172/2168-9547.1000e105)

Copyright: © 2013 McAlear MA. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Organism	Regulon	No. of genes	No. of adj. genes	P value
<i>S. cerevisiae</i>	Ribosomal Proteins	180	24	4.4×10^{-5}
	Ribosome and rRNA biogenesis	188	28	3.9×10^{-7}
	Carbohydrate metabolism	91	9	5.5×10^{-4}
	Nitrogen metabolism	86	8	9.9×10^{-4}
<i>C. albicans</i>	Ribosomal Proteins	118	25	1.4×10^{-14}
	Ribosome and rRNA biogenesis	168	46	2.3×10^{-15}
<i>A. nidulans</i>	Ribosomal Proteins	132	16	1.5×10^{-7}
<i>C. elegans</i>	Ribosomal Proteins	86	6	1.1×10^{-1}
<i>D. melanogaster</i>	Ribosomal Proteins	165	6	1.1×10^{-1}

Table 1: Examples of co-regulated gene set containing high levels of adjacent gene pairs.

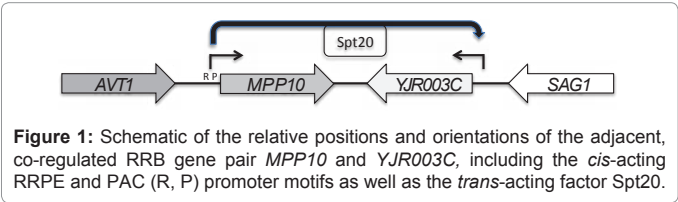


Figure 1: Schematic of the relative positions and orientations of the adjacent, co-regulated RRB gene pair *MPP10* and *YJR003C*, including the *cis*-acting RRPE and PAC (R, P) promoter motifs as well as the *trans*-acting factor Spt20.

promoter abrogated the characteristic regulated repression of *MPP10* in response to heat shock and other stressors. What was surprising, however, was that these minor promoter motif substitutions also abrogated the regulated repression of the adjacent *YJR003C* gene, even though *YJR003C* is transcribed in the opposite (convergent) direction from the *MPP10* gene and from a promoter that lies some 4 kbp away. The *YJR003C* promoter does not contain matches to either the PAC or the RRPE motifs, yet its regulated expression mimics that of *MPP10* and the other members of the RRB regulon. Furthermore, when we disrupted the immediate adjacency of the *MPP10* and *YJR003C* genes by inserting the 3 kbp *URA3 KANr* pCORE reporter cassette between them, the regulated expression of *YJR003C* was disturbed, but the *MPP10* gene was regulated normally. The importance of immediate adjacency for maintaining transcriptional co-regulation of the two genes is reflected in the observation that across the budding yeast genome in no case did we observe a case whereby two otherwise co-regulated RRB (or RP) genes were separated by a non-RRB (or non-RP) gene. The members of the respective regulons were found as either immediately adjacent gene pairs, or they were scattered as isolated genes. Interestingly, we observed that the insertion of the *LEU2* gene between *MPP10* and *YJR003C* did not interfere with the gene co-regulation when the cells are grown in YPD media (i.e. conditions in which *LEU2* is repressed) but it does disrupt the regulated expression of *YJR003C* when the cells are grown in SC- Leu media (i.e. conditions in which *LEU2* is expressed). The influence of the *MPP10* promoter was limited, in the sense that it could not impart heat shock repression onto the transplanted *LEU2* gene. Since a similar insertion of the RNA pol III driven tRNA-Thr gene between *MPP10* and *YJR003C* was found not to interfere with the co-regulated expression, the transcriptional coupling mechanism appears to be RNA pol II specific. Further mutational analysis of this sort will help determine the relative *cis*-acting sequence, spatial and orientation constraints under which adjacent gene co-regulation is achieved.

