Synexpression groups in eukaryotes

Christof Niehrs & Nicolas Pollet

Division of Molecular Embryology, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany

In 1960, Jacob and Monod described the bacterial operon, a cluster of functionally interacting genes whose expression is tightly coordinated. Global expression analysis has shown that the highly coordinate expression of genes functioning in common processes is also a widespread phenomenon in eukaryotes. These sets of co-regulated genes, or 'synexpression groups', show a striking parallel to the operon, and may be a key determinant facilitating evolutionary change leading to animal diversity.

Differential gene activity largely accounts for the regulation of the genetic program underlying cell differentiation and metazoan development. Turning genes on and off at the right time determines cell fate, metabolic state and division. The development of DNA micro-arrays for the simultaneous monitoring of thousands of transcripts constitutes a breakthrough that has allowed global insights into gene expression. Now that the bigger picture is emerging, the expression data reveal a high degree of order in the genetic program: a tight coordination of the expression of groups of genes functioning in a common process. This phenomenon was initially discovered 40 years ago in the bacterial operon, but the concept was dismissed for eukaryotes because of their different mode of gene regulation.

Here we review groups of co-regulated eukaryotic genes and discuss their functional significance and relationship to other aspects of the modular genetic architecture. We focus on the evolutionary implications of coordinate gene expression because excellent reviews on application and results[1–3], methodological[4–6] and bioinformatic[7] aspects of DNA micro-arrays are available.

Identification of synexpression groups

In a typical gene-expression monitoring experiment using DNA micro-arrays, transcript levels are measured in cells after experimental treatment at various time points to yield an expression profile. Where such analyses have been carried out with sufficient temporal resolution, a correlation between the expression profiles of genes functioning in the same process has been found. In yeast, where expression of the entire genome can be monitored, 17 genes were induced by overexpression of Yap-1, a transcription factor conferring resistance to various noxic agents[8]. Of these, 9 genes encode different kinds of oxido-reductases that are thought to act in detoxification. The transcriptional program of yeast after sporulation[9] and during cell cycle[10,11] has been analysed by the same technique. In these studies, distinct temporal expression profiles were observed, and co-expression of genes involved in a common process, such as replication, chromosome pairing and mitosis, was a recurring theme (Table 1).

The response of starved human fibroblasts to serum addition has also been followed over several hours. A striking observation was not only the induction of genes involved in wound healing and the G2/M transition, but in particular the coordinate expression of five genes involved in cholesterol biosynthesis (Figs 1a, 2a)[12]. Groups of co-regulated genes were also identified by producing a high-resolution temporal expression profile during spinal cord development[13]. Using polymerase chain reaction with reverse transciption (RT-PCR) analysis, five basic waves of expression were discovered, and a number of genes acting in GABA (γ-aminobutyric acid) signalling mapped to one particular expression profile (Figs 1b, 2b).

Although this suggests that the co-expression of genes involved in the same biological process is a widespread phenomenon, in many expression profiling studies, the functional correlation of member genes is often very imperfect because of noise from false-positive genes that happen to be co-regulated. For example, the immediate early genes *Krox20* and *junB* were co-induced in a global-expression profiling experiment after growth factor treatment[14]; however, these genes have very different expression patterns and function during early mouse embryogenesis. This 'noise' is due to both the limited number of samples constituting the expression profiles, and the fact that expression profiles are typically clustered under one specific experimental condition, for example, diauxic shift, sporulation or serum addition, which each addresses only one sector of the regulatory repertoire of a cell.

Two global expression analyses have reported a very tight correlation between function and expression[15,16]. In one, a cluster analysis was carried out on all yeast genes, aggregating data of about 80 experiments from diauxic shift, cell cycle, sporulation, temperature and reducing shocks. Under these diverse conditions, 10 clusters of an exceptionally tight correlation between expression profile and function was apparent[15] (Table 1). The largest cluster is formed by more than 100 genes involved in protein synthesis —either ribosomal proteins or translation factors. Tight correlation was also observed for genes encoding components of large molecular complexes such as the proteasome or chromatin, and for genes encoding components of the glycolytic pathway (Table 1).

