A green light to expression in time and space

Susan E Mango

The latest genome-scale study in worms examines spatiotemporal gene expression patterns by automated GFP imaging.

© 2007 Nature Publishing Group http://www.nature.com/naturebiotechnology

In 1888, Heinrich Wilhelm Gottfried von Waldeyer-Hartz unwittingly began the 'omics revolution by coining the term 'chromosome' to describe colored (chromo) bodies (soma) that could be stained and visualized. Chromosome begat genome, which was followed by proteome, transcriptome, orfeome, interactome, metabolome and, most recently, the localizome—a representation of global gene expression throughout an organism. Writing in this issue, Dupuy *et al.*[1] present a high-throughput approach for visualizing a partial localizome through development in *C. elegans*. The resulting gene expression profiles were subsequently used for cluster analysis to suggest potential functional groups of genes.

The year 1998 was an important one for *C. elegans*. The genome sequence was completed[2], and RNA interference was discovered[3]. These two events, coupled with the development of green fluorescent protein (GFP) to monitor expression, set the stage for global studies of gene function and expression[4–8]. Typically, to determine when and where a gene is expressed, GFP reporter constructs are introduced into *C. elegans* as semi-stable pseudochromosome arrays. Dupuy *et al.* simplify the task of imaging GFP reporter constructs by using a special flow cytometer, the complex object parametric analysis and sorter instrument (Biosorter).

The Biosorter can handle large samples such as an intact *C. elegans* worm and generates expression profiles along the anterior-posterior axis. Because the size of a worm corresponds to its developmental stage, the expression profiles also reveal when in development a GFP reporter is activated. Dupuy *et al.* averaged and aligned the profiles of multiple worms to generate a 'chronogram' that provides temporal and spatial information on GFP expression for a particular reporter (**Fig. 1**). The advantages of this approach are its speed and the fact that thousands of animals can be analyzed for each GFP reporter. These features enabled the authors to obtain expression profiles for 876 genes, from larval development to adulthood.

Susan E. Mango is at the Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah, Salt Lake City, Utah 84112, USA.
email: susan.mango@hci.utah.edu

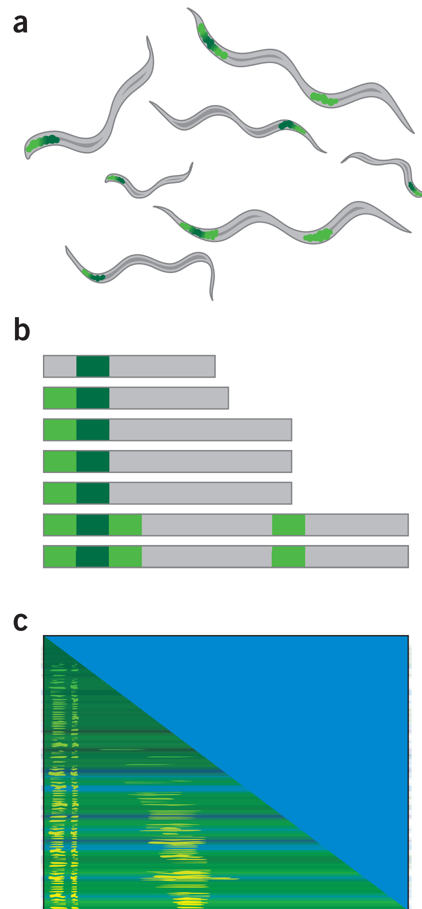

Figure 1 Chronogram assembly. (**a**) Worms expressing GFP. (**b**) Processing of the worms by the Biosorter determines fluorescence along the anterior-posterior axis. (**c**) Alignment and averaging generates a chronogram of expression from larval stages to adulthood.

The chronograms were used to identify genes with similar expression profiles, which could reveal proteins involved in a common developmental process or biochemical pathway. In one approach, pairs of chronograms were compared using Pearson's correlation coefficient. Those chronograms with a high correlation coefficient were more likely to share similar expression patterns and, in some cases, similar GO terms. This analysis was extended to generate trees by hierarchical clustering, similar to the clustering method used for microarray studies. Genes in closely related subtrees shared expression domains and often biological features as well. For example, genes expressed in embryos according to their chronograms were also enriched for the GO terms 'DNA metabolism' and 'mitotic cell cycle'. Surprisingly, however, genes in subtrees did not share similar RNAi phenotypes.

