SMP: RNA

Project: High-throughput High Resolution Gene Expression Analysis in Mid-gestation Mouse Embryos

Bernhard Herrmann - Max Planck Institute for Molecular Genetics, Berlin - herrmann@molgen.mpg.de

Introduction

Understanding the pathogenesis of human disease is an essential prerequisite for developing efficient pharmaceutical compounds and therapies. Many human diseases arise by distortions in signalling mechanisms of complex nature. Signal cascades are no isolated entities, but interconnected. Moreover, there is only a fairly small number of signal cascades which are utilized over and over again in a rather large number of processes in the adult organism as well as in embryonic development. They are modulated by many modifiers resulting in various cellular responses. To date it is not clear, except for a few known cases, how signal cascades are interconnected and what distortions exactly lead to disease development. For that reason it is important to study signalling mechanisms, including cellular readouts, globally.

One important step in the functional analysis of genes involved in cellular communication is the identification of the place of gene action, that is, the cellular context in which the gene is activated. Whole-mount in situ hybridisation of mouse embryos is a suitable tool to identify the expression patterns of genes during mid-gestation embryonic development, a stage which is characterized by extensive signalling and rapid cellular changes (about 40% of the genes specifically expressed at that stage are involved in transcriptional regulation or signalling; ref 1). Mid-gestation mouse embryos are ideally suited to study a multitude of signal pathways in vivo and identify genes involved in such pathways. I want to stress here that the embryo is not merely a transition stage, but the basis on which the adult organism has been built during evolution. Thus, the embryo teaches us how the adult organism is designed and working.

Project Status

In the German Human Genome Project we have analysed more than 10.000 cDNAs for expression in mid-gestation mouse embryos (ref. 1, and unpublished data). More than 1000 cDNAs showed a restricted expression pattern marking a multitude of tissue types and organ anlagen (Fig. 1). Such data allow the identification of functional groups (e.g. co-expression groups) of genes involved in the same process, which can be subjected to functional analysis. Among many others, 152 orthologs of human chromosome 21 genes have been analysed.[2] The latter analysis provided a number of candidates for the pathogenesis of Down's syndrome caused by trisomy 21.

Nevertheless, our database still does not cover all genes encoded by the genome. Therefore, we need to continue analysing gene expression in mouse embryos in order to complete the molecular anatomy of the mid-gestation mouse embryo.

Due to progress in identification of the genes encoded by the human and the mouse genomes and their assignment to gene ontology groups we are now in the position to select genes for analysis on the basis of their belonging to functional groups. Our focus is on transcriptional regulators, including genes involved in chromatin control, and signalling molecules. In addition, we utilize expression profiling on Affymetrix Gene Chips and bioinformatics analyses to identify genes, which have not been analysed yet. In this way we plan to complete the transcriptomes of stage E9.5 to E11.5 of mouse embryos.

Several thousand genes will be analysed for expression in whole mount mid-gestation mouse embryos (stage E9-E11), including expression analysis, digital image acquisition, image description and processing for storage in the MAMEP (Molecular Analysis of the Mouse Embryo Project) database established by us in DHGP-II. Over 400 genes have been analysed already in the first half of 2005, since the beginning of NGFN2.

In addition, we will set up a system for producing 3-D images (OPT: ref. 3) and start a 3-D image database for high-resolution gene expression data in mid-gestation mouse embryos. This will greatly improve usability of such data in the future.

