Comparative biology: beyond sequence analysis

Itay Tirosh[1], Yonatan Bilu[1] and Naama Barkai[1,2]

Comparative analysis is a fundamental tool in biology. Conservation among species greatly assists the detection and characterization of functional elements, whereas inter-species differences are probably the best indicators of biological adaptation. Traditionally, comparative approaches were applied to the analysis of genomic sequences. With the growing availability of functional genomic data, comparative paradigms are now being extended also to the study of other functional attributes, most notably the gene expression. Here we review recent works applying comparative analysis to large-scale gene expression datasets and discuss the central principles and challenges of such approaches.

Addresses
[1] Department of Molecular Genetics, Weizmann Institute of Science, 76100 Rehovot, Israel
[2] Department of Physics of Complex Systems, Weizmann Institute of Science, 76100 Rehovot, Israel

Corresponding author: Tirosh, Itay (itay.tirosh@weizmann.ac.il), Bilu, Yonatan (yonatan.bilu@weizmann.ac.il) and Barkai, Naama (naama.barkai@weizmann.ac.il)

Current Opinion in Biotechnology 2007, **18**:371–377

This review comes from a themed issue on
Systems biology
Edited by Alan Aderem

Available online 10th August 2007

0958-1669/$ – see front matter

DOI 10.1016/j.copbio.2007.07.003

Introduction

One of the surprising discoveries of modern biology is the strong conservation of protein sequences, as well as cellular mechanisms, across evolution. Some of our metabolic genes, for example, display strong sequence and functional similarities to their bacterial counterparts. Moreover, central features of core cellular processes such as cell cycle progression, DNA replication or transcription are conserved from yeast to human. Indeed, this extensive conservation had motivated the use of model organisms as means for studying conserved processes that are more difficult to assay in complex systems. Sequence similarity, in particular, has emerged as a key tool in predicting functional properties. In fact, most annotations of newly sequenced genomes, including gene prediction, gene function or regulatory elements, are based on similarity with other sequences whose functions have been described (e.g. [1–3]).

Whereas most comparative studies focus on conserved properties as means for characterizing functional elements, inter-species differences are also of interest. These differences are arguably the best indicators of evolutionary history and provide much information about species-specific adaptations. Identifying such differences is thus central to our search for what makes us human and how biological diversity is generated.

Technological advances over the past decade have led to the accumulation of genome-scale data describing not only gene sequence but also functional properties including gene expression, protein–protein interactions, and the binding of transcription factors to DNA. Such data are now available for multiple organisms, and their complexity presents a new set of challenges to comparative analysis. For example, unlike sequence information, most functional properties are condition-dependent, a property that needs to be accounted for during inter-species comparisons. Furthermore, sequence analysis is typically gene-specific and compares specific sets of homologs. Functional properties, on the contrary, often reflect the integrated function of multiple genes, calling for novel methods that allow network-centered rather than gene-centered comparisons. Finally, functional genomic data generated with current technologies suffer from high levels of noise and therefore need to be filtered in order to obtain valid conclusions.

In this review we focus on recent studies that employ comparative methods to analyze gene expression data. We describe the different approaches employed in such an analysis and highlight the remaining challenges.

Comparative analysis of condition-specific gene expression

An important element in a gene's function is the spatial and temporal pattern by which it is expressed. For example, diverse cell types are generated by extensive modifications in the expression of the same set of genes. In recent years, microarray technology has facilitated thousands of experiments that characterized genome-wide expression levels under a wide variety of conditions. Such data are now available for diverse organisms, providing a rich resource for comparative studies.

The initial motivation for comparing gene expression across species was to identify conserved expression programs that are likely to be functionally important. This was done by assembling compatible expression datasets from different organisms and searching for similar patterns. For example, McCarroll *et al.* applied this approach

to study the expression changes that occur during aging [4••], comparing gene expression in middle-aged flies and worms to their young adult counterparts. Notably, age-related changes in gene expression were found to be significantly correlated between flies and worms; in both species, for example, oxidative metabolism genes and cellular transport genes are repressed during aging. Similarly, conserved expression programs for embryonic development were also identified.