We used *Xenopus* embryos for expression monitoring by whole mount *in situ* hybridization to identify differentially expressed genes. Of the 273 genes identified, 38 could be grouped into four gene clusters that each share a very distinctive, complex expression pattern[16]. An example is the growth factor bone morphogenetic protein 4 (BMP-4) group (Figs 1c, 2c), members of which are expressed dorsally in the eye, heart, tailbud and lateral plate mesoderm of tailbud-stage *Xenopus* embryos. This group consists of seven members which all encode components of the BMP signalling pathway, as studied in early dorso–ventral patterning of mesoderm (Fig. 2c).

The largest group (25 members) identified in *Xenopus* is the chromatin group. These genes are characterized by their repression in tissues becoming postmitotic; and most encode chromatin proteins (for example, histones or HMG proteins), or proteins indirectly interacting with chromatin such as ornithin decarboxylase, a key enzyme in spermidin synthesis (Table 1).

We proposed the term 'synexpression groups' to designate sets of genes that share a complex 'spatial' expression pattern (multiple tissues), and that function in the same process[16,17]. However, we propose that the term could be extended to groups of genes that share complex 'temporal' (multiple time points) or 'experimental' (multiple experimental conditions) expression profiles, because the currently used term 'gene cluster' is ill defined: traditionally, it refers to physically linked genes and is therefore misleading in the context of co-expression. Also, synexpression groups allow the prediction of gene function with high precision, unlike many noise-laden co-expressed gene clusters, and they probably confer emergent properties during evolution (see below).

Notably, synexpression refers to groups of genes co-expressed

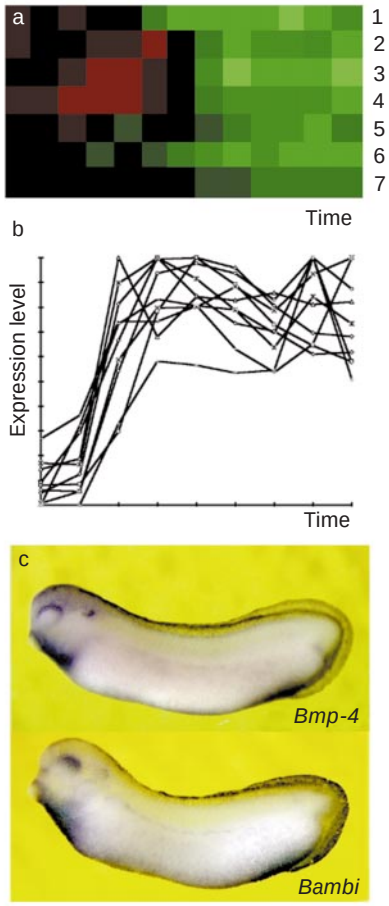

Figure 1 Identification of synexpression groups. **a,** Cholesterol biosynthesis group. Cluster image resulting from large-scale gene expression analysis using DNA micro-arrays in the response of human fibroblasts to serum. Each line depicts the temporal expression pattern of a gene. HMG CoA reductase (1); IPP isomerase (triplicate) (2–4); farnesyl-diphosphate farnesyl transferase (5); squalene epoxidase (6); cytochrome P450 lanosterol 14-α-demethylase (7). Red levels, fold repression; black levels, no change; green levels, fold induction[12]. **b**, RT-PCR analysis of genes expressed in developing rat spinal cord reveals a group of co-expressed genes defining a GABA synexpression group[13]. **c**, Wholemount *in situ* hybridization of BMP-4 and BAMBI in *Xenopus* tailbud embryos, showing the expression of members of the BMP-4 synexpression group[16]. Expression patterns are highly similar, for example in dorsal eye, heart, proctodeum and lateral plate mesoderm.

