Mechanisms of Development 122 (2005) 365–439

www.elsevier.com/locate/modo

An atlas of differential gene expression during early *Xenopus* embryogenesis

Nicolas Pollet*, Nadja Muncke, Barbara Verbeek, Yan Li[1], Ursula Fenger, Hajo Delius, Christof Niehrs

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

Received 21 March 2004; received in revised form 11 October 2004; accepted 7 November 2004
Available online 30 December 2004

Abstract

We have carried out a large-scale, semi-automated whole-mount in situ hybridization screen of 8369 cDNA clones in *Xenopus laevis* embryos. We confirm that differential gene expression is prevalent during embryogenesis since 24% of the clones are expressed non-ubiquitously and 8% are organ or cell type specific marker genes. Sequence analysis and clustering yielded 723 unique genes displaying a differential expression pattern. Of these, 18% were already described in *Xenopus*, 47% have homologs and 35% are lacking significant sequence similarity in databases. Many of them encode known developmental regulators. We classified 363 of the 723 genes for which a Gene Ontology annotation for molecular function could be attributed and found 'DNA binding' and 'enzyme' the most represented terms. The most common protein domains encoded in these embryonic, differentially expressed genes are the homeobox and RNA Recognition Motif (RRM). Fifty-nine putative orthologs of human disease genes, and 254 organ or cell specific marker genes were identified. Markers were found for nasal placode and archenteron roof, organs for which a specific marker was previously unavailable. Markers were also found for novel subdomains of various other organs. The tissues for which most markers were found are muscle and epidermis. Expression of cell cycle regulators fell in two classes, containing proliferation-promoting and anti-proliferative genes, respectively. We identified 66 new members of the BMP4, chromatin, endoplasmic reticulum, and karyopherin synexpression groups, thus providing a first glimpse of their probable cellular roles. Cluster analysis of tissues to measure tissue relatedness yielded some unorthodox affinities besides expectable lineage relationships. In conclusion, this study represents an atlas of gene expression patterns, which reveals embryonic regionalization, provides novel marker genes, and makes predictions about the functional role of unknown genes.

Keywords: *Xenopus laevis*; Synexpression; Screening; In situ hybridization; Expressed sequence tags; Partial cDNA sequencing; Functional genomics; Regionalization; Database; cDNA sequencing; Expression profiles; Pattern formation; Transcriptional regulation; Amphibian

1. Introduction

The nuclear genomic DNA of various eukaryotes has been sequenced and the speed at which sequence information is generated in model organisms is growing. So far, 11 animal genomes have been completed or drafted (including man) and the majority of the projects under way are on vertebrate species. However, functional gene analysis is not keeping pace with sequencing progress. For example, from the recently completed sequence of human chromosome 14, a quarter of annotated genes have been classified as putative. Furthermore, Gene Ontology annotation is available for only 59% of the 30,000 or so mouse genes (Okazaki et al., 2002). Even if a molecular function can be tentatively assigned for the majority of the remaining genes by sequence homology, their biological role is typically unknown.

Modern developmental biology forms an important part of postgenomic basic research by describing the function and interactions of genes and gene products during development.

* Corresponding author. Present address: Laboratoire de Transgenèse et Génétique des Amphibiens, CNRS UMR 8080, IBAIC Bat 447 Université Paris Sud, F-91405 Orsay, France. Tel.: +33 169 156 816; fax: +33 169 157 273.
E-mail address: nicolas.pollet@ibaic.u-psud.fr (N. Pollet).
[1] Present address: Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, OH 45229-3039, USA.

0925-4773/$ - see front matter © 2004 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.mod.2004.11.009

Transcriptome analysis is an important part of this effort and various high-throughput techniques, such as microarrays, SAGE or RT-PCR are being employed (Velculescu et al., 1995; Wen et al., 1998; Holloway et al., 2002). However, these techniques have a limited resolution and are unable to reveal complex spatial expression patterns so characteristic of many developmental control genes. Therefore, whole mount in situ hybridisation (WISH) screens have and are being carried out, with the aim to catalogue genes with differential expression. These screens identify marker genes for organ/cells of interest, reveal the logic of progressive embryonic patterning, identify candidate developmental control genes and uncover synexpression groups, i.e. genetic modules composed of genes that share both a complex expression pattern and a biological process in which they function (Niehrs and Pollet, 1999). One chief advantage of WISH screening is the immediate access to the cDNA for functional analysis by mRNA microinjection and Morpholino antisense oligonucleotide knock-down (Smith and Harland, 1991; Nutt et al., 2001).