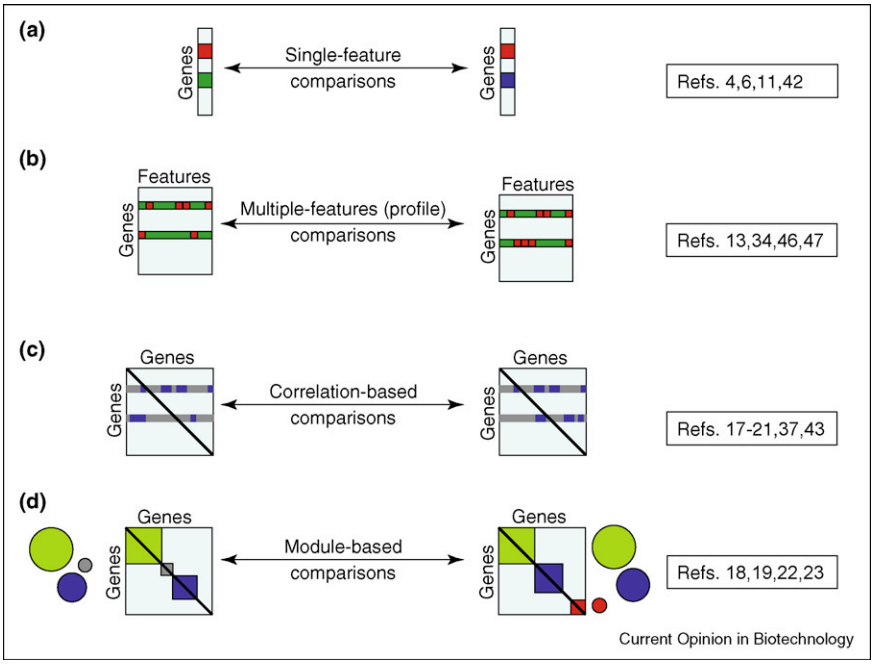
The comparison of distantly related organisms is useful for identifying conserved expression programs. Worms and flies, for example, had shared their last common ancestor about 1 billion years ago, so that the few remaining similarities between their highly divergent expression programs are likely to represent important processes. This extensive diversification, however, limits the utility of these comparisons for studying the principles of evolutionary changes of gene expression.

To overcome this limitation, subsequent studies have compared the expression programs of evolutionarily related organisms at identical conditions [5,6•,7–12,13••,14] (Figure 1a and b). Rifkin *et al.* [11], for example, compared the expression levels of several species and strains of fruit fly undergoing metamorphosis and classified genes into different evolutionary modes: conserved, evolving by drift and evolving by species-specific adaptation. Similar studies compared the expression levels of different species of yeasts [13••], flies [10], plants [8], and primates [6•,7]. A general insight derived from these studies is that genes differ greatly in the evolutionary plasticity of their expression profile. Specifically, gene expression divergence was found to be correlated with multiple properties, such as the number of protein–protein interactions [15], promoter type [13••], sex-biased expression [10], and whether or not they had undergone duplications [9,16].

Comparative analysis of co-expression

A major difficulty in comparing expression data between organisms is that gene expression is not static and changes depending on the external conditions. As was described above, this fact can be accounted for by considering compatible gene expression data. This approach, however, severely limits the data that can be used for comparative analysis, as only a small fraction of the available

Figure 1



Types of comparative analyses: from single-features to whole networks comparisons. Comparative analysis can be performed with different approaches depending on the type of data and choice of methods. Four different approaches are illustrated with increasing complexity; in each case, squares represent the datasets for two organisms, colors represent different values, and examples are given for conservation (top) and divergence (bottom). **(a)** In the simplest case, a single value (e.g. mRNA abundance or codon usage) is compared [4••,6•,11,42•]. **(b)** Multiple values for each gene are compared, resulting in a profile (expression across various conditions or tissues) [13••,34,46,47]. **(c)** Multiple values (as in **(b)**) can also be used to calculate correlations among the genes profiles. Pairwise correlations (e.g. co-expression) are compared to analyze interactions among genes [21], and correlations with all other genes (a row in the matrix) are compared to analyze the relative position of a gene in the network [17,37,43]. **(d)** Correlations can be used to define modules of functionally related genes, and the modular structure of different networks may be compared [18,19,22•,23•,25••]. Three modules are illustrated in each network: the green module is conserved across the networks; the blue and grey modules in the left network correspond to a single module (blue) in the right network (e.g. [23•]); and the red module is found only in the right network (e.g. [19]). Several references are given as examples for each scenario.

data in different species is comparable. Moreover, even if conditions seem equivalent, evolutionary distant organisms may react to them differently, since they do not necessarily perceive them as identical.