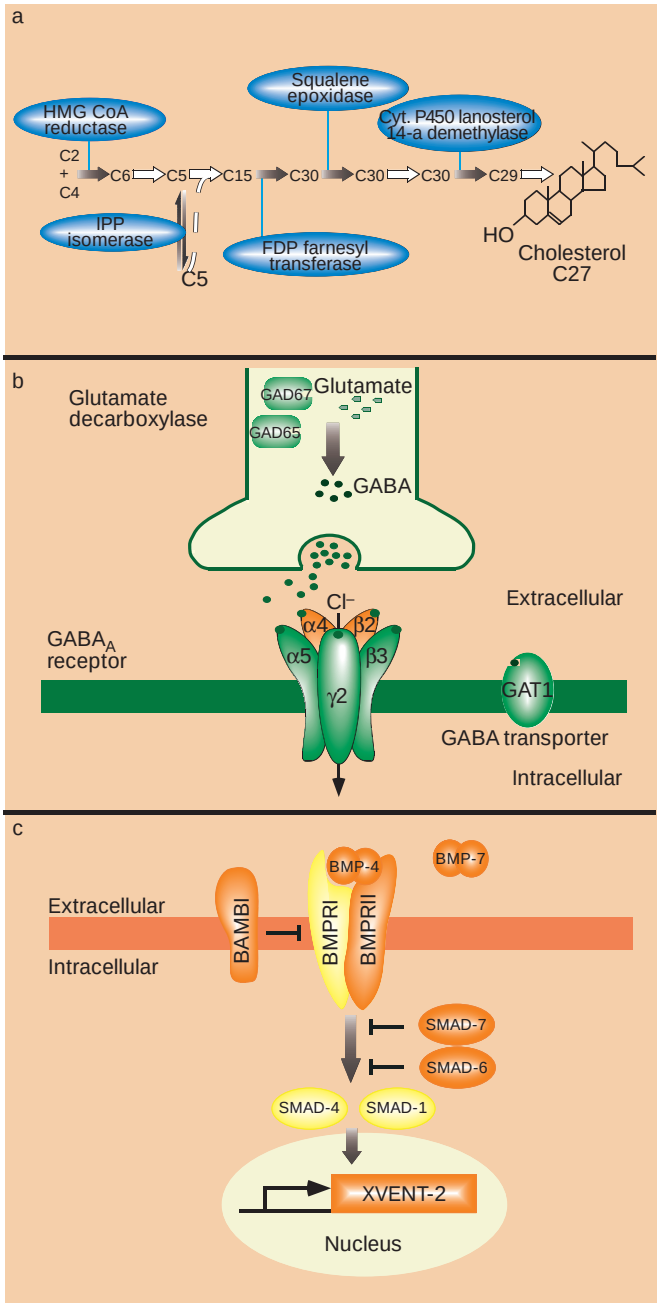

Figure 2 Synexpression groups. **a**, The cholesterol biosynthesis synexpression group and pathway, with co-expressed enzymes (see Fig. 1a) indicated (blue). C(*n*), carbon backbones. Filled and empty arrows, one or multiple enzymatic reactions, respectively. Data from ref. 12. **b**, The GABA synexpression group, with co-expressed components of GABA signal transmission (see Fig. 1b) indicated (green). Glutamate decarboxylase (isoforms GAD65 and GAD67) catalyses the decarboxylation of glutamate into the inhibitory neurotransmitter GABA, which binds to the GABA$_A$/benzodiazepine receptor, a ligand-gated chloride channel. The pentameric GABA$_A$ receptor is composed of multiple chain subtypes. GAT1 terminates the action of GABA by re-uptake into presynaptic terminals. For simplicity, not all GABA$_A$ receptor subtypes are shown. Data from ref. 13. **c**, The BMP-4 synexpression group and signalling pathway, with co-expressed components of the BMP signal transduction pathway (Fig. 1c) indicated (orange). BMP-4 and BMP-7 are ligands for type I and type II BMP transmembrane receptors. Upon ligand binding, type II receptors phosphorylate SMAD-1 transcriptional activators which complex with SMAD-4 and enter the nucleus, activating transcription of target genes such as XVENT-2. Inhibitory SMADs such as SMAD-6 and SMAD-7, as well as BAMBI[41], interfere with signal transduction. Data from refs 16 and 38–41.

under diverse conditions and hence their identification typically requires expression profiling at high resolution. This is naturally the case with *in situ* hybridization, in which the expression of one gene is profiled over thousands of samples (that is, different cells). However, screening by *in situ* hybridization is only practical for a few hundred transcripts per week. The advantage of DNA micro-arrays is that the experiment is carried out with thousands of genes simultaneously; the drawback is the limited number of samples constituting the expression profile (typically less than 20). In practical terms, it is not useful to consider genes induced in one or a few samples as a synexpression group, they may be co-regulated only under these conditions[14]. Similarly, muscle-specific genes are not necessarily members of the same synexpression group, as these genes may be involved in diverse, yet muscle-specific biological processes (for example, respiration, contraction and cell adhesion). A gene that is member of one synexpression group cannot be part of another.

