

Deciphering the regulatory code

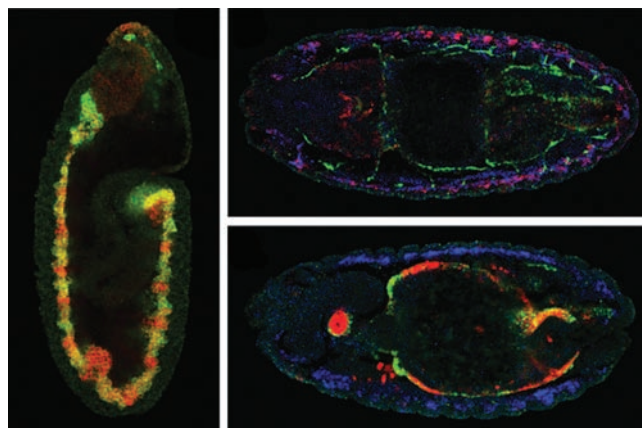
EMBL scientists take new approach to predict gene expression

Heidelberg, 5 November 2009 - Embryonic development is like a well-organised building project, with the embryo's DNA serving as the blueprint from which all construction details are derived. Cells carry out different functions according to a developmental plan, by expressing, i.e. turning on, different combinations of genes. These patterns of gene expression are controlled by transcription factors: molecules which bind to stretches of DNA called cis-regulatory modules (CRMs), and, once bound, switch the relevant genes on or off. Thanks to scientists at the European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany, it is now possible to accurately predict when and where different CRMs will be active. The study, published today in *Nature*, is a first step towards forecasting the expression of all genes in a given organism and demonstrates that the genetic regulation that is crucial for correct embryonic development is more flexible than previously thought.

Through an interdisciplinary collaboration between biologist Robert P. Zinzen, computer scientist Charles Girardot and statistician Julien Gagneur, a novel, integrated approach was possible. They combined detailed experimental data about where and when transcription factors are binding to CRMs with a computational approach, and were able to forecast CRM activity.

"Going from global binding data to CRM activity was a big challenge in the field – one which we have now begun to overcome", says Eileen Furlong, who headed the study.

Using a comprehensive, systematic approach, the scientists identified and recorded the binding profiles – i.e. the combinations of transcription factors binding at different times and places – of approximately 8000 CRMs involved in regulating muscle development in the fruit fly *Drosophila*. The activity of a number of such CRMs had been previously studied, and the EMBL team used this information to group them into classes according to the type of muscle and developmental stages they were active in. The scientists then trained a computer to unravel the binding profiles for each of these groups, and search the 8000 newly identified CRMs for ones whose binding profiles fitted that picture. Such CRMs were predicted to have similar activity patterns, implying they



© Furlong/EMBL

These fluorescence microscopy images of fruit fly embryos demonstrate that the scientists' computer predictions were correct. As predicted, during the early stages of development (left) a CRM called 1070 is active (red) in the mesoderm (green) – the tissue which will give rise to all muscle types. At a later developmental stage (right, top), the same CRM is active (red/pink) in the embryo's body wall muscle (blue), but not in its gut muscle (green). At the same time (right, bottom), another CRM, called 5570 (red), drives development in the gut muscle (green) but not in the body wall muscle (blue).

are involved in regulating the development of the same muscle type.

When the scientists tested their predictions experimentally, the results were not only accurate but also enlightening. It turns out that the regulatory code, in which one binding profile leads to one pattern of CRM activity, is actually not that straightforward. CRMs with strikingly different binding profiles can have similar patterns of activity. This plasticity was unexpected, but makes sense in evolutionary terms, the researchers say. The fact that different combinations of transcription factors, or binding codes, can regulate the same developmental process means that even if some transcription factors or CRMs change or are lost during an organism's evolution, it can still develop a gut muscle, for instance.

"What's exciting for me is that this study shows that it is possible to predict when and where genes are expressed, which is a crucial first step towards understanding how regulatory networks drive development", Furlong concludes. ●

Source Article

Zinzen, R.P., Girardot, C., Gagneur, J., Braun, M. & Furlong, E.E.M. Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature*, 5 November 2009.

Contact:

Sonia Furtado, EMBL Press Officer, Heidelberg, Germany, Tel: +49 6221 387 8263, www.embl.org, sonia.furtado@embl.de

About EMBL

The European Molecular Biology Laboratory is a basic research institute funded by public research monies from 20 member states (Austria, Belgium, Croatia, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom) and associate member state Australia. Research at EMBL is conducted by approximately 80 independent groups covering the spectrum of molecular biology. The Laboratory has five units: the main Laboratory in Heidelberg, and Outstations in Hinxton (the European Bioinformatics Institute), Grenoble, Hamburg, and Monterotondo near Rome. The cornerstones of EMBL's mission are: to perform basic research in molecular biology; to train scientists, students and visitors at all levels; to offer vital services to scientists in the member states; to develop new instruments and methods in the life sciences and to actively engage in technology transfer activities. EMBL's International PhD Programme has a student body of about 170. The Laboratory also sponsors an active Science and Society programme. Visitors from the press and public are welcome.

Policy regarding use

EMBL press and picture releases including photographs, graphics, movies and videos are copyrighted by EMBL. They may be freely reprinted and distributed for non-commercial use via print, broadcast and electronic media, provided that proper attribution to authors, photographers and designers is made. High-resolution copies of the images can be downloaded from the EMBL web site: www.embl.org