To overcome these limitations, an alternative approach was suggested, where the compared entities are patterns of gene co-expression, rather than the expression itself (Figure 1c and d). That is, co-expression between each gene pair is defined on the basis of the available expression data in each organism and is then compared among different organisms. These comparisons focus on the similarity and differences in the relative positions of orthologous genes within their expression networks and can thus be employed even when the conditions of different species are not comparable. Indeed, this approach enabled the comparative analysis of thousands of expression datasets that are available for different model organisms [17–21].

Consider for example, two genes that are similarly expressed within one organism across various conditions. This co-expression could be meaningful, indicating some functional relationship between the two genes, but could also be accidental, without biological implications. Conservation of this co-expression relationship between organisms could indicate its biological significance. Indeed, recent studies have shown that co-expression tends to be conserved among organisms and that co-expression relationships that are conserved are more likely to indicate similarity in function than co-expression that is restricted to just one organism [17–21]. Higher-order organization, such as modules of co-expressed genes, were also compared to provide a more robust approach for genome-wide comparisons [18,19]. Since such modules are often functional entities, their comparison enables direct functional interpretations [18,19,22•,23•,24] (Figure 1d).

In addition to the analysis of conserved properties, additional studies attempted to define changes in the connectivity of the transcriptional network that could indicate functional adaptation. Ihmels *et al.* compared the high-order connectivity in the transcriptional network of the model yeast *Saccharomyces cerevisiae* with that of the pathogenic yeast *Candida albicans* [23•,25••]. Several conserved modules of co-expressed genes were identified, including two modules associated with cytosolic versus mitochondrial ribosomes. Interestingly, while the division of these two modules was conserved, the relationship between them differed dramatically. In *C. albicans* the two modules were co-expressed, while in *S. cerevisiae* they were not [25••]. Indeed, this large-scale expression difference is linked to the metabolic adaptation of *S. cerevisiae* to anaerobic growth, rendering the oxidative metabolism dispensable for rapid growth. Importantly, this study also identified a *cis*-regulatory element in the promoters of both modules in *C. albicans* whereas it was identified only in the promoters of the cytosolic module of *S. cerevisiae*. This suggests that the observed expression differences resulted from massive regulatory changes, whereby *cis*-regulatory elements were lost from promoters of dozens of *S. cerevisiae* mitochondrial genes but conserved in the promoters of cytosolic genes coding for proteins involved in ribosomal biogenesis.