In summary, tight spatio-temporal co-expression of genes functioning in the same cell-biological process is a widespread phenomenon in eukaryotes. The proteins encoded by genes of synexpression

Table 1 Examples of synexpression groups

Synexpression group (model system)	No. of genes	Method	Significance	References
DNA structure				
Histone (yeast)	8, 9	DNA chip	Chromatin structure	10, 15, 11
Chromatin (*Xenopus*)	25	*In situ* hybridization	Cell cycle regulated DNA scaffold components	16
Protein biosynthesis				
Ribosome and translation (yeast)	112	DNA chip	Protein biosynthesis	15
ER-import (*Xenopus*)	7	*In situ* hybridization	Components of ER/secretion machinery	16
mRNA splicing (yeast)	14	DNA chip	Gene expression	15
Cell cycle				
Spindle pole body assembly (yeast)	11	DNA chip	Cell division	15
Early (I) induction (yeast)	62	DNA chip	Sporulation, entry in meiosis	9
MCM complex (yeast)	34	DNA chip	DNA replication	15, 11
CIB2 (yeast)	35	DNA chip	Mitosis	11
CLN2 (yeast)	76	DNA chip	DNA replication	11
MAT (yeast)	13	DNA chip	Mating	11
G_2/M transition (yeast)	9	DNA chip	DNA replication	12
Mitochondrial function				
Mitochondrial ribosome (yeast)	22	DNA chip	Mitochondrial protein biosynthesis	15
Mitochondrial ATP Synthesis (yeast)	15	DNA chip	ATP Synthesis	15
Mitochondrial TCA cycle (yeast)	16	DNA chip	Mitochondrial metabolism	15
Metabolism				
Glycolysis (yeast)	17	DNA chip	Carbon metabolism	15
Methionine biosynthesis (yeast)	20	DNA chip	Sulfur amino acid metabolism	11
Cholesterol biosynthesis (human fibroblasts)	5	DNA chip	Membrane biosynthesis	12
Cell signalling				
BMP-4 (*Xenopus*)	7	*In situ* hybridization	BMP signalling	38,39,16,40
Delta1 (*Xenopus*)	5	*In situ* hybridization	Delta–Notch signalling	16
GABA (rat)	9	RT-PCR	GABA neurotransmission	13
Others				
Yap-1 response (yeast)	17	DNA chip	Resistance to oxidative stress	8
Proteasome (yeast)	27	DNA chip	Protein degradation	15

Groups of coordinately expressed genes with tight correlation between expression profiles and role of the genes in a defined process were selected. ER, endoplasmic reticulum. TCA, tricarboxylic acid.

groups are not necessarily homologous (for example, receptor, ligand and transcription factor in the BMP-4 synexpression group; Fig. 2c) and hence co-expression is not the consequence of gene duplication. The close correlation between expression pattern and gene function in synexpression implies that high-resolution expression profiling allows the assembly of genes into groups that define molecular pathways, and allows predictions to be made about the roles of constituent members with unknown function.

Such prognoses have already proved successful[9,18,19], and although the groups of genes did not reveal a genetic hierarchy nor a biochemical protein function, they permitted the formulation of specific and testable hypotheses which were rapidly confirmed experimentally. A similar route is being taken in microbial genetics, in which DNA sequencing of entire bacterial genomes has identified new operons and unknown gene functions using this 'guilt-by-association' paradigm[20]. Furthermore, as at least three synexpression groups contribute to mitochondrial function (ATP synthesis, trichloroacetic acid cycle and mitochondrial ribosome[15]; Table 1), synexpression groups may reveal subroutines in biological programmes.

Synexpression groups versus operons

The first example of tight integration of gene expression and function was the *lac* operon, a transcriptional unit of genes functioning in a common metabolic pathway in bacteria utilizing lactose[21]. The genes encode *lacZ*, a hydrolase cleaving lactose, *lacY*, a galactoside permease and *lacA*, a galactoside transacetylase. The co-expression of these genes is achieved by their organization in a linear transcription unit that is regulated by a repressor. In prokaryotes, genes of many pathways are organized into operons, and remarkably there appear to be operons for cell division and protein export[20,22], similar to the respective yeast and *Xenopus* synexpression groups (Table 1). Estimates are that only a minority of genes is organized in operons (13%)[23] or synexpression groups (5–10%)[15,16].