WISH screens have been previously reported in nematode, mouse, zebrafish, medaka-fish, *Ciona* and *Xenopus* (Bettenhausen and Gossler, 1995; Birchall et al., 1995; Gawantka et al., 1998; Kopczynski et al., 1998; Neidhardt et al., 2000; Kudoh et al., 2001; Nguyen et al., 2001; Quiring et al., 2004). An alternative method to identify differentially expressed genes is the random insertion of reporter constructs by transgenesis (Gossler et al., 1989). Such gene trap screens have the advantage to produce mutations and to allow live recording of the expression pattern. However, the difficulty to identify the targeted transcription unit and the difference of half-lives between the reporter and the endogenous mRNA make WISH screen the tool of choice for systematic studies.

We have previously carried out a pilot WISH screen in *Xenopus* using randomly picked cDNA clones from an early neurula stage library (Gawantka et al., 1998), and described 273 unique differentially expressed genes. Here we have extended this screen using a combination of cDNA libraries of different embryonic stages and describe the pattern and partial cDNA sequence of 723 unique, differentially expressed genes.

2. Results

2.1. Overview

To uncover a broad sprectrum of differentially expressed genes, we performed the WISH screen on cDNAs from different embryonic libraries to minimize the redundancy while maximising novel gene discovery (Table 1). First, clones were randomly picked either from a stage 13 library (‘random st13’), or a library from embryos hyperdorsalised with LiCl at 32 cell stage and harvested at stage 13 (‘random st13 LiCl’) and sequenced afterwards. Hyperdorsalisation is known to trigger the overexpression of genes expressed in dorsal structures, such as the Spemann organizer. Second, publicly available, partially sequenced clones from the mentioned stage 13 LiCl (‘EST st13 LiCl’) or from a stage 24 library (‘EST st24’), of the *Xenopus* EST IMAGE (Klein et al., 2002) collection were screened without eliminating redundant clones. Third, a non-redundant set of 2431 clones was screened, which derived from clustering of 40824 ESTs and 3465 known *Xenopus* gene sequences (‘EST clustered’). All WISH expression pictures, detailed expression annotation and sequence data are available electronically in the axeldb database http://www.dkfz.de/molecular_embryology/axeldb.htm (Pollet et al., 2000b). The accession numbers are given in Table S1 and in axeldb.

From 18048 bacterial clones that were initially picked as PCR templates, 8369 passed PCR amplification and RNA probe synthesis and were screened semi-automatically by WISH using a robot. Clones corresponding to differentially expressed genes at any of the examined stages (st10,5 13 and 30) were selected and the expression pattern was confirmed and documented after at least one additional round of WISH. Sixty-three percent of the 8369 clones were screened using the random strategy and 37% from ESTs. To estimate the proportion of clones with wrong assignment of sequence in public EST database, we resequenced 16 rearrayed clones and found two assignment errors.

The highest recovery of clones with restricted expression patterns (Table 1, ratio *R*/*S*) was from the randomly picked cDNA libraries. This may be due to biological reasons,

Table 1
Overall statistics

Libraries	*S*=No. of clones WISH screened	*R*=No. of clones with restricted patterns	No. of genes	Ratio *R*/*S*	No. of marker genes	
					No.	% of unique
RANDOM st13	3182	854	239	0.27	69	28.9
RANDOM st13 LiCl	2050	625	262	0.30	60	22.9
EST st13 LiCl	315	28	21	0.09	5	23.8
EST st24	391	109	86	0.28	35	39.8
EST clustered	2431	413	248	0.17	101	40.7
Total	8369	2029	723	0.24	254	35.1

The number of clones screened per library is presented with the number of corresponding unique genes and marker genes.