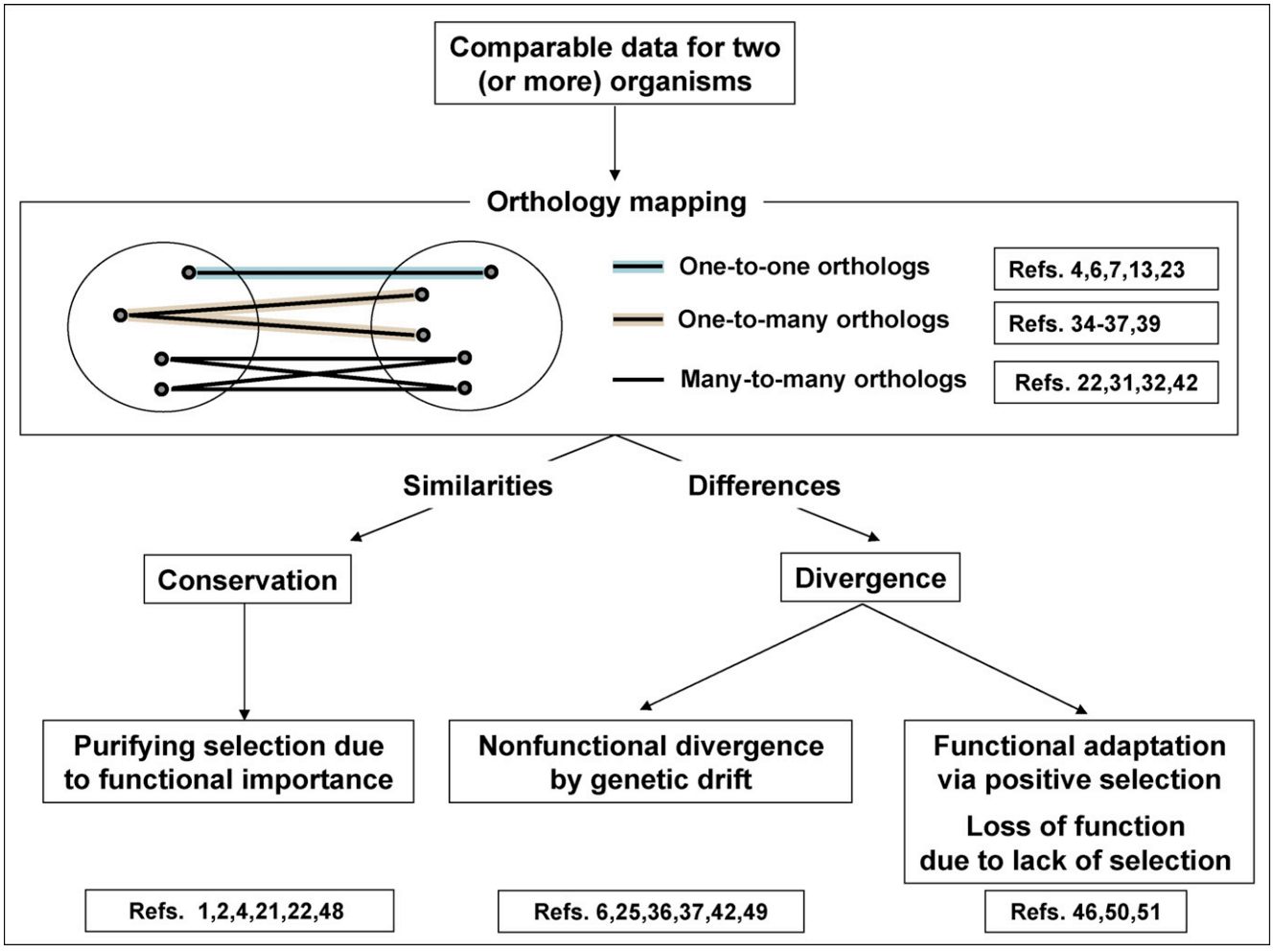
This example demonstrates that evolution of co-expression can also be reflected by the presence of common *cis*-regulatory elements in the promoters of co-expressed genes [24,26,27,28•,29]. In fact, other studies have tried to characterize evolutionary dynamics of gene expression also in the absence of gene expression information. Thus, Gasch *et al.* have systematically searched for common *cis*-regulatory elements at promoters of 14 yeast species [26]. Many of the *S. cerevisiae cis*-regulatory elements were found to be conserved in multiple yeast species suggesting that co-expression is conserved. Divergence of *cis*-regulation was also observed, including species-specific *cis*-regulatory elements, alterations in the set of genes that contain a specific *cis*-regulatory element, and changes in the sequence of the binding sites for a specific transcription factor (RPN4). It should be noted that although these and other examples [24,27] imply evolutionary divergence in co-expression patterns, it may well be that co-expression is conserved despite the change in *cis*-regulatory sequences or mechanisms; this possibility was in fact demonstrated in the case of the yeast ribosomal genes [28•], which are co-expressed both in *S. cerevisiae* and in *S. pombe*, but their co-expression is achieved by different mechanisms. Examination of additional yeast species suggested a gradual shift from regulation by one transcription factor to another, with evolutionarily intermediate species regulated by both [28•].

Duplicated genes and orthology relationships

For any comparison to be meaningful, one must first decide what is being compared and identify the common elements of the compared objects (Figure 2). In comparing data of different organisms, the basic elements are typically pairs of one-to-one orthologous genes, which rely on a mapping of orthology. Most comparative studies employ a strict definition for one-to-one orthology by considering, for example, syntenic or reciprocal best sequence matches and excluding ambiguities that arise from species-specific duplications or losses [30]. This approach is useful for comparison of closely related organisms or to obtain the most reliable orthology pairs.