Operons are not generally found in eukaryotes. Genes are not transcribed as polycistronic messenger RNA but are individually regulated by complex promoters and are arranged on chromosomes without correlation regarding their function or expression, although there are exceptions[24]. Hence, co-expression is unlikely to be achieved by *cis*-regulation of physically associated transcription units, as many genes of a synexpression group are located on different chromosomes. Synexpression is more probably coordinated by one or a few *trans*-acting factors regulating common promoter elements. Such promoter elements and their candidate transcriptional regulators have been identified in yeast[8,9,11]. For example, a single ECB *cis*-regulatory element is sufficient to confer early G1 expression of *cdc6*, *cdc46* and *swi4* through the MCM1 *trans*-acting factor[25]. This simple regulatory mode contrasts with the majority of higher eukaryotic genes with complex promoters that are controlled by numerous *trans*-acting factors[26].

In the operon, the lac repressor serves simultaneously as the transcriptional regulator and as a 'sensor'[27] that integrates cellular signals, in this case the availability of lactose, to orchestrate expression of a battery of structural genes. Although such simple modes may also operate in synexpression groups, it can be envisaged that the site of integration of complex signals is made upstream of transcriptional regulators, for example, through growth factors.

Evolutionary implications

The concept of synexpression groups implies that within individual member genes, there is co-evolution of function and expression, in other words, co-evolution of promoter and coding sequences. Hence, there must be selective pressure promoting this type of genetic organization. There are two modes that may explain the adaptability of synexpression groups, one benefiting the individual organism, the other its capacity to evolve. We will only outline the possible consequences of modular gene expression on evolution.

For the individual organism, operons and synexpression groups

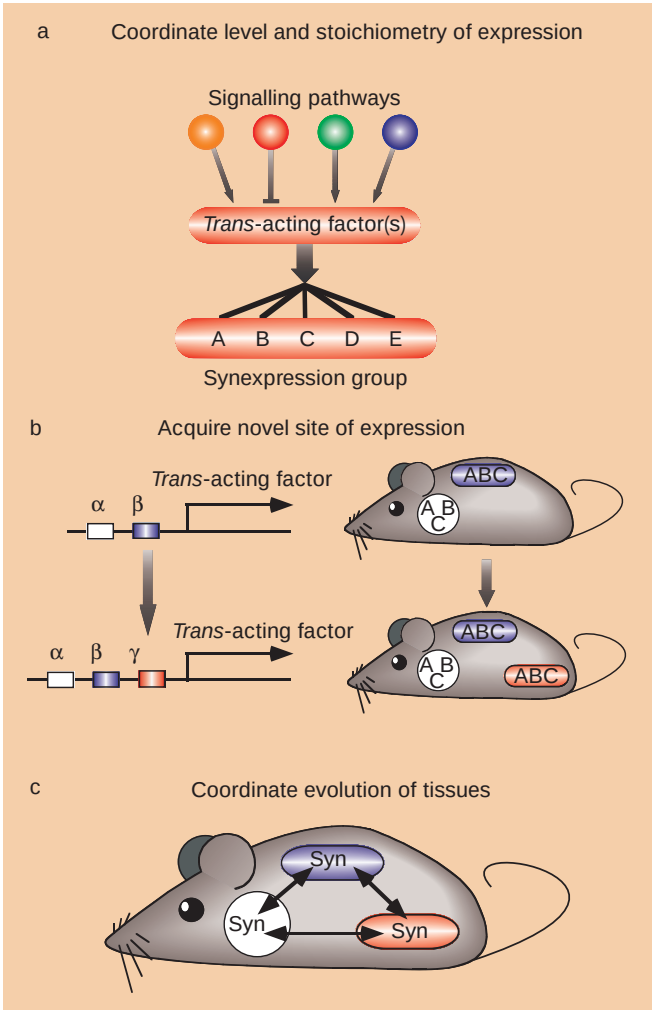

Figure 3 Adapative advantages of synexpression groups. **a**, A battery of interacting genes (A–E) is co-controlled through common *cis*-regulatory elements by a *trans*-acting factor (or a network thereof) that is regulated by inputs from various signalling pathways. Common regulation of expression of functionally interacting genes allows co-regulation of expression level, for example, upon the appropriate stimulus, and promotes synthesis of individual gene products in stoichiometric amounts. More complex interactions between individual members of synexpression groups (cross-talk and feedback mechanisms) are possible but are not shown for simplicity. **b**, A battery of interacting genes (A–C) is co-regulated by a *trans*-acting factor, which in turn is controlled by tissue- (or stage-) specific promoter elements, α and β. Acquisition of promoter element, γ, allows deployment of the entire synexpression group in a new site. Note that this enhanced integration is paralleled by an increased genetic constraint due to the coupling of genetic modules. Loss of the individual elements α, β or γ may lead to rudimentation of the respective organ. Synexpression groups may regulate each other leading to highly complex gene networks. **c**, Tissues sharing synexpression groups (Syn) will experience evolutionary covariation, which leads to physiological integration and to co-evolution of the linked tissues. The consequences are enhanced rate of evolution and increased pleiotropy and genetic constraints (see refs 33 and 42).