Inspection of chronograms revealed that many reporters, once activated, remained on throughout development. GFP is a relatively stable protein, which may obscure fluctuations in gene expression. Alternatively, continual GFP expression may reflect the natural state. With the exception of the vulva and germ line, the morphology and anatomy of *C. elegans* larvae do not undergo dramatic changes, and expression patterns may be relatively stable in many tissues. Future studies using destabilized GFP, *in situ* or microarray analysis may help resolve these issues.

The major limitation of the analysis, however, is the expression patterns themselves. One complication is the Biosorter, which cannot identify individual cells. Thus, expression in tissues such as 'intestine' and 'body wall muscle' are virtually impossible to distinguish without additional, traditional imaging. Moreover, the Biosorter is not as sensitive as some other imaging approaches, which may help explain why Dupuy *et al.* obtained data for only 54% of their GFP reporters, compared with 92% in another high-throughput study in which 366 promoters were analyzed[9].

A second complication is the accuracy of the GFP reporters. The reporters rely on fusions between upstream intergenic sequences and GFP. However, *C. elegans* regulatory elements often reside within introns[5], and an estimated 2.7% of genes are predicted to have internal promoters[10]. Absence of intronic regulatory sequences can produce faulty ectopic expression (from lack of a repressor element) or absence of normal expression (from lack of an enhancer). The current estimate is that up to 50% of genes may depend on internal or distant *cis*-regulatory sequences that would be missing from the reporters under study[9]. Post-transcriptional regulation of translation or mRNA stability through the 3′ untranslated region can also contribute to regulation, but 3′ untranslated regions were not included in this study. These problems should be relatively easy to solve in the future, given the small size of most *C. elegans* genes.

The inclusion of vector sequences upstream of GFP can pose another problem, by introducing cryptic regulatory elements such as artificial foregut or intestinal enhancers[9]. The presence of a cryptic enhancer may explain why over a third of GFP reporters analyzed by Dupuy *et al.* were expressed in the foregut. A simple solution would be to remove vector sequences from the DNAs used to generate the pseudochromosome arrays. Germline expression poses yet another hurdle, because exogenous DNA sequences, including GFP reporters, are often silenced in the germ line[11].

A third complication of the GFP expression patterns is their complexity. Genes are rarely expressed in one cell type or one developmental stage, complicating computational analysis. For example, if two genes are expressed in a common tissue but each is also expressed robustly elsewhere, their correlation coefficients will drop, and the relationship of one to another may go unrecognized. This problem is particularly acute for transcription factors, which often work combinatorially in diverse cell types[12].

The approach of Dupuy *et al.* provides an interesting starting point for high-throughput expression studies from larval development to adulthood. Automated analysis of GFP reporters in living animals is a powerful technique that will complement the current approaches of microarray and *in situ* analysis. A challenge for the future will be to refine the chronograms by increasing the imaging sensitivity and cellular resolution. Recently, Bao *et al.*[13] developed an automated imaging approach that can trace single cells in embryos. Although this method has been tested only in embryos and has a lower throughput than Dupuy *et al.*'s approach, if combined with more accurate GFP constructs, it may solve some of the problems outlined here. As these approaches become established, an exciting prospect will be to observe changes in expression under different conditions. The current study was conducted with wild-type hermaphrodites at the onset of starvation. Modifying environmental cues, genotype or gender will elucidate how the localizome changes under different conditions.

COMPETING INTERESTS STATEMENT
The author declares no competing financial interests.