Ultimately, however, orthology is not a one-to-one relationship, but, as a result of gene duplications that occurred since the compared species have diverged, a many-to-many relationship. Thus, for comparative studies to attain a full view of the homology relations among genes (or the compared elements in general), orthologous families, rather than orthologous pairs, need to be considered [31–33]. With the accumulation of fully sequenced

Figure 2



Principles of comparative analysis. Comparative analysis typically starts by collecting comparable data for two or more organisms. To compare the datasets, an orthology mapping and the type of orthology comparison have to be determined. Three types of comparisons are shown: (i) many-to-many, which considers all potential orthology relationships [22•,31,32,42•]; (ii) one-to-one, which considers only the best match of each gene and in some cases excludes ambiguities (i.e. the middle gene in the right circle) [4••,6•,7,13••,23•]; (iii) one-to-many, which focuses on gene duplication and losses [34–37,39]. After the compared objects have been determined, their inter-species similarities are examined. Objects with significant similarity can be identified, which suggests that they were evolutionarily conserved.This conservation may be interpreted as the result of purifying selection and therefore as an indication for functional importance [1,2,4••,21,22•,48•]. Conversely, objects with significant differences are likely to be evolutionarily divergent. This divergence may be associated with either a functional change, being the result of positive selection or lack of selection [6•,25••,36,37,42•,49], or functionally neutral, being the result of random drift [46,50,51]. Several references are given as examples for each scenario.

genomes, it is becoming possible to track the evolutionary history of individual genes in a large-scale manner.

The diversification of gene function following duplications has itself been the focus of many studies that compared pairs of duplicated genes with their single orthologs [34–39,40•]. These studies were largely facilitated by the finding of an ancient whole genome duplication event (WGD) [39,41], which produced hundreds of duplicated pairs that often diverged in function. Divergence of these duplicate pairs was studied by comparing their sequence [36,38,39] and expression profiles [34,35,37] with that of their single orthologs in a species whose genome did not undergo duplication. These studies revealed evidences for divergence of duplicates by both *subfunctionalization* (e.g. [35,38]) and *neofunctionalization* (e.g. [36,37,39]). Furthermore, the retention of duplicates following the WGD has also been analyzed by comparing synteny between post-WGD yeast species [40•]. Interestingly, retention of alternative copies in different species was found for many duplicates, and this was proposed to facilitate reproductive isolation and speciation.

Concluding remarks

We described different approaches for comparative analysis of functional genomics data (Figure 1). The most

straightforward approach is to consider a single feature, characterizing each gene (e.g. expression levels under a specific condition) and compare it among orthologous genes. This approach is easy to interpret and the simplest to apply. It can also be easily extended to profiles of multiple features (e.g. expression levels under several conditions) to capture a broader scope of gene function. One limitation of this approach, however, is that it requires comparable datasets describing the behavior of the different organisms at seemingly identical conditions and as such cannot rely on the wealth of data already available. One possible solution is to focus on within-species relationships between a set of genes and compare these relationships rather than properties of the individual genes. For example, instead of comparing expression profiles across species, various studies have compared the resulting co-expression relationships. Notably, this approach is not specific to co-expression, but can also be applied to any data type that may define a similarity measure between genes (e.g. protein–protein interactions [22•]).

Comparative analysis has typically been focused on the conserved features and had only begun to examine inter-species differences. The main motivation for identifying differences is their role in adaptation. However, the majority of evolutionary differences are believed to merely reflect neutral drift. Thus, a central challenge is to distinguish the differences associated with adaptive changes from the neutral ones (Figure 2). Several studies were indeed successful in connecting inter-species differences in gene expression to functional adaptation [25••]. However, a general theoretically driven framework for distinguishing adaptive changes is greatly missing. One approach to overcome this problem is to employ a systems-level view, and identify coherent differences in large sets of functionally related genes (modules; e.g. [18,23•]). Such coherent differences are unlikely to occur by chance and thus probably carry functional consequences.

In this review we have focused on the comparative analysis of gene expression data. Comparative methods have also been applied to other data types, including protein–protein interaction networks of yeast, flies, and worms [22•], or gene translation efficiencies of orthologous yeast genes [42•]. One of the remaining challenges is the integration of different data types, which is becoming particularly important as additional data types are being accumulated. Different functional properties often co-evolve and complement one another, and thus their combined analysis increases the power of comparative analysis and reveals additional insights [28•,37,43,44•,45]. This was best exemplified by the recent analysis of cell cycle regulation in *S. cerevisiae*, *S. pombe*, *A. thaliana*, and *H. sapiens* [44•], where the conserved regulation of functional complexes could only be revealed through the integration of protein–protein interactions and gene co-expression data. Indeed, whereas different genes are cell cycle regulated in these organisms, these genes represent different subunits of the same cellular complexes.