may be adaptive because they allow the co-expression of a set of genes at high level only when required, for example, the *lac* operon in the presence of galactose or the detoxification group in the presence of poisons. Apart from energetic economy, interacting gene products frequently need to assemble stoichiometrically or may require co-translation for forming a complex[20], which is promoted by co-expression (Fig. 3a). Therefore, components of supramolecular complexes will probably be organized in synexpression groups, as is the case for the proteasome, ribosomal machinery and the chromatin complex (Table 1).

The other major reason why synexpression groups may be adaptive is because they enhance the modularity of the genetic potential and promote rapid evolution. The primary source leading to animal diversity is not thought to be the differences in gene products but the differences in the networks in which the gene products are connected during development. The regulatory linkages that make the function of gene products or whole genetic modules contingent on extracellular conditions are thought to be a major factor facilitating evolutionary change[27,28–30]. The concept of such genetic modules, also called syntagms or gene cassettes, was developed by *Drosophila* geneticists when it was realized that embryogenesis can be broken down into discrete elementary operations, such as segmentation, neurogenesis or appendage formation. These processes often encompass the same genetic pathways at multiple stages and multiple places of development[31,30], for example, the mitogen-activated protein kinase pathway used in the development of *Drosophila* eye, *Caenohrabditis elegans* vulva and *Xenopus* mesoderm[32]. By definition, every synexpression group is a gene cassette or a subset thereof.

Compartmentalizing an organism's genetic potential in synexpression groups may produce evolutionary change requiring few mutational steps. In bacteria, for example, the horizontal mode of gene transfer may drive the evolution of operons: during conjugation only parts of the genome may be exchanged, and thus physical proximity of interacting genes promotes transfer of complete functional units[22]. A similar reasoning applies to eukaryotes and metazoa. First, through acquisition or loss of a single region-specific promoter element controlling relevant *trans*-acting factors, an entire synexpression battery may be re-deployed in a new developmental context (time or site of expression) or be connected to other regulatory pathways (Fig. 3b). An enormous complexity of regulatory networks can be built with this model, assuming that higher order *trans*-acting factors integrate multiple synexpression groups. In the classic Britten–Davidson model of coordinate gene regulation, these and other adaptive features of integrated gene expression have been considered, although the proposed mechanism for co-expression involving regulatory antisense RNAs has not been borne out[27].

Furthermore, it may be adaptive to tie certain tissues sharing the same synexpression group genetically together because it allows co-evolution of those tissues, including properties such as size, shape and enzyme composition[3] (Fig. 3c). For example, the precursors of fore- and hindlimbs were genetically linked together early in their history by the common expression of *Hox* genes. This may explain why digits arose at the same time in hand and foot; these appendages could evolve together presumably because their development was brought under the control of a common enhancer for the *Hox* complex[34]. However, although it is reasonable that the genetic linkage of appendages may lead to adaptive co-diversification of these functionally related structures, it is less obvious which properties are shared by the central nervous system and developing somites that makes it advantageous to have them co-controlled by the Delta–Notch pathway[16] (Table 1). Possibly, it is the need to select one or a few cells from an equivalent group to take on one particular cell fate, which is promoted by the lateral inhibition mediated by the Notch pathway[35]. However, a combination of tissues expressing a synexpression group may just reflect evolutionary history—the chance acquisition of a functional module proving to be adaptive.

Only a subset of functionally interacting genes is found in synexpression groups. For example, some but not all genes involved in methionine biosynthesis are part of the respective synexpression group[11], and *SMAD-1* and *SMAD-4* which also participate in BMP-4 signalling are ubiquitously expressed during early embryogenesis[36,37] and thus are not part of the BMP-4 group. One possible explanation for such non-selected genes is that they are

evolutionarily younger and may not have had enough time to be recruited. Another reason may be that non-recruited components serve in multiple processes (for example, SMAD-4 in BMP, TGF-β, as well as activin pathways) and are not tightly coupled to any one synexpression group. This suggests that a gene involved in multiple pathways may rarely be a member of a synexpression group.