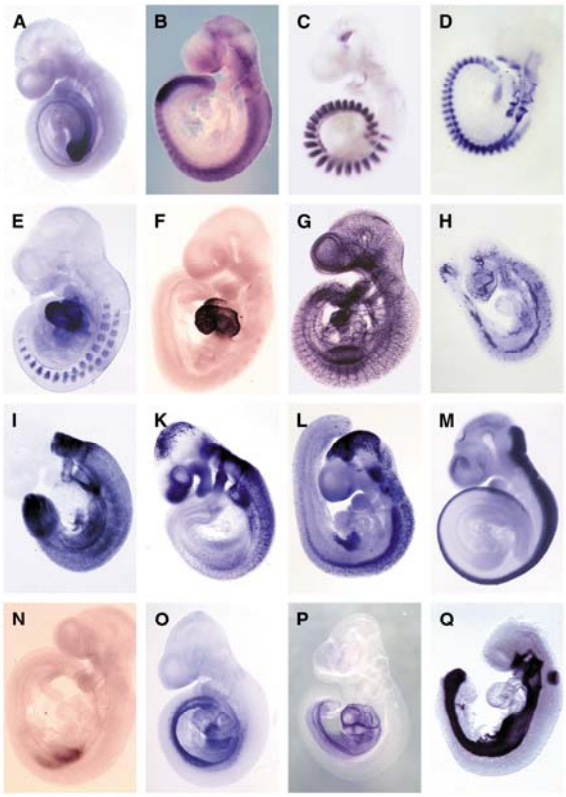

Fig 1: *Expression analysis in whole mount E9.5 mouse embryos. Main sites of expression are A tailbud, B presomitic mesoderm and neural tube, C somites, D sclerotome and pharyngeal endoderm, E heart and myotomes, F heart, G blood vessel endothelium, H blood cells, I hind brain and trunk, K neural crest cells, L neuroblasts, M dorsal hindbrain and spinal cord, N forelimb buds, O urogenital ridge and lateral mesoderm, P mesothelium, Q gut.*

Gene expression analysis identifies functionally related genes involved in mesenchyme formation, a process promoting metastasis formation.

Among the genes showing a restricted expression pattern in the embryo we have, among thousands of unrelated genes identified several hundred genes expressed during mesoderm formation. In this process, epithelial proliferating stem cells are induced to undergo a cellular switch to form mesenchymal cells under control of signalling molecules. This differentiation process is very similar to the formation of

mesenchymal cells from epithelial proliferating tumours, a prerequisite for metastasis formation as observed in human patients. Identical (orthologous) genes are involved in both processes, mesenchyme formation in mouse embryos and metastasis formation in human tumors. Thus, the mouse embryo can serve as in vivo model to dissect and finally understand mesenchyme formation on a molecular level. Such data are bound to further our knowledge of metastasis formation in the near future and will support the development of new drugs targeted against cancer cells.

High-resolution gene expression data form the basis for constructing regulatory networks controlling tissue and organ development

The expression analysis in whole mouse embryos allows us to identify co-expression groups of genes involved in the same process such as organ formation or tissue development. For instance, genes controlling the formation and patterning of the brain or the heart, or as mentioned before, the formation of mesenchyme from epithelial stem cells, can be collected and as a group subjected to further analyses. That means, the expression data serve as tool to spot functionally related genes in a myriad of unrelated genes, and to select genes from such collections for functional analyses.

The SMP_RNA project is tightly connected to several projects in NGFN-2 aiming at understanding the regulatory networks controlling tissue and organ development. Regulatory networks are constructed on the basis of several data sets. In our case, expression data form the basis, promoter analyses on hundreds of related genes in silico, in vitro and finally in vivo (SMP_DNA) extend these data to derive functional relationships between genes expressed during the same process in the same or neighbouring cells. Finally, predicted genetic interactions between these genes are investigated in vivo in mutant embryos produced by knock-down mutagenesis (SMP_RNAi). Functional analyses in vivo are essential for complementing and controlling network descriptions based on in silico data.

Outlook

Gene expression data form the basis for the global analysis of regulatory networks controlling tissue and organ formation in the embryo, In the first step such regulatory networks are developed on transcriptional read-outs. These need to be extended to protein networks in the future, aiming at a global understanding of the networks and the elements controlling differentiation processes as well as leading to disease when impaired.

*Lit.: **1.** Neidhardt L et al. Large scale screen for genes controlling embryogenesis, using high-throughput gene expression analysis in mouse embryos. Mech. Development 2000; 98: 77-93 **2.** Gitton Y et al. A gene expression map of human chromosome 21 orthologs in the mouse. Nature 2002; 420: 586-590 **3**. Sharpe J et al. Optical Projection Tomography as a tool for 3D microscopy and gene expression studies. Science 2002; 296: 541-545*