More than three decades ago Theodosius Dobzhansky had claimed that ''nothing in biology makes sense except in the light of evolution.'' Comparative analysis has proven to be a successful methodology for turning this notion into a practical framework, and with the growing wealth of functional genomic data, this framework promises to indeed turn the light on the complexity of biological systems.

Acknowledgements

This work was supported by grants from the Kahn fund for Systems Biology at the Weizmann Institute of Science, the Tauber fund and the Israeli Ministry of Science. This work was partially funded by the UniNet EC NEST consortium contract number 12990 to YB.

References and recommended reading

Papers of particular interest, published within the annual period of the review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Cliften P, Sudarsanam P, Desikan A, Fulton L, Fulton B, Majors J, Waterston R, Cohen BA, Johnston M: **Finding functional features in Saccharomyces genomes by phylogenetic footprinting**. *Science* 2003, **301**:71-76.

2. Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES: **Sequencing and comparison of yeast species to identify genes and regulatory elements**. *Nature* 2003, **423**:241-254.

3. **Insights into social insects from the genome of the honeybee Apis mellifera**. *Nature* 2006, **443**:931–949.

4. •• McCarroll SA, Murphy CT, Zou S, Pletcher SD, Chin CS, Jan YN, Kenyon C, Bargmann CI, Li H: **Comparing genomic expression patterns across species identifies shared transcriptional profile in aging**. *Nat Genet* 2004, **36**:197-204.
This study demonstrates how comparison of analogous expression datasets from distant organisms can be used to identify conserved expression programs. In particular, conserved expression programs were identified during aging of flies and worms.

5. Fay JC, McCullough HL, Sniegowski PD, Eisen MB: **Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cerevisiae***. *Genome Biol* 2004, **5**:R26.

6. • Gilad Y, Oshlack A, Smyth GK, Speed TP, White KP: **Expression profiling in primates reveals a rapid evolution of human transcription factors**. *Nature* 2006, **440**:242-245.
Comparison of gene expression levels in liver cells among humans, chimpanzees, orangutans, and rhesus macaques using a multi-species array. Both conserved and divergent genes were identified and analyzed.

7. Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, Weiss G, Lachmann M, Paabo S: **Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees**. *Science* 2005, **309**:1850-1854.

8. Kliebenstein DJ, West MA, van Leeuwen H, Kim K, Doerge RW, Michelmore RW, St Clair DA: **Genomic survey of gene expression diversity in *Arabidopsis thaliana***. *Genetics* 2006, **172**:1179-1189.

9. Landry CR, Oh J, Hartl DL, Cavalieri D: **Genome-wide scan reveals that genetic variation for transcriptional plasticity in yeast is biased towards multicopy and dispensable genes**. *Gene* 2006, **366**:343-351.

10. Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL: **Sex-dependent gene expression and evolution of the *Drosophila transcriptome***. *Science* 2003, **300**:1742-1745.

11. Rifkin SA, Kim J, White KP: **Evolution of gene expression in the *Drosophila melanogaster* subgroup**. *Nat Genet* 2003, **33**:138-144.

12. Sartor MA, Zorn AM, Schwanekamp JA, Halbleib D, Karyala S, Howell ML, Dean GE, Medvedovic M, Tomlinson CR: **A new method to remove hybridization bias for interspecies comparison of global gene expression profiles uncovers an association between mRNA sequence divergence and differential gene expression in Xenopus**. *Nucl Acids Res* 2006, **34**:185-200.

13. •• Tirosh I, Weinberger A, Carmi M, Barkai N: **A genetic signature of interspecies variations in gene expression**. *Nat Genet* 2006, **38**:830-834.
Comparison of gene expression among four yeast species subjected to several environmental stresses identified a common signature (TATA-box) in divergent genes from various studies and organisms. This study demonstrates the comparison of expression data from closely related organisms at identical conditions.

14. Yvert G, Brem RB, Whittle J, Akey JM, Foss E, Smith EN, Mackelprang R, Kruglyak L: **Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors**. *Nat Genet* 2003, **35**:57-64.