Once a gene is recruited into a synexpression group it should be very stable because of the connectivity in multiple tissues, and we propose that members of synexpression groups will be conserved even between different animal phyla. Indeed, there are chromatin synexpression groups both in yeast and in *Xenopus* (Table 1); and in the homologous *Drosophila* dpp and *Xenopus* Bmp-4 pathways, inhibitory SMADs (*dad*; *SMAD6/7*) are co-expressed with their ligands. In contrast, the region-specific deployment of individual synexpression groups by changing the regulation of a sensor[27], may be subject to much greater variability because it requires few genetic changes.

Conclusion

Synexpression groups reflect the functional compartmentalization of the eukaryote genome and have a striking parallel to the prokaryote operon. DNA micro-array expression analysis is clearly the most efficient way to identify more synexpression groups, but a future challenge will be to carry out profiling over hundreds or thousands of conditions to map the genomic programme at high resolution. We will probably encounter synexpression groups that make 'no sense' at first glance, in other words, groups in which the functions of the co-expressed genes have no common denominator. These might be particularly interesting, as they might reveal an otherwise hidden logic of cellular regulation. Where synexpression groups couple seemingly unrelated tissues, we might search for adaptive reasons for their integration. The hardwired correspondence between gene function and regulation shown by synexpression groups unveils a degree of order that was previously unsuspected and provides us with new opportunities to study genetic networks. □