1. Dupuy, D. *et al. Nat. Biotechnol.* **25**, 663–668 (2007).
2. *C. elegans* Sequencing Consortium. *Science* **282**, 2012–2018 (1998).
3. Fire, A. *et al. Nature* **391**, 806–811 (1998).
4. Kamath, R.S. *et al. Nature* **421**, 231–237 (2003).
5. Gaudet, J. & Mango, S.E. *Science* **295**, 821–825 (2002).
6. Kim, S.K. *et al. Science* **293**, 2087–2092 (2001).
7. Li, S. *et al. Science* **303**, 540–543 (2004).
8. Baugh, L.R. *et al. Development* **130**, 889–900 (2003).
9. Reece-Hoyes, J.S. *et al. BMC Genomics* **8**, 27 (2007).
10. Choi, J. & Newman, A.P. *Dev. Biol.* **296**, 537–544 (2006).
11. Schaner, C.E. & Kelly, W.G. *Germline Chromatin in WormBook* doi/10.1895/wormbook.1.129.1 (ed. The *C. elegans* Research Community) (2006).
12. Gaudet, J. *et al. PLoS Biol.* **2**, e352 (2004).
13. Bao, Z. *et al. Proc. Natl. Acad. Sci. USA* **103**, 2707–2712 (2006).

Membrane proteins up for grabs

Gunnar von Heijne

Protein designers have produced transmembrane helices that interact specifically with target membrane proteins.

Attempts at *de novo* protein design have so far been restricted to small, water-soluble proteins. But why not try membrane proteins? This simple question has now been asked for the first time by DeGrado and colleagues[1] in a recent paper in *Science*. Using a clever *in silico*-design algorithm that tries to optimize shape complementarity between a target transmembrane α-helix and a designed helix, the authors succeeded in constructing three helices that bind with high affinity and specificity to two closely related integrin target helices and even specifically activate one or the other integrin *in vivo*. These results have interesting implications for membrane protein design and suggest new strategies for drug development.

Gunnar von Heijne is at the Center for Biomembrane Research, Department of Biochemistry and Biophysics, Stockholm University, SE-10691 Stockholm, Sweden. e-mail: gunnar@dbb.su.se

The work by DeGrado and colleagues is the culmination of 15 years of studies of the sequence patterns that drive the dimerization of transmembrane helices. Pioneering work in Don Engelman's lab implicated the so-called GX_3G motif as the main determinant of homodimerization of glycophorin A[2,3], a protein with a single transmembrane helix found in the plasma membrane of red blood cells. Structure determination by NMR later confirmed the hypothesis that the glycine residues allow exceptionally tight packing between the two glycophorin A transmembrane helices, leading to the formation of a very stable right-handed coiled-coil structure[4]. GX_3G motifs are found in many (but certainly not all) transmembrane helices[5], and their special status is underscored by studies in which stably interacting helices have been selected *in vivo* from large, randomized protein libraries[6].

Last year, DeGrado's group categorized different helix-helix packing configurations found in known membrane-protein structures[7]. A generalization of the right-handed, GX_3G-mediated helix dimer—the "parallel GAS_{Right}" motif—was the third most common configuration found, after the "antiparallel GAS_{Left}" and "antiparallel GAS_{Right}" motifs. All in all, five well-defined motifs were found to account for three-quarters of all helix-helix pairs in the structures examined.

DeGrado and colleagues have now taken the next logical step, namely to use the insights gained from the previous studies to guide the development of a simple and apparently very successful algorithm for designing interacting pairs of transmembrane helices (**Fig. 1**). In short, once the target helix and the desired interaction motif have been chosen, all helix-helix pairs with this motif are pulled out from the database of membrane protein structures. The target helix is then 'threaded' onto one of the two helices in the identified helix pairs, creating a set of design templates composed of the 'threaded' target helix and the peptide backbone of a partner helix that becomes the template for the new "CHAMP" helix (CHAMP stands for "computed helical anti-membrane protein").

In the next step, a side-chain rotamer library is used to systematically decorate the CHAMP-helix backbone until a satisfactory packing with the target helix is obtained. Finally, the completely lipid-exposed residues on the CHAMP helix are selected randomly from a set of hydrophobic residues (leucine, isoleucine, valine, alanine, phenylalanine). In general, this procedure results in a number of candidate CHAMP helices for a given target helix, and the final selection among the candidates is made based on the uniformity of packing and the absence of large voids in the helix-helix interface.

To test their approach, DeGrado and colleagues chose two closely related integrin transmembrane helices, α_{IIb} and α_v, as their targets. Both contain a "parallel GAS_{Right}" motif, and both can form heterodimers with

© 2007 Nature Publishing Group http://www.nature.com/naturebiotechnology