15. Lemos B, Meiklejohn CD, Hartl DL: **Regulatory evolution across the protein interaction network**. *Nat Genet* 2004, **36**:1059-1060.

16. Gu Z, Rifkin SA, White KP, Li WH: **Duplicate genes increase gene expression diversity within and between species**. *Nat Genet* 2004, **36**:577-579.

17. Berg J, Lassig M: **Cross-species analysis of biological networks by Bayesian alignment**. *Proc Natl Acad Sci USA* 2006, **103**:10967-10972.

18. Bergmann S, Ihmels J, Barkai N: **Similarities and differences in genome-wide expression data of six organisms**. *PLoS Biol* 2004, **2**:E9.

19. Stuart JM, Segal E, Koller D, Kim SK: **A gene-coexpression network for global discovery of conserved genetic modules**. *Science* 2003, **302**:249-255.

20. Tirosh I, Barkai N: **Computational verification of protein–protein interactions by orthologous co-expression**. *BMC Bioinformatics* 2005, **6**:40.

21. Snel B, van Noort V, Huynen MA: **Gene co-regulation is highly conserved in the evolution of eukaryotes and prokaryotes**. *Nucl Acids Res* 2004, **32**:4725-4731.

22. • Sharan R, Suthram S, Kelley RM, Kuhn T, McCuine S, Uetz P, Sittler T, Karp RM, Ideker T: **Conserved patterns of protein interaction in multiple species**. *Proc Natl Acad Sci USA* 2005, **102**:1974-1979.
Comparison of genome-wide PPI data using ''network alignment'', and prediction of conserved complexes and pathways.

23. • Ihmels J, Bergmann S, Berman J, Barkai N: **Comparative gene expression analysis by differential clustering approach: application to the *Candida albicans* transcription program**. *PLoS Genet* 2005, **1**:e39.
Describes a general framework for comparative analysis that focuses on the modular network structure. This framework is applied to compare two yeast species (*S. cerevisiae* and *C. albicans*) and reveals various similarities and pheonotypic-related differences.

24. Lozada-Chavez I, Janga SC, Collado-Vides J: **Bacterial regulatory networks are extremely flexible in evolution**. *Nucl Acids Res* 2006, **34**:3434-3445.

25. •• Ihmels J, Bergmann S, Gerami-Nejad M, Yanai I, McClellan M, Berman J, Barkai N: **Rewiring of the yeast transcriptional network through the evolution of motif usage**. *Science* 2005, **309**:938-940.
Comparison of the regulatory networks from two yeast species reveals a large-scale difference in the regulation of the mitochondrial ribosomal proteins. This difference was generated by the loss of dozens of regulatory motifs and is implicated in the evolution of the capacity of yeast species for anaerobic growth. This study demonstrates the comparison of co-expression networks and their modular structure.

26. Gasch AP, Moses AM, Chiang DY, Fraser HB, Berardini M, Eisen MB: **Conservation and evolution of *cis*-regulatory systems in ascomycete fungi**. *PLoS Biol* 2004, **2**:e398.

27. Hershberg R, Margalit H: **Co-evolution of transcription factors and their targets depends on mode of regulation**. *Genome Biol* 2006, **7**:R62.

28. • Tanay A, Regev A, Shamir R: **Conservation and evolvability in regulatory networks: the evolution of ribosomal regulation in yeast**. *Proc Natl Acad Sci USA* 2005, **102**:7203-7208.
Integration of expression data and promoter sequences from several yeast species revealed that extensive regulatory evolution underlies the conserved co-expression of ribosomal proteins.

29. Bilu Y, Barkai N: **The design of transcription-factor binding sites is affected by combinatorial regulation**. *Genome Biol* 2005, **6**:R103.

30. Frazer KA, Elnitski L, Church DM, Dubchak I, Hardison RC: **Cross-species sequence comparisons: a review of methods and available resources**. *Genome Res* 2003, **13**:1-12.

31. Li L, Stoeckert CJ Jr, Roos DS: **OrthoMCL: identification of ortholog groups for eukaryotic genomes**. *Genome Res* 2003, **13**:2178-2189.

32. Remm M, Storm CE, Sonnhammer EL: **Automatic clustering of orthologs and in-paralogs from pairwise species comparisons**. *J Mol Biol* 2001, **314**:1041-1052.