1. Brown, P. O. & Botstein, D. Exploring the new world of the genome with DNA microarrays. *Nat. Genet.* **21,** 33–37 (1999).
2. Gerhold, D., Rushmore, T. & Caskey, C. T. DNA chips: promising toys have become powerful tools. *Trends Biochem. Sci.* **24,** 168–173 (1999).
3. Lipshutz, R. J., Fodor, S. P., Gingeras, T. R. & Lockhart, D. J. High density synthetic oligonucleotide arrays. *Nat. Genet.* **21,** 20–24 (1999).
4. Bowtell, D. D. Options available—from start to finish—for obtaining expression data by microarray. *Nat. Genet.* **21,** 25–32 (1999).
5. Cheung, V. G. *et al.* Making and reading microarrays. *Nat. Genet.* **21,** 15–19 (1999).
6. Duggan, D. J., Bittner, M., Chen, Y., Meltzer, P. & Trent, J. M. Expression profiling using cDNA microarrays. *Nat. Genet.* **21,** 10–14 (1999).
7. Bassett, D. E., Eisen, M. B. & Boguski, M. S. Gene expression informatics—it's all in your mine. *Nat. Genet.* **21,** 51–55 (1999).
8. DeRisi, J. L., Vishwanath, R. I. & Brown, P. O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278,** 680–686 (1997).
9. Chu, S *et al.* The transcriptional program of sporulation in budding yeast. *Science* **282,** 699–705 (1998).
10. Cho, R. J. *et al.* A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell* **2,** 65–73 (1998).
11. Spellman, P. T. *et al.* Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* **9,** 3273–97 (1998).
12. Vishwanath, R. I. *et al.* The transcriptional program in the response of human fibroblasts to serum. *Science* **283,** 83–87 (1999).
13. Wen, X. *et al.* Large-scale temporal gene expression mapping of central nervous system development. *Proc. Natl Acad. Sci. USA* **95,** 334–339 (1998).
14. Fambrough, D., McClure, K., Kazlauskas, A. & Lander, E. S. Diverse signaling pathways activated by growth factor receptors induce broadly overlapping, rather than independent, sets of genes. *Cell* **97,** 727–741 (1999).
15. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95,** 14863–14868 (1998).
16. Gawantka, V. *et al.* Gene expression screening in *Xenopus* identifies molecular pathways, predicts gene function and provides a global view of embryonic patterning. *Mech. Dev.* **77,** 95–141 (1998).
17. Niehrs, C. Gene-expression screens in vertebrate embryos: more than meets the eye. *Genes Funct.* **1,** 229–231 (1997).
18. Onichtchouk, D. *et al.* The Xvent-2 homeobox gene is part of the BMP-4 signaling pathway controling dorso–ventral patterning of *Xenopus* mesoderm. *Development* **122,** 3045–3053 (1996).
19. Jen, W. -C., Gawantka, V., Pollet, N., Niehrs, C. & Kintner, C. Periodic repression of Notch pathway genes governs the segmentation of *Xenopus* embryos. *Genes Dev.* **13,** 1486–1499 (1999).
20. Dandekar, T., Snel, B., Huynen, M. & Bork, P. Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem. Sci.* **273,** 324–328 (1998).
21. Jacob, F. The operon—25 years later. *C. R. Acad. Sci. Paris* **320,** 199–206 (1997).
22. Lawrence, J. G. Selfish operons and speciation by gene transfer. *Trends Microbiol.* **5,** 355–359 (1997).
23. Blattner, F. R. *et al.* The complete genome sequence of Escherichia coli K-12. *Science* **277,** 1453–1474 (1997).
24. Blumenthal, T. Gene clusters and polycistronic transcription in eukaryotes. *BioEssays* **20,** 480–487 (1998).
25. McInerny, C. J., Partridge, J. F., Mikesell, G. E., Creemer, D. P. & Breeden, L. L. A novel Mcm1-dependent element in the SWI4, CLN3, CDC6, and CDC47 promoters activates M/G1-specific transcription. *Genes Dev.* **11,** 1277–1288 (1997).
26. Yuh, C. H., Bolouri, H. & Davidson, E. H. Genomic *cis*-regulatory logic: experimental and computational analysis of a sea urchin gene. *Science* **279,** 1896–1902 (1998).
27. Britten, R. J. & Davidson, E. H. Gene regulation for higher cells: a theory. *Science* **165,** 349–357 (1969).
28. Gerhart, J. & Kirschner, M. *Cells, Embryos and Evolution* (Blackwell, Malden, 1997).
29. Duboule, D. & Wilkins, A. S. The evolution of 'bricolage'. *Trends Genet.* **14,** 54–59 (1998).
30. Huang, F. Syntagms in development and evolution. *Int. J. Dev. Biol.* **42,** 487–494 (1998).
31. Jan, Y. N. & Jan, L. Y. Functional gene cassettes in development. *Proc. Natl Acad. Sci. USA* **90,** 8305–8307 (1993).
32. Nishida, E. & Gotoh, Y. The MAP kinase cascade is essential for diverse signal transduction pathways. *Trends Biochem. Sci.* **18,** 128–131 (1993).
33. Cheverud, J. M. Developmental integration and the evolution of pleiotropy. *Am. Zool.* **36,** 44–50 (1996).
34. Shubin, N., Tabin, C. & Carroll, S. Fossils, genes and the evolution of animal limbs. *Nature* **388,** 639–648 (1997).
35. Artavanis-Tsakonas, S., Rand, M. D. & Lake, R. J. Notch signaling: cell fate control and signal integration in development. *Science* **284,** 770–776 (1999).
36. Lagna, G., Hata, A., Hemmati-Brivanlou, A. & Massague, J. Partnership between DPC4 and SMAD proteins in TGF-beta signalling pathways. *Nature* **383,** 832–836 (1996).
37. Meersman, G. *et al.* The C-terminal domain of Mad-like signal transducers is sufficient for biological activity in vivo and transcriptional activation. *Mech. Develop.* **61,** 127–140 (1997).
38. Bhushan, A., Chen, Y. & Vale, W. SMAD7 inhibits mesoderm formation and promotes neural cell fate in *Xenopus* embryos. *Dev. Biol.* **200,** 260–268 (1998).
39. Frisch, A. & Wright, C. V. E. XBMPRII, a novel *Xenopus* type II receptor mediating BMP signalling in embryonic tissues. *Development* **125,** 431–442 (1998).
40. Hata, A., Lagna, G., Massague, J. & Hemmati-Brivanlou, A. SMAD6 inhibits BMP/SMAD1 signaling by specifically competing with the SMAD4 tumor suppressor. *Genes Dev.* **12,** 186–197 (1998).
41. Onichtchouk, D. *et al.* Silencing of TGF-β signalling by the pseudoreceptor BAMBI. *Nature* **400,** 480–485 (1999).
42. Wagner, G. P. in *Advances in Artificial Life* (eds Moran, F., Moreno, A., Merelo, J. J. & Chacon, P.) 317–328 (Springer, Berlin, 1995).

Acknowledgements

We thank D. Duboule and W. Pyerin for critical reading of the manuscript.

Correspondence should be addressed to C. N.