The identification of the relevant *trans*-factors that mediate adjacent gene co-regulation is also proceeding. There is a reasonable expectation that the co-regulation of the adjacent gene pairs may be related to changes in the status of the local chromatin, and potentially with the modification or displacement of local nucleosomes. As such, we have been screening chromatin modifier mutants for phenotypes

consistent with the proper regulation of the *MPP10* gene, but with defects in the regulated expression of *YJR003C*. We have found such a mutant in the non-essential *SPT20* gene (M. McAlear, unpublished results). *SPT20* is a component of the SAGA complex [10], a multi-subunit chromatin modifier that has been shown to play an important role in gene regulation [11], and nucleosome acetylation, methylation and deubiquitination. In *SPT20* mutants, the regulated heat shock induced repression of *YJR003C* becomes uncoupled from that of *MPP10*, although it remains to be determined exactly what relevant biochemical activity is being disturbed. Likewise, the identification and characterization of other *trans*-acting factors will be important for uncovering the exact mechanism of co-regulation.

The suggestion that one of the strategies that cells use to co-regulate large sets of functionally related genes is to place them across the genome as discreet, immediately adjacent pairs, is supported by several other observations, particularly in budding yeast. A computational analysis of whole-genome expression profiles from *S. cerevisiae* revealed that adjacent genes exhibited highly correlated expression profiles, even in cases where only member of the pair contained the relevant upstream activator sequence [12]. More recently, an analysis of the genome-wide deletion collections revealed that a substantial fraction of the assigned phenotypes were likely erroneously attributed to individual gene deletions because they arose from defects associated with the activity of the neighboring gene [13]. This so called Neighboring Gene Effect (NGE) could be revealing the interdependence between subsets of adjacent genes for their proper transcriptional control. Fifty years later, these observations would seem familiar to Jacob and Monod, even though the detailed mechanisms that they described for controlling the expression of tandem, polycistronic operons are more relevant for prokaryotes. Understanding the mechanisms whereby adjacent gene pairs of all orientations are transcriptionally co-regulated in eukaryotes is no less important.

References

1. Pardee AB, Jacob F, Monod J (1959) The genetic control and cytoplasmic expression of "Inducibility" in the synthesis of β -galactosidase by *E. coli*. J Mol Biol 1: 165-178.
2. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, et al. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131: 861-872.
3. Lashkari DA, DeRisi JL, McCusker JH, Namath AF, Gentile C, et al. (1997) Yeast microarrays for genome wide parallel genetic and gene expression analysis. Proc Natl Acad Sci U S A 94: 13057-13062.
4. Hughes JD, Estep PW, Tavazoie S, Church GM (2000) Computational identification of cis-regulatory elements associated with groups of functionally related genes in *Saccharomyces cerevisiae*. J Mol Biol 296: 1205-1214.
5. Wade C, Shea KA, Jensen RV, McAlear MA (2001) EBP2 is a member of the yeast RRB regulon, a transcriptionally coregulated set of genes that are required for ribosome and rRNA biosynthesis. Mol Cell Biol 21: 8638-8650.
6. Wade CH, Umbarger MA, McAlear MA (2006) The budding yeast rRNA and ribosome biosynthesis (RRB) regulon contains over 200 genes. Yeast 23: 293-306.
7. Lippman SI, Broach JR (2009) Protein kinase A and TORC1 activate genes for ribosomal biogenesis by inactivating repressors encoded by Dot6 and its homolog Tod6. Proc Natl Acad Sci U S A 106: 19928-19933.
8. Johnston M, Davis RW (1984) Sequences that regulate the divergent GAL1-GAL10 promoter in *Saccharomyces cerevisiae*. Mol Cell Biol 4: 1440-1448.
9. Arnone JT, Robbins-Pianka A, Arace JR, Kass-Gergi S, McAlear MA (2012) The adjacent positioning of co-regulated gene pairs is widely conserved across eukaryotes. BMC Genomics 13: 546.
10. Arnone JT, McAlear MA (2011) Adjacent gene pairing plays a role in the coordinated expression of ribosome biogenesis genes *MPP10* and *YJR003C* in *Saccharomyces cerevisiae*. Eukaryot Cell 10: 43-53.

-
11. Sterner DE, Grant PA, Roberts SM, Duggan LJ, Belotserkovskaya R, et al. (1999) Functional organization of the yeast SAGA complex: distinct components involved in structural integrity, nucleosome acetylation, and TATA-binding protein interaction. Mol Cell Biol 19: 86-98.
 12. Huisinga KL, Pugh BF (2004) A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in *Saccharomyces cerevisiae*. Mol Cell 13: 573-585.
 13. Cohen BA, Mitra RD, Hughes JD, Church GM (2000) A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. Nat Genet 26: 183-186.