33. Sharan R, Ideker T: **Modeling cellular machinery through biological network comparison**. *Nat Biotechnol* 2006, **24**:427-433.

34. Huminiecki L, Wolfe KH: **Divergence of spatial gene expression profiles following species-specific gene duplications in human and mouse**. *Genome Res* 2004, **14**:1870-1879.

35. Chain FJ, Evans BJ: **Multiple mechanisms promote the retained expression of gene duplicates in the tetraploid frog *Xenopus laevis***. *PLoS Genet* 2006, **2**:e56.

36. Byrne KP, Wolfe KH: **Consistent patterns of rate asymmetry and gene loss indicate widespread neofunctionalization of yeast genes after wholegenome duplication**. *Genetics* 2006.

37. Tirosh I, Barkai N: **Comparative analysis indicates regulatory neofunctionalization of yeast duplicates**. *Genome Biol* 2007, **8**:R50.

38. Woolfe A, Elgar G: **Comparative genomics using Fugu reveals insights into regulatory subfunctionalization**. *Genome Biol* 2007, **8**:R53.

39. Kellis M, Birren BW, Lander ES: **Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae***. *Nature* 2004, **428**:617-624.

40. • Scannell DR, Byrne KP, Gordon JL, Wong S, Wolfe KH: **Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts**. *Nature* 2006, **440**:341-345.
Analysis of gene-loss following the ancient WGD. It is demonstrated that alternative copies were retained in different organisms and this differential gene-loss is proposed to underlie speciation events.

41. Dehal P, Boore JL: **Two rounds of whole genome duplication in the ancestral vertebrate**. *PLoS Biol* 2005, **3**:e314.

42. • Man O, Pilpel Y: **Differential translation efficiency of orthologous genes is involved in phenotypic divergence of yeast species**. *Nat Genet* 2007, **39**:415-421.
Comparison of codon usage (and the associated translation efficiencies) among nine yeast species. It is shown that differences in translation efficiencies among orthologs are consistent with known phenotypic differences.

43. Dutilh BE, Huynen MA, Snel B: **A global definition of expression context is conserved between orthologs, but does not correlate with sequence conservation**. *BMC Genomics* 2006, **7**:10.

44. • Jensen LJ, Jensen TS, de Lichtenberg U, Brunak S, Bork P: **Co-evolution of transcriptional and post-translational cell-cycle regulation**. *Nature* 2006, **443**:594-597.

Comparison of cell cycle regulation among *S. cerevisiae*, *S. pombe*, *A. thaliana*, and *H. sapiens*. It is shown that different genes are cell cycle regulated in each organism, but these roughly correspond to activation of the same complexes. Furthermore, changes in transcription regulation have co-evolved with changes in post-transcription regulation (i.e. protein phosphorylation).

45. Kopelman NM, Lancet D, Yanai I: **Alternative splicing and gene duplication are inversely correlated evolutionary mechanisms**. *Nat Genet* 2005, **37**:588-589.

46. Yanai I, Graur D, Ophir R: **Incongruent expression profiles between human and mouse orthologous genes suggest widespread neutral evolution of transcription control**. *Omics* 2004, **8**:15-24.

47. Liao BY, Zhang J: **Evolutionary conservation of expression profiles between human and mouse orthologous genes**. *Mol Biol Evol* 2006, **23**:530-540.

48. • Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K, Lander ES, Kellis M: **Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals**. *Nature* 2005, **434**:338-345.

Analysis of conserved non-coding sequences among human, mouse, rat, and dog, discovers numerous known and novel regulatory motifs. This work illustrates the power of evolutionary conservation as a method to identify functional elements.

49. Andolfatto P: **Adaptive evolution of non-coding DNA in Drosophila**. *Nature* 2005, **437**:1149-1152.

50. Khaitovich P, Weiss G, Lachmann M, Hellmann I, Enard W, Muetzel B, Wirkner U, Ansorge W, Paabo S: **A neutral model of transcriptome evolution**. *PLoS Biol* 2004, **2**:E132.

51. Yanai I, Korbel JO, Boue S, McWeeney SK, Bork P, Lercher MJ: **Similar gene expression profiles do not imply similar tissue functions**. *Trends Genet* 2006, **22